Tensor-Based Multi-Modal Multi-Target Regression for Alzheimer's Disease Prediction

Jun Yu[1], Benjamin Zalatan[1], Yong Chen[2], Li Shen[2], and Lifang He[1,*]

[1]*Department of Computer Science and Engineering, Lehigh University*, Bethlehem, PA, U.S.A.
{juy220, bjz222, lih319}@lehigh.edu
[2]*Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania*, Philadelphia, PA, U.S.A.
{ychen123, li.shen}@pennmedicine.upenn.edu

Abstract—The assessment of Alzheimer's Disease (AD) progression via the analysis of physical changes within the brain has attracted great interest from the fields of healthcare, computational medicine, and machine learning alike. Recent studies have demonstrated that using both multi-modal data and multiple AD assessment scores in a predictive model can better reflect pathological characteristics and enhance prediction performance. However, using such high-dimensional structure information to model inter-correlation between multiple targets remains a challenging task. In this paper, we propose a Tensor-based Multi-modal Multi-Target Regression (TMMTR) method for AD detection and prediction, which enables simultaneously modeling multilinear structure information as well as intrinsic inter-target correlations in a general learning framework. We also investigate the tensor-structured sparsity that supports the interpretability of our prediction. Experiments conducted on the ADNI dataset validate the superior performance of our method when compared to other state-of-the-art methods.

Index Terms—Alzheimer's disease, multi-modal, multi-target regression, tensor, interpretability

I. INTRODUCTION

Alzheimer's Disease (AD) is a neurodegenerative disorder that is the most common cause of dementia [1]. Those inflicted by AD often experience memory deterioration and cognitive decline that increases in severity over time [2]. In order to effectively slow disease progression and prevent the onset of more severe symptoms, the investigation of computer-aided prediction and diagnosis of the prodromal stage of AD, also known as Mild Cognitive Impairment (MCI), has gained popularity in recent years due to its effectiveness and potential for refinement. Over the last decade, several studies [3]–[5] have shown great promise in AD prediction by analyzing Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) scans. Previous research has also shown that the prediction of AD-related clinical scores, e.g. Mini-Mental State Examination (MMSE), can be used to determine the state of cognitive decline in patients. Regression models [6]–[8] are extensively explored to estimate these continuous clinical scores from brain-imaging scans of identifiable biomarkers. In contrast to categorical predictive models that focus on classification, regression models can produce granular and specific predictions on continuous variables rather than values in terms of quantized classes. In most cases, we can solve a classification problem by constructing multiple binary classifiers using logistic regressions [9].

It has been found that using biomarkers across multiple modalities achieves better predictive performance for MCI/AD assessment than conventional single-modal learning as it can contain complementary information of clinical relevance [10]–[15]. In some cases, the large amount of data that comes with using multiple modalities may introduce undesirable noise, which could hinder the learning process. Along with this, conventional multi-modal fusing techniques, such as principal component analysis (PCA) [16] or kernel-based methods [17], can ignore data structure and waste informative features, thus leading to mediocre performance with little interpretability. Accordingly, feature reduction [18], feature selection [19] and feature clustering [20] are generally used to mitigate these issues. However, all existing methods above rely on a single target to generate inter-correlations between biomarkers from different modalities, which could easily lead to overfitting on high-dimensional training data.

Recent AD prediction methods [21]–[25] have begun to use Multi-Target Regression (MTR) methods to find inter-target relations and cross-modal connections simultaneously. The Sparse Multi-Task Regression and Feature selection (SMART) method [21] has been shown to be effective at recognizing the interconnected structures in neuroimaging data. To overcome the limitation of using $\ell_{2,1}$ norm in SMART that could cause the model to learn identical regions across different modalities, the Multi-Task Sparse Group Lasso (MT-SGL) method [23] encourages individual feature selection coupled with group lasso. Furthermore, the Robust Multi-Label Transfer Feature Learning (rMLTFL) method [24] applies the $\ell_{2,1}$ norm on both the feature selection and modality transfer to model inter-target relationships on unrelated source modalities. However, these sparse MTR models seek interpretability from vector-level space, which can achieve acceptable performance when selecting effective factors, but ignore the intrinsic structure within high-dimensional tensor data. Deep models such as neural networks [22], [25] and graph convolutional networks [26] generally outperform conventional sparse models by using more complicated feature mappings, but can become non-interpretable as they are prone to make decisions in a latent feature space.

In this paper, we propose the Tensor-based Multi-modal

---

* Corresponding author.

Multi-Target Regression (TMMTR) algorithm to bridge the gap of learning interpretable features from tensor-structured multi-modal data. In particular, we jointly regress three clinical scores—AD Severity (ADS) score, AD Assessment Scale–Cognitive 13-item (ADAS-Cog 13), and MMSE—from three different imaging modalities: voxel-based morphometry on MRI (VBM-MRI), ^{18}F-fluorodeoxyglucose PET (FDG-PET) and 18-F florbetapir PET (AV45-PET). The main contributions are threefold:

- TMMTR leverages the tensor-structured information inherent in high-dimensional multi-modal data and inter-target correlation among related responses to improve learning performance.
- We present an efficient divide-and-conquer algorithm to solve this tensor-based sparse model and investigate tensor-level sparsity in the pursuit of finding multilinear relationships.
- To verify the consistent performance of TMMTR on tensor-structured data, we conduct extensive experiments on tensors of different dimensionalities. The results show that our method can compete with state-of-the-art methods in both performance and interpretability.

II. MATERIALS AND METHODS

A. Data Acquisition and Preprocessing

The data used in this study was acquired from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.adni-info.org). We collected data from a total of 692 non-Hispanic Caucasian participants including 182 Cognitively Normal (CN) controls, 75 patients with Significant Memory Concern (SMC), 217 patients with Early Mild Cognitive Impairment (EMCI), 184 patients with Late Mild Cognitive Impairment (LMCI), and 97 AD patients. For each patient, we gathered neuroimage data in the VBM-MRI, FDG-PET and AV45-PET modalities, as well as three clinical scores of ADS, ADAS-Cog 13 and MMSE.

The multi-modal image data was aligned to a single visit for each patient. The structural MRI scans were preprocessed with voxel-based morphometry (VBM) using the SPM software [27]. All scans were aligned to a T1-weighted template image. They were then segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) maps, normalized to the standard Montreal Neurological Institute (MNI) space as $2 \times 2 \times 2$ mm^3 voxels, and smoothed with an 8 mm full-width at half-maximum (FWHM) kernel. The FDG-PET and AV45-PET scans were also registered to the same MNI space using SPM. The MarsBaR toolbox [28] was used to group voxels into 116 regions of interest (ROIs) defined in [28]. ROI-level measures were calculated by averaging the measures of all voxels within each ROI.

