
Proceedings Paper:
Fan, X., Zhou, M., Zhang, Y. et al. (3 more authors) (2025) Adaptive multi-cognitive objective temporal task approach for predicting AD progression. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE BIBM, 03-06 Dec 2024, Lisbon, Portugal. Institute of Electrical and Electronics Engineers (IEEE) , pp. 1933-1938. ISBN 979-8-3503-8623-3

https://doi.org/10.1109/bibm62325.2024.10822758

Adaptive Multi-Cognitive Objective Temporal Task Approach for Predicting AD Progression

Xuanhan Fan
School of Software
Yunnan University
Kunming, China
fanxuanhan@gmail.com

Menghui Zhou*
Department of Computer Science
The University of Sheffield
Sheffield, UK
menghui.zhou@sheffield.ac.uk

Yu Zhang
Department of Computer Science
The University of Sheffield
Sheffield, UK
yzhang489@sheffield.ac.uk

Jun Qi
Department of Computing
Xi'an JiaoTong-Liverpool University
Suzhou, China
Jun.Qi@xjtlu.edu.cn

Yun Yang
School of Software
Yunnan University
Kunming, China
yangyun@ynu.edu.cn

Po Yang*
Department of Computer Science
The University of Sheffield
Sheffield, UK
po.yang@sheffield.ac.uk

***Abstract*—As the population rapidly ages, Alzheimer's disease (AD), the most common form of dementia, urgently requires the identification of reliable structural brain biomarkers and the development of effective therapeutic strategies. Multiple multi-task learning (MTL) paradigms have been developed to enhance model generalization by sharing information between tasks to predict AD progression and accurately identify MRI-associated biomarkers. Unlike previous MTL approaches that consider only a single kind of cognitive score to predict the complicated AD progression over time, we have developed an innovative MTL method to deal with various cognitive scores simultaneously, with each focusing on different aspects of patient cognition. To effectively capture the intricate associations among different cognitive scores at multiple time points, we first propose an Adaptive Multiple Cognitive Objective Temporal (AMCOT) task-relationship binding penalty mechanism. This mechanism adaptively reveals temporal correlations between various cognitive scores at different time points and uses these relationships to predict cumulative disease progression accurately. To select the most informative MRI features in AD progression, we consider integrating the sparse group Lasso into our model. Our algorithms are designed to handle large datasets efficiently. Empirical evaluation on the Alzheimer's disease dataset shows that our approach significantly outperforms existing state-of-the-art algorithms in both overall and individual task performance. Additionally, we applied stability selection techniques to identify stable MRI biomarkers and analyzed their temporal patterns to gain insights into AD progression. The implementation source can be found at https://github.com/XuanhanFan/MTL-AMCOT-BB.**

***Index Terms*—Alzheimer's Disease, disease progression, cognitive score,multi-task learning, Adaptive Multi-Cognitive Objective Temporal Task, identification of biomarkers**

I. Introduction

Alzheimer's Disease (AD) [1] is a widespread neurodegenerative disorder marked by severe dementia. The International World Alzheimer's Disease Report [2] states that 75% of dementia cases remain undiagnosed. With an aging global population, the number of individuals with dementia could rise to 152.8 million by 2050. Early intervention is vital as AD is irreversible and currently incurable, yet definitive diagnosis often depends on invasive procedures like brain biopsies or autopsies, hindering early diagnosis and treatment. Research into AD biomarkers and developing predictive models is essential. Various cognitive scoring tools, such as the Clinical Dementia Rating Scale-Sum of the Boxes (CDR-SB) [3], the Mini-Mental State Examination (MMSE) [4], the Alzheimer's Disease Assessment Scale-cognition sub-scale (ADAS-Cog) [5], and the Rey Auditory Verbal Learning Test (RAVLT) [6], are crucial for assessing cognitive function and daily activities. These instruments, combined with AD-related MRI biomarkers [1], play a critical role in early diagnosis and progression prediction.

Advanced machine learning techniques are increasingly applied in modelling MRI features and cognitive scores to enhance AD research through disease classification, survival analysis, and regression modelling. Traditional single-task regression methods often neglect correlations across multiple future cognitive assessments despite AD's consistent pathology over time. Multi-task learning (MTL) [12] addresses this by analyzing multiple cognitive assessments across different time points, capturing intrinsic correlations especially in cases where feature numbers exceed sample sizes, thereby boosting model performance. While significant strides have been made in using MTL for predicting disease progression, challenges persist in accurately identifying and modelling these task correlations effectively.

