

Tensor Completion for Alzheimer's Disease Prediction From Diffusion Tensor Imaging

Gou, Yixin ; Liu, Yipeng ; He, Fei ; Hunyadi, Borbála; Zhu, Ce

DOI

[10.1109/TBME.2024.3365131](https://doi.org/10.1109/TBME.2024.3365131)

Publication date

2024

Document Version

Final published version

Published in

IEEE Transactions on Biomedical Engineering

Citation (APA)

Gou, Y., Liu, Y., He, F., Hunyadi, B., & Zhu, C. (2024). Tensor Completion for Alzheimer's Disease Prediction From Diffusion Tensor Imaging. *IEEE Transactions on Biomedical Engineering*, 71(7), 2211-2223. <https://doi.org/10.1109/TBME.2024.3365131>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

<https://www.openaccess.nl/en/you-share-we-take-care>

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

Tensor Completion for Alzheimer's Disease Prediction From Diffusion Tensor Imaging

Yixin Gou , Yipeng Liu , Senior Member, IEEE, Fei He , Graduate Student Member, IEEE, Borbála Hunyadi , and Ce Zhu , Fellow, IEEE

Abstract—Objective: Alzheimer's disease (AD) is a slowly progressive neurodegenerative disorder with insidious onset. Accurate prediction of the disease progression has received increasing attention. Cognitive scores that reflect patients' cognitive status have become important criteria for predicting AD. Most existing methods consider the relationship between neuroimages and cognitive scores to improve prediction results. However, the inherent structure information in interrelated cognitive scores is rarely considered. **Method:** In this article, we propose a relation-aware tensor completion multitask learning method (RATC-MTL), in which the cognitive scores are represented as a third-order tensor to preserve the global structure information in clinical scores. We combine both tensor completion and linear regression into a unified framework, which allows us to capture both inter and intra modes correlations in cognitive tensor with a low-rank constraint, as well as incorporate the relationship between biological features and cognitive status by imposing a regression model on multiple cognitive scores. **Result:** Compared to the single-task and state-of-the-art multi-task algorithms, our proposed method obtains the best results for predicting cognitive scores in terms of four commonly used metrics. Furthermore, the overall performance of our method in classifying AD progress is also the best. **Conclusion:** Our results demonstrate the effectiveness of the proposed framework in fully exploring the global structure information in cognitive scores. **Significance:** This study introduces a novel concept of leveraging tensor completion to assist in disease diagnoses, potentially offering a solution to the issue of data scarcity encountered in prolonged monitoring scenarios.

Manuscript received 9 August 2023; revised 20 November 2023; accepted 6 February 2024. Date of publication 13 February 2024; date of current version 20 June 2024. This work was supported in part by the National Key R&D Program of China under Grant 2023YFC3806003, in part by Medico-Engineering Cooperation Funds from University of Electronic Science and Technology of China under Grant ZYGX2021YGLH215, and in part by the National Natural Science Foundation of China under Grant 62171088 and Grant 62020106011. (Corresponding author: Yipeng Liu.)

Yixin Gou, Fei He, and Ce Zhu are with the School of Communication and Information Engineering, University of Electronic Science and Technology of China, China.

Yipeng Liu is with the School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: yipengliu@uestc.edu.cn).

Borbála Hunyadi is with the Department of Microelectronics, TU Delft, Mekelweg 4, Delft CD 2628, The Netherlands.

This article has supplementary downloadable material available at <https://doi.org/10.1109/TBME.2024.3365131>, provided by the authors.

Digital Object Identifier 10.1109/TBME.2024.3365131

Index Terms—Alzheimer's disease, multitask learning, regression model, tensor completion.

I. INTRODUCTION

ALZHEIMER'S disease (AD) is a progressive neurological disorder, which mainly affects cognitive function and memory capacity of the brain [1]. The symptoms of AD typically start with mild memory loss and difficulties with daily activities, as the disease progresses, it can lead to language impairment, emotional lability, many behavioral problems [2], and eventually death. As the condition worsens, the patient is often progressively disconnected from family or society [3]. Unfortunately, the etiology of AD remains unclear, and effective cures are yet to be discovered. These circumstances impose significant monetary and psychological strains on both patients and their families [4]. Consequently, understanding the progression of AD has become a research priority [5].

AD can only be accurately diagnosed by brain biopsy or autopsy [6]. Fortunately, several studies have identified a strong association between patterns of brain atrophy and AD progression [7]. Therefore, cognitive assessment tests can be used to evaluate the cognitive status, which in turn can partially indicate the disease state [8], [9]. Such cognitive assessment tests include the mini-mental state examination (MMSE) [10], the Alzheimer's disease assessment scale cognitive score (ADAS) [11], and the Rey auditory verbal learning test (RAVLT) [12]. Since disease intervention is more effective at an early stage, there is an urgent need to refine the prediction of AD progression based on clinical scores and to identify the most predictive biomarkers of the progression. As neuroimages can provide more sensitive early biomarkers than other technologies, many studies use imaging data to predict cognitive scores, including magnetic resonance imaging (MRI) [13], [14], [15] and diffusion tensor imaging (DTI) [16], [17], [18]. In order to study the progression of the disease, there is a need to track pathophysiological changes using the cognitive scores as explained above. However, conducting long-term tests on a large group of individuals can be a resource-intensive and time-consuming endeavor. Moreover, the inherent difficulty of regularly following patients - i.e. patients missing follow-up visits or dropping out from the study - can lead to irregular data collection or even missing records.

Due to these challenges, biological markers based multitask learning (MTL) regression models have been proposed to predict

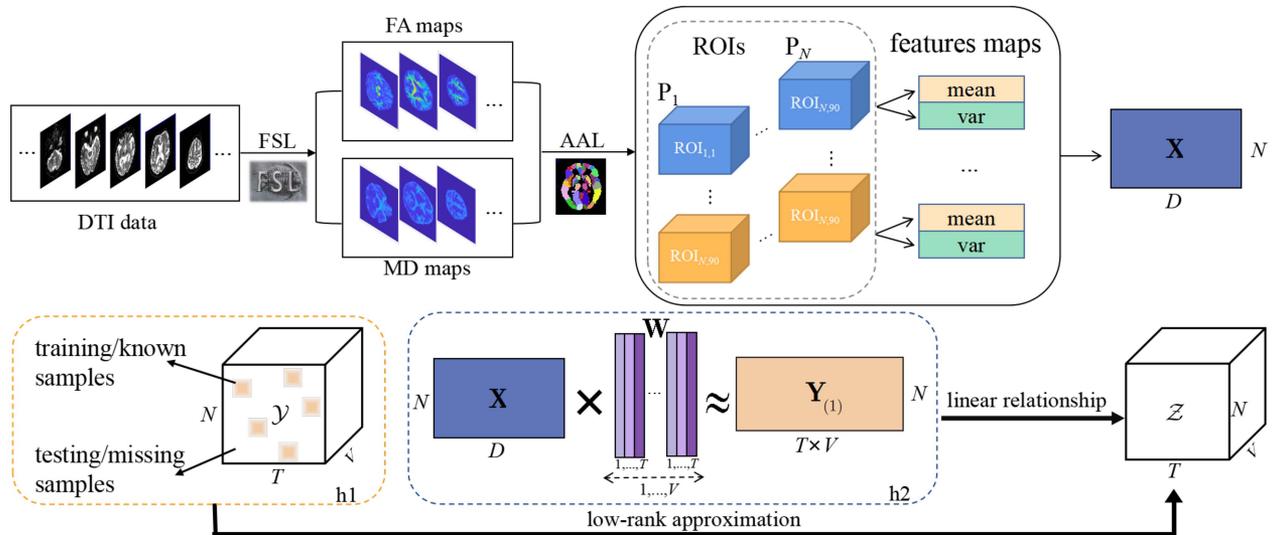


Fig. 1. Overall framework of the proposed method. The first line shows the process of DTI data processing to get the input feature \mathbf{X} . The disease progression can be predicted by assuming the tensor representation of cognitive scores is low-rank (hypothesis 1), and \mathcal{Y} is the observed data. Furthermore, the relationship between image data and cognitive status may provide additional information to improve the performance (hypothesis 2), which is constrained by the minimization of $\|\mathbf{X}\mathbf{W} - \mathbf{Y}_{(1)}\|_F$. By combining these two modules, cognitive scores are predicted and preserved in \mathcal{Z} .

cognitive scores from neuroimaging data. Two families of MTL problems have been widely considered: cross-sectional analysis and longitudinal analysis [19]. In cross-sectional analysis, cognitive scores are assumed to be related to each other, so the task models are constrained to predict multiple cognitive scores at a given time point [20], [21], [22]. In contrast, longitudinal analysis pays more attention to temporal changes in AD progression, which assumes that the clinical scores are correlated across multiple time steps. Thus, multi-task learning in this case means simultaneously learning the same score at different time points [23], [24], [25], [26], in which, the correlation between different cognitive scores is ignored.

Although MTL regression has shown good performance in AD prediction, existing studies focus on *inter-correlations*, i.e. correlations within one aspect, and neglect the *intra-correlations*, i.e. correlations across these two modes. Therefore the tasks are simply represented as a single index, resulting in the loss of additional inherent structure in cognitive scores, which is similar to the dataset mentioned in [27]. Furthermore, although multiple clinical scores reflect different aspects of cognitive status, the underlying pathology is the same. Therefore, we hypothesize that it is beneficial to incorporate this correlation in our works. Besides, for the problem of missing items, the classic solution either deletes the entire samples or replaces them with the mean value, which significantly reduces the number of training samples and introduces the pseudo-labels.

In this paper, we propose a relation-aware tensor completion multitask learning method, denoted as RATC-MTL. To preserve the global structure information of the clinical scores, we organize the scores of an individual patient in a matrix with two indices: one running index for the different types of cognitive scores, and one index for measurement time points. Then, these patient matrices are stacked behind each other to form a third-order tensor, possibly with missing entries when

a score is not available. Then, the cognitive score prediction can be formulated as a tensor completion problem. As shown in Fig. 1, the framework of RATC-MTL consists of two parts, i.e., low-rank completion on the cognitive score tensor and multitask regression module defined on unfolding cognitive scores matrix. This combination allows us to effectively explore the global consistency underlying the cognitive scores and the relationship between biomarkers and cognitive status.

In addition, among traditional methods, the use of MRI-based biomarkers is more common, while DTI-based studies so far have focused on predicting categorical variables in classification tasks [28], [29]. However, DTI is a powerful tool for early diagnosis of AD and differential diagnosis of other dementia [30]. Therefore, the relationship between DTI biomarkers and cognitive scores is estimated in this paper.