B. Tensor Construction

Tensors are defined as the higher-order generalizations of vectors and matrices, which can be regarded as 1-mode and 2-mode tensors respectively. Tensors are commonly used in medical research to represent multi-dimensional and multi-relational data. In this study, we represent tensor data with three different sizes: (1) $\mathbf{116 \times 3}$. We concatenate 116 ROI feature values from all three imaging modalities along an additional modality dimension in the order of VBM-MRI, FDG-PET and AV45-PET, which results in a two-dimensional tensor of size 116×3 for each patient. (2) $\mathbf{116 \times 116}$. We construct a ROI-to-ROI connectivity matrix based on the pairwise similarity of ROIs, obtaining a two-dimensional tensor of size 116×116 for each patient. For each ROI, we concatenate the features from each of the three modalities into a three-dimensional vector, denoted as $\mathbf{r}_i \in \mathbb{R}^3, i = 1, 2, \cdots, 116$. We then use a K-Nearest Neighbor (KNN) graph [29] to construct a connectivity matrix via the Gaussian similarity function, i.e., $\mathbf{X}_{ij} = \exp(-\frac{\|\mathbf{r}_i - \mathbf{r}_j\|^2}{\sigma^2})$, where σ is a user defined parameter specifying width. In our study, we simply set $\sigma = 1$ and consider $K = 1, 2, \cdots, 116$ as a hyper-parameter. When $K = 116$, the tensor is fully connected. (3) $\mathbf{116 \times 116 \times 3}$. We construct three ROI-to-ROI connectivity matrices using the same method specified for the 116×116 tensor, one for each modality, and concatenate them along an additional modality dimension. This results in a three-dimensional tensor of size $116 \times 116 \times 3$ for each patient.

C. Tensor-Based Multi-Modal Multi-Target Regression

In this section, we will formalize the TMMTR method and present an efficient algorithm to perform it. Fig. 1 shows the learning pipeline of our TMMTR method, where we use tensors as input data to regress multiple targets simultaneously.

1) Notation and Tensor Algebra: We begin with the notations and definitions of tensor operations. For more details, the readers can refer to [30]. We denote tensors by calligraphic letters (e.g., $\mathcal{X}, \mathcal{W}$), matrices by bold uppercase letters (e.g., $\mathbf{X}, \mathbf{W}$), vectors by bold lowercase letters (e.g., $\mathbf{x}, \mathbf{w}$), and scalars by lowercase letters (e.g., x, w). Any M-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_M}$, we denote $x_{i_1, \cdots, i_M}$ as the $(i_1, \cdots, i_M)$-th entry, where $i_1 \in \{1, \cdots, I_1\}, \cdots, i_M \in \{1, \cdots, I_M\}$.

Definition 1 (Inner product): The inner product of two tensors $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_M}, \mathcal{B} \in \mathbb{R}^{I_1 \times \cdots \times I_M}$ is defined as $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1=1}^{I_1} \cdots \sum_{i_M=1}^{I_M} a_{i_1, \cdots, i_M} b_{i_1, \cdots, i_M}$.

Definition 2 (Outer/Tensor product): The outer product of two tensors $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_M}, \mathcal{B} \in \mathbb{R}^{I_{M+1} \times \cdots \times I_N}$ is defined as $\mathcal{A} \otimes \mathcal{B} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, of which the $(i_1, \cdots, i_N)$-th entry is given by $a_{i_1, \cdots, i_M} b_{i_{M+1}, \cdots, i_N}, i_1 \in \{1, \cdots, I_1\}, \cdots, i_N \in \{1, \cdots, I_N\}$.

Definition 3 (n-mode product): The n-mode product of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_M}$ with a vector $\mathbf{b} \in \mathbb{R}^{I_m}$ is defined as $\mathcal{A} \times_m \mathbf{b} \in \mathbb{R}^{I_1 \times \cdots I_{m-1} \times I_{m+1} \cdots \times I_M}$, of which the $(i_1, \cdots i_{m-1}, i_{m+1}, \cdots, i_M)$-th entry is given by $\sum_{i_m=1}^{I_m} a_{i_1, \cdots, i_M} b_{i_m}$.

Definition 4 (CP-decomposition): A tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_M}$ can be decomposed as a weighted sum of rank-1 tensors as $\mathcal{X} = \sum_{r=1}^{R} \sigma_r \mathbf{x}_r^{(1)} \otimes \cdots \otimes \mathbf{x}_r^{(M)}$, where $\sigma_r \in \mathbb{R}, \mathbf{x}_r^{(m)} \in \mathbb{R}^{I_m}$, and $\|\mathbf{x}_r^{(m)}\|_2 = 1$ hold for every $r \in \{1, \cdots, R\}, m \in \{1, \cdots, M\}$.
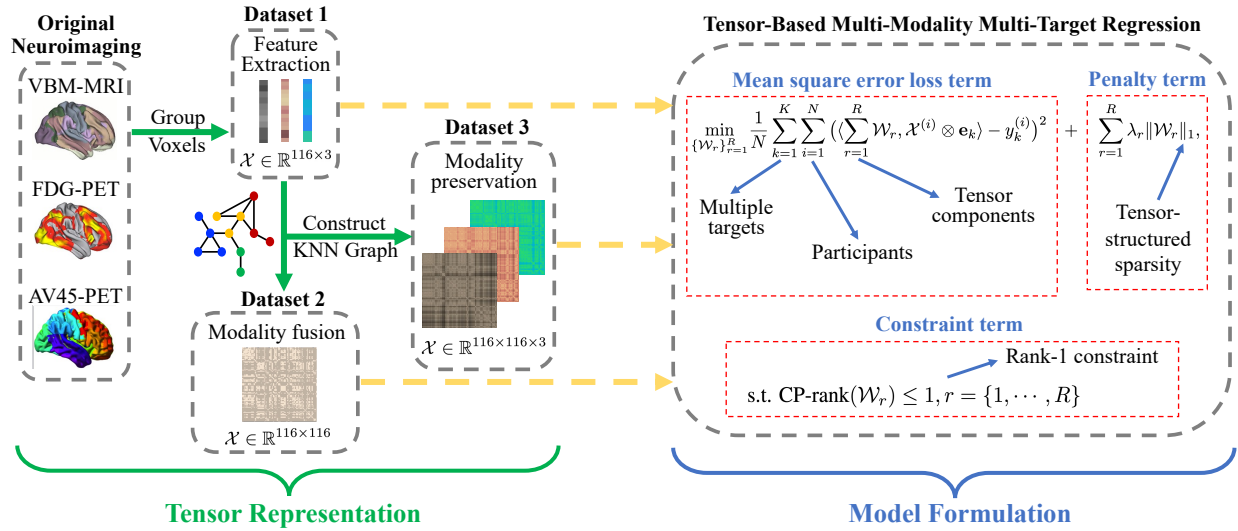
Fig. 1. The learning pipeline of our proposed method. The left part shows the data preprocessing stage, and the right gives a detailed illustration of TMMTR.

Definition *5 (CP-rank):* The tensor rank[1] of $\mathcal{A}$, denoted by CP-rank($\mathcal{A}$), is the smallest number R such that the CP-decomposition is exact.

2) Problem Formalization: Suppose that the dataset includes N sample pairs $\{(\mathcal{X}^{(i)}, \mathbf{y}^{(i)})\}, i = 1, 2, \cdots, N$, where $\mathcal{X}^{(i)} \in \mathbb{R}^{D_1 \times \cdots \times D_M}$ is the multi-modal data tensor, and its response vector $\mathbf{y}^{(i)}$ contains K different response targets, i.e., $\mathbf{y}^{(i)} = [y_1^{(i)}, y_2^{(i)}, \cdots, y_K^{(i)}]^\top$.

