



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/242541/>

Version: Accepted Version

Article:

Fan, X., Zhou, M., Zhang, Y. et al. (2026) Beyond single scores: A multi-cognitive objective learning for AD progression prediction. Pattern Recognition, 180 (Part D). 114348. ISSN: 0031-3203

<https://doi.org/10.1016/j.patcog.2026.114348>

© 2026 The Authors. Except as otherwise noted, this author-accepted version of a journal article published in Pattern Recognition is made available via the University of Sheffield Research Publications and Copyright Policy under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Beyond single scores: A multi-cognitive objective learning for AD Progression prediction

Xuanhan Fan^{a,b}, Menghui Zhou^{c,*}, Yu Zhang^c, Jun Qi^d, Yun Yang^a, Po Yang^{a,c,**}

^a*Yunnan Key Laboratory of Software Engineering, School of Software, Yunnan University, Chengong District, Kunming, Yunnan, 650504, China*

^b*School of Medical Science and Engineering, Beijing Institute of Technology, No. 5 South Zhongguancun Street, Haidian District, Beijing, 100081, China*

^c*Department of Computer Science, University of Sheffield, , Sheffield, , S1 4DP, UK*

^d*Department of Computing, Xi'an Jiaotong-Liverpool University, 111 Ren'ai Road, Suzhou, Jiangsu, 215123, China*

Abstract

Amidst a rising global incidence of Alzheimer's Disease (AD) and an aging population, the search for predictive structural brain biomarkers and effective treatments is critical. Traditional multi-task learning (MTL) methods for predicting AD progression often focus on single cognitive scores across multiple time points. In contrast, our innovative approach, the Adaptive Multi-Cognitive Objective Temporal (AMCOT) mechanism, predicts AD progression by leveraging multiple cognitive assessments. This method not only integrates diverse cognitive data but also captures temporal correlations between different cognitive tasks, enhancing the model's ability to forecast disease progression. Unlike conventional MTL approaches that focus on isolated cognitive scores or single-time series predictions, AMCOT dynamically identifies temporal correlations across various cognitive-objective scores at different time points. We developed a novel MTL methodology using sparse group lasso techniques to pinpoint biomarkers linked to cognitive assessments accurately. A robust algorithm designed for large datasets significantly outperforms existing models in both overall and task-specific performance. Moreover, we applied stability selection to determine stable MRI

*Corresponding author. Email: menghui.zhou@sheffield.ac.uk.

**Corresponding author. Email: po.yang@sheffield.ac.uk.

Email addresses: fanxuanhan@gmail.com (Xuanhan Fan), menghui.zhou@sheffield.ac.uk (Menghui Zhou), yzhang489@sheffield.ac.uk (Yu Zhang), Jun.Qi@xjtlu.edu.cn (Jun Qi), yangyun@ynu.edu.cn (Yun Yang), po.yang@sheffield.ac.uk (Po Yang)

biomarkers, analyzing their temporal patterns to deepen our understanding of AD progression. This comprehensive approach offers significant advancements in predictive accuracy and biomarker identification for AD. The implementation source can be found at <https://github.com/XuanhanFan/AMCOT>.

Keywords:

Alzheimer’s Disease, disease progression, cognitive score, multi-task learning, identification of biomarkers

1. Introduction

Alzheimer’s disease (AD) [1] is a prevalent neurodegenerative disorder characterized by severe dementia symptoms. According to the World Alzheimer’s Disease International Report [2], 75% of global dementia cases are undiagnosed. As populations age, the incidence of dementia is expected to rise sharply, potentially affecting 152.8 million people by 2050. AD is irreversible, and no cure exists, making early intervention crucial. Definitive diagnosis often requires invasive procedures like brain biopsies or autopsies, limiting early diagnosis and treatment efforts. Consequently, identifying biomarkers of AD progression and developing predictive models are essential. Previous studies have employed various cognitive assessment tools such as the Clinical Dementia Rating Scale-Sum (CDR-SB) [3], Mini-Mental State Examination (MMSE) [4], Alzheimer’s Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) [5], Rey Auditory Verbal Learning Test (RAVLT) [6], Functional Activity Questionnaire (FAQ) [7], and Trail Making Test Part B (TRABSCOR) [8]. These tools [9] aid clinicians in monitoring cognitive function and daily activities and, when combined with AD-specific MRI biomarkers [10], improve early diagnosis and progression prediction. Moreover, advanced machine learning techniques are increasingly applied to analyze MRI [11] and cognitive data, enhancing AD research toward more accurate and operational approaches.