Physicians utilize cognitive tests and MRI scans to evaluate patients' cognitive status and tailor treatment plans, especially for AD. Acknowledging the need for comprehensive assessment, we introduced a novel MTL model that harnesses correlations across multiple cognitive states over time. Our model, which uses an adaptive multi-target temporal task matrix approach as depicted in Fig.1, merges various cognitive and temporal aspects to enhance prediction accuracy. This

Corresponding author: Menghui Zhou, Po Yang

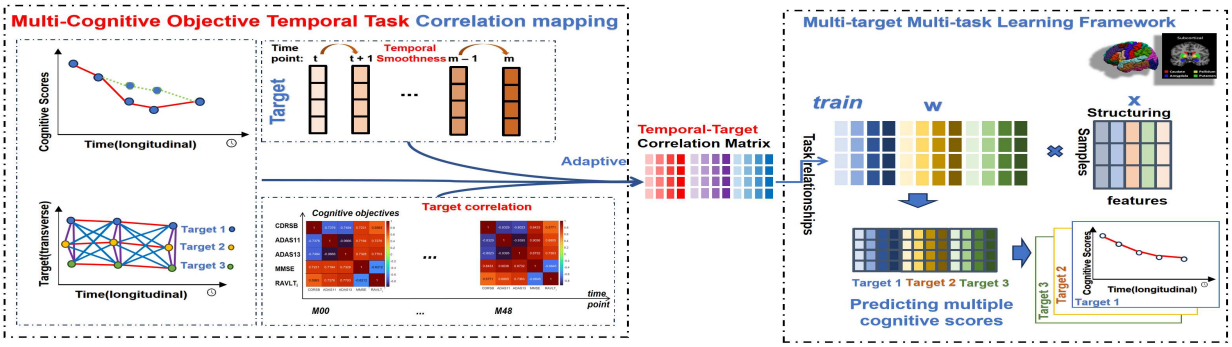

Fig. 1. The AMCOT method flowchart integrates measurements of target scores and temporal smoothness. During the temporal-target mapping phase, we merge temporal smoothness and correlations among various targets at different times to create a temporal-target correlation matrix. The mapping of target correlations is depicted in Fig 2.

approach aligns cognitive scores with disease progression at multiple points, improving both the predictive accuracy and interpretability of our findings.

We summarize our contributions as follows:

- **Adaptive MTL Framework**: The MTL-AMCOT model we developed adaptively integrates multiple cognitive scores and temporal data, accurately reflecting AD progression and outperforming multiple baseline methods i in accuracy and robustness.
- **Optimization Enhancements**: We enhanced the Accelerated Proximal Gradient Method (APM) [9] by incorporating the Barzilai-Borwein (BB) step [20] and line search to address high iteration numbers and slow convergence in large-scale AD data, boosting computational efficiency.
- **Interpretable and Effective Methodology**: Unlike existing deep learning approaches with limited interpretability, our method provides clear insights and effectively uses multi-target cognitive scores to identify stable MRI biomarkers over time, offering significant clinical value for AD biomarker identification.

II. RELATED WORK

Zhou J et al. [16] introduced the TGL model, an MTL approach for longitudinal disease analysis using a common feature set across all time points, but not accounting for biomarker variability. The cFSGL method [23], utilizing a sparse group Lasso penalty [19], refines this by enabling task-specific feature selection and maintaining temporal smoothness, indicating minor score changes over time [16] [23]. Romeo L et al. [24] developed a spatio-temporal MTL approach with a graph-based framework for diabetes complications, focusing on local temporal dependencies. The Longitudinal Stability Adjustment (LSA) [25] addresses global temporal correlations by integrating long-term progression data. Zhou M et al. [26] and Liu X et al. [27] presented AutoTR and an MTL model using the Laplacian sparse group Lasso, respectively, emphasizing temporal dynamics without considering multi-cognitive correlations. Conversely, Liang W et al. [28] proposed BGP-MTFL for exploring multi-task relationships in cognitive scoring, excluding progression and temporal dynamics.

Deep neural networks, especially RNNs [17] and LSTMs, effectively process time-series data by capturing long-term dependencies. LSTMs manage complex multivariate patterns through gating mechanisms. Yet, traditional missing value techniques like zero or mean padding can impede performance. Adaptive imputation models, such as Nguyen et al. [10] forward-filling with MinimalRNN and Liang W et al. [11] end-to-end deep MTL framework for progression forecasting, are addressing these limitations. Despite advancements, the non-interpretability of deep neural networks [15] remains a significant barrier in healthcare. In contrast, traditional MTL methods provide better interpretability by grouping MRI features by brain regions, enhancing feature correlation integration and ROI clarity, and aiding feature selection through various penalties [21] [33] [34] [18].