In general, each target indicates a specific task and is associated with a linear regression function f_k as

$$f_k(\mathcal{X}_i) = \langle \mathcal{W}^k, \mathcal{X}_i \rangle + \varepsilon_{i,k}, i = 1, \cdots, N, k = 1, \cdots, K, \quad (1)$$

where $\varepsilon_{i,k}$ is the error term associated with i-th sample and its corresponding k-th target, and $\mathcal{W}^k$ is the coefficient tensor for the prediction of k-th target. To measure the distance between predicted values and their corresponding ground-truth responses, we introduce K different loss functions for each target. Accordingly, the preliminary learning framework can be formulated as follows:

$$\min \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{N} \mathcal{L}_k \left(\langle \mathcal{W}^k, \mathcal{X}^{(i)} \rangle, y_k^{(i)} \right) + \lambda \Omega(\mathcal{W}), \quad (2)$$

where $\mathcal{L}_k$ is the specific loss function for the corresponding k-th response target, and $\mathcal{W} = [\mathcal{W}^1, \cdots, \mathcal{W}^K] \in \mathbb{R}^{D_1 \times \cdots \times D_M \times D_{M+1}}$ is the stacking of K coefficient tensors, where $D_{M+1} = K$. We denote Ω and λ as the regularized term and balancing hyper-parameter, respectively.

3) Objective Function: In our method, we hope to model these K linear relationships collectively. It is essential to learn the multilinear map for both the inter-target correlations and input-target relationships. We appropriately introduce one-hot encodings $\mathbf{e}_k = [\underbrace{0, \cdots, 0}_{k-1}, 1, 0, \cdots, 0]^\top \in \mathbb{R}^K$ as the indicator

vector for the k-th target, $k = 1, \cdots, K$. At the same time, we employ Mean Squared Error (MSE) to compute the loss of all response targets. Problem (2) is reformulated as:

$$\min \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{N} \left(\langle \mathcal{W}, \mathcal{X}^{(i)} \otimes \mathbf{e}_k \rangle - y_k \right)^2 + \lambda \Omega(\mathcal{W}), \quad (3)$$

where $\mathcal{X}^{(i)} \otimes \mathbf{e}_k$ increases the dimensionality of the input tensor via zero padding (See **Definition** 2), and the higher order coefficient tensor $\mathcal{W} \in \mathbb{R}^{D_1 \times \cdots \times D_M \times K}$ has a more compact tensor structure more suitable for decomposition.

To utilize the high-dimensional structure and correlations within the tensor coefficients $\mathcal{W}$, we propose a sparse, low-rank tensor regression model that decomposes the tensor coefficients as follows:

$$\min_{\{\mathcal{W}_r\}_{r=1}^R} \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{N} \left(\langle \sum_{r=1}^{R} \mathcal{W}_r, \mathcal{X}^{(i)} \otimes \mathbf{e}_k \rangle - y_k^{(i)} \right)^2 +$$

$$\sum_{r=1}^{R} \lambda_r \|\mathcal{W}_r\|_1, \text{ s.t. CP-rank}(\mathcal{W}_r) \leq 1, r = \{1, \cdots, R\}, \quad (4)$$

where $r \in \{1, \cdots, R\}$ is the sequential number of the rank-1 terms $\mathcal{W}_r = \sigma_r \mathbf{w}_r^{(1)} \otimes \cdots \otimes \mathbf{w}_r^{(M+1)}$, i.e., CP-rank($\mathcal{W}_r$) = 1. Thus we have $\|\mathcal{W}_r\|_1 = \|\sigma_r \mathbf{w}_r^{(1)} \otimes \cdots \otimes \mathbf{w}_r^{(M+1)}\|_1 = \sigma_r \prod_{m=1}^{M+1} \|\mathbf{w}_r^{(m)}\|_1$. λ_r is the corresponding hyper-parameter for each rank-1 component to balance the loss term and penalty term.

Due to the ℓ_1-norm regularizers acting on the decomposed rank-1 tensors, the solution of problem (4) prefers a sparse model where few parameters are non-zero, thus resulting in a better interpretability. The zero-elements in $\mathcal{W}$ indicate that the corresponding features are negligible in the prediction responses.

4) Solution: A common way to solve problem (4) is to utilize the alternating least squares (ALS) algorithm [30] to compute CP-decomposition with a fixed R. However, this has an undesirable computational cost as R must be readjusted

---

[1]This is different from matrix rank, though there are multiple ways to define the rank of a tensor. In this paper, we define it based on CP-decomposition and call it CP-rank.

many times, meaning that problem (4) must be repeated over many iterations. Here we adopt a divide-and-conquer strategy [31] to sequentially solve the following sparse rank-1 estimation problems, which can automatically estimate λ_r and easily tune R.

$$\min_{\mathcal{W}_r} \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{N} \left(\langle \mathcal{W}_r, \mathcal{X}^{(i)} \otimes \mathbf{e}_k \rangle - y_{r,k}^{(i)} \right)^2 + \lambda_r \|\mathcal{W}_r\|_1,$$
$$\text{s.t. CP-rank}(\mathcal{W}_r) \leq 1, \ r = \{1, \cdots, R\}, \quad (5)$$

where r is the sequential number of the rank-1 terms and $\mathbf{y}_r^{(i)} = [y_{r,1}^{(i)}, \cdots, y_{r,k}^{(i)}]^\top$ is the current residue of the target value with

$$\mathbf{y}_r^{(i)} := \begin{cases} \mathbf{y}^{(i)}, & \text{if } r = 1 \\ \mathbf{y}_{r-1}^{(i)} - \sum_{k=1}^{K} \langle \mathcal{X}^{(i)} \otimes \mathbf{e}_k, \mathcal{W}_{r-1} \rangle, & \text{otherwise,} \end{cases}$$

where $\mathcal{W}_{r-1}$ is the estimated rank-1 tensor in the $(r-1)$-th step. The final estimator can be recovered as $\mathcal{W} = \sum_{r=1}^{R} \mathcal{W}_r$. In this manner, we only seek a single rank-1 tensor at a time and then deflate to find further rank-1 tensors from the target residuals. Furthermore, we decompose the tensor component $\mathcal{W}_r$ (See **Definition** 4) and reformulate problem (5) as:

$$\min_{\{\mathbf{w}_r^{(m)}\}_{m=1}^{M+1}} \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{N} \left(\sigma_r \mathcal{X}^{(i)} \otimes \mathbf{e}_k \times_1 \mathbf{w}_r^{(1)} \times_2 \cdots \times_{M+1} \right.$$
$$\left. \mathbf{w}_r^{(M+1)} - y_{r,k}^{(i)} \right)^2 + \sigma_r \lambda_r \prod_{m=1}^{M+1} \|\mathbf{w}_r^{(m)}\|_1,$$
$$\text{s.t. } \sigma_r \geq 0, \|\mathbf{w}_r^{(m)}\|_1 = 1, m \in \{1, \cdots, M+1\}. \quad (6)$$

To solve this problem, we resort to an alternating convex search (ACS) approach [32]. Specifically, a group of variables $(\sigma_r, \mathbf{w}_r^{(m)})$ is alternately optimized with others fixed. Let $\hat{\mathbf{w}}_r^{(m)} = \sigma_r \mathbf{w}_r^{(m)}$ and $\mathbf{c}_{r,m}^{(i)} = \mathcal{X}^{(i)} \otimes \mathbf{e}_k \times_1 \mathbf{w}_r^{(1)} \times_2 \cdots \times_{m-1} \mathbf{w}_r^{(m-1)} \times_{m+1} \cdots \times_{M+1} \mathbf{w}_r^{(M+1)}$ to simplify this sub-problem formulation as

$$\min_{\hat{\mathbf{w}}_r^{(m)}} \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{N} \left(\hat{\mathbf{w}}_r^{(m)\top} \mathbf{c}_{r,m}^{(i)} - y_{r,k}^{(i)} \right)^2 + \lambda_r \|\hat{\mathbf{w}}_r^{(m)}\|_1,$$
$$\text{s.t. } \sigma_r \geq 0, \|\mathbf{w}_r^{(m)}\|_1 = 1, m \in \{1, \cdots, M+1\}. \quad (7)$$

The sub-problem above is a standard LASSO [33], which we solve using a boosting strategy proposed in [34]. Algorithm 1 summarizes the structure of our stagewise procedures.