In comparison to existing studies, our work makes the following contributions:

- 1) We propose a novel multi-task learning method that uses a tensor completion algorithm to model the disease progression. By this, we can not only leverage the longitudinal and the horizontal correlations present in individual indices but also explore the global association among different indices (i.e. tensor modes).
- 2) We integrate regression and completion into a unified framework. The former incorporates the close relationship between neuroimages and cognitive status, while the latter explores the global correlation of multiple aspects in cognitive scores at the same time.

The remainder of this paper is organized as follows. Section II gives the notations and preliminaries. Section III presents a review of related works on predicting cognitive scores. The proposed approach is outlined in Section IV. Details regarding the dataset, experimental results, and their discussion are presented in Section V. Finally, Section VI concludes the paper.

II. NOTATION AND PRELIMINARIES

A. Notations

For brevity, we represent scalars by lowercase letters, e.g., x , vectors by bold lowercase letters, e.g., \mathbf{x} , matrices by bold uppercase letters, e.g., \mathbf{X} , and high-order tensor by calligraphic uppercase letters, e.g., \mathcal{X} . For matrix $\mathbf{X} \in \mathbb{R}^{I \times J}$, its j -th column is represented as \mathbf{x}_j and its i, j -th element can be written as x_{ij} . A K -th-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_K}$, its element is represented by $\mathcal{X}(i_1, \dots, i_K)$, where its index typically ranges from 1 to their capitalized version, e.g., $i_k = 1, \dots, I_k$. The inner product of two tensors $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{I_1 \times \dots \times I_K}$ is defined by $\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i_1, \dots, i_K} \mathcal{X}(i_1, \dots, i_K) \mathcal{Y}(i_1, \dots, i_K)$, and the squared Frobenius norm of tensor \mathcal{X} is $\|\mathcal{X}\|_{\text{F}}^2 = \langle \mathcal{X}, \mathcal{X} \rangle$.

B. Preliminaries on Tensor Operations

Definition 1. (Mode- k unfolding): For $k \in [K]$, the mode- k unfolding matrix of a K -th-order tensor \mathcal{X} is defined by $\mathbf{X}_{(k)} \in \mathbb{R}^{I_k \times J_k}$, where $J_k = \prod_{i=1, i \neq k}^K I_i$, with its entries are $\mathbf{X}_{(k)}(i_k, \overline{i_1 \dots i_{k-1} i_{k+1} \dots i_K}) = \mathcal{X}(i_1, \dots, i_N)$, $\overline{i_1 \dots i_K} = i_1 + (i_2 - 1)I_1 + (i_3 - 1)I_1 I_2 + \dots + (i_K - 1)I_1 I_2 \dots I_{K-1}$. And its opposite operation ‘Fold $_k$ ’ is defined as $\mathcal{X} = \text{Fold}_k(\mathbf{X}_{(k)})$.

Definition 2. (Tensor trace norm) [31]: For a K -th-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_K}$, its trace norm is defined as:

$$\|\mathcal{X}\|_* = \sum_{k=1}^K \alpha_k \|\mathbf{X}_{(k)}\|_*, \quad (1)$$

where $\alpha_k, k = 1, \dots, K$ are scales that $\alpha_k > 0$ and $\sum_{k=1}^K \alpha_k = 1$, the nuclear norm (trace norm) of matrix $\mathbf{X}_{(k)} \in \mathbb{R}^{I_k \times J_k}$, $J_k = \prod_{i=1, i \neq k}^K I_i$ is denoted as $\|\mathbf{X}_{(k)}\|_* = \sum_{r_k=1}^{R_k} \sigma_{r_k}(\mathbf{X}_{(k)})$, where R_k is the rank of $\mathbf{X}_{(k)}$, $\sigma_{r_k}(\mathbf{X}_{(k)})$ is the r_k -th singular value of matrix $\mathbf{X}_{(k)}$.

Definition 3. (Singular value threshold operator) [32]: Let $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ represent the singular value decomposition (SVD) for matrix \mathbf{X} , the ‘‘threshold’’ operator $\text{SVT}_{\tau}(\mathbf{X})$ is defined as

$$\text{SVT}_{\tau}(\mathbf{X}) = \mathbf{U}\mathbf{\Sigma}_{\tau}\mathbf{V}^T, \quad (2)$$

where $\mathbf{\Sigma}_{\tau} = \text{diag}(\max(\sigma_k - \tau), 0)$.

III. RELATED WORKS

Regarding the prediction of cognitive scores and identification of relevant biomarkers, multitask learning as an effective approach is widely used in this field, which assumes there is an inherent correlation between different data records obtained from the same individuals and is also the focus of this paper. Let $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T \in \mathbb{R}^{N \times D}$ represent the input biological features, which is consisted of N subjects with d dimensional features. $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]^T \in \mathbb{R}^{N \times M}$ is the corresponding cognitive scores, and M denotes the number of tasks. The regularized multitask regression model between \mathbf{X} and \mathbf{Y} can be written as:

$$\min_{\mathbf{W}} \mathcal{L}(\mathbf{Y}, \mathbf{X}, \mathbf{W}) + \lambda \mathfrak{R}(\mathbf{W}), \quad (3)$$

where $\mathbf{W} \in \mathbb{R}^{D \times M}$ is the weight matrix that indicates the implication between input features and output values, $\mathcal{L}(\cdot, \cdot)$ is the loss function, and $\mathfrak{R}(\cdot)$ is the regularization term that use to prevent overfitting and add prior knowledge on model.

In regularized multitask learning, one of the key issues is to build learning models that can capture prior task correlation knowledge. Two kinds of commonly used analysis incorporate different correlations. One considers the variation in cognitive scores across different time points, which is known as longitudinal analysis. Specifically, the task number of formulation (3) is $M = T$ and each \mathbf{w}_j corresponds to a time point. Some works assumed a small subset of biomarkers is predictive of the disease progression, and multiple regression models from different time points satisfy the smoothness property, thus the temporal group lasso regularization was used. Additionally, considering the problem of missing target values, the unified optimization model can be extended as:

$$\min_{\mathbf{W}} \|\mathbf{P} \odot (\mathbf{Y} - \mathbf{X}\mathbf{W})\|_{\text{F}}^2 + \lambda_1 \|\mathbf{W}\|_{\text{Q}}^2 + \lambda_2 \|\mathbf{W}\mathbf{H}\|_{\text{Q}}^2 + \lambda_3 \|\mathbf{W}\|_{2,1}, \quad (4)$$

where \mathbf{P} is used to indicate missing target values, $\mathbf{p}_{ij} = 0$ if the target value of sample i is missing at the j -th task, and $\mathbf{p}_{ij} = 1$ otherwise, \odot represents the elementwise operator that $\mathbf{A} \odot \mathbf{B} = \mathbf{a}_{ij} \mathbf{b}_{ij}$, $\|\mathbf{W}\|_{\text{Q}} = \sqrt[Q]{\sum_{d,m} |\mathbf{w}_{dm}|^Q}$, $\|\mathbf{W}\|_{2,1} = \sum_{d=1}^D \sqrt{\sum_{m=1}^M \mathbf{w}_{dm}^2}$, $\lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_3 \geq 0$ are the regularization parameters that control the generalization error, temporal smoothness and feature selection respectively. In [23], $\text{Q} = 2$ and $\|\mathbf{W}\mathbf{H}\|_{\text{F}}^2 = \sum_{t=1}^T \|\mathbf{w}_t - \mathbf{w}_{t+1}\|_2^2$ is a temporal smoothness term, where $\mathbf{H} \in \mathbb{R}^{T \times T-1}$ is defined as $\mathbf{h}_{ij} = 1$ if $i = j$, $\mathbf{h}_{ij} = -1$ if $i = j + 1$, and $\mathbf{h}_{ij} = 0$ for others. To better capture the correlation of the biomarkers, Zhou et al. [24] utilized both ℓ_1 -norm and $\ell_{2,1}$ -norm to achieve common feature selection for multiple tasks and specific feature selection for each task at the same time, which is known as the convex fused sparse group lasso (cFSGL) and the optimization problem of cFSGL is the (4) with $\text{Q} = 1$. Moreover, other types of regularization methods, such as group $\ell_{2,1}$ -norm and subspace structure penalty are used in [26] to explore the correlation of tasks and the task relatedness with shared subspace.

Since low rank is another approach applied in MTL, which assumes the relatedness of multiple tasks may lead to the correlation in corresponding task parameters, thus implying the low rankness of model parameter [33], has been studied in AD prediction. Chen et al. [34] proposed a robust multitask learning algorithm that uses trace norm to capture the task relation and simultaneously identifies the irrelevant tasks through group sparse structure, while the trace norm is a natural choice when estimating data with low-rank structure [35]. The optimization model can be formulated as:

$$\min_{\mathbf{L}, \mathbf{S}} \|\mathbf{Y} - \mathbf{X}(\mathbf{L} + \mathbf{S})\|_{\text{F}}^2 + \lambda_1 \|\mathbf{L}\|_* + \lambda_2 \|\mathbf{S}\|_{1,2} \quad (5)$$

where $\mathbf{W} = \mathbf{L} + \mathbf{S}$, \mathbf{L} and \mathbf{S} are the low-rank structure and the group-sparse structure, respectively. Moreover, H. Wang et al. also imposed the trace norm on model parameters to explore the task correlations at different time points [36]. Nie et al. [37]

adopted a non-convex low-rank regularization to explore the shared information. To explore the correlation information between biomarkers and cognitive scores, Jiang et al. [38] proposed a correlation-aware sparse and low-rank constrained model, whose regularizer is comprised of $\ell_{2,1}$ -norm, ℓ_1 -norm and trace norm.

In contrast, the cross-sectional analysis is another MTL framework that takes the interrelation between different scores into account, the task number of formulation (3) is $M = V$ and each \mathbf{w}_j corresponds to a cognitive score. The formulation of cross-sectional analysis is akin to that of longitudinal analysis, with the exception of focusing on different task-relatedness, such as [21] also utilize $\ell_{2,1}$ -norm and ℓ_1 -norm to select features. In [22], an SVM-based method was proposed to jointly predict different clinical scores and $\ell_{2,1}$ -norm used to constrain that all tasks share a common set of features. And [20] uses the Bayesian regression method to model multiple medical variables and capture their inherent correlations.

Although the above-mentioned studies have performed outstanding performance, they only consider the relationship between neuroimages and cognitive status, ignoring the high-order interrelated structure within cognitive scores, and may have limitations in getting the optimal results. To solve this, we propose a novel formulation that simultaneously incorporates biomarkers and high-order correlation in cognitive scores.