Deep learning methods face interpretability challenges and rely heavily on large amounts of data [12, 13]. Despite significant correlations between cognitive assessments at different time points due to disease progression, traditional single-task re-

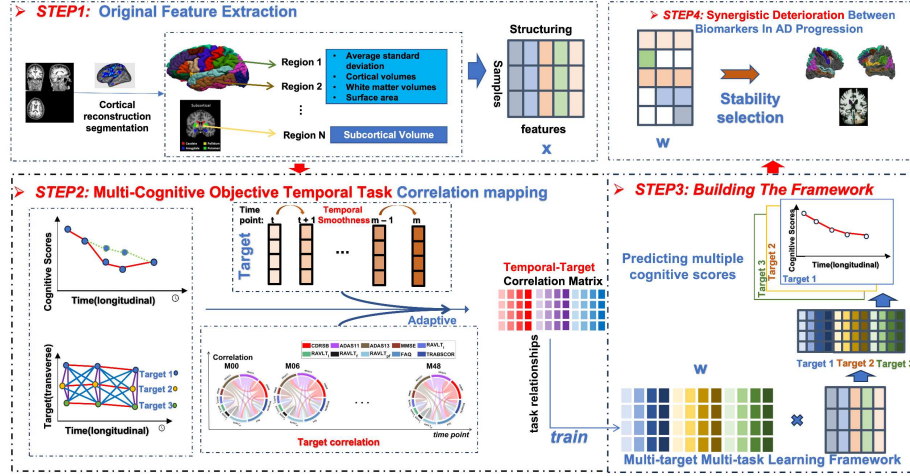


Fig 1: Flowchart of the AMCOT method, which integrates target scores and temporal smoothness into a cohesive multi-task learning framework. This method combines temporal dynamics with correlations between targets at each time point to establish a target-temporal correlation matrix. The temporal progression of these associations and their representation through chord diagrams are detailed in Fig. 2.

gression models tend to ignore temporal correlations in cognitive states. Multi-task learning (MTL) frameworks [14, 15] address this issue by analysing these temporal relationships, significantly improving model performance. The core principle of MTL is to improve model generalisation by exploiting relationships between related tasks [16]. Recent advances have shown that MTL can effectively predict disease progression, but challenges remain in identifying task relevance and developing effective models. Yet these frameworks predominantly treat each cognitive instrument as an independent prediction target, ignoring a clinically important asymmetry: MMSE scores *decrease* with disease progression whereas ADAS-Cog 13 scores *increase*, yet both reflect the same underlying neurodegeneration and exhibit systematic co-deterioration. A model predicting either score in isolation cannot exploit this complementary cross-score structure.

Clinically, physicians rely on diverse cognitive tests and MRI scans to tailor AD treatment plans, yet existing MTL methods often overlook the evolving longitudinal dependencies among multiple cognitive assessment scores simultaneously. To address this gap, we propose a new MTL framework that incorporates multi-target temporal

domain correlations. Specifically, we implement a temporal smoothness strategy to regularise consecutive cognitive score predictions and construct an adaptive correlation matrix via Pearson coefficients to capture inter-score dependencies at each time point (Fig. 1).

Given the nonsmooth regularization term in the objective function, which renders it non-differentiable, we adopt the widely used ADMM algorithm [17, 18]. This algorithm partitions the original problem into subproblems that can be solved in parallel across multiple computing nodes and exhibits robust parameter tuning (e.g., penalty parameters). Building on this property, we introduce a distinctive dynamic adjustment strategy for the quadratic penalty coefficient [19], specifically designed to mitigate the slow convergence and excessive iteration issues frequently encountered in ADMM. Extensive evaluations on the latest Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset [20] confirm the approach’s superior predictive performance and robustness, especially under stringent demands for high accuracy and rapid convergence.

We summarize our contributions as follows:

- We propose MTL-AMCOT, an innovative multi-task learning framework that overcomes the limitations of prior longitudinal studies [14, 21, 22, 23, 24] on AD, which typically consider only a single cognitive score at a time. Our approach adaptively integrates multiple cognitive objectives along with their temporal dynamics, enabling a more comprehensive modeling of the multidimensional evolution of cognitive states. Crucially, MTL-AMCOT explicitly exploits the complementary directional asymmetry among scores (e.g., MMSE *decreases* while ADAS-Cog 13 *increases* with disease progression), a structure that prior single-score MTL methods ignore.
- We designed an ADMM-based dynamic adjustment strategy for the quadratic penalty coefficient, designed to alleviate the high iteration count and slow convergence commonly observed in MTL optimization on large-scale AD datasets, particularly during hyperparameter tuning. Compared to conventional ADMM methods without this strategy, our approach achieves a 43.61% reduction in the average number of iterations and improves computational efficiency by 19.22%.

- To enhance model decision transparency and reliability, we emphasize interpretability and leverage interrelations among multiple scoring tasks. Analysis of MRI biomarkers from different target scores shows that our model highlights both well-established key regions (e.g., hippocampus, entorhinal cortex, middle temporal lobe) and potential biomarkers requiring further medical validation (e.g., isthmus of the cingulate, inferior parietal lobule, cerebellar cortex). Notably, the model’s focus on cerebrospinal fluid (CSF) aligns with recent findings[10] on cognitive recovery, offering deeper insight into AD pathogenesis.

Organization:The structure of the remainder of this paper is as follows: Section 2 reviews related literature. Section 3 introduces our approach for managing multi-objective temporal task relationships. Section 4 details the relevant optimization algorithms. Section 5 presents experimental results. Section 6 concludes the paper.