5) Complexity Analysis: Based on the forward and backward steps in the **SURF** algorithm [34], both are merged into the convex optimization problem as follows:

$$(\hat{m}, \hat{i}_{\hat{m}}) = \arg\max_{(m, i_m)} \pm 2 \sum_{i=1}^{N} \mathbf{1}_{i_m}^\top \mathbf{c}_{r,m}^{(i)} \left(\sum_{k=1}^{K} y_k^{(i)} - \hat{\mathbf{w}}_r^{(m)\top} \mathbf{c}_{r,m}^{(i)} \right)$$
$$- \epsilon \sum_{m=1}^{I_m} \sum_{i=1}^{N} \mathbf{c}_{r,m}^{(i)\top} \mathbf{c}_{r,m}^{(i)},$$
$$\text{s.t. } i_m \in \{1, \cdots, I_m\}, m \in \{1, \cdots, M+1\} \quad (8)$$

Algorithm 1 Solution of TMMTR problem in Eq. (4)

Input: Multi-modal tensor pairs $\{(\mathcal{X}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^{N}$, and a small step size ϵ.
Output: Coefficient tensor $\mathcal{W}$.
 1: Initialize R with a constant, $\mathbf{y}_1^{(i)} = \mathbf{y}^{(i)}, i \in \{1, \cdots, N\}$.
 2: **for** $r = 1, \cdots, R$ **do**
 3:    **for** $m = 1, \cdots, M+1$ **do**
 4:      Initialize $\mathbf{w}_r^{(1)}, \cdots, \mathbf{w}_r^{(m-1)}, \mathbf{w}_r^{(m+1)}, \cdots, \mathbf{w}_r^{(M+1)}$.
 5:      Compute $\mathbf{c}_{r,m}^{(i)} = \mathcal{X}^{(i)} \otimes \mathbf{e}_k \times_1 \mathbf{w}_r^{(1)} \times_2 \cdots \times_{m-1}$ $\mathbf{w}_r^{(m-1)} \times_{m+1} \cdots \times_{M+1} \mathbf{w}_r^{(M+1)}, i = 1, \cdots, N$.
 6:      Run **SURF**(ϵ) in [34] to solve problem (7).
 7:    **end for**
 8:    $\mathcal{W}_r = \hat{\mathbf{w}}_r^{(1)} \otimes \cdots \otimes \hat{\mathbf{w}}_r^{(m)} \otimes \cdots \otimes \hat{\mathbf{w}}_r^{(M+1)}$.
 9:    $\mathbf{y}_{r+1}^{(i)} = \mathbf{y}_r^{(i)} - \sum_{k=1}^{K} \langle \mathcal{W}_r, \mathcal{X}^{(i)} \otimes \mathbf{e}_k \rangle$.
10: **end for**
11: $\mathcal{W} = \sum_{r=1}^{R} \mathcal{W}_r$.

---

where $\mathbf{1}_{i_m} = [\underbrace{0, \cdots, 0}_{i_m - 1}, 1, 0, \cdots, 0]^\top \in \mathbb{R}^{I_m}$, which is designed to select the corresponding entry of $\mathbf{c}_{r,m}^{(i)}, i \in \{1, \cdots, N\}$. Problem (8) can be efficiently solved, and the computational complexity is dominated by the second term $\epsilon \sum_{m=1}^{I_m} \sum_{i=1}^{N} \mathbf{c}_{r,m}^{(i)\top} \mathbf{c}_{r,m}^{(i)}$ under $\mathcal{O}$-notation. In the context of tensors, we want a more elaborate computational analysis that considers all tensor modes. According to the direct conclusion in [34] and the adaptive problem (8), the computational complexity of the 6-th line in **Algorithm 1** is $\mathcal{O}\big(N(\sum_{m \neq \hat{m}}^{M+1}(\prod_{s \neq m, \hat{m}}^{M+1} I_s + 5I_m) + 2I_{\hat{m}} + 2K)\big)$, and the 5-th is $\mathcal{O}\big(N \sum_{m \neq \hat{m}}^{M+1}(\prod_{s \neq m, \hat{m}}^{M+1} I_s + 5I_m)\big)$. If we further consider the presence of two for loops, the total computational complexity of **Algorithm 1** is $\mathcal{O}\big(NR(M+1)(\sum_{m \neq \hat{m}}^{M+1}(\prod_{s \neq m, \hat{m}}^{M+1} I_s + 5I_m) + I_{\hat{m}} + K)\big)$. Compared with ACS algorithm that requires $\mathcal{O}\big(NR(M+1) \prod_{m=1}^{M+1} I_m\big)$ [35] computational cost to solve same problem, our proposed learning algorithm is faster.

For space complexity, the indicator vector $\mathbf{e}_k$ increases the dimensionality of the data tensor $\mathcal{X}^{(i)}$ through the tensor product operation ($\otimes$). The large quantity of zeros caused by padding results in heavy space consumption, but we reduce this consumption by utilizing properties of a rank-1 tensor's inner product. In particular, for the two rank-1 tensors $\mathcal{X} = \mathbf{x}^{(1)} \otimes \cdots \otimes \mathbf{x}^{(M)}$ and $\mathcal{Y} = \mathbf{y}^{(1)} \otimes \cdots \otimes \mathbf{y}^{(M)}$, it holds that:

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \prod_{m=1}^{M} \langle \mathbf{x}^{(m)}, \mathbf{y}^{(m)} \rangle. \quad (9)$$

We can then simplify the computation process of $\langle \mathcal{W}_r, \mathcal{X}^{(i)} \otimes \mathbf{e}_k \rangle$ by eliminating the last mode:

$$\langle \mathcal{W}_r, \mathcal{X}^{(i)} \otimes \mathbf{e}_k \rangle$$
$$= \langle \mathbf{w}_r^{(1)} \times_2 \cdots \times_{M+1} \mathbf{w}_r^{(M+1)}, \mathcal{X}^{(i)} \otimes \mathbf{e}_k \rangle$$
$$= (\mathbf{e}_k^\top \mathbf{w}_r^{(M+1)}) \langle \mathcal{X}^{(i)}, \mathbf{w}_r^{(1)} \otimes \cdots \otimes \mathbf{w}_r^{(M)} \rangle$$
$$= w_{r,k}^{(M+1)} \langle \mathcal{X}^{(i)}, \mathbf{w}_r^{(1)} \otimes \cdots \otimes \mathbf{w}_r^{(M)} \rangle, \quad (10)$$