IV. METHODS

To the best of our knowledge, when it comes to the prediction of Alzheimer's disease from neuroimages, most studies rely on a regression framework, and tensor completion has not been applied. Even though a plethora of applications have leveraged its effectiveness, such as image recovery [39], traffic data analysis [40], and multitask learning [41]. Although there exist several differences between cognitive scores, they both reveal a certain aspect of the disease, which means that multiple clinical scores are interrelated. Most existing works do not make full use of this prior knowledge to exploit low-rank structure on cognitive scores, yet this is a natural underlying assumption after theoretical analysis. Inspired by these, our model leverages both biomarkers and global correlation by utilizing matrix and tensor representation simultaneously.

A. Problem Formulation

For an MTL problem with M tasks, each task is associated with two (or more) modes, and the data can be organized as a tensor $\mathcal{Y} \in \mathbb{R}^{N \times I_1 \times \dots \times I_{K-1}}$, where $M = \prod_{k=1}^{K-1} I_k$. In the context of AD, $\mathcal{Y} \in \mathbb{R}^{N \times T \times V}$ denotes that there are N subjects, and each subject has V different cognitive scores at all T time points, thus the total task in this condition is $M = VT$. To jointly predict different cognitive scores of new subjects at multiple time points, the traditional optimization model can only be applied to the unfolding matrix of cognitive tensor \mathcal{Y} , which means $\|\mathbf{X}\mathbf{W} - \mathbf{Y}_{(1)}\|_F$ is constructed for optimization. However, this formulation may suffer from the limitation that unfolding a tensor into matrix form will lose the high-order structural characteristics, affecting its exploration of global consistency.

Considering the correlation among different clinical scores exists in both inter-mode and intra-modes, which means that \mathcal{Y} should be of low-rank, we take this prior knowledge into account. Furthermore, during data collection, the cognitive scores are missing at some time points, thus \mathcal{Y} may be incomplete. A commonly used strategy is either removing all patients with missing items or filling the missing items with mean value, which significantly affects the model performance. We consider extending the model with missing target values. To solve the limitation, low-rank tensor completion as a solution ensures the exploration of global correlation. Then the optimization formulation can be written as:

$$\begin{aligned} \min_{\mathbf{W}, \mathcal{Z}} \text{rank}(\mathcal{Z}) + \frac{\lambda_1}{2} \|\mathfrak{P}_{\mathbb{O}}(\mathbf{X}\mathbf{W} - \mathbf{Z}_{(1)})\|_F^2 + \lambda_2 \mathfrak{R}(\mathbf{W}), \\ \text{s. t. } \mathcal{Z}_{\mathbb{O}} = \mathcal{Y}_{\mathbb{O}}. \end{aligned} \quad (6)$$

where \mathcal{Z} is the predicted tensor and $\mathfrak{P}_{\mathbb{O}}$ denotes the random sampling operator, which is defined by

$$\mathfrak{P}_{\mathbb{O}}(\mathcal{T}) = \begin{cases} \mathcal{T}(i_1, \dots, i_K), & i_1, \dots, i_K \in \mathbb{O} \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

where \mathbb{O} denotes the sample index set which contains the known indexes. We utilize the known entries which are worked as training data to predict the other observations.

Note that there are many kinds of tensor ranks, such as Tucker rank [42], tensor ring rank [43], and so on. To facilitate the algorithm solution, Tucker rank is considered. And the Tucker rank of K -th-order tensor \mathcal{Z} is denoted by $\text{rank}_{\text{T}}(\mathcal{Z})$, and its elements are the ranks of corresponding modes unfolding matrix, i.e., $\text{rank}_{\text{T}}(\mathcal{Z}) = (\text{rank}(\mathbf{Z}_{(1)}), \dots, \text{rank}(\mathbf{Z}_{(K)}))$. In addition, to identify the relevant features and incorporate the local temporal correlation, ℓ_1 -norm, $\ell_{2,1}$ -norm, and fused lasso penalty are used, which is similar to cFSLG, and our model is to solve the following optimization problem:

$$\begin{aligned} \min_{\mathbf{W}, \mathcal{Z}} \text{rank}(\mathcal{Z}) + \frac{\lambda_1}{2} \|\mathfrak{P}_{\mathbb{O}}(\mathbf{X}\mathbf{W} - \mathbf{Z}_{(1)})\|_F^2 + \lambda_2 \|\mathbf{W}\|_1 \\ + \lambda_3 \|\mathbf{R}\mathbf{W}^T\|_1 + \lambda_4 \|\mathbf{W}\|_{2,1} \\ \text{s. t. } \mathcal{Z}_{\mathbb{O}} = \mathcal{Y}_{\mathbb{O}}, \end{aligned} \quad (8)$$

where $\mathbf{R} = \mathbf{H}^T$, defined in (4).

The formulation of RATC-MTL explores both global consistency contained in cognitive scores and the linear relationship that exists between neuroimage and cognitive status. The missing value of \mathcal{Y} can be predicted by both image data and low-rank properties. The detailed solution for problem (8) will be given as follows.

B. Optimization Algorithm

The objective function of the problem (8) can be solved by splitting it into several subproblems. And the pseudocodes are summarized in Algorithm 1.

1) Update \mathbf{W} : When fixing other variables, the subproblem of \mathbf{W} is:

$$\min_{\mathbf{W}} \frac{1}{2} \|\mathfrak{P}_{\mathbb{O}}(\mathbf{X}\mathbf{W} - \mathbf{Y}_{(1)})\|_F^2 + \lambda_2 \|\mathbf{W}\|_1$$

$$+ \lambda_3 \|\mathbf{R}\mathbf{W}^T\|_1 + \lambda_4 \|\mathbf{W}\|_{2,1}. \quad (9)$$

Notably, to prevent the performance degradation of the model caused by learning with pseudo-labels, the update of \mathbf{W} only uses the training data which is achieved by sampling operator $\mathfrak{P}_\mathcal{O}$. Thus, \mathbf{W} can be updated at first. This problem is the same as cFSGL, which can be solved by accelerated gradient method [24], [44].

2) **Update \mathcal{Y}** : The subproblem of \mathcal{Y} is:

$$\begin{aligned} \min_{\mathcal{Z}} \text{rank}(\mathcal{Z}) + \frac{\lambda_1}{2} \|\mathbf{X}\mathbf{W} - \mathbf{Z}_{(1)}\|_{\mathbb{F}}^2, \\ \text{s. t. } \mathcal{Z}_\mathcal{O} = \mathcal{Y}_\mathcal{O}, \end{aligned} \quad (10)$$

where \mathbf{W} is the weight matrix learned above, which allows relation-aware tensor completion for disease progression in the following steps. Moreover, the optimization problem (10) is nonconvex since minimizing the rank function is NP-hard [45], a common approach is to use the trace norm instead, which is the tightest convex envelop for the rank function. Then the objective function can be relaxed to its convex one as:

$$\begin{aligned} \min_{\mathbf{W}, \mathcal{Z}} \|\mathcal{Z}\|_* + \frac{\lambda_1}{2} \|\mathfrak{P}_\mathcal{O}(\mathbf{X}\mathbf{W} - \mathbf{Z}_{(1)})\|_{\mathbb{F}}^2, \\ \text{s. t. } \mathcal{Z}_\mathcal{O} = \mathcal{Y}_\mathcal{O}. \end{aligned} \quad (11)$$

According to Definition 2, the equivalent formulation of the problem (11) can be obtained as:

$$\begin{aligned} \min_{\mathbf{W}, \mathcal{Z}, \mathcal{A}_1, \dots, \mathcal{A}_K} \sum_{k=1}^K \alpha_k \|\mathbf{A}_{k(k)}\|_* + \frac{\lambda_1}{2} \|\mathbf{X}\mathbf{W} - \mathbf{Z}_{(1)}\|_{\mathbb{F}}^2 \\ \text{s. t. } \mathcal{Z}_\mathcal{O} = \mathcal{Y}_\mathcal{O}, \mathcal{A}_k = \mathcal{Z}, k = 1, \dots, K. \end{aligned} \quad (12)$$

where $\mathcal{A}_k \in \mathbb{R}^{N \times I_1 \times \dots \times I_{K-1}}$, $k = 1, \dots, K$ are auxiliary tensors.

The augmented Lagrangian function of the problem (12) is defined as:

$$\begin{aligned} \mathfrak{L}_p(\mathcal{Z}, \mathcal{A}_1, \dots, \mathcal{A}_K, \mathcal{T}_1, \dots, \mathcal{T}_K) \\ = \frac{\lambda_1}{2} \|\mathbf{X}\mathbf{W} - \mathbf{Z}_{(1)}\|_{\mathbb{F}}^2 + \sum_{k=1}^K \alpha_k \|\mathbf{A}_{k(k)}\|_* \\ + \langle \mathcal{A}_k - \mathcal{Z}, \mathcal{T}_k \rangle + \frac{\rho}{2} \|\mathcal{A}_k - \mathcal{Z}\|_{\mathbb{F}}^2, \end{aligned} \quad (13)$$

under the constraint $\mathcal{Z}_\mathcal{O} = \mathcal{Y}_\mathcal{O}$. where ρ is a positive penalty scalar and $\mathcal{T}_k \in \mathbb{R}^{N \times I_1 \times \dots \times I_{K-1}}$, $k = 1, \dots, K$ is the Lagrange multiplier. The problem (13) can be further split into several subproblems within the alternating direction method of multipliers framework [46].

• **Update \mathcal{A}_k** : The subproblem of \mathcal{A}_k is:

$$\min_{\mathcal{A}_k} \sum_{k=1}^K \alpha_k \|\mathbf{A}_{k(k)}\|_* + \frac{\rho}{2} \|\mathcal{A}_k - \mathcal{Z} + \frac{\mathcal{T}_k}{\rho}\|_{\mathbb{F}}^2, \quad (14)$$

Then it can be divided into 3 subproblems:

$$\min_{\mathbf{A}_{k(k)}} \alpha_k \|\mathbf{A}_{k(k)}\|_* + \frac{\rho}{2} \|\mathbf{A}_{k(k)} - \mathbf{Z}_{(k)} + \frac{\mathbf{T}_{k(k)}}{\rho}\|_{\mathbb{F}}^2, \quad (15)$$

Algorithm 1: Relation-Aware Tensor Completion Multitask Learning.

Input: $\mathbf{X} \in \mathbb{R}^{N \times D}$, $\mathcal{Y} \in \mathbb{R}^{N \times I_1 \times \dots \times I_{K-1}}$ and index set \mathcal{O} , $\mathbf{W} \in \mathbb{R}^{D \times M}$, $M = \prod_{k=1}^{K-1} I_k$, $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \rho$.