III. METHODS

A. Multi-task Learning

In the temporal domain, each cognitive measure score at a specific time is considered a separate task, creating an MTL problem with t tasks. For each task $i \in \{1, \ldots, t\}$, there is a set of samples (X_i^o, y_i^o) where $X_i^o \in \mathbb{R}^{n_i \times p}$ represents the input data and $y_i^o \in \mathbb{R}^{n_i}$ the output data for cognitive goals o. We denote all input data across t tasks as $X^o = [X_1^o, \ldots, X_t^o]$ and all output data as $Y^o = [y_1^o, \ldots, y_t^o]$. The regression parameter matrices for these tasks are represented by $W^O = [w_1^j, \ldots, w_t^O] \in \mathbb{R}^{p \times (t \times O)}$. Each row in X reflects all characteristics of a patient for the i-th task, while each column relates to a specific MRI biomarker at baseline. The model uses a squared loss function defined as $L(Y, X, W) = \frac{1}{2} \sum_{i=1}^{t} \sum_{j=1}^{O} \|X_i^j W_i^j - Y_i^j\|_2^2$.

B. Temporal Domain Task Relationships

In the temporal domain, inter-task relationships correlate AD progression with time. Progression i is defined as $\delta y_i = y_i - y_{i+1}$, where y_i^o is the cognitive score at time i for target o in set O. Due to the slow, subtle initial progression of AD, δy_i is typically small, indicating temporal smoothness. We assume minimal differences between consecutive cognitive

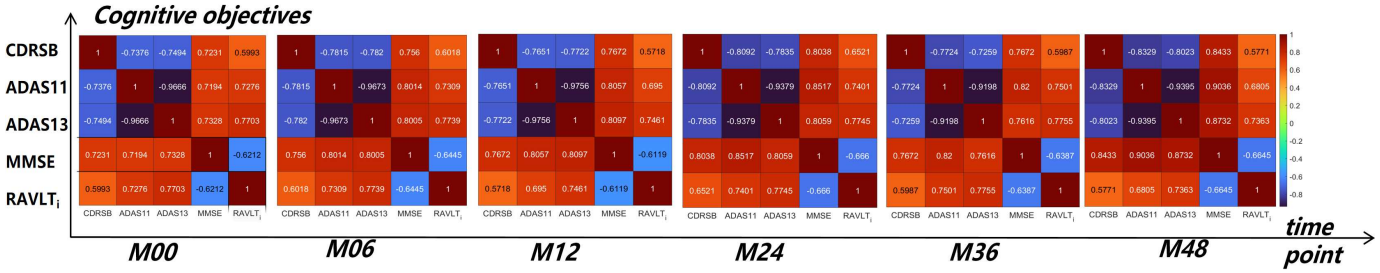

Fig. 2. Variation in correlational information between scores on different cognitive objectives at different time points.

scores, correlating these with model variations. The penalty term for temporal task regularization is defined as follows:

$$|y_{i+1}^o - y_i^o| = |Xw_{i+1}^o - Xw_i^o| = |X\left(w_{i+1}^o - w_i^o\right)| \quad (1)$$

C. Target Score Correlation

To predict AD progression, we use a set of cognitive target scores $\mathbf{O}$ to model different target tasks concurrently. We normalized cognitive scores to z-scores to address scale disparities, unifying the data. Using the Pearson correlation coefficient, we developed a correlation matrix to examine the relationships among cognitive score tasks, uncovering notable temporal variations in correlations, depicted in Fig. 2. This analysis prompted the creation of an adaptive temporal correlation matrix that dynamically captures the evolving interrelations among cognitive scores over time.

$$c_\tau^{i,j} = \frac{cov(y_\tau^i, y_\tau^j)}{\sigma_{y_\tau^i}\sigma_{y_\tau^j}} \quad (2)$$

where $i, j \in O$. y_n^i represents the cognitive scores for the n samples of task i, the scores for other tasks are approximated by the correlation coefficients $c(i,j)$ as shown by: $y_n^i \approx c(i,j) \cdot y_n^j$. Based on these correlation coefficients, we construct multiple target correlation matrices at time τ, denoted as M_τ.

$$M_\tau = \begin{bmatrix} c_\tau^{1,1} & -c_\tau^{1,2} & \cdots & -c_\tau^{1,o} \\ -c_\tau^{2,1} & c_\tau^{2,2} & \cdots & -c_\tau^{2,o} \\ \vdots & \vdots & \ddots & \vdots \\ -c_\tau^{o,1} & -c_\tau^{o,2} & \cdots & c_\tau^{o,o} \end{bmatrix} \quad (3)$$

D. Temporal-Target correlation and AMCOT Penalty

We created a task relationship matrix R using temporal smoothness and target association matrices, as detailed in Algorithm 1. Through the target time correlation matrix, we can more effectively monitor and evaluate the condition changes and cognitive decline rate of AD patients. Therefore, we have the following AMCOT penalty $\|\Re\|$:

$$\begin{aligned} \|\Re\| &= \|WR\| \\ &= |\sum_{\tau=1}^{T}(W^i - M_\tau^{(i,j)} \cdot W^j) + \sum_{\tau=1}^{T-1}(W_{\tau+1} - W_\tau)| \end{aligned} \quad (4)$$