Feature Tensor	Assessment	Metrics	SMART [21]	MT-SGL [23]	rMLTFL [24]	DBN-based MTL [25]	GCN [26]	TMMTR
116×3	ADS	RMSE ↓	0.331 ± 0.018	0.338 ± 0.023	0.335 ± 0.016	0.324 ± 0.014	N/A	$\mathbf{0.307 \pm 0.009}$
		Sparsity↑	0.799 ± 0.013	0.759 ± 0.016	0.678 ± 0.009	N/A	N/A	$\mathbf{0.966 \pm 0.005}$
	ADAS-Cog 13	RMSE↓	0.168 ± 0.023	0.144 ± 0.033	$\mathbf{0.141 \pm 0.029}$	0.146 ± 0.025	N/A	0.145 ± 0.019
		Sparsity↑	0.835 ± 0.012	0.773 ± 0.023	0.713 ± 0.005	N/A	N/A	$\mathbf{0.986 \pm 0.004}$
	MMSE	RMSE↓	0.151 ± 0.017	0.152 ± 0.018	0.151 ± 0.020	0.149 ± 0.016	N/A	$\mathbf{0.142 \pm 0.011}$
		Sparsity↑	0.735 ± 0.016	0.698 ± 0.049	0.641 ± 0.013	N/A	N/A	$\mathbf{0.969 \pm 0.002}$
	Total	RMSE↓	0.403 ± 0.020	0.398 ± 0.023	0.394 ± 0.020	0.386 ± 0.021	N/A	$\mathbf{0.368 \pm 0.010}$
		Sparsity↑	0.790 ± 0.015	0.743 ± 0.030	0.677 ± 0.010	N/A	N/A	$\mathbf{0.976 \pm 0.004}$
116×116	ADS	RMSE↓	0.337 ± 0.015	0.334 ± 0.016	0.329 ± 0.014	0.332 ± 0.019	$\mathbf{0.302 \pm 0.012}$	0.328 ± 0.010
		Sparsity↑	0.963 ± 0.012	0.941 ± 0.013	0.862 ± 0.011	N/A	N/A	$\mathbf{0.998 \pm 0.000}$
	ADAS-Cog 13	RMSE↓	0.156 ± 0.029	0.152 ± 0.032	0.152 ± 0.029	0.155 ± 0.029	0.154 ± 0.013	$\mathbf{0.148 \pm 0.031}$
		Sparsity↑	0.981 ± 0.014	0.969 ± 0.010	0.893 ± 0.021	N/A	N/A	$\mathbf{0.999 \pm 0.000}$
	MMSE	RMSE ↓	0.174 ± 0.030	0.164 ± 0.031	0.161 ± 0.030	0.160 ± 0.024	0.194 ± 0.012	$\mathbf{0.153 \pm 0.016}$
		Sparsity↑	0.931 ± 0.013	0.920 ± 0.012	0.839 ± 0.009	N/A	N/A	$\mathbf{0.997 \pm 0.000}$
	Total	RMSE↓	0.411 ± 0.019	0.402 ± 0.024	0.397 ± 0.022	0.400 ± 0.021	$\mathbf{0.391 \pm 0.019}$	$\mathbf{0.391 \pm 0.021}$
		Sparsity↑	0.958 ± 0.013	0.943 ± 0.012	0.865 ± 0.013	N/A	N/A	$\mathbf{0.998 \pm 0.000}$
$116 \times 116 \times 3$	ADS	RMSE↓	0.338 ± 0.025	0.328 ± 0.026	0.326 ± 0.021	0.334 ± 0.028	0.306 ± 0.011	$\mathbf{0.273 \pm 0.010}$
		Sparsity↑	0.994 ± 0.003	0.986 ± 0.004	0.966 ± 0.009	N/A	N/A	$\mathbf{1.000 \pm 0.000}$
	ADAS-Cog 13	RMSE ↓	0.157 ± 0.031	0.153 ± 0.032	0.158 ± 0.031	0.172 ± 0.035	0.149 ± 0.012	$\mathbf{0.141 \pm 0.013}$
		Sparsity↑	0.997 ± 0.002	0.991 ± 0.003	0.979 ± 0.005	N/A	N/A	$\mathbf{1.000 \pm 0.000}$
	MMSE	RMSE↓	0.172 ± 0.021	0.169 ± 0.026	0.154 ± 0.021	0.185 ± 0.028	0.193 ± 0.010	$\mathbf{0.146 \pm 0.013}$
		Sparsity↑	0.989 ± 0.005	0.965 ± 0.011	0.945 ± 0.012	N/A	N/A	$\mathbf{1.000 \pm 0.000}$
	Total	RMSE↓	0.411 ± 0.031	0.399 ± 0.032	0.394 ± 0.030	0.419 ± 0.034	0.391 ± 0.018	$\mathbf{0.378 \pm 0.010}$
		Sparsity↑	0.993 ± 0.003	0.981 ± 0.006	0.963 ± 0.009	N/A	N/A	$\mathbf{1.000 \pm 0.000}$

where $w_{r,k}^{(M+1)}$ is the k-th entry element of the last decomposed mode vector $\mathbf{w}_r^{(M+1)}$, which can be directly indexed by the arrangement of targets without extra space consumption. By this trick, we cut down the space complexity of the training stage from $\mathcal{O}\left(K \prod_{m=1}^M I_m\right)$ to $\mathcal{O}\left(K + \prod_{m=1}^M I_m\right)$.

III. RESULTS

A. Experimental Settings

Baselines. To show the validity of our proposed method, we compared our method with: (1) three sparse learning methods, namely the Sparse Multi-Task Regression and Feature selection (SMART) [21], Multi-Task Sparse Group Lasso (MT-SGL) [23], and Robust Multi-Label Transfer Feature Learning (rMLTFL) [24] methods, and (2) two deep models, namely Deep Belief Network-based Multi-Task Learning (DBN-based MTL) [25] and a Graph Convolutional Neural Network (GCN) [26].

Implementation Details. For model validation, subjects were split into training and test sets with a ratio of 5:1. The hyperparameters of all methods were optimized using 5-fold cross validation on the training set. For the SMART, MT-SGL, and rMLTFL methods, the regularisation parameters before different norms were selected from a grid search in a much wider range of $\{10^{-6}, 10^{-5}, \cdots, 10^1\}$ in the slow-start phase, and then switched to a narrower and more granular grid search range of $\{0.1, 0.2, \cdots, 0.9\}$ to obtain the best performance. In DBN-based MTL, after the feature extraction stage by PCA method, there are two hidden layers in the DBM module with

18 and 9 neurons respectively, where the dropout rate in the final layer is traversed in the range of $\{0.1, 0.2, \cdots, 0.9\}$. For the GCN, there were two convolutional layers followed by three fully-connected layers. The first two fully-connected layers were each followed by batch normalization and a dropout of 0.8. In our proposed method, the rank R is incremented from 1 to 70 with a step size of 1. All methods used traverse the hyper-parameter K in the KNN graph from 1 to 116 with step size of 1.

B. Experimental Results

To assess the level of disease progression from different perspectives, we take three common clinical scores (ADS, ADAS-Cog 13, and MMSE) into account. In order to fairly measure each of these targets, we first transform ADS into quantitative values (CN-1, SMC-2, EMCI-3, LMCI-4, and AD-5), and then normalize all of assessment scores to the range of $[0, 1]$, with higher scores indicating greater severity of cognitive impairments.