Initialization: $\mathcal{Z}_\mathcal{O} = \mathcal{Y}_\mathcal{O}$, $\mathcal{A}_k = \mathcal{Z}$, $\mathcal{T}_k = 0$, $k = 1, \dots, K$.

update \mathbf{W} according to [24].

for iter = 1 **to** Maxiter **do**

$\hat{\mathcal{Z}} = \mathcal{Z}$;

for $k = 1$ **to** K **do**

update \mathcal{A}_k via (16);

update \mathcal{T}_k via (19);

end for

update \mathcal{Z} via (18), and note that each time \mathcal{Z} is updated, the known training labels should keep constant, which is a common operation in completion methods;

if $\frac{\|\hat{\mathcal{Z}} - \mathcal{Z}\|_{\mathbb{F}}}{\|\hat{\mathcal{Z}}\|_{\mathbb{F}}} < \text{tol}$ **then**
break;

end if

end for

Output: predicted tensor $\hat{\mathcal{Z}}$.

and the solution of $\mathbf{A}_{k(k)}$ is:

$$\mathbf{A}_{k(k)} = \mathbf{S}\mathbf{V}\mathbf{T}_\tau \left(\mathbf{Z}_{(k)} - \frac{\mathbf{T}_{k(k)}}{\rho} \right), \quad (16)$$

where $\tau = \frac{\alpha_k}{\rho}$ is a parameter for thresholding.

• **Update \mathcal{Z}** : Fix other variables, the subproblem on \mathcal{Z} is:

$$\begin{aligned} \min_{\mathcal{Z}} \sum_{k=1}^K \langle \mathcal{A}_k - \mathcal{Z}, \mathcal{T}_k \rangle + \frac{\rho}{2} \|\mathcal{A}_k - \mathcal{Z}\|_{\mathbb{F}}^2 \\ + \frac{\lambda_1}{2} \|\mathbf{X}\mathbf{W} - \mathbf{Z}_{(1)}\|_{\mathbb{F}}^2, \\ \text{s. t. } \mathcal{Z}_\mathcal{O} = \mathcal{Y}_\mathcal{O}. \end{aligned} \quad (17)$$

Since this objective function is convex and differentiable, \mathcal{Z} can be updated by

$$\mathcal{Z} = \text{Fold}_1 \left(\lambda_1 \mathbf{X}\mathbf{W} + \sum_{k=1}^3 (\rho \mathbf{A}_k + \mathbf{T}_k)_{(1)} \right), \quad (18)$$

and keep known training labels unchanged.

• **Update \mathcal{T}_k** : The Lagrange multiplier can be updated by

$$\mathcal{T}_k = \mathcal{T}_k + \rho(\mathcal{A}_k - \mathcal{Z}) \quad (19)$$

V. EXPERIMENT AND RESULTS

A. Data

The clinical data used in this research were all collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI began in 2003, funded as a private-public partnership, and is a longitudinal multicenter

study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of Alzheimer's disease, where the data of subjects are obtained by conducting a long-term following test, and recorded every half year. The first time when the subject performs the screening becomes Baseline (BL), and the duration starting from BL to the follow-up visit is denoted by "M.", such as "M06" denotes 6 months after the first screening. Notably, it has been facilitating the scientific evaluation of diverse data for medical diagnosis, including MRI, positron emission tomography (PET), clinical neuropsychological assessments, and other biomarkers, see www.adni-info.org for up-to-date information.

In this paper, we use the baseline biomarkers from DTI as input features, which is a promising imaging technique that has greatly helped to identify white matter regions affected by AD in its early stages [47]. And the diffusion tensor images, from the ADNI website, used for this research were preprocessed using FSL software [48], including:

- 1) transformed the images into Nifti format;
- 2) extracted the gradient directions and b -values;
- 3) correction for eddy current and head motion;
- 4) skull-stripping using the brain extraction tool.

After preprocessing, a single diffusion tensor was fitted at each voxel in the image by DTifit, fraction anisotropy (FA) and mean diffusivity (MD) were calculated [49], [50]. Then FA and MD maps need to be carefully aligned to a group-wise space, we achieved this by means of the tract-based spatial statistics (TBSS) algorithm implemented in FSL [51].

By applying the automated anatomical labeling (AAL) [52], the brain space was partitioned into 90 brain regions of interest (ROIs) for each image. Then the mean and variance of FA and MD of all 90 regions are calculated, and there are a total of 360 ($90 \times 2 \times 2$) features. In this study, we assume that there are no missing values in input features.

B. Evaluation Metrics

To quantitatively assess the predictive performance of our model, we utilized the root mean square error (rMSE) and correlation coefficient (CC) as the assessment metrics. Moreover, for the overall performance, normalized mean square error (nMSE), which is widely utilized in multitask learning [53], and weighted R-value (wR), which is utilized in the medical field to address the progression of AD [54].

$$\text{rMSE}(\mathbf{y}, \hat{\mathbf{y}}) = \sqrt{\frac{\|\mathbf{y} - \hat{\mathbf{y}}\|_2^2}{n}} \quad (20)$$

$$\text{CC}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{\text{cov}(\mathbf{y}, \hat{\mathbf{y}})}{\sigma(\mathbf{y})\sigma(\hat{\mathbf{y}})} \quad (21)$$

$$\text{nMSE}(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{\sum_{m=1}^M \|\mathbf{Y}_m - \hat{\mathbf{Y}}_m\|_2^2 / \sigma(\mathbf{Y}_m)}{\sum_{m=1}^M n_m} \quad (22)$$

$$\text{wR}(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{\sum_{m=1}^M \text{corr}(\mathbf{Y}_m, \hat{\mathbf{Y}}_m) N_m}{\sum_{m=1}^M n_m} \quad (23)$$

For rMSE and CC, \mathbf{y} is the ground truth of the target for a signal task (a time point or a cognitive score), and $\hat{\mathbf{y}}$ is the

TABLE I
THE DESCRIPTION OF COGNITIVE SCORES USED IN THIS STUDY

Cognitive Score	Description
ADAS 11 ²	Alzheimer's Disease Assessment Scale with 11 items
ADAS 13 ^{1,2}	Alzheimer's Disease Assessment Scale with 13 items
MMSE ^{1,2}	Mini-Mental State Examination
MOCA ²	Montreal Cognitive Assessment
CDRSB ²	Clinical Dementia Rating Sum of Boxes
BOSNAM ²	Total number correct
DIGIT ²	Digit symbol substitution
RAVLT TOTAL ²	Total score of the first 5 learning trails
RAVLT TOTB ²	Immediately after the fifth learning trial
RAVLT TOT6 ^{1,2}	Trail 6 Total number of words recalled
RAVLT T30 ^{1,2}	30 min delay total number of words recalled
RAVLT RECOG ^{1,2}	30 min delay recognition
FLU ANIM ^{1,2}	Animal Total score
FLU VEG ²	Vegetable Total score
CLOCK DRAW ²	Clock drawing
CLOCK COPY ²	Clock copying
DSPAN For ²	Digit span forward
DSPAN BAC ²	Digit span backward
TRAILS A ²	Trail making test A score
TRAILS B ²	Trail making test B score
DSS ¹	Disease Severity Score: 1-CN,2-SMC,3-EMCI,4-LMCI,5-AD

corresponding predicted value. For nMSE and wR, \mathbf{Y}_m represents the ground truth of target for task m , $m = 1, \dots, M$, $\hat{\mathbf{Y}}_m$ represents the corresponding predicted result. n is the number of task samples, cov is the covariance between two vectors, σ is the standard deviation of the vector, and corr is the correlation coefficient between two vectors. It is worth noting that for regression method $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{W}$ after learning the parameter \mathbf{W} , but for our work $\hat{\mathbf{y}} = \hat{\mathbf{z}}$, and we utilize corresponding elements of $\hat{\mathbf{y}}$ to calculate these metrics. For nMSE and rMSE, a smaller value indicates better performance, whereas, for wR and CC, a larger value represents improved performance.

C. Compared Algorithms

In this paper, we conduct two groups of experiments on ADNI data to validate the proposed algorithm: a simulated situation and a real situation. These two experiments were all designed to jointly predict multiple cognitive scores at multiple time points using baseline DTI images. To compare the prediction performance, we select seven state-of-the-art algorithms, including single-task method and multi-task learning methods which incorporate the task prior correlation knowledge.

- Lasso regression (Lasso) [55];
- Temporal Group Lasso (TGL) [23];
- Convex Fused Sparse Group Lasso (cFSGL) [24];
- Non-convex Fused Sparse Group Lasso (nFSGL) [25];
- Robust Multitask Learning (RMTL) [34];
- Trace norm Multitask Learning (Trace) [56];
- Non-Convex Calibrated Multitask Learning (NC-CMTL) [37].

Cognitive scores used in this paper are listed in Table I. As matrix \mathbf{P} in (4) is used to deal with incomplete target data, we employ this operation in all comparison methods like [57]. And all input data have been normalized by z-scored.

TABLE II
BASELINE DEMOGRAPHIC INFORMATION OF SUBJECTS USED IN SIMULATION EXPERIMENTS

class	Number(F/M)	Age(years)	Edu(years)
AD	34(10/24)	76.44±7.23	15.70±2.83
CN	47(24/23)	76.89±6.89	16.89±2.38
EMCI	93(39/54)	75.97±7.56	15.86±2.68
LMCI	49(21/28)	74.14±6.79	15.91±2.73
SMC	12(7/5)	75.30±3.89	16.84±2.10

D. Simulation Experiments

In this situation, we use 235 subjects who have high-quality images at BL and seven cognitive scores well-documented at three continuous time points (BL, M06, and M12), which ensures these samples have accurate ground truth to evaluate the results. This population consists of 34 participants with Alzheimer's disease (AD), 47 cognitively normal subjects (CN), 93 early mild cognitive impairment (EMCI), 49 late mild cognitive impairment (LMCI), and 12 significant memory concern (SMC), the demographic information of these subjects at baseline are shown in Table II, including the female and male number of subjects corresponding to each category, age, and education. And the data size in this experiment is $\mathbf{X} \in \mathbb{R}^{235 \times 360}$, $\mathcal{Y} \in \mathbb{R}^{235 \times 3 \times 7}$.

To compare with existing methods, we conducted a two-fold analysis on them, given that our method enables simultaneous prediction of multiple scores across different time points. (1) Longitudinal analysis, simultaneously predicts one cognitive score at all time points. (2) Cross-sectional analysis, predict all seven clinical variables together at a one-time step. Note that in these two situations, the experiment settings of matrix-based methods are the same as [24] and [19], which is to build a prediction model for each target. Specifically, they use baseline DTI features to predict one cognitive score at multiple time points and run corresponding methods seven times for all seven clinical scores in longitudinal analysis. Similarly, for the cross-sectional study, the baseline biomarkers are also used to predict all cognitive scores at one-time points and run three times for all time steps.