2. RELATED WORK

2.1. Multiple Cognitive Scores for Assessing the Progression of AD

Several cognitive assessment instruments have been widely used to evaluate different aspects of cognitive decline in AD, including the Clinical Dementia Rating Scale-Sum of Boxes (CDRSB) [3], the Alzheimer’s Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) [5], the Mini-Mental State Examination (MMSE) [4], the Rey Auditory Verbal Learning Test (RAVLT) [6], the Functional Activities Questionnaire (FAQ) [7], and the Trail Making Test Part B (TRABSCOR) [8]. These instruments assess complementary cognitive domains, including global cognition, memory, executive function, and activities of daily living.

To improve sensitivity to early cognitive changes, Skinner et al. [25] proposed ADAS-Cog-Plus by augmenting ADAS-Cog with measures of executive functioning and functional ability. This development highlights an important observation: no single cognitive score can fully characterise the multifaceted nature of cognitive deterioration in AD. Instead, different cognitive assessments capture complementary and evolving manifestations of disease progression. Consequently, jointly modelling multiple cog-

nitive scores has the potential to provide a more comprehensive representation of AD progression and improve longitudinal prediction.

2.2. Multi-task Learning Methods for AD Progression Prediction

Multi-task learning (MTL) has emerged as an effective framework for modelling longitudinal disease progression by exploiting relationships among related prediction tasks. Zhou et al. [21] proposed the Temporal Group Lasso (TGL), which assumes that different time points share a common set of predictive biomarkers and performs joint feature selection across all longitudinal tasks. However, TGL cannot adequately capture temporal variations in biomarker relevance. To address this limitation, cFSGL [26] introduced a sparse group Lasso regularisation strategy [27], enabling both shared and task-specific feature selection.

Most early longitudinal MTL methods are built upon the temporal smoothness assumption [14, 21, 26], which assumes that cognitive states and model parameters change gradually between consecutive follow-up visits. Along this direction, Longitudinal Stability Adjustment (LSA) [22] and Automatic Temporal Relationship (AutoTR) [24] was proposed to adaptively learn global temporal dependencies from longitudinal observations. Similarly, Liu et al. [23] incorporated a Laplacian sparse group Lasso model to capture weighted dependencies among tasks through a graph-regularised formulation. Despite their methodological differences, these approaches primarily focus on relationships within the temporal domain and treat each cognitive assessment as an independent prediction target.

Recognising that different cognitive scores often reflect complementary aspects of the same underlying neurodegenerative process, Liang et al. [28] proposed the Bi-Graph guided self-paced Multi-Task Feature Learning (BGP-MTFL) framework to jointly model multiple cognitive objectives and exploit inter-task relationships. Nevertheless, existing approaches either model temporal relationships or exploit correlations among cognitive objectives, but rarely consider that the relationships between different cognitive assessments may themselves evolve over time during disease progression. Consequently, a unified framework capable of simultaneously modelling temporal dynamics and the evolving interdependencies among multiple cognitive objectives re-

mains largely unexplored.

2.3. Deployment of Machine Learning Systems for AD Prediction

Recent deep learning approaches have substantially advanced AD prediction by leveraging large-scale multimodal datasets and highly expressive architectures, including structured knowledge alignment for diagnosis [29], hybrid federated learning for multi-centre detection [30], generative adversarial frameworks with foundation models for risk prediction [31], and contrast-enhanced hierarchical attention networks on structural MRI [32]. Owing to their ability to automatically learn hierarchical representations directly from raw MRI, PET, and clinical data, deep learning methods have achieved impressive performance on benchmark datasets [33, 34, 35]. These capabilities make deep learning particularly attractive for automated diagnostic pipelines and large-scale cloud deployment.

Nevertheless, clinical deployment involves considerations that extend beyond predictive performance and computational efficiency alone. Longitudinal AD cohorts frequently exhibit distribution shifts arising from variations in imaging scanners, acquisition protocols, patient populations, and irregular follow-up schedules. Although deep learning methods often achieve state-of-the-art performance under in-distribution settings, their reliability and robustness under out-of-distribution scenarios remain difficult to characterise and monitor in real-world clinical practice [36]. Furthermore, despite considerable advances in explainable artificial intelligence, the highly parameterised and nonlinear nature of deep neural networks makes their decision-making processes inherently difficult to fully interpret from a clinical perspective. Such limited transparency may impede physician trust, hinder model validation, and ultimately restrict routine clinical adoption [37, 38]. More recently, the rapid progress of large language models has stimulated growing interest in incorporating agentic AI into clinical diagnostic workflows, which is increasingly viewed as a promising direction for the next generation of AI for healthcare [39]. However, substantial barriers remain before such systems can be deployed in real-world settings. In particular, their safe and effective integration into healthcare systems requires adaptive oversight mechanisms and flexible deployment pathways [40].