In order to avoid randomness and receive reliable results, we repeat the experiments over five trials with different, consistent training/testing splits for each method and report the mean and standard deviation. For the quantitative performance evaluation, we employed two metrics: (1) Root Mean Squared Prediction Error (RMSE), measuring the deviation between the ground truth response and the predicted values, and (2) sparsity of coefficients, defined as the ratio of zero elements relative to the total number of elements in the model coefficients.
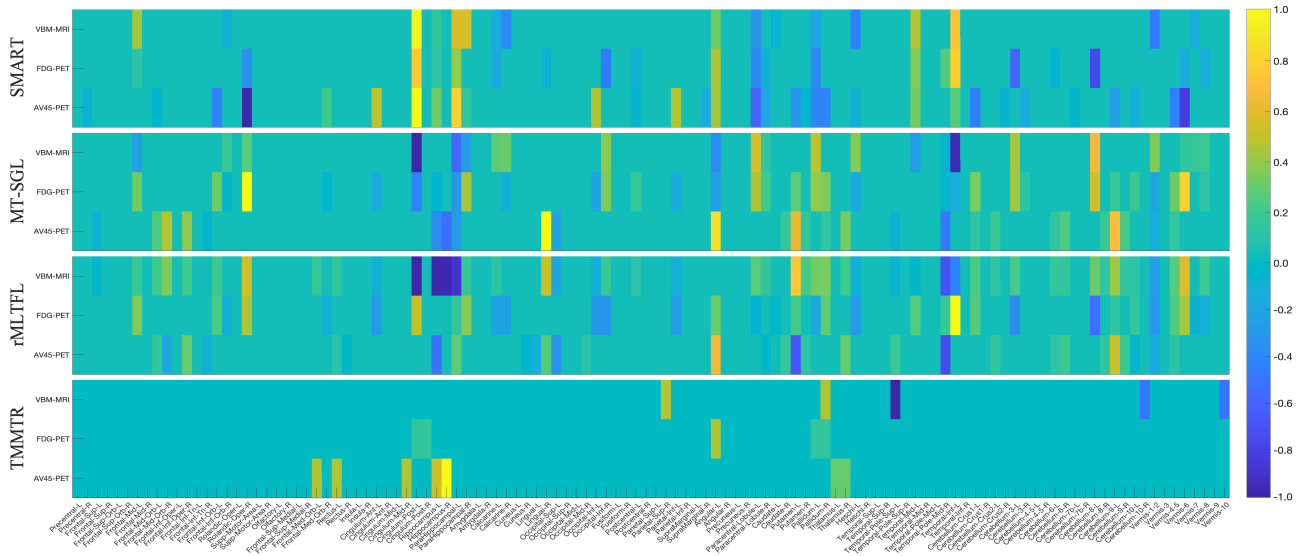
Fig. 2. Comparison of coefficient weights in terms of each imaging modality across five trials. Each row corresponds to a feature selection method: SMART, MT-SGL, rMLTFL, and our proposed TMMTR (from top to bottom). Within each panel, there are three rows corresponding to each of the three imaging modalities (i.e., VBM-MRI, FDG-PET, and AV45-PET) and the 116 brain regions defined in [28].

(a) SMART

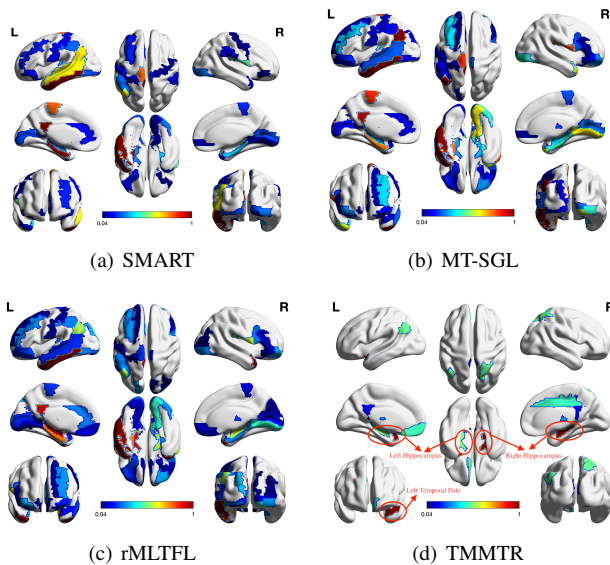(b) MT-SGL

(c) rMLTFL

(d) TMMTR

Fig. 3. The colormaps of 116 ROIs on the physical brains to the corresponding sparse solutions of each feature selection method. Each method shows the full eight-brain views, in which the first row from left to right are lateral view of left hemisphere, topside, lateral view of right hemisphere, the second row from left to right are medial view of left hemisphere, bottom side, medial view of right hemisphere, and the third row are frontal side and backside.

TABLE I shows the performance comparison of six methods. It is clear that the proposed method outperforms almost all of the competing methods in terms of both RMSE and sparsity. In addition:

- Our method's prediction performance improves as the dimensionality of input data grows, while the baseline methods tend to remain stagnant. This confirms that TMMTR can learn tensor-structure information, such as ROI-to-ROI connectivity and cross-modality interaction.

In contrast, all of the baseline methods struggle to handle higher dimensional data since vectorization in the model design leads to the loss of structural information.

- The GCN model consistently outperformed the other baselines and approached the results of TMMTR in terms of total loss, as it is better at capturing connectivity information than the other baselines. However, our proposed method shows elevated performance for each target, which means the inter-target correlation is learned even more strongly.

- TMMTR produces both the best predictive performance and highest sparsity, which validates that our method provides a performance improvement along with a higher interpretability.

C. Interpretability

To further analyse the effectiveness of feature selection in TMMTR compared with other baseline methods, we explore the most discriminative regions of the 116×3 (ROIs$\times$ Modalities) tensor using learned coefficients corresponding to each feature. Fig. 2 visualizes the average coefficient weights of four feature selection methods across five trials. Our method achieves a more sparse solution and pays relatively balanced attention to all three modalities, which demonstrates the learning of tensor structure on the modality-level. The top 10 brain regions selected by our method are: Hippocampus-R, Temporal-Pole-Sup-L, Hippocampus-L, Frontal-Med-Orb-L, Cingulum-Mid-R, Rectus-L, Parietal-Sup-R, Angular-L, Pallidum-R, Thalamus-L. Most of these selected regions are known to be highly related to AD and MCI in previous studies [36]–[40]. From this visualization, it is apparent that the $\ell_{2,1}$ regularizer in SMART prefers the model to select similar features across all of the different modalities. MT-SGL

TABLE II
ABLATION STUDY OF MTR USED IN TMMTR METHOD. RESULTS ARE
SHOWN AS THE MEAN VALUES AND STANDARD DEVIATION (MEAN ± STD)
ACROSS FIVE TRIALS.

Feature Tensor	Assesment	Metrics	TMSTR	TMMTR
116×3	ADS	RMSE↓	0.316 ± 0.016	$\mathbf{0.307 \pm 0.009}$
		Sparsity↑	0.963 ± 0.007	$\mathbf{0.966 \pm 0.005}$
	ADAS-Cog 13	RMSE↓	0.165 ± 0.032	$\mathbf{0.145 \pm 0.019}$
		Sparsity↑	0.983 ± 0.005	$\mathbf{0.986 \pm 0.004}$
	MMSE	RMSE↓	0.216 ± 0.019	$\mathbf{0.142 \pm 0.011}$
		Sparsity↑	0.964 ± 0.003	$\mathbf{0.969 \pm 0.002}$
116×116	ADS	RMSE↓	$\mathbf{0.314 \pm 0.020}$	0.328 ± 0.010
		Sparsity↑	0.997 ± 0.001	$\mathbf{0.998 \pm 0.000}$
	ADAS-Cog 13	RMSE↓	0.160 ± 0.012	$\mathbf{0.148 \pm 0.031}$
		Sparsity↑	$\mathbf{0.999 \pm 0.000}$	$\mathbf{0.999 \pm 0.000}$
	MMSE	RMSE↓	0.194 ± 0.021	$\mathbf{0.153 \pm 0.016}$
		Sparsity↑	0.997 ± 0.001	$\mathbf{0.997 \pm 0.000}$
$116 \times 116 \times 3$	ADS	RMSE↓	0.281 ± 0.011	$\mathbf{0.273 \pm 0.010}$
		Sparsity↑	0.999 ± 0.001	$\mathbf{1.000 \pm 0.000}$
	ADAS-Cog 13	RMSE↓	0.143 ± 0.013	$\mathbf{0.141 \pm 0.013}$
		Sparsity↑	0.999 ± 0.001	$\mathbf{1.000 \pm 0.000}$
	MMSE	RMSE↓	0.184 ± 0.015	$\mathbf{0.146 \pm 0.011}$
		Sparsity↑	0.999 ± 0.000	$\mathbf{1.000 \pm 0.000}$

and rMLTFL overcome this disadvantage but lose sparsity in the process.