To simulate the missing situation, we artificially introduce the sampling rate (SR) on cognitive scores. Considering various missing scenarios, we set SR to vary from 10% to 80%. Upon giving SR, the corresponding proportion of a particular cognitive score at a certain time point is chosen as known labels. We apply this operation to all cognitive scores across three-time steps. Then the samples corresponding to these labels constitute the training dataset. Subsequently, the remaining samples are evenly partitioned into validation and testing sets. Notably, the cross-validation method used in this part is a specific version of the Monte Carlo cross-validation (MCCV), in which the data are partitioned K times into disjoint train, test, and validate subsets. The key distinction between MCCV and standard k -fold cross-validation is that in MCCV the different train subsets are chosen randomly and need not be disjoint [58]. MCCV has shown its superiority in many applications [59], [60]. However, due to the randomness of the method, it is also possible to

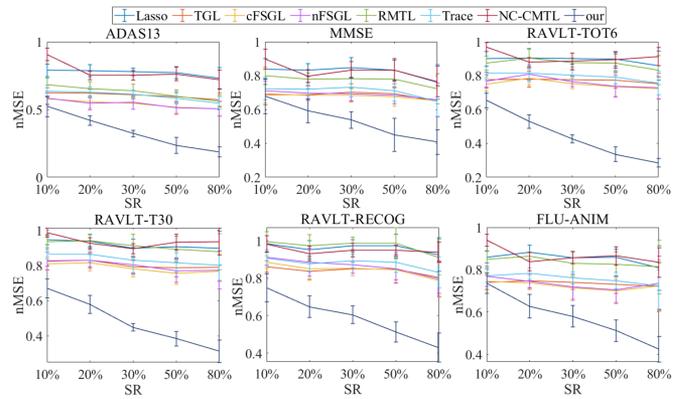


Fig. 2. Comparison of different methods on six cognitive scores with different SR in Longitudinal analysis.

introduce more noise, and the stability is relatively poor. We mitigate this problem to some extent through multiple repetitions, which are measured between computational cost and model accuracy, and we set $K = 10$ in our experiments. In the end, we build the model on training data, evaluate on testing set, and repeat this process 10 times, the average of all 10 metrics was used to evaluate the performance of our method. Since the parameters of the optimization problem (8) need to be selected, we use the validation set to select these parameters in grid search loops, where the range of each parameter varies from 0.01 to 1000. In the end, we chose $\lambda_1 = 0.01$, $\lambda_2 = 0.01$, $\lambda_3 = 0.1$, $\lambda_4 = 0.01$, and $\rho = 0.1$. The experimental results are as follows.

1) Performance in the Longitudinal Study: The results of the longitudinal analysis with 20% SR are shown in Tables III and IV and use boldfaced to represent the best result, underline to indicate suboptimal results.

From these, we can draw the following conclusions: First, compared with single-task learning (Lasso), MTL can utilize the temporal correlation among tasks and have better performance in longitudinal AD prediction. Second, compared with other MTL algorithms, the proposed method consistently shows the best performance on overall prediction measurements. More specifically, the nMSE measurement of DSS and MMSE achieves a decline of 28.88% and 13.37% compared with the suboptimal values, and the wR increased by 17.39% and 26.41% respectively. These results verify the effectiveness of the proposed method to simultaneously explore correlations between multiple modalities in clinical scores and incorporate the medical features.

Fig. 2 shows the comparison curves of different methods in predicting six clinical scores with SR ranging from 10% to 80%. As the number of training samples increases, all methods show an improvement in their respective performance. It shows that our method shows the best results across different SRs, which further proves the power of the proposed algorithm.

An interesting phenomenon is that the prediction performance of ADAS-13 is better than other scores, which is the same as [19]. The possible reason is that ADAS-13 exhibits more pronounced temporal correlations than other tests. What's more, it is also

TABLE III
PERFORMANCE COMPARISON OF LONGITUDINAL ANALYSIS IN TERMS OF RMSE AND CC ON SEVEN COGNITIVE SCORES AT THREE-TIME POINTS

Method	BL		M06		M12	
	rMSE	CC	rMSE	CC	rMSE	CC
Target:DSS						
Lasso	0.9028±0.0539	0.4336±0.0774	0.9022±0.0602	0.4422±0.0776	0.8869±0.0548	0.4717±0.0536
TGL	0.7992±0.0382	0.5964±0.0479	0.7846±0.0332	0.6223±0.0443	0.7839±0.0406	0.6267±0.0420
cFSGL	0.7440±0.0316	0.6722±0.0381	0.7408±0.0423	0.6717±0.0430	0.7367±0.0364	0.6812±0.0428
nFSGL	0.7496±0.0292	0.6734±0.0330	0.7428±0.0488	0.6744±0.0440	0.7460±0.0363	0.6789±0.0412
RMTL	0.8209±0.0491	0.5778±0.0631	0.8288±0.0543	0.5662±0.0529	0.8005±0.0445	0.6100±0.0505
Trace	0.7860±0.0413	0.6157±0.0510	0.7861±0.0378	0.6262±0.0456	0.7723±0.0453	0.6496±0.0434
NC-CMTL	0.8740±0.0424	0.5402±0.0635	0.8766±0.0409	0.5659±0.0549	0.8762±0.0613	0.5804±0.0496
ours	0.3928±0.0206	0.8142±0.0292	0.4094±0.0288	0.8128±0.0259	0.3888±0.0275	0.8270±0.0351
Target:ADAS-13						
Lasso	0.8027±0.0397	0.4636±0.0686	0.8439±0.0474	0.5280±0.0529	1.0213±0.0520	0.4557±0.0536
TGL	0.7093±0.0334	0.6119±0.0514	0.7470±0.0367	0.6645±0.0486	0.9111±0.0470	0.6117±0.0439
cFSGL	0.6879±0.0422	0.6528±0.0543	0.7008±0.0326	0.7091±0.0463	0.8529±0.0497	0.6682±0.0438
nFSGL	0.6927±0.0429	0.6585±0.0520	0.6946±0.0322	0.7171±0.0443	0.8361±0.0573	0.6840±0.0456
RMTL	0.7265±0.0476	0.5935±0.0720	0.7612±0.0365	0.6446±0.0534	0.9403±0.0545	0.5747±0.0659
Trace	0.7083±0.0350	0.6162±0.0536	0.7593±0.0426	0.6565±0.0482	0.9180±0.0486	0.6057±0.0490
NC-CMTL	0.7570±0.0332	0.5823±0.0525	0.8536±0.0576	0.5829±0.0501	1.0041±0.0617	0.5616±0.0457
ours	0.3626±0.0271	0.8031±0.0257	0.3971±0.0192	0.8175±0.0371	0.4594±0.0324	0.8290±0.0243
Target:MMSE						
Lasso	0.7692±0.0433	0.3142±0.0843	0.8724±0.0657	0.3407±0.0492	1.0759±0.0698	0.3869±0.0697
TGL	0.7067±0.0444	0.5617±0.0693	0.7864±0.0616	0.6281±0.0398	0.9771±0.0571	0.4881±0.0582
cFSGL	0.7243±0.0480	0.5583±0.0598	0.7627±0.0492	0.6469±0.0455	0.9726±0.0540	0.5132±0.0538
nFSGL	0.7426±0.0519	0.5533±0.0613	0.7617±0.0523	0.6496±0.0507	0.9749±0.0562	0.5199±0.0540
RMTL	0.7482±0.0420	0.4969±0.0689	0.8270±0.0534	0.5675±0.0375	1.0519±0.0584	0.4042±0.0709
Trace	0.7168±0.0355	0.5528±0.0703	0.8107±0.0701	0.6063±0.0406	0.9996±0.0629	0.4527±0.0661
NC-CMTL	0.7485±0.0348	0.5483±0.0628	0.8787±0.0813	0.5833±0.0540	1.0365±0.0536	0.4649±0.0740
ours	0.4393±0.0386	0.6533±0.0355	0.4654±0.0592	0.7222±0.0524	0.5524±0.0461	0.6860±0.0507
Target:RAVLT-TOT6						
Lasso	0.9506±0.0553	0.3142±0.0843	0.9256±0.0601	0.3407±0.0492	0.9536±0.0575	0.3869±0.0697
TGL	0.8820±0.0433	0.4465±0.0634	0.8530±0.0489	0.4858±0.0605	0.8920±0.0431	0.4936±0.0638
cFSGL	0.8942±0.0342	0.4692±0.0465	0.8552±0.0431	0.5059±0.0498	0.8846±0.0354	0.5169±0.0529
nFSGL	0.9105±0.0290	0.4660±0.0393	0.8700±0.0453	0.5009±0.0495	0.8936±0.0351	0.5158±0.0552
RMTL	0.9517±0.0424	0.3640±0.0553	0.9277±0.0576	0.3733±0.0553	0.9459±0.0565	0.4339±0.0617
Trace	0.8966±0.0491	0.4179±0.0729	0.8846±0.0531	0.4263±0.0582	0.9035±0.0487	0.4812±0.0716
NC-CMTL	0.9225±0.0418	0.4091±0.0755	0.9186±0.0543	0.4443±0.0781	0.9550±0.0350	0.4682±0.0629
ours	0.4597±0.0276	0.7205±0.0672	0.4629±0.0405	0.7083±0.0449	0.4569±0.0329	0.7546±0.0375
Target:RAVLT-T30						
Lasso	0.9603±0.0237	0.3201±0.0558	0.9036±0.0350	0.3280±0.0751	1.0476±0.0657	0.2990±0.0634
TGL	0.9050±0.0264	0.4336±0.0433	0.8457±0.0338	0.4538±0.0600	0.9887±0.0641	0.4156±0.0470
cFSGL	0.8967±0.0334	0.4760±0.0438	0.8425±0.0290	0.4925±0.0482	0.9754±0.0577	0.4569±0.0354
nFSGL	0.9025±0.0385	0.4802±0.0429	0.8551±0.0313	0.4908±0.0504	0.9820±0.0566	0.4590±0.0329
RMTL	0.9579±0.0199	0.3651±0.0443	0.8985±0.0395	0.3672±0.0823	1.0581±0.0530	0.3254±0.0435
Trace	0.9172±0.0282	0.4055±0.0506	0.8675±0.0370	0.4028±0.0734	1.0111±0.0672	0.3689±0.0606
NC-CMTL	0.9519±0.0359	0.3893±0.0487	0.8947±0.0316	0.4088±0.0872	1.0461±0.0620	0.1925±0.0715
ours	0.4743±0.0248	0.7120±0.0300	0.4358±0.0285	0.6965±0.0376	0.5139±0.0445	0.7188±0.0579
Target:RAVLT-RECOG						
Lasso	0.9550±0.0684	0.1760±0.0733	0.9606±0.0557	0.3172±0.0624	1.0089±0.0436	0.3422±0.0320
TGL	0.9081±0.0646	0.3118±0.0684	0.8917±0.0466	0.4646±0.0530	0.9419±0.0375	0.4586±0.0461
cFSGL	0.9265±0.0732	0.3310±0.0761	0.8870±0.0390	0.4834±0.0619	0.9503±0.0447	0.4687±0.0465
nFSGL	0.9495±0.0734	0.3210±0.0771	0.9006±0.0391	0.4765±0.0644	0.9681±0.0537	0.4590±0.0604
RMTL	0.9627±0.0651	0.2273±0.0665	0.9725±0.0494	0.3275±0.0506	1.0229±0.0452	0.3613±0.0370
Trace	0.9167±0.0635	0.2698±0.0762	0.9215±0.0552	0.4083±0.0581	0.9706±0.0445	0.4059±0.0456
NC-CMTL	0.9246±0.0492	0.2903±0.0859	0.9644±0.0605	0.4322±0.0588	1.0078±0.0448	0.4273±0.0544
ours	0.4999±0.0504	0.5778±0.0321	0.5086±0.0537	0.6426±0.0668	0.5216±0.0376	0.6719±0.0655
Target:FLU-ANIM						
Lasso	0.9517±0.0616	0.3367±0.0719	0.8872±0.0392	0.4352±0.0337	0.9723±0.0766	0.3573±0.0853
TGL	0.8742±0.0618	0.4870±0.0703	0.8154±0.0480	0.5622±0.0451	0.8996±0.0616	0.4841±0.0534
cFSGL	0.8723±0.0606	0.5015±0.0629	0.7955±0.0560	0.5897±0.0515	0.9026±0.0569	0.5010±0.0466
nFSGL	0.8766±0.0571	0.5065±0.0596	0.7949±0.0587	0.5943±0.0531	0.9102±0.0555	0.5031±0.0496
RMTL	0.9452±0.0468	0.3759±0.0752	0.8646±0.0468	0.4877±0.0326	0.9720±0.0732	0.3930±0.0748
Trace	0.8923±0.0598	0.4600±0.0664	0.8411±0.0416	0.5232±0.0314	0.9143±0.0685	0.4585±0.0669
NC-CMTL	0.9208±0.0668	0.4848±0.0753	0.8827±0.0388	0.5237±0.0331	0.9388±0.0744	0.4912±0.0648
ours	0.5224±0.0504	0.6067±0.0578	0.4788±0.0270	0.6676±0.0686	0.4866±0.0315	0.6813±0.0385