E. One Novel MTL Method

The L_1-norm promotes sparsity in the coefficient matrix. Extending this, the $L_{2,1}$-norm, defined as $\|W\|_{2,1} = \sum_j \sqrt{\sum_i w_{i,j}^2}$, where $w_{i,j}$ represents the matrix element at row i and column j, aids in selecting common biomarkers across all tasks. We introduce a novel MTL framework with AMCOT penalty (MTL-AMCOT) defined as follows:

$$\begin{aligned} \min_W \frac{1}{2}\sum_{i=1}^{t}\sum_{j=1}^{O}\|X_i^j W_i^j - Y_i^j\|_2^2 + \lambda_1 \sum_{j=1}^{O}\|W^j\|_1 \\ + \lambda_2 \sum_{j=1}^{O}\|W^j\|_{2,1} + \lambda_3\|\Re\|_1. \\ \text{s.t.} \Re = WR. \end{aligned} \quad (5)$$

Where $\lambda_1, \lambda_2, \lambda_3$ are fine-tuned parameters.

IV. OPTIMIZATION ALGORITHMS

In this section, we develop an optimization algorithm that improves upon the APM approach using a two-point step gradient method to solve the objective function of MTL-AMCOT.

A. The APM-Based Algorithm

We use the APM. Due to its fast convergence as a class of first-order methods [13], APM has been widely used to solve MTL problems. Its form is as follows:

$$\min_W \mathcal{L}(W) = f(W) + g(W), \quad (6)$$

Where $f(W)$ is a smooth convex function and $g(W)$ is a non-smooth convex function. APM is built on two sequences, the search point $\{S_k\}$ and the approximation point $\{W_k\}$. S_k is a linear combination of W_{k-1} and W_k. Since $g(W)$ is a non-smooth convex function, we need to solve for the approximation point. The approximation point W_i is given by:

$$S_{k+1}^o = W_k^o + \alpha_k(W_k^o - W_{k-1}^o) \quad (7)$$

$$W_{k+1}^o = \pi(S_k^o - \eta_k \nabla f(S_k^o)) \quad (8)$$

where α_k is the momentum factor and $\pi(V)$ is the proximal operator of V, with η_k representing the step size. Updating η_k in the APM depends on specific conditions to ensure

feasible domain convergence. The appropriate η_k must satisfy the following inequality:

$$f(W_k) \leq \sum_{j=1}^{O} f_\eta(W_k^j, S_k^j) \tag{9}$$
$$= f(S_k) + \langle \nabla f(S_k), W_k - S_k \rangle + \frac{1}{2\eta_k} \|W_k - S_k\|_F^2.$$

However, the update process can elevate computational costs. A key step in our APM-based [22] algorithm involves calculating the proximal operator that integrates three non-smooth penalty terms, reformulated as follows:

$$\pi(V^j) = \arg\min_W \sum_{j=1}^{O} \frac{1}{2} \|W^j - V^j\|_F^2 + \lambda_1 \sum_{j=1}^{O} \|W^j\|_1$$
$$+ \lambda_2 \sum_{j=1}^{O} \|W^j\|_{2,1} + \lambda_3 \|\Re^T\|_1. \tag{10}$$

where $V^j = S_i^j - \frac{1}{\eta_k} f'(S_i^j)$. In Eq. 10, each row of W is obviously independent.

Incorporating the BB step size into the APM, we dynamically adjust the step size η_k based on changes in the gradient. Define $s_{k-1} = W_k - W_{k-1}$ as the difference between solutions from consecutive iterations, and $g_{k-1} = f'(S_k) - f'(S_{k-1})$ as the gradient difference. To optimize η_k, it is calculated as follows to minimize the norm $\|\Delta s - \eta_k \Delta g\|^2$:

$$\eta_k = \frac{\langle s_{k-1}, g_{k-1} \rangle}{\langle g_{k-1}, g_{k-1} \rangle} \tag{11}$$

Alternatively, to minimize $\|\eta_k \Delta s - \Delta g\|^2$:

$$\eta_k = \frac{\langle s_{k-1}, s_{k-1} \rangle}{\langle s_{k-1}, g_{k-1} \rangle} \tag{12}$$

To prevent abnormal extremes in step size due to erratic gradient changes, η_k is constrained within specified bounds:

$$\eta_{\text{BB}} = \min(\max(\eta_k, \eta_{\min}), \eta_{\max}) \tag{13}$$

where $\eta_{\min}$ and $\eta_{\max}$ are the predetermined limits for the step size, ensuring stability and efficiency in convergence.