As shown in Fig. 3, we further use BrainNet Viewer [41] to visualize the brain structure and highlight the regions that the four sparse methods (SMART, MT-SGL, rMLTFL, and TMMTR) rely on to make their predictions. It can be seen that our TMMTR is more sparse and uses more relevant ROIs (as marked in Fig. 3(d)).

D. Ablation Study

We conducted an ablation study to validate the effectiveness of the MTR used in our method. TABLE II reports the comparison of performance between TMMTR and its variation version—Tensor-based Multi-modal Single-Target Regression (TMSTR). It was found that the joint modeling process of multiple targets can effectively learn inter-correlation between them and thus enhance the performance for each target. We also observed that sparsity is consistently improved without decreasing learning performance, which means that a more sparse structure is worthy of exploration as feature redundancy is common in different neuroimaging modalities.

E. Hyperparameter Analysis

We also investigated the influence of two important hyperparameters in our TMMTR method: the number of neighbors K in the connectivity tensor construction stage and the rank R to constrain the low-rank property of our model. From Fig. 4, we found that the higher rank R consistently achieved better performance with a gradual decrease in sparsity, which is intuitive as a more sparse structure is of lower rank. Using $R = 60$ can uphold a high sparsity and performance, though learning performance tends to stay stable as R increases. For the selection of K value, the best results are consistently achieved with the fully-connected graph (i.e., $K = 116$), which indicates that our method can make full use of high-order relations among ROIs and different modalities.
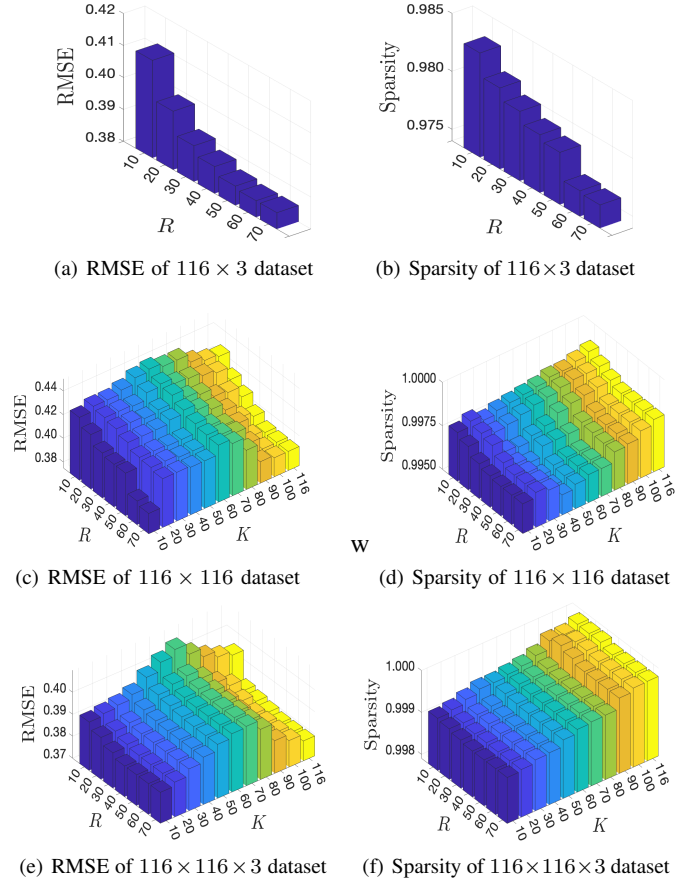

(a) RMSE of 116×3 dataset

(b) Sparsity of 116×3 dataset

(c) RMSE of 116×116 dataset

(d) Sparsity of 116×116 dataset

(e) RMSE of $116 \times 116 \times 3$ dataset

(f) Sparsity of $116 \times 116 \times 3$ dataset

Fig. 4. Influence of two hyperparameters (i.e., R, K) on the performance of TMMTR over tensors with different dimensionalities. (a) and (b) The influence of CP-rank R over tensor data $\mathcal{X} \in \mathbb{R}^{116 \times 3}$; (c) and (d) the influence of the K value in KNN graph and tensor CP-rank R over tensor data $\mathcal{X} \in \mathbb{R}^{116 \times 116}$; (e) and (f) the influence of the K value in the KNN graph and tensor CP-rank R over tensor data $\mathcal{X} \in \mathbb{R}^{116 \times 116 \times 3}$.

IV. CONCLUSION

In our proposed algorithm, TMMTR, the cross-modal connections, inter-target correlation, and input-target relationship are utilized simultaneously to successfully identify biomarkers related to AD with a high tensor-structured sparsity and interpretability. We investigated three different clinical scores using three different tensor representations for each. Extensive experiments validated a higher predictive performance than other state-of-the-art methods. Our approach is of wide general interest as it can be generalized to other diseases when high-dimensional data is available. In the future work, we hope to extend it to jointly model multiple tasks into a general tensor-based framework.

V. ACKNOWLEDGEMENTS

REFERENCES

[1] Md Sahab Uddin and Ghulam Md Ashraf. Introductory chapter: Alzheimer's disease—the most common cause of dementia. In *Advances in Dementia Research*. IntechOpen, 2018.

[2] Michel Goedert and Maria Grazia Spillantini. A century of alzheimer's disease. *science*, 314(5800):777–781, 2006.

[3] Jorge Samper-González, Ninon Burgos, Simona Bottani, Sabrina Fontanella, Pascal Lu, Arnaud Marcoux, Alexandre Routier, Jérémy Guillon, Michael Bacci, Junhao Wen, et al. Reproducible evaluation of classification methods in alzheimer's disease: Framework and application to mri and pet data. *NeuroImage*, 183:504–521, 2018.

[4] Muhammad Tanveer, Bharat Richhariya, Riyaj Uddin Khan, Ashraf Haroon Rashid, Pritee Khanna, Mukesh Prasad, and CT Lin. Machine learning techniques for the diagnosis of alzheimer's disease: A review. *ACM TOMM*, 16(1s):1–35, 2020.

[5] Carlo Fabrizio, Andrea Termine, Carlo Caltagirone, and Giulia Sancesario. Artificial intelligence for alzheimer's disease: Promise or challenge? *Diagnostics*, 11(8):1473, 2021.

[6] Simon Duchesne, Anna Caroli, Cristina Geroldi, Giovanni B Frisoni, and D Louis Collins. Predicting clinical variable from mri features: application to mmse in mci. In *MICCAI*, pages 392–399. Springer, 2005.

[7] Yong Fan, Daniel Kaufer, and Dinggang Shen. Joint estimation of multiple clinical variables of neurological diseases from imaging patterns. In *IEEE ISBI: From Nano to Macro*, pages 852–855, 2010.

[8] Nguyen Thanh Duc, Seungjun Ryu, Muhammad Naveed Iqbal Qureshi, Min Choi, Kun Ho Lee, and Boreom Lee. 3d-deep learning based automatic diagnosis of alzheimer's disease with joint mmse prediction using resting-state fmri. *Neuroinformatics*, 18(1):71–86, 2020.