Bold values represent the best results, and underlined values represent the suboptimal results.

TABLE IV
PERFORMANCE COMPARISON OF LONGITUDINAL ANALYSIS IN TERMS OF NMSE AND WR ON SEVEN COGNITIVE SCORES

method	DSS	ADAS-13	MMSE	RAVLT-TOT6	RAVLT-T30	RAVLT-RECOG	FLU-ANIM
nMSE:							
Lasso	0.8130±0.0410	0.7858±0.0450	0.8331±0.0281	0.9046±0.0563	0.9376±0.0384	0.9514±0.0486	0.8820±0.0327
TGL	0.6280±0.0204	0.6187±0.0311	0.6884±0.0353	0.7795±0.0477	0.8307±0.0377	0.8363±0.0413	0.7484±0.0344
cFSGL	0.5543±0.0328	0.5561±0.0454	0.6850±0.0337	0.7840±0.0387	0.8154±0.0446	0.8510±0.0626	0.7391±0.0350
nFSGL	0.5631±0.0415	0.5478±0.0505	0.6973±0.0409	0.8082±0.0337	0.8302±0.0473	0.8859±0.0749	0.7461±0.0341
RMTL	0.6746±0.0310	0.6531±0.0537	0.7802±0.0350	0.9026±0.0551	0.9398±0.0369	0.9733±0.0599	0.8648±0.0147
Trace	0.6157±0.0291	0.6280±0.0363	0.7201±0.0404	0.8141±0.0501	0.8650±0.0440	0.8770±0.0422	0.7819±0.0268
NC-CMTL	0.7720±0.0436	0.7528±0.0313	0.7972±0.0387	0.8804±0.0274	0.9253±0.0336	0.9317±0.0380	0.8380±0.0439
ours	0.3942±0.0451	0.4200±0.0350	0.5934±0.0707	0.5297±0.0406	0.5781±0.0506	0.6470±0.0578	0.6267±0.0558
wR:							
Lasso	0.4494±0.0339	0.4825±0.0434	0.4239±0.0262	0.3464±0.0438	0.3152±0.0302	0.2781±0.0397	0.3769±0.0374
TGL	0.6155±0.0245	0.6292±0.0297	0.5586±0.0215	0.4741±0.0434	0.4337±0.0307	0.4119±0.0403	0.5112±0.0339
cFSGL	0.6750±0.0274	0.6768±0.0330	0.5719±0.0238	0.4966±0.0270	0.4749±0.0269	0.4286±0.0450	0.5307±0.0243
nFSGL	0.6756±0.0276	0.6867±0.0322	0.5732±0.0274	0.4936±0.0251	0.4766±0.0287	0.4199±0.0495	0.5345±0.0229
RMTL	0.5844±0.0245	0.6041±0.0399	0.4888±0.0311	0.3895±0.0294	0.3525±0.0334	0.3056±0.0367	0.4189±0.0310
Trace	0.6308±0.0225	0.6260±0.0327	0.5366±0.0238	0.4408±0.0454	0.3920±0.0387	0.3615±0.0435	0.4810±0.0313
NC-CMTL	0.5628±0.0299	0.5753±0.0336	0.5312±0.0242	0.4391±0.0467	0.3873±0.0425	0.3829±0.0452	0.5000±0.0396
ours	0.8179±0.0212	0.8166±0.0086	0.6857±0.0285	0.7270±0.0382	0.7094±0.0190	0.6298±0.0348	0.6516±0.0361

Bold values represent the best results, and underlined values represent the suboptimal results.

TABLE V
PERFORMANCE COMPARISON OF CROSS-SECTIONAL ANALYSIS IN TERMS OF NMSE AND WR AT THREE-TIME POINTS

method	BL	M06	M12
nMSE:			
Lasso	0.8482±0.0298	0.8272±0.0242	0.9371±0.0292
TGL	0.8162±0.0258	0.8088±0.0185	0.8999±0.0238
cFSGL	0.7653±0.0358	0.7545±0.0260	0.8501±0.0391
nFSGL	0.8629±0.0269	0.8671±0.0222	0.9368±0.0281
RMTL	0.8065±0.0328	0.7790±0.0312	0.9113±0.0339
Trace	0.7351±0.0240	0.7209±0.0183	0.8000±0.0218
NC-CMTL	0.7098±0.0995	0.6562±0.0202	0.7022±0.0200
ours	0.5254±0.0234	0.5240±0.0263	0.5741±0.0299
wR:			
Lasso	0.3603±0.0273	0.3983±0.0231	0.3846±0.0215
TGL	0.3902±0.0247	0.4145±0.0255	0.4122±0.0236
cFSGL	0.4479±0.0299	0.4821±0.0265	0.4741±0.0289
nFSGL	0.3334±0.0460	0.3441±0.0399	0.3694±0.0499
RMTL	0.4299±0.0260	0.4636±0.0259	0.4436±0.0176
Trace	0.4850±0.0248	0.5105±0.0224	0.5110±0.0187
NC-CMTL	0.5158±0.0886	0.5728±0.0209	0.5940±0.0169
ours	0.7069±0.0202	0.7222±0.0229	0.7295±0.0230

Bold values represent the best results, and underlined values represent the suboptimal results.

observed that in the longitudinal study, using the temporal correlation as a prior regularization is more effective than low-rank constraints. To statistically evaluate the distinctions between the proposed method and other approaches, we conduct the paired sample t-test in this part and apply the Bonferroni correction [61] to account for multiple comparisons. Our findings reveal that, in comparison to other methods, our approach demonstrates a consistently increased wR ($p < 0.001$), and decreased nMSE ($p < 0.001$). Moreover, we can also observe enhanced CC ($p < 0.05$) and decreased rMSE ($p < 0.001$) among all time points. For more details, see Appendix A.

2) Performance in the Cross-Sectional Analysis: The cross-sectional study's nMSE and wR with SR = 20% are

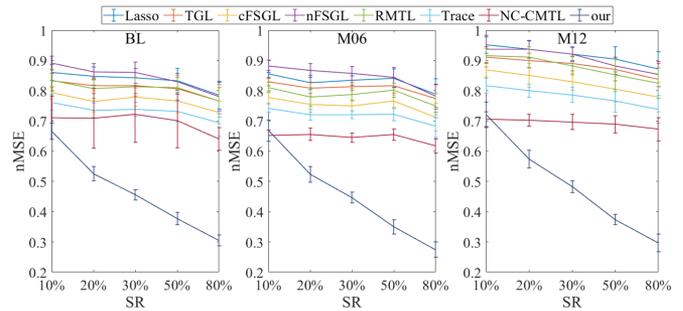


Fig. 3. Comparison of different methods at three-time points with different SR in cross-sectional analysis.

reported in Table V, while Fig. 3 indicates the comparison curves of different methods with varying SRs at three distinct periods.

Based on these results, we can derive several observations. First, in this experiment, multitask learning is still competitive with single-task learning which justifies the correlations between different cognitive scores are also useful information, and it is reasonable to incorporate such prior knowledge in associated studies. Second, the proposed method shows the best performance at all time points, for instance, the nMSE measurement achieves 25.97%, 19.93%, and 18.24% reduction compared to NC-CMTL at three-time points respectively, and wR values have increased by 27.03%, 20.68%, and 18.57% respectively. These results further validate the effectiveness of our framework. Third, from Tables IV and V, Low-rank methods are more suitable for cross-sectional analysis, this may due to the temporal correlation are not prominent in the horizontal direction and is difficult to apply in such formulation. Additionally, imposing low-rank constraints on clinical scores can be a useful way to explore the underlying structural information. Furthermore, the same paired sample t-test is conducted in this experiment. The results show that our method consistently indicates significant increases in wR ($p < 0.001$) compared to other methods,

TABLE VI

PERFORMANCE COMPARISON BY USING UNFOLDED DATA IN TERMS OF NMSE AND WR AT THREE-TIME POINTS

method	BL	M06	M12
nMSE:			
Lasso	0.8482±0.0298	0.8272±0.0242	0.9371±0.0292
TGL	0.7211±0.0272	0.7143±0.0206	0.8095±0.0168
cFSGL	0.6857±0.0307	0.6607±0.0271	0.7605±0.0299
nFSGL	0.6810±0.0331	0.6465±0.0273	0.7424±0.0351
RMTL	0.7669±0.0333	0.7324±0.0271	0.8647±0.0348
Trace	0.6749±0.0245	0.6656±0.0211	0.7482±0.0254
NC-CMTL	0.5981±0.0203	0.6019±0.0228	0.6487±0.0216
ours	0.5254±0.0234	0.5140±0.0263	0.5741±0.0331
wR:			
Lasso	0.3603±0.0273	0.3983±0.0231	0.3846±0.0215
TGL	0.4948±0.0301	0.5234±0.0252	0.5091±0.0281
cFSGL	0.5350±0.0251	0.5680±0.0251	0.5459±0.0266
nFSGL	0.5458±0.0242	0.5836±0.0231	0.5634±0.0276
RMTL	0.4758±0.0267	0.5106±0.0200	0.4842±0.0187
Trace	0.5446±0.0236	0.5622±0.0229	0.5549±0.0239
NC-CMTL	0.6137±0.0234	0.6173±0.0207	0.6329±0.0205
ours	0.7069±0.0202	0.7222±0.0229	0.7295±0.0230

Bold values represent the best results, and underlined values represent the suboptimal results.

as well as decreased nMSE ($p < 0.001$). For more details see Appendix A.