In contrast, MTL-AMCOT was explicitly designed around clinically meaningful concepts that are already familiar to physicians. The proposed framework operates directly on anatomically interpretable MRI biomarkers and routinely collected cognitive assessments while explicitly modelling temporal progression and interdependencies among cognitive objectives. Every model coefficient can be directly associated with a specific brain region and cognitive score, enabling transparent biomarker-level analyses and longitudinal stability investigations. More broadly, intrinsically interpretable machine learning methods have been successfully deployed in numerous medical applications. For example, LASSO, one of the most influential statistical machine learning methods, has been successfully applied to dyslipidaemia and cardiovascular disease risk assessment [41] and to predicting interbody cage subsidence following OLIF surgery [42]. Therefore, rather than competing directly with recent deep learning approaches, MTL-AMCOT represents a complementary method that prioritises transparency, robustness, and clinical interpretability, all of which are particularly important for real-world deployment and decision-support systems in AD research.

3. METHODS

3.1. Problem Formulation

We formulate AD progression prediction as a MTL problem in which each cognitive objective at each longitudinal time point constitutes one task. Table 1 summarises the notation used throughout this paper.

Table 1: Summary of key notation.

Symbol	Description	Symbol	Description
t	Number of longitudinal time points	O	Set of cognitive objectives (scores)
$o \in O$	A single cognitive objective	$i \in \{1, \dots, t\}$	Time-point index
n_i	Number of subjects at time point i	p	Number of MRI features (biomarkers)
$X_i^o \in \mathbb{R}^{n_i \times p}$	Feature matrix	$y_i^o \in \mathbb{R}^{n_i}$	Cognitive score vector
$W_i^o \in \mathbb{R}^p$	Regression weight vector	$W^O \in \mathbb{R}^{p \times (t O)}$	Full weight matrix
$\mathbf{H} \in \mathbb{R}^{(t \times (t-1))}$	Temporal difference matrix	M_τ	Adaptive target-correlation matrix
$\mathfrak{R}$	AMCOT penalty term	$\lambda_1, \lambda_2, \lambda_3$	Regularisation hyperparameters

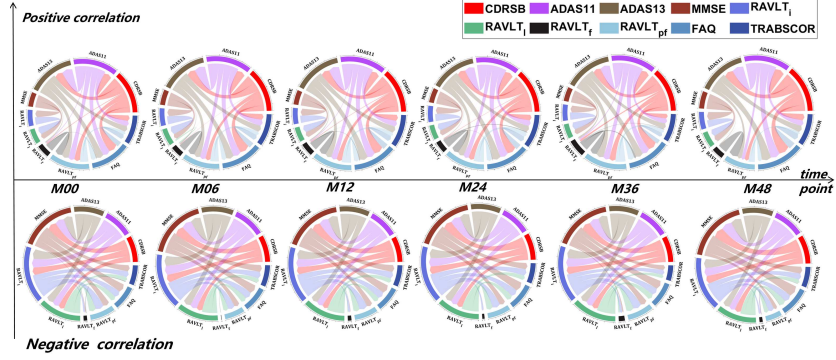


Fig 2: Variation in correlational information between scores on different cognitive objectives at different time points.

The squared loss over all tasks and objectives is

$$\mathcal{L}(Y, X, W) = \frac{1}{2} \sum_{i=1}^t \sum_{o \in O} \|X_i^o W_i^o - Y_i^o\|_2^2. \quad (1)$$

The input X_i^o contains baseline MRI features; each row corresponds to one subject and each column to one biomarker. The output Y_i^o is the vector of cognitive scores for objective o at a future time point.

3.2. Temporal Domain Task Relationships–Temporal Smoothness Regularisation

Given the gradual progression of AD, task-specific model parameters at adjacent follow-up time points are expected to change smoothly. Accordingly, we impose an ℓ_1 penalty on the differences between regression weights at successive time points:

$$\Omega_{\text{temp}}(W^o) = \sum_{i=1}^{t-1} \|w_{i+1}^o - w_i^o\|_1 = \|W^o \mathbf{H}\|_1, \quad (2)$$

where $\mathbf{H} \in \mathbb{R}^{t \times (t-1)}$ denotes the temporal difference matrix, with $H_{ij} = -1$ if $i = j$, $H_{ij} = 1$ if $i = j + 1$, and $H_{ij} = 0$ otherwise. This regularisation promotes local temporal smoothness and is consistent with the slow, progressive nature of AD, as well as prior MTL formulations [21, 14].

3.3. Adaptive Target-Temporal Correlation and AMCOT Penalty

Different cognitive target scores are pivotal for monitoring the progression of AD. We utilise a comprehensive set of cognitive objectives $\mathbf{O}$:

$$\mathbf{O} = \left\{ \begin{array}{l} \text{CDRSB, ADAS11, ADAS13, MMSE, TRABSCOR,} \\ \text{RAVLT_immediate, RAVLT_learning, FAQ,} \\ \text{RAVLT_forgetting, RAVLT_perc_forgetting} \end{array} \right\} \quad (3)$$

To address scale differences, all targets are standardised via z -scores. We then compute the Pearson correlation between objective pairs at each time point τ :

$$c_{\tau}^{i,j} = \frac{\text{cov}(y_{\tau}^i, y_{\tau}^j)}{\sigma_{y_{\tau}^i} \sigma_{y_{\tau}^j}}, \quad i, j \in \mathbf{O}. \quad (4)$$