[9] Anil Rao, Ying Lee, Achim Gass, and Andreas Monsch. Classification of alzheimer's disease from structural mri using sparse logistic regression with optional spatial regularization. In *IEEE EMBS*, pages 4499–4502, 2011.

[10] Bo Cheng, Mingxia Liu, Heung-Il Suk, Dinggang Shen, and Daoqiang Zhang. Multimodal manifold-regularized transfer learning for mci conversion prediction. *Brain imaging and behavior*, 9(4):913–926, 2015.

[11] Biao Jie, Daoqiang Zhang, Bo Cheng, Dinggang Shen, and Alzheimer's Disease Neuroimaging Initiative. Manifold regularized multitask feature learning for multimodality disease classification. *Human brain mapping*, 36(2):489–507, 2015.

[12] Xiaoke Hao, Xiaohui Yao, Jingwen Yan, Shannon L Risacher, Andrew J Saykin, Daoqiang Zhang, and Li Shen. Identifying multimodal intermediate phenotypes between genetic risk factors and disease status in alzheimer's disease. *Neuroinformatics*, 14(4):439–452, 2016.

[13] Mingxia Liu, Jun Zhang, Pew-Thian Yap, and Dinggang Shen. View-aligned hypergraph learning for alzheimer's disease diagnosis with incomplete multi-modality data. *Medical image analysis*, 36:123–134, 2017.

[14] Kichang Kwak, Hyuk Jin Yun, Gilsoon Park, Jong-Min Lee, Alzheimer's Disease Neuroimaging Initiative, et al. Multi-modality sparse representation for alzheimer's disease classification. *Journal of Alzheimer's disease*, 65(3):807–817, 2018.

[15] Haifeng Chen, Weikai Li, Xiaoning Sheng, Qing Ye, Hui Zhao, Yun Xu, and Feng Bai. Machine learning based on the multimodal connectome can predict the preclinical stage of alzheimer's disease: a preliminary study. *European Radiology*, 32(1):448–459, 2022.

[16] M López, J Ramírez, JM Górriz, D Salas-Gonzalez, I Alvarez, F Segovia, and CG Puntonet. Automatic tool for alzheimer's disease diagnosis using pca and bayesian classification rules. *Electronics letters*, 45(8):389–391, 2009.

[17] MM López, J Ramírez, JM Górriz, I Álvarez, D Salas-Gonzalez, F Segovia, and R Chaves. Svm-based cad system for early detection of the alzheimer's disease using kernel pca and lda. *Neuroscience Letters*, 464(3):233–238, 2009.

[18] Tao Zhou, Mingxia Liu, Kim-Han Thung, and Dinggang Shen. Latent representation learning for alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data. *IEEE T-MI*, 38(10):2411–2422, 2019.

[19] Xiaofeng Zhu, Heung-Il Suk, and Dinggang Shen. Multi-modality canonical feature selection for alzheimer's disease diagnosis. In *MICCAI*, pages 162–169. Springer, 2014.

[20] Yanbei Liu, Lianxi Fan, Changqing Zhang, Tao Zhou, Zhitao Xiao, Lei Geng, and Dinggang Shen. Incomplete multi-modal representation learning for alzheimer's disease diagnosis. *Medical Image Analysis*, 69:101953, 2021.

[21] Hua Wang, Feiping Nie, Heng Huang, Shannon Risacher, Chris Ding, Andrew J Saykin, and Li Shen. Sparse multi-task regression and feature selection to identify brain imaging predictors for memory performance. In *IEEE ICCV*, pages 557–562, 2011.

[22] Xiaoqian Wang, Xiantong Zhen, Quanzheng Li, Dinggang Shen, and Heng Huang. Cognitive assessment prediction in alzheimer's disease by multi-layer multi-target regression. *Neuroinformatics*, 16(3):285–294, 2018.

[23] Xiaoli Liu, André R Goncalves, Peng Cao, Dazhe Zhao, Arindam Banerjee, Alzheimer's Disease Neuroimaging Initiative, et al. Modeling alzheimer's disease cognitive scores using multi-task sparse group lasso. *Computerized Medical Imaging and Graphics*, 66:100–114, 2018.

[24] Bo Cheng, Mingxia Liu, Daoqiang Zhang, and Dinggang Shen. Robust multi-label transfer feature learning for early diagnosis of alzheimer's disease. *Brain imaging and behavior*, 13(1):138–153, 2019.

[25] Nianyin Zeng, Han Li, and Yonghong Peng. A new deep belief network-based multi-task learning for diagnosis of alzheimer's disease. *Neural Computing and Applications*, pages 1–12, 2021.

[26] Tzu-An Song, Samadrita Roy Chowdhury, Fan Yang, Heidi Jacobs, Georges El Fakhri, Quanzheng Li, Keith Johnson, and Joyita Dutta. Graph convolutional neural networks for alzheimer's disease classification. In *IEEE 16th ISBI*, pages 414–417, 2019.

[27] John Ashburner and Karl J Friston. Voxel-based morphometry—the methods. *Neuroimage*, 11(6):805–821, 2000.

[28] Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289, 2002.

[29] Ulrike Von Luxburg. A tutorial on spectral clustering. *Stat. Comput.*, 17(4):395–416, 2007.

[30] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.

[31] Anh-Huy Phan, Petr Tichavský, and Andrzej Cichocki. Tensor deflation for candecomp/parafac—part i: Alternating subspace update algorithm. *IEEE TSP*, 63(22):5924–5938, 2015.

[32] Kun Chen, Kung-Sik Chan, and Nils Chr Stenseth. Reduced rank stochastic regression with a sparse singular value decomposition. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(2):203–221, 2012.

[33] Robert Tibshirani. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Series B Stat. Methodol.*, 58(1):267–288, 1996.

[34] Lifang He, Kun Chen, Wanwan Xu, Jiayu Zhou, and Fei Wang. Boosted sparse and low-rank tensor regression. *NeurIPS*, 31, 2018.

[35] Alex Pereira da Silva, Pierre Comon, and Andre Lima Ferrer de Almeida. Rank-1 tensor approximation methods and application to deflation. *arXiv preprint arXiv:1508.05273*, 2015.

[36] AT Du, Nea Schuff, D Amend, MP Laakso, YY Hsu, WJ Jagust, K Yaffe, JH Kramer, B Reed, D Norman, et al. Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 71(4):441–447, 2001.

[37] Judith Miklossy. Alzheimer's disease-a neurospirochetosis. analysis of the evidence following koch's and hill's criteria. *Journal of neuroinflammation*, 8(1):1–16, 2011.

[38] Emma J Bubb, Claudia Metzler-Baddeley, and John P Aggleton. The cingulum bundle: anatomy, function, and dysfunction. *Neuroscience & Biobehavioral Reviews*, 92:104–127, 2018.

[39] Paul J Mattis, Martin Niethammer, Wataru Sako, Chris C Tang, Amir Nazem, Marc L Gordon, Vicky Brandt, Vijay Dhawan, and David Eidelberg. Distinct brain networks underlie cognitive dysfunction in parkinson and alzheimer diseases. *Neurology*, 87(18):1925–1933, 2016.

[40] Andrei G Vlassenko, Tammie LS Benzinger, and John C Morris. Pet amyloid-beta imaging in preclinical alzheimer's disease. *Biochimica et Biophysica Acta*, 1822(3):370–379, 2012.

[41] Mingrui Xia, Jinhui Wang, and Yong He. Brainnet viewer: a network visualization tool for human brain connectomics. *PloS one*, 8(7):e68910, 2013.