From the above, we can observe that our method consistently exhibits optimal results, and there is a large difference between the matrix-based methods and the tensor-based method. Therefore, we want to further investigate the reason for this difference. Namely, we identified two possible reasons: (1) the tensor-based method uses more data. That is the matrix-based methods only consider one correlation and handle a limited number of tasks, i.e. T or V , but our approach involves $T * V$ tasks, thereby leveraging a piece of more abundant information. (2) the low-rank tensor structure carries a lot of important information. To investigate how each of these factors affects prediction performance, we take the cross-sectional analysis as an example, and do the following experiment: We unfolded the tensor to a matrix along mode-1, and let the matrix-based methods solve the equation $\|\mathbf{X}\mathbf{W} - \mathbf{Y}_{(1)}\|_F$ with corresponding regularization terms. Then compare it with our tensor-based solution. This experiment compares the results when all data are used but without the low-rank tensor assumption. We set other experiment settings to be consistent with those in Table V and report the results in Table VI to make a direct comparison. In addition, the paired t-test is also conducted on these obtained results, the test results show increased wR ($p < 0.001$), decreased nMSE ($p < 0.001$) in comparison to other methods, as shown in Appendix A. These results can reveal some interesting points:

- 1) As the number of tasks increases, the predictive performance of matrix-based methods improves. This enhancement is likely attributed to the presence of correlations not only among multiple cognitive scores at the same time point but also across different time points. Consequently, our approach leverages various forms of correlation and more abundant data, contributing to the improvement in model performance.

TABLE VII

BASELINE DEMOGRAPHIC INFORMATION OF SUBJECTS AND MISSING RATE IN REAL SITUATION EXPERIMENTS

Class	Number(F/M)	Age(years)	Edu(years)	Missing rate(1/2)
AD	62(21/41)	75.00±8.06	15.43±2.72	48.23%/37.02%
CN	74(42/32)	75.79±7.29	17.02±2.41	42.43%/38.38%
EMCI	98(39/59)	75.91±7.60	15.86±2.67	24.34%/21.28%
LMCI	50(21/29)	74.18±6.73	16.00±2.76	20.30%/21.90%

- 2) Despite the improved performance of matrix-based methods as shown in Table VI, our approach still demonstrates the optimal results. The prominent reason behind this lies in the low-rank tensor assumption for cognitive scores. This assumption facilitates the exploration of the global structural information within cognitive scores, leading to enhanced predictive accuracy.

E. Real Experiments

In the simulation experiments, the missing condition of data is artificially designed. To verify the performance of our method in dealing with the real missing situation on the ADNI dataset, we process the data including (1) deleting the samples without baseline DTI records, (2) deleting samples without labels, (3) deleting the samples whose entries are missed more than 10%. Finally, 284 subjects which consist of 62 participants with Alzheimer's disease (AD), 74 cognitively normal subjects (CN), 98 early mild cognitive impairment (EMCI), and 50 late mild cognitive impairment (LMCI) are used. It is noteworthy that, due to the limited number of subjects in the SMC class, we omit subjects from the SMC class. Apart from the above data, these subjects' corresponding 20 cognitive measurements at three-time points in the real situation are analyzed as labels. Then the data structures are $\mathbf{X} \in \mathbb{R}^{284 \times 360}$, $\mathcal{Y} \in \mathbb{R}^{284 \times 3 \times 20}$. Statistical information of these subjects at BL and the missing ratio of M06 (1) and M12 (2) is recorded in Table VII. The missing rates of different cognitive scores are shown in Appendix B. Evidently, the complexity of data incompleteness increases in real-world scenarios. Disparities in data missing rates exist among different time points, different disease progression, and different cognitive scores.

In this part, we retain the incomplete situation of authentic data. The known entries are used as the training set, and the remaining samples are partitioned into two equivalent sets, with one for validation and the other for testing. In the absence of ground truth under true missing conditions, five binary classification tasks, which utilize the predicted cognitive scores as classification criteria, are conducted to indirectly verify the accuracy of our model. Due to the efficiency of support vector machine (SVM) in small sample classification tasks [62], [63], an SVM classifier with a radial basis function (RBF) kernel is adopted in our experiments. The penalty coefficient of the RBF kernel is set based on the empirical value and remains constant during the whole experiment. The validation method of classification experiments is the leave-one-out which can exploit data to the fullest. The same classifier is used for all comparison methods, and prediction accuracy (ACC) and sensitivity (SEN) are used

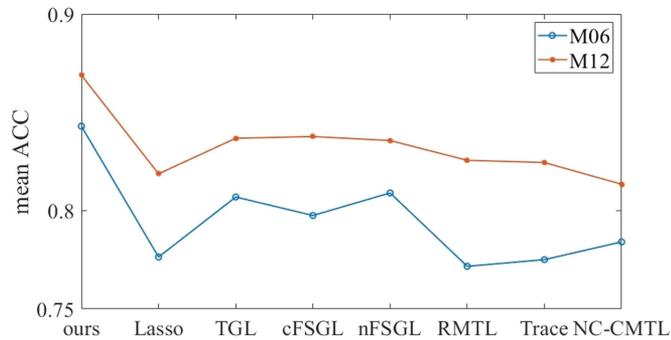


Fig. 4. Mean accuracy comparison of different methods at M06 and M12.

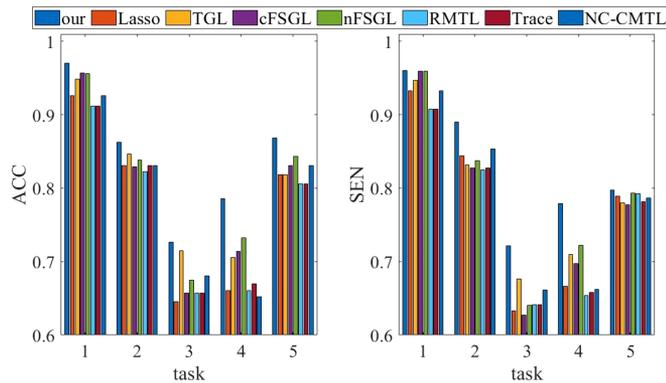


Fig. 5. ACC and SEN of different methods at M06.

as evaluation metrics. When employing the predicted cognitive scores as classification features, the statistical disparities among different categories also constitute a crucial metric. Thus for each classification task, we perform group t-tests on the features across different categories to validate the statistical significance of our method. For simplicity, different numbers are used to indicate specific classification tasks, and 1-‘CN vs. AD’, 2-‘CN vs. LMCI’, 3-‘CN vs. EMCI’, 4-‘AD vs. LMCI’, 5-‘AD vs. EMCI’.

Similar to the simulation one, the parameters of the optimization problem (8) in this experiment are also selected in the grid search loop that $\lambda_1 = 0.01, \lambda_2 = 0.01, \lambda_3 = 0.01, \lambda_4 = 100$, and $\rho = 10$. Also, the parameters of the comparison methods are also adjusted to the optimum.

As Fig. 4 shows the overall classification accuracy for five binary classification tasks, the mean ACC of our proposed method is 84.29% and 86.88% at M06 and M12, which yield the best performance. Compared to the suboptimal method, our method improves by 4.02% and 3.69% at two-time points respectively. Due to the higher overall missing rate at M06, M12 exhibits superior results in comparison.

For further analysis, Fig. 5 shows the comparison results of ACC and SEN on all binary classification tasks at M06, and Fig. 6 shows the corresponding results at M12. Excepting the SEN value of the first classification task at M12, the proposed method shows the state-of-the-art ones in other values, especially, the proposed framework has a significant advantage in ‘AD vs.

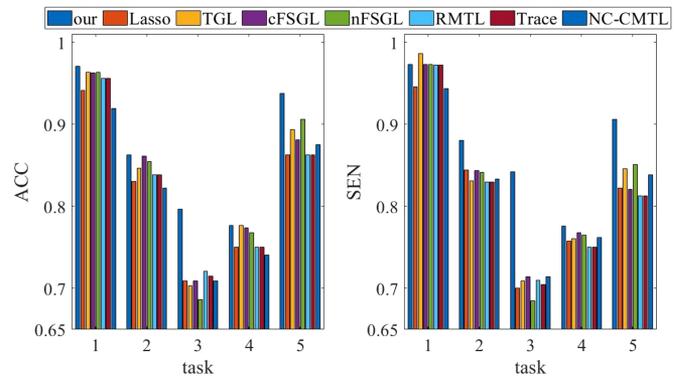


Fig. 6. ACC and SEN of different methods at M12.

LMCI’ ($p < 0.05$) task at M06 and ‘CN vs. EMCI’ ($p < 0.05$) task at M12. Detailed statistical analysis results are provided in Appendix C.

In addition, there are also some interesting phenomena in this experiment. In the simulation experiments, our method shows a significant advantage, but the classification results of different methods are similar. The possible reason for the discrepancy between these two results may be: (1) The missing situation is different. In the real situation, some cognitive scores’ missing rate is much higher than others, such as DSPAN For and DIGIT are missing in almost all patients, and only 7 CN subjects retain the values. (2) A large RMSE or CC may not translate to a large classification difference, which may be due to the choice of the classifier not being very well. To substantiate our hypothesis, we further conduct a series of experiments where we carefully consider the difference between cognitive scores for participants with different disease categories at different time points. The details are shown in Appendix D.