V. EXPERIMENTAL RESULT

Our study utilized the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [14], collecting longitudinal measurements semi-annually or annually up to 48 months. Focusing on early disease stages, we excluded cases lacking initial MRI records, removed features failing quality control, and imputed missing values using mean substitution, resulting in a dataset of 314 features across six time points. We employ 90% of the data for training and utilize 10-fold cross-validation to select the regularization parameter. Model performance is evaluated using root mean squared error (rMSE) for task-specific regression, normalized mean squared error (nMSE) for overall performance, and weighted r-value (wR) [7] to seek lower rMSE and nMSE and higher wR values.

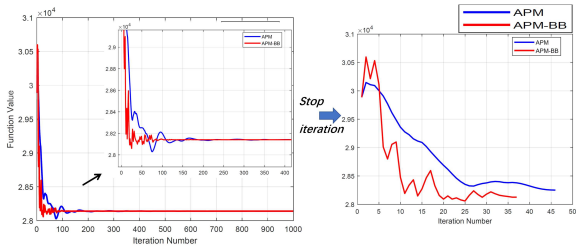

Fig. 3. The convergence situation of using BB step size based on APM algorithms.

A. Comparison of Accelerated vs. Non-Accelerated Algorithms and Model Analysis

We assessed our algorithm on the ADNI dataset in MATLAB, comparing performance under parameters $\lambda_1 = 0.1$, $\lambda_2 = 1000$, $\lambda_3 = 500$, and with or without the BB step size in APM. Figure 3 illustrates that our method hastens the initial loss reduction and attains closer convergence in the final stages. We extensively compared our proposed method with various leading models: RMTL [8], TGL [16], cFSGL [23], FL_SGL [27], LSA [25], AutoTR [26], MinimalRNN [10], and LSTM [10], across scenarios from baseline (M00) to M48, detailed in Table I.

Our proposed AMCOT method has shown remarkable performance in predicting cognitive scores. It achieved the lowest nMSE (0.515 ± 0.018) and the highest wR (0.701 ± 0.010) for CDRSB scores, significantly outperforming other comparative methods. AMCOT also excelled in predicting ADAS11 scores, with the lowest nMSE (0.468 ± 0.036) and the highest wR (0.733 ± 0.023), proving its superiority over Lasso, Ridge, and other MTL models. For ADAS13 scores, AMCOT achieved the best nMSE (0.428 ± 0.008) and the highest wR (0.758 ± 0.006). In predicting MMSE scores, AMCOT obtained the best nMSE (0.524 ± 0.023) and the highest wR (0.695 ± 0.016), demonstrating its excellent predictive ability and robustness with complex data.

Figure 4 shows single-task rMSEs for various models over time. LSTM and MinimalRNN, though powerful, lack interpretability and rely heavily on initial cognitive scores, limiting long-term predictions. Their accuracy declines due to incomplete datasets and a focus on temporal dynamics over feature correlations. In contrast, MTL leverages inter-task correlations and handles incomplete data better. AutoTR and LSA perform poorly due to their restriction to six-time points and LSA's suboptimal task relevance setting ($\alpha = 0.2$). RMTL's inconsistent performance suggests outlier tasks affect robustness. AMCOT outperforms other MTL methods with lower nMSE and higher wR, effectively using multiple cognitive scores and temporal tasks. It dynamically adjusts weights to minimize irrelevant information.

B. Multiple Targets Stabilize Temporal Patterns of Biomarkers

The MTL-AMCOT model integrates multiple cognitive scores to identify stable biomarker patterns over time, employing longitudinal stability selection [23]. We identified 25 stable