As shown in Fig. 2, significant temporal variations in correlations (both positive and negative) exist among different cognitive objectives. These coefficients populate the adaptive correlation matrix at each time point:

$$M_{\tau} = \begin{bmatrix} c_{\tau}^{1,1} & -c_{\tau}^{1,2} & \cdots & -c_{\tau}^{1,o} \\ -c_{\tau}^{2,1} & c_{\tau}^{2,2} & \cdots & -c_{\tau}^{2,o} \\ \vdots & \vdots & \ddots & \vdots \\ -c_{\tau}^{o,1} & -c_{\tau}^{o,2} & \cdots & c_{\tau}^{o,o} \end{bmatrix} \quad (5)$$

The sign convention — positive diagonal, negated off-diagonal — ensures that WM_{τ} yields pairwise residuals $W^i - c_{\tau}^{i,j} W^j$ for each objective pair (i, j) , which are stacked across pairs and time points to form the association matrix R (Algorithm 1).

Combining temporal smoothness (Eq. (2)) with the adaptive inter-objective correlation, the AMCOT penalty is:

$$\|\mathfrak{R}\| = \|WR\| = \sum_{\tau=1}^T (W^i - M_{\tau}^{(i,j)} \cdot W^j) + \sum_{\tau=1}^{T-1} (W_{\tau+1} - W_{\tau}), \quad (6)$$

where R is the temporal–target association matrix constructed in Algorithm 1. The first term penalises inconsistency between weight vectors of correlated objectives at the same time point; the second term (cf. Eq. (2)) penalises large changes across consecutive time points. The full derivation of the penalty term $\mathfrak{R}$ is provided in Appendix B. The regularisation term $\|WR\|$ thus penalises inter-task prediction inconsistency weighted by correlation strength (Fig. 3).

Numerical illustration. Consider two objectives (CDRSB, ADAS13) at two time points (M00, M06), with observed Pearson correlations $c_{M00}^{CDRSB, ADAS13} = 0.7494$ and $c_{M06}^{CDRSB, ADAS13} = 0.782$. The per-time-point blocks are:

$$M_{M00}^{(2 \times 2)} = \begin{bmatrix} 1 & -0.749 \\ -0.749 & 1 \end{bmatrix}, \quad M_{M06}^{(2 \times 2)} = \begin{bmatrix} 1 & -0.782 \\ -0.782 & 1 \end{bmatrix}.$$

Stacking the target-consistency blocks and the temporal difference $\mathbf{H}$ yields the association matrix:

$$R^{(4 \times 4)} = \begin{bmatrix} 1 & 0 & \frac{1}{2} \cdot 0.749 & 0 \\ -1 & 1 & -\frac{1}{2} \cdot 0.749 & -\frac{1}{2} \cdot 0.782 \\ \frac{1}{2} \cdot 0.749 & 0 & 1 & 0 \\ -\frac{1}{2} \cdot 0.749 & -\frac{1}{2} \cdot 0.782 & -1 & 1 \end{bmatrix}.$$

When w^{CDRSB} deviates from $0.749 w^{ADAS13}$ at M00, or from $0.782 w^{ADAS13}$ at M06, the corresponding off-diagonal entries in R generate a large penalty, directing the optimiser to maintain weight vectors consistent with the evolving correlation structure across time. The sign and magnitude of each R entry thus translate empirical correlations directly into a structured regularisation signal.

Remark on correlation measure. The inter-target relationship in AMCOT is modelled via Pearson correlation, which is straightforward to estimate and integrates directly into $\mathbf{M}_\tau$ without additional computational overhead. As this linear measure may be insufficient for weakly coupled or nonlinear objective pairs, nonlinear alternatives such as distance correlation are identified as a future extension (see Section 6).

3.4. One Novel MTL Method — MTL-AMCOT

We employ sparse group Lasso regularisation — combining the ℓ_1 -norm for individual-feature sparsity and the $\ell_{2,1}$ -norm for group sparsity across tasks — to enable both joint and task-specific biomarker selection [27]. The AMCOT penalty (Eq. (6)) further encodes the adaptive temporal–target structure. A novel MTL framework with AMCOT penalty (MTL-AMCOT) is defined as:

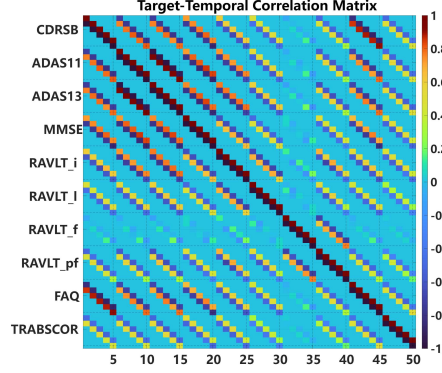


Fig 3: The target-temporal task relationship association matrix R used by AMCOT. Each 10×10 block corresponds to one time point; diagonal blocks encode temporal smoothness weights, and off-diagonal blocks encode the Pearson-correlation-weighted inter-target penalties $M_\tau^{(i,j)}$. Colour scale: red = positive, blue = negative association. Target associations visibly evolve across time points, motivating the adaptive construction of M_τ .