VI. CONCLUSION

In summary, this article focuses on studying the disease progression of AD by baseline imaging markers and high-order correlation in cognitive scores. Based on existing approaches, we propose a relation-aware tensor completion multitask learning formulation, in which the prediction of cognitive status can be modeled as a tensor completion problem. This approach not only allows us to explore both inter-mode and intra-mode correlations within cognitive score simultaneously but is also a natural choice for addressing the issue of missing target values. Furthermore, we also take the linear relationship between neuroimages and cognitive scores into account. This method gives an idea of how to integrate the regression and completion models in AD prediction. Specifically, in this framework, tensor completion explores the underlying high-order structural information of labels, while the regression model incorporates the close relationship between neuroimaging and cognitive scores. In the end, we validate the effectiveness of the proposed method on the ADNI dataset, compared with traditional models for AD prediction, our algorithm shows better performance in both

missing situations. We also validate our algorithm in a more reasonable situation and discuss corresponding results.

Although the above results have been shown, there are still some limitations in this study. We only use the mean and variance of FA and MD features in DTI data as input biomarkers, suggesting that more complex and accurate feature designs could further enhance the model performance. Additionally, this article only considers the DTI data as input features, and other biomarker types such as PET and CSF may help to extend knowledge on AD.

REFERENCES

- [1] W. W. Barker et al., "Relative frequencies of Alzheimer disease, lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the state of florida brain bank," *Alzheimer Dis. Assoc. Disord.*, vol. 16, no. 4, pp. 203–212, 2002.
- [2] Q. Lin et al., "Resting-state functional connectivity predicts cognitive impairment related to Alzheimer's disease," *Front. Aging Neurosci.*, vol. 10, 2018, Art. no. 94.
- [3] H. W. Querfurth and F. M. LaFerla, "Alzheimer's disease," *New England J. Med.*, vol. 362, no. 4, pp. 329–344, 2010.
- [4] Alzheimer's Association, "2016 Alzheimer's disease facts and figures," *Alzheimer's Dement.*, vol. 12, no. 4, pp. 459–509, 2016.
- [5] F. Márquez and M. A. Yassa, "Neuroimaging biomarkers for alzheimer's disease," *Mol. Neurodegener.*, vol. 14, pp. 1–14, 2019.
- [6] J. R. Petrella et al., "Neuroimaging and early diagnosis of alzheimer disease: A look to the future," *Radiol.*, vol. 226, no. 2, pp. 315–336, 2003.
- [7] C. Misra et al., "Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: Results from ADNI," *Neuroimage*, vol. 44, no. 4, pp. 1415–1422, 2009.
- [8] S. E. O'Bryant et al., "Staging dementia using clinical dementia rating scale sum of boxes scores: A Texas Alzheimer's research consortium study," *Arch. Neurol.*, vol. 65, no. 8, pp. 1091–1095, 2008.
- [9] H. Wang et al., "Sparse multi-task regression and feature selection to identify brain imaging predictors for memory performance," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 557–562.
- [10] T. N. Tombaugh and N. J. McIntyre, "The mini-mental state examination: A comprehensive review," *J. Amer. Geriatrics Soc.*, vol. 40, no. 9, pp. 922–935, 1992.
- [11] R. C. Mohs et al., "Development of cognitive instruments for use in clinical trials of antedementia drugs: Additions to the Alzheimer's disease assessment scale that broaden its scope," *Alzheimer Dis. Assoc. Disord.*, vol. 11, pp. 13–21, 1997.
- [12] M. Schmidt et al., *Rey Auditory Verbal Learning Test: A Handbook*, vol. 17, Los Angeles, CA, USA: Western Psychological Services, 1996.
- [13] C. R. Jack Jr. et al., "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Reson. Imag. Official J. Int. Soc. Magn. Reson. Med.*, vol. 27, no. 4, pp. 685–691, 2008.
- [14] B. Schmand et al., "Value of neuropsychological tests, neuroimaging, and biomarkers for diagnosing Alzheimer's disease in younger and older age cohorts," *J. Amer. Geriatrics Soc.*, vol. 59, no. 9, pp. 1705–1710, 2011.
- [15] C. Lian et al., "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 880–893, Apr. 2020.
- [16] C. D. Mayo et al., "Relationship between DTI metrics and cognitive function in Alzheimer's disease," *Front. Aging Neurosci.*, vol. 10, 2019, Art. no. 436.
- [17] K. Oishi et al., "DTI analyses and clinical applications in Alzheimer's disease," *J. Alzheimer's Dis.*, vol. 26, no. s3, pp. 287–296, 2011.
- [18] B. Lei et al., "Self-calibrated brain network estimation and joint non-convex multi-task learning for identification of early Alzheimer's disease," *Med. Image Anal.*, vol. 61, 2020, Art. no. 101652.
- [19] S. Tang et al., "Dual feature correlation guided multi-task learning for Alzheimer's disease prediction," *Comput. Biol. Med.*, vol. 140, 2022, Art. no. 105090.
- [20] Y. Fan, D. Kaufer, and D. Shen, "Joint estimation of multiple clinical variables of neurological diseases from imaging patterns," in *Proc. IEEE Int. Symp. Biomed. Imag. Nano Macro*, 2010, pp. 852–855.
- [21] H. Wang et al., "Sparse multi-task regression and feature selection to identify brain imaging predictors for memory performance," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 557–562.
- [22] D. Zhang et al., "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease," *NeuroImage*, vol. 59, no. 2, pp. 895–907, 2012.
- [23] J. Zhou et al., "A multi-task learning formulation for predicting disease progression," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2011, pp. 814–822.
- [24] J. Zhou et al., "Modeling disease progression via multi-task learning," *NeuroImage*, vol. 78, pp. 233–248, 2013.
- [25] J. Zhou et al., "Modeling disease progression via fused sparse group lasso," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2012, pp. 1095–1103.
- [26] P. Cao et al., "Sparse shared structure based multi-task learning for MRI based cognitive performance prediction of Alzheimer's disease," *Pattern Recognit.*, vol. 72, pp. 219–235, 2017.
- [27] B. Romera-Paredes et al., "Multilinear multitask learning," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1444–1452.
- [28] T. Maggipinto et al., "DTI measurements for alzheimer's classification," *Phys. Med. Biol.*, vol. 62, no. 6, pp. 2361–2375, 2017.
- [29] A. Kumar et al., "Diffusion tensor imaging based white matter changes and antioxidant enzymes status for early identification of mild cognitive impairment," *Int. J. Neurosci.*, vol. 129, no. 3, pp. 209–216, 2019.
- [30] J. L. Fu et al., "The value of diffusion tensor imaging in the differential diagnosis of subcortical ischemic vascular dementia and Alzheimer's disease in patients with only mild white matter alterations on T2-weighted images," *Acta Radiologica*, vol. 53, no. 3, pp. 312–317, 2012.
- [31] J. Liu et al., "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.
- [32] J. F. Cai et al., "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [33] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 12, pp. 5586–5609, Dec. 2022.
- [34] J. Chen et al., "Integrating low-rank and group-sparse structures for robust multi-task learning," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2011, pp. 42–50.
- [35] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.
- [36] H. Wang et al., "High-order multi-task feature learning to identify longitudinal phenotypic markers for Alzheimer's disease progression prediction," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1277–1285.
- [37] F. Nie et al., "Calibrated multi-task learning," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2018, pp. 2012–2021.
- [38] P. Jiang et al., "Correlation-aware sparse and low-rank constrained multi-task learning for longitudinal analysis of Alzheimer's disease," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1450–1456, Jul. 2019.
- [39] Y. Liu, Z. Long, and C. Zhu, "Image completion using low tensor tree rank and total variation minimization," *IEEE Trans. Multimedia*, vol. 21, no. 2, pp. 338–350, Feb. 2019.
- [40] X. Chen et al., "Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 12301–12310, Aug. 2022.
- [41] K. Wimalawarne et al., "Multitask learning meets tensor factorization: Task imputation via convex optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2825–2833.
- [42] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
- [43] L. Yuan et al., "Rank minimization on tensor ring: An efficient approach for tensor decomposition and completion," *Mach. Learn.*, vol. 109, pp. 603–622, 2020.
- [44] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, vol. 87. Berlin, Germany: Springer Science & Business Media, 2003.
- [45] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 49–95, 1996.
- [46] S. Boyd et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
- [47] W. S. Tae et al., "Current clinical applications of diffusion-tensor imaging in neurological disorders," *J. Clin. Neurol.*, vol. 14, no. 2, pp. 129–140, 2018.

- [48] M. Jenkinson et al., "FSL," *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.
- [49] P. J. Basser et al., "MR diffusion tensor spectroscopy and imaging," *Biophysical J.*, vol. 66, no. 1, pp. 259–267, 1994.
- [50] D. Le Bihan et al., "Diffusion tensor imaging: Concepts and applications," *J. Magn. Reson. Imag. Official J. Int. Soc. Magn. Reson. Med.*, vol. 13, no. 4, pp. 534–546, 2001.
- [51] S. M. Smith et al., "Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data," *Neuroimage*, vol. 31, no. 4, pp. 1487–1505, 2006.
- [52] N. Tzourio-Mazoyer et al., "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.
- [53] A. Argyriou et al., "Convex multi-task feature learning," *Mach. Learn.*, vol. 73, pp. 243–272, 2008.
- [54] C. M. Stonnington et al., "Predicting clinical scores from magnetic resonance scans in Alzheimer's disease," *Neuroimage*, vol. 51, no. 4, pp. 1405–1413, 2010.
- [55] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc.: Ser. B (Methodol.)*, vol. 58, no. 1, pp. 267–288, 1996.
- [56] S. Ji and J. Ye, "An accelerated gradient method for trace norm minimization," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 457–464.
- [57] S. Tabarestani et al., "A distributed multitask multimodal approach for the prediction of Alzheimer's disease in a longitudinal study," *NeuroImage*, vol. 206, 2020, Art. no. 116317.
- [58] P. Burman, "A comparative study of ordinary cross-validation, v -fold cross-validation and the repeated learning-testing methods," *Biometrika*, vol. 76, no. 3, pp. 503–514, 1989.
- [59] J. Shao, "Linear model selection by cross-validation," *J. Amer. Stat. Assoc.*, vol. 88, no. 422, pp. 486–494, 1993.
- [60] P. Smyth, "Clustering using Monte Carlo cross-validation," in *Proc. 2nd Int. Conf. Knowl. Discov. Data Mining*, 1996, pp. 26–133.
- [61] E. W. Weisstein, "Bonferroni correction," 2004. [Online]. Available: <https://mathworld.wolfram.com/>
- [62] V. N. Vapnik, "Pattern recognition using generalized portrait method," *Automat. Remote Control*, vol. 24, no. 6, pp. 774–780, 1963.
- [63] A. Vabalas et al., "Machine learning algorithm validation with a limited sample size," *PLoS One*, vol. 14, no. 11, 2019, Art. no. e0224365.