TABLE I
COMPARISON OF LONGITUDINAL COGNITIVE SCORES PREDICTING MTL METHODS

SCORE	Metric	TGL	cFSGL	AutoTR	RMTL	LSA	FL-SGL	MinimalRNN	LSTM	AMCOT (Ours)
CDRSB	nMSE	0.546 ± 0.026	0.535 ± 0.027	0.683 ± 0.017	0.672 ± 0.013	0.798 ± 0.044	0.579 ± 0.020	0.957 ± 0.086	0.956 ± 0.245	**0.515 ± 0.018**⋆
	wR	0.681 ± 0.018	0.688 ± 0.019	0.548 ± 0.019	0.553 ± 0.017	0.523 ± 0.019	0.656 ± 0.016	0.215 ± 0.133	0.226 ± 0.122	**0.701 ± 0.010**⋆
ADAS11	nMSE	0.486 ± 0.021	0.477 ± 0.021	0.516 ± 0.026	0.526 ± 0.032	0.607 ± 0.039	0.523 ± 0.019	0.482 ± 0.110	0.493 ± 0.139	**0.468 ± 0.036**⋆
	wR	0.724 ± 0.012	0.729 ± 0.012	0.708 ± 0.015	0.695 ± 0.019	0.655 ± 0.021	0.695 ± 0.014	0.718 ± 0.084	0.716 ± 0.183	**0.733 ± 0.023**⋆
ADAS13	nMSE	0.461 ± 0.023	0.453 ± 0.022	0.498 ± 0.013	0.505 ± 0.008	0.595 ± 0.018	0.492 ± 0.014	0.484 ± 0.092	0.490 ± 0.131	**0.428 ± 0.008**⋆
	wR	0.737 ± 0.014	0.742 ± 0.014	0.720 ± 0.005	0.707 ± 0.005	0.664 ± 0.006	0.715 ± 0.010	0.722 ± 0.066	0.720 ± 0.183	**0.758 ± 0.006**⋆
MMSE	nMSE	0.551 ± 0.023	0.540 ± 0.025	0.594 ± 0.029	0.596 ± 0.027	0.692 ± 0.033	0.600 ± 0.014	0.740 ± 0.126	0.708 ± 0.202	**0.524 ± 0.023**⋆
	wR	0.676 ± 0.014	0.685 ± 0.015	0.659 ± 0.014	0.642 ± 0.021	0.600 ± 0.017	0.640 ± 0.010	0.508 ± 0.138	0.510 ± 0.158	**0.695 ± 0.016**⋆
RAVLT_i	nMSE	0.600 ± 0.020	0.581 ± 0.017	0.615 ± 0.022	0.611 ± 0.021	0.715 ± 0.021	0.638 ± 0.014	0.616 ± 0.116	0.623 ± 0.176	**0.576 ± 0.038**⋆
	wR	0.637 ± 0.014	0.650 ± 0.012	0.634 ± 0.014	0.625 ± 0.018	0.568 ± 0.013	0.604 ± 0.011	0.669 ± 0.074	**0.676 ± 0.175**	0.653 ± 0.027⋆

***Note**: bold font is used to mark the best result for the average of each indicator, while a star indicates that the indicator has a better generalization in the case of a better average.*

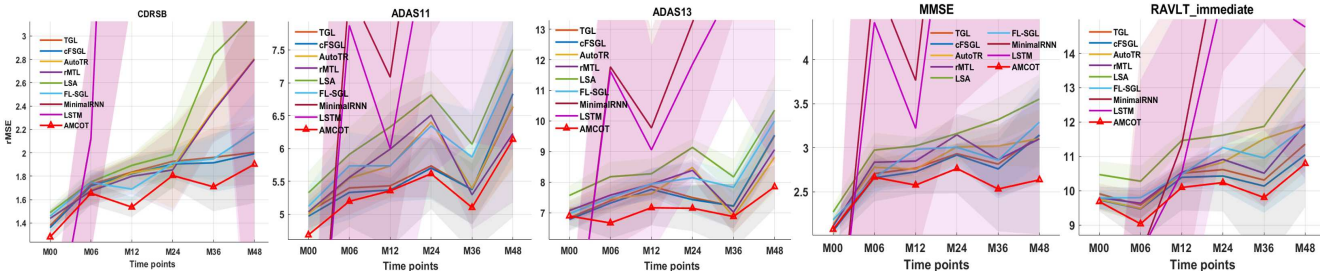

Fig. 4. The average root Mean Squared Error (rMSE) compared to the baseline model is presented. Our method (AMCOT) involves multiple target cognitive scores. Therefore, when comparing single-target score tasks, we focus on optimizing the single-target scores.

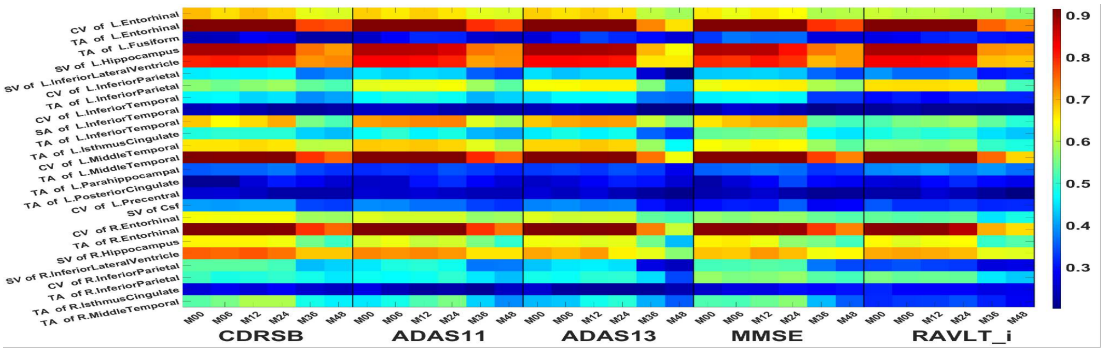

Fig. 5. The stability vectors of the 25 stable MRI features generated by AMCOT. Longitudinal stability selection was performed by combining ten target scores. The larger the value, the more stable the feature.