$$\min_W \frac{1}{2} \sum_{i=1}^t \sum_{j=1}^O \|X_i^j W_i^j - Y_i^j\|_2^2 + \lambda_1 \sum_{j=1}^O \|W^j\|_1 + \lambda_2 \sum_{j=1}^O \|W^j\|_{2,1} + \lambda_3 \|\mathfrak{R}\|_1.$$

$$\text{s.t. } \mathfrak{R} = WR. \quad (7)$$

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

4. OPTIMIZATION ALGORITHMS

In this section, we design a quadratic penalty strategy incorporating an adaptive dynamically tuned ADMM to optimize the objective function-solving process for MTL-AMCOT.

4.1. The ADMM-Based Algorithm

According to (7), it can be equivalently transformed into the following constrained problem:

$$\min_{W, \Theta, \mathfrak{R}} \sum_{j=1}^O \frac{1}{2} \|XW^j - Y^j\|_F^2 + \lambda_1 \|\Theta\|_1 + \lambda_2 \|\Theta\|_{2,1} + \lambda_3 \|\mathfrak{R}\|_1$$

$$\text{s.t. } W = \Theta, \mathfrak{R} = WR. \quad (8)$$

Unlike conventional MTL models where tasks decouple across time and target domains, MTL-AMCOT introduces coupling through the AMCOT penalty, which we handle via Cholesky decomposition of the resulting positive-definite system for numerical stability.

Algorithmic flexibility is further enhanced by a dynamic adjustment of the quadratic penalty coefficient combined with a linearisation technique, yielding an inexact ADMM scheme. The corresponding augmented Lagrangian (Eq. 8) is defined as follows.

$$\begin{aligned} L_\rho(W, \Theta, \mathfrak{R}, U, V) = & \sum_{j=1}^O \frac{1}{2} \|XW^j - Y^j\|_F^2 + \lambda_1 \|\Theta\|_1 + \lambda_2 \|\Theta\|_{2,1} + \lambda_3 \|\mathfrak{R}\|_1 \\ & + \text{Tr}(U^T(W - \Theta)) + \frac{\rho}{2} \|W - \Theta\|_F^2 + \text{Tr}(V^T(\mathfrak{R} - WR)) + \frac{\rho}{2} \|\mathfrak{R} - WR\|_F^2. \end{aligned} \quad (9)$$

Update for W :

$$\begin{aligned} W^{k+1} = & \arg \min_W \frac{1}{2} \|XW - Y\|_F^2 + \text{Tr}(C^T(W - \Theta)) + \frac{\rho}{2} \|W - \Theta\|_F^2 \\ & + \text{Tr}(D^T(WR - \mathfrak{R})) + \frac{\rho}{2} \|WR - \mathfrak{R}\|_F^2. \end{aligned} \quad (10)$$

The closed-form update for W is obtained by setting the gradient of Eq. (10) to zero (full derivation in [Appendix D](#)):

$$X^T XW + \rho W + \rho WF = X^T Y - U + \rho \Theta - VG + \rho \mathfrak{R}G, \quad (11)$$

where $F = HH^T$ and $G = H^T$.

In the ADMM framework, iterative updates for the variables Θ , $\mathfrak{R}$, U , and V are critical for algorithmic convergence. The update rules are adapted from the standard ADMM to fit our model's specific regularization and structural requirements. Here, we briefly present the update equation for each variable:

1. Update for Θ : The update for variable Θ at the $k + 1$ iteration is governed by the optimization problem:

$$\Theta^{(k+1)} = \arg \min_{\Theta} \frac{1}{2} \|\Theta - (W^{(k)} + \frac{U^{(k)}}{\rho})\|_F^2 + \frac{1}{\rho} (\lambda_1 \|\Theta\|_1 + \lambda_2 \|\Theta\|_{2,1}) \quad (12)$$

2. Update for $\mathfrak{R}$: Similarly, the update for $\mathfrak{R}$ integrates a sparsity-inducing ℓ_1 -norm, formulated as:

$$\mathfrak{R}^{(k+1)} = \arg \min_{\mathfrak{R}} \frac{1}{2} \|\mathfrak{R} - (W^{(k)}R + \frac{V^{(k)}}{\rho})\|_F^2 + \frac{\lambda_3}{\rho} \|\mathfrak{R}\|_1 \quad (13)$$

3. Update for U : The update for U incorporates a dual ascent step, which is crucial for adjusting the dual variables appropriately:

$$U^{(k+1)} = U^{(k)} + \tau \rho (W^{(k+1)} - \Theta^{(k+1)}) \quad (14)$$

where τ is a step-size parameter that influences the rate of convergence of the dual updates.

4. Update for V : Finally, the update for D is specified as:

$$V^{(k+1)} = V^{(k)} + \tau \rho (W^{(k+1)}R - \mathfrak{R}^{(k+1)}) \quad (15)$$

This equation aligns with the dual update strategy, facilitating the alignment between the primal and dual variables.

4.2. Dynamic Adjustment of Quadratic Penalty Coefficients

Static choices of ρ lead to an inherent trade-off: a large ρ reduces primal residuals quickly but slows dual convergence, while a small ρ has the opposite effect. To balance both, we dynamically adjust ρ at each iteration based on the primal and dual residuals:

$$r^k = (W^k - \Theta^k) + (W^k R - \mathfrak{R}^k), \quad (16)$$

$$s^k = (\Theta^k - \Theta^{k-1}) + (\mathfrak{R}^k - \mathfrak{R}^{k-1})R^\top. \quad (17)$$

If the two residuals are imbalanced by more than a factor of μ , ρ is scaled up or down accordingly:

$$\rho^{k+1} = \begin{cases} \gamma_p \rho^k, & \text{if } \|r^k\| > \mu \|s^k\|, \\ \rho^k / \gamma_d, & \text{if } \|s^k\| > \mu \|r^k\|, \\ \rho^k / 2, & \text{otherwise.} \end{cases} \quad (18)$$

Each change to ρ requires recalculating the cached Cholesky decomposition. When the loss function begins to oscillate after initial stabilisation, a backtracking strategy reverts to the previous iteration parameters, halves ρ , and suspends further dynamic

Table 2: One-at-a-time sensitivity analysis of the ADMM dynamic penalty hyperparameters on the CDRSB prediction task. For each scan, the other two hyperparameters are fixed at their selected values.

μ scan					γ_p scan					γ_d scan				
Val.	Iter.	nMSE	wR	Time	Val.	Iter.	nMSE	wR	Time	Val.	Iter.	nMSE	wR	Time
1.5	416	0.5084	0.7049	7.0	1.2	139	0.5096	0.7051	3.7	1.2	416	0.5084	0.7049	6.9
2.0	416	0.5084	0.7049	7.0	1.5	161	0.5089	0.7054	3.3	1.5	416	0.5084	0.7049	6.9
5.0	416	0.5084	0.7049	6.9	2.0*	416	0.5084	0.7049	6.9	2.0*	416	0.5084	0.7049	6.9
10.0*	416	0.5084	0.7049	6.9	5.0	1000 [†]	0.8565	0.5748	15.7	5.0	416	0.5084	0.7049	7.0
20.0	416	0.5084	0.7049	6.9	10.0	1000 [†]	0.5069	0.7057	15.3	10.0	416	0.5084	0.7049	6.9

Note: * Selected value used in the main experiments. [†] Reached the maximum iteration budget and was therefore not regarded as a reliably converged configuration. For the μ scan, $\gamma_p = 2$ and $\gamma_d = 2$ were fixed; for the γ_p scan, $\mu = 10$ and $\gamma_d = 2$ were fixed; for the γ_d scan, $\mu = 10$ and $\gamma_p = 2$ were fixed.

adjustment until stable convergence is confirmed. The full procedure is outlined in Algorithm 2.

The hyperparameters μ , γ_p , and γ_d were selected via a one-at-a-time sensitivity analysis on CDRSB (Table 2). μ and γ_d are entirely insensitive: all tested values yield identical performance (416 iterations, nMSE = 0.5084), so $\mu^* = 10$ and $\gamma_d^* = 2$ are adopted following [43]. γ_p is the only parameter with meaningful influence: small values (≤ 1.5) under-adjust ρ and yield higher nMSE; large values (≥ 5) cause divergence or maxIter termination. $\gamma_p^* = 2$ achieves the lowest nMSE among all convergent configurations and is therefore selected.

4.2.1. Complexity of the ADMM-based Algorithm

Each iteration requires T Cholesky decompositions of matrices in $\mathbb{R}^{p \times p}$, giving a per-iteration cost of $O(T \times (np^2 + p^3))$ [17, 18]. Combined with the $O(1/\epsilon)$ iteration count, the overall complexity is $O(1/\epsilon \times T \times (np^2 + p^3))$. Empirical validation with a fixed 500-iteration budget confirms fitted exponents of $O(p^{1.96})$ and $O(T^{1.69})$ (Table 3): the former closely matches the dominant $O(np^2)$ term; the slightly super-linear T scaling reflects the growth of $\mathbf{H}$ and $F = \mathbf{H}\mathbf{H}^T$ beyond a strict linear multiplier.

Table 3: Empirical runtime scaling under a fixed 500-iteration budget (mean $\pm$ std over 3 runs).

p scaling ($T = 6$)		T scaling ($p = 314$)	
Value	Time (s)	Value	Time (s)
50	0.09 ± 0.01	2	0.51 ± 0.02
100	0.25 ± 0.02	3	0.98 ± 0.05
150	0.69 ± 0.01	4	1.59 ± 0.06
200	1.08 ± 0.02	5	2.39 ± 0.01
250	2.01 ± 0.01	6*	3.22 ± 0.01
314*	3.22 ± 0.04		

Note: * Setting used in the main experiments. Fitted exponents are approximately $O(p^{1.96})$ and $O(T^{1.69})$.