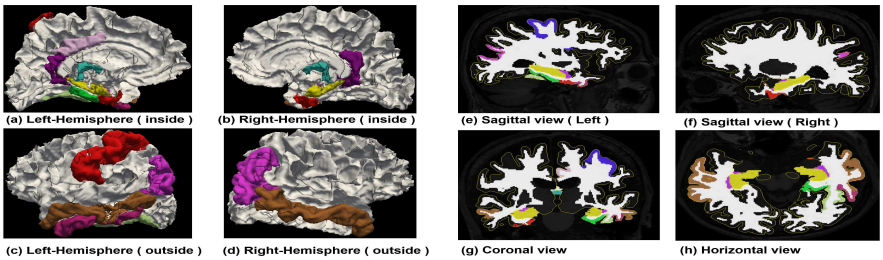

Fig. 6. AMCOT uses stability selection to choose the brain maps with the highest ROIs. (a) - (d) are the selected cortical ROIs, and (e) - (g) are the selected subcortical ROIs.

biomarkers across all targets and time points, particularly in the entorhinal cortex, hippocampus, inferior lateral ventricle, and middle temporal gyrus (see Fig. 5). Notable features include the volume and thickness of the left hippocampus and bilateral thickness of the entorhinal cortex. The left hippocampus volume [31], linked to early AD memory loss, and dysfunction in the entorhinal and inferotemporal cortices, are correlated with memory and cognitive decline. Enlarged in-

ferior lateral ventricles serve as sensitive markers for MCI and AD progression. Additionally, research on cerebrospinal fluid (CSF), including findings that young CSF [30] may restore memory in aged mice, highlights potential therapeutic benefits. These biomarkers are validated across multiple studies [29]-[31]. Figure 6 illustrates stable MRI features identified by the AMCOT model from various anatomical perspectives, including medial and lateral views of both hemispheres, and sagittal, coronal, and horizontal orientations. The colored areas denote ROIs with high stability confirmed through longitudinal analyses across multiple target scores.

VI. CONCLUSION

In this paper, we explore AD progression prediction using baseline MRI features and cognitive scores at five future time points, introducing the MTL-AMCOT model designed to capture associations in a Multi-Cognitive Objective-Temporal Domain matrix. This model, incorporating sparse group Lasso, outperformed several baselines across cognitive targets in the AD dataset. To address non-smooth and biconvex functions, we employed an efficient APM optimization with the Barzilai-Borwein step size, reducing iterations and aiding hyperparameter tuning. In the future, we will explore applying AMCOT to other neurodegenerative diseases, such as Parkinson's disease, and non-neurological conditions, such as depression, to validate its broad applicability and effectiveness.

ACKNOWLEDGMENT

This research was supported by the National Natural Science Foundation of China (No.62301452) and the National Natural Science Foundation of China (No. 62061050).

REFERENCES

[1] Scheltens P, De Strooper B, Kivipelto M, et al. Alzheimer's disease[J]. The Lancet, 2021, 397(10284): 1577-1590.

[2] Gauthier S, Webster C, Servaes S, et al. World Alzheimer Report 2022: Life after diagnosis: Navigating treatment, care and support[J]. London, England: Alzheimer's Disease International, 2022: 25-26.

[3] Hughes C P, Berg L, Danziger W, et al. A new clinical scale for the staging of dementia[J]. The British journal of psychiatry, 1982, 140(6): 566-572.

[4] Folstein M F, Robins L N, Helzer J E. The mini-mental state examination[J]. Archives of general psychiatry, 1983, 40(7): 812-812.

[5] Rosen W G, Mohs R C, Davis K L. A new rating scale for Alzheimer's disease[J]. The American journal of psychiatry, 1984, 141(11): 1356-1364.

[6] Schmidt M. Rey auditory verbal learning test[M]. Los Angeles: Western Psychological Services, 1996.

[7] Stonnington C M, Chu C, Klöppel S, et al. Predicting clinical scores from magnetic resonance scans in Alzheimer's disease[J]. Neuroimage, 2010, 51(4): 1405-1413.

[8] Gong P, Ye J, Zhang C. Robust multi-task feature learning[C]//Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. 2012: 895-903.

[9] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. SIAM journal on imaging sciences, 2009, 2(1): 183-202.

[10] Nguyen M, He T, An L, et al. Predicting Alzheimer's disease progression using deep recurrent neural networks[J]. NeuroImage, 2020, 222: 117203.

[11] Liang W, Zhang K, Cao P, et al. Rethinking modeling Alzheimer's disease progression from a multi-task learning perspective with deep recurrent neural network[J]. Computers in Biology and Medicine, 2021, 138: 104935.

[12] Zhang Y, Yang Q. A survey on multi-task learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 34(12): 5586-5609.

[13] Wright J, Ma Y. High-dimensional data analysis with low-dimensional models: Principles, computation, and applications[M]. Cambridge University Press, 2022.

[14] Jack Jr C R, Bernstein M A, Fox N C, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods[J]. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2008, 27(4): 685-691.

[15] Aya Abdelsalam Ismail, Mohamed Gunady, Hector Corrada Bravo,and Soheil Feizi. Benchmarking deep learning interpretability in time series predictions. Advances in neural information processing systems, 33:6441–6452, 2020.

[16] Zhou J, Yuan L, Liu J, et al. A multi-task learning formulation for predicting disease progression[C]//Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. 2011: 814-822.

[17] Ghazi M M, Nielsen M, Pai A, et al. Training recurrent neural networks robust to incomplete data: Application to Alzheimer's disease progression modeling[J]. Medical image analysis, 2019, 53: 39-46.

[18] Zhou M, Wang X, Liu T, et al. Integrating Visualised Automatic Temporal Relation Graph into Multi-Task Learning for Alzheimer's Disease Progression Prediction[J]. IEEE Transactions on Knowledge and Data Engineering, 2024.

[19] Simon N, Friedman J, Hastie T, et al. A sparse-group lasso[J]. Journal of computational and graphical statistics, 2013, 22(2): 231-245.

[20] Barzilai J, Borwein J M. Two-point step size gradient methods[J]. IMA journal of numerical analysis, 1988, 8(1): 141-148.

[21] Cao P, Shan X, Zhao D, et al. Sparse shared structure based multi-task learning for MRI based cognitive performance prediction of Alzheimer's disease[J]. Pattern Recognition, 2017, 72: 219-235.

[22] Li H, Lin Z. Accelerated proximal gradient methods for nonconvex programming[J]. Advances in neural information processing systems, 2015, 28.

[23] Zhou J, Liu J, Narayan V A, et al. Modeling disease progression via multi-task learning[J]. NeuroImage, 2013, 78: 233-248.

[24] Romeo L, Armentano G, Nicolucci A, et al. A Novel Spatio-Temporal Multi-Task Approach for the Prediction of Diabetes-Related Complication: a Cardiopathy Case of Study[C]//IJCAI. 2020: 4299-4305.

[25] Zhou M, Zhang Y, Liu T, et al. Multi-task Learning with Adaptive Global Temporal Structure for Predicting Alzheimer's Disease Progression[C] Proceedings of the 31st ACM International Conference on Information & Knowledge Management. 2022: 2743-2752.

[26] Zhou M, Yang P. Automatic temporal relation in multi-task learning[C]/Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023: 3570-3580.

[27] Liu X, Cao P, Gonçalves A R, et al. Modeling alzheimer's disease progression with fused laplacian sparse group lasso[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2018, 12(6): 1-35.

[28] Liang W, Zhang K, Cao P, et al. Exploiting task relationships for Alzheimer's disease cognitive score prediction via multi-task learning[J]. Computers in Biology and Medicine, 2023, 152: 106367.

[29] Desikan R S, Fischl B, Cabral H J, et al. MRI measures of temporoparietal regions show differential rates of atrophy during prodromal AD[J]. Neurology, 2008, 71(11): 819-825.

[30] Iram T, Kern F, Kaur A, et al. Young CSF restores oligodendrogenesis and memory in aged mice via Fgf17[J]. Nature, 2022, 605(7910): 509-515.

[31] Mu Y, Gage F H. Adult hippocampal neurogenesis and its role in Alzheimer's disease[J]. Molecular neurodegeneration, 2011, 6: 1-9.

[32] Igarashi K M. Entorhinal cortex dysfunction in Alzheimer's disease[J]. Trends in neurosciences, 2023, 46(2): 124-136.

[33] Wang X, Zhang Y, Zhou M, et al. Spatio-temporal similarity measure based multi-task learning for predicting alzheimer's disease progression using mri data[C]//2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2023: 940-943.

[34] Zhou M, Zhang Y, Yang Y, et al. Robust temporal smoothness in multi-task learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2023, 37(9): 11426-11434.

[35] Wang X, Zhou M, Zhang Y, et al. Empirical Analysis of Regularised Multi-Task Learning for Modelling Alzheimer's Disease Progression[C]//2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2023: 4444-4451.