5. Experimental Result

In this study, we repeated experiments ten times using 90% of the data for training, supplemented by 10-fold cross-validation to refine regularization parameters. We normalized all data using z-scores. The dataset is derived from the ADNI [20]. We excluded patients without baseline MRI data, removed MRI images failing quality standards, and imputed missing features using mean values. After preprocessing, the dataset consisted of 314 features across six-time points. Due to high attrition rates in the ADNI Long-Term Research Project, our analysis was restricted to the initial six time points to concentrate on the early stages of AD and its implications for interventions. The progressive reduction in sample size from $n = 1,532$ at baseline to $n = 289$ at M48 (Table 4) reflects selective attrition, whereby patients with more severe cognitive decline are disproportionately lost to follow-up; the implications of this for model bias and future mitigation strategies are discussed in Section 6.

Table 4: Demographic information of subjects at different time points

Attribute	Baseline (M00)	M06	M12	M24	M36	M48
Sample Size	1532	1317	1375	1099	432	289
(AD, MCI, CN)	(315, 817, 400)	(265, 669, 383)	(366, 768, 241)	(132, 653, 314)	(0, 277, 155)	(0, 172, 117)
(Female, Male)	(688, 844)	(570, 747)	(606, 769)	(487, 612)	(169, 263)	(127, 162)
APOE4 (0, 1, 2)	(795, 567, 160)	(675, 493, 149)	(721, 508, 146)	(597, 399, 103)	(259, 142, 31)	(175, 92, 22)

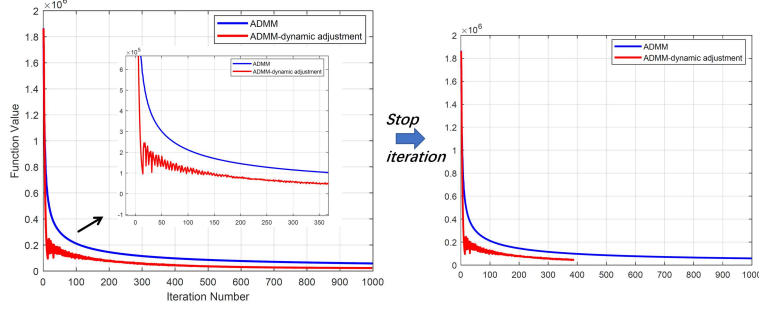


Fig 4: The convergence behavior of ADMM algorithms with a dynamically adjusted quadratic penalty coefficient.

5.1. Comparison of accelerated algorithms and non-accelerated algorithms

We compared dynamic and static ADMM in MATLAB on ADNI with $\lambda_1 = 10$, $\lambda_2 = 100$, $\lambda_3 = 100$; trends were consistent across other parameter settings. Furthermore, we applied stopping and backtracking strategies to mitigate oscillations near convergence points. As shown in Fig. 4, these strategies expedited the reduction of the loss function in early iterations and aided in achieving closer convergence upon termination. Although the dynamic adjustment increases per-iteration overhead due to frequent recomputation of the Cholesky decomposition, it substantially improves overall convergence. As shown in Table 5, across a broad range of hyperparameter configurations the dynamic strategy boosts iteration efficiency by 43.61% and reduces execution time by 19.22%, supporting its effectiveness for complex multi-objective MTL.

Table 5: Comparison of ADMM Performance With and Without Dynamic Adjustment

Metric	Dynamic	Non-Dynamic	Improvement
Average Iterations	255,663.7	453,376.5	43.61% ↑
Average Execution Time (s)	7,616.59	9,428.66	19.22% ↑

Note: Results are based on 10-fold cross-validation with hyperparameters (10^{-2} , 10^{-1} , 1, 10, 50, 100, 500, 1000).

5.2. Assessment Metrics

For our quantitative evaluation, we utilized Root Mean Squared Error (rMSE) to measure the discrepancy between predicted and actual cognitive scores for task-specific

Table 6: Comparison of various methods for predicting the progression of longitudinal cognitive scores in AD

Data	Metric	TGL	cFSGL	AutoTR	RMTL	LSA
CDRSB	nMSE	0.546 ± 0.026	0.535 ± 0.027	0.683 ± 0.017	0.672 ± 0.013	0.798 ± 0.044
	wR	0.681 ± 0.018	0.688 ± 0.019	0.548 ± 0.019	0.553 ± 0.017	0.523 ± 0.019
ADAS11	nMSE	0.486 ± 0.021	0.477 ± 0.021	0.516 ± 0.026	0.526 ± 0.032	0.607 ± 0.039
	wR	0.724 ± 0.012	0.729 ± 0.012	0.708 ± 0.015	0.695 ± 0.019	0.655 ± 0.021
ADAS13	nMSE	0.461 ± 0.023	0.453 ± 0.022	0.498 ± 0.013	0.505 ± 0.008	0.595 ± 0.018
	wR	0.737 ± 0.014	0.742 ± 0.014	0.720 ± 0.005	0.707 ± 0.005	0.664 ± 0.006
MMSE	nMSE	0.551 ± 0.023	0.540 ± 0.025	0.594 ± 0.029	0.596 ± 0.027	0.692 ± 0.033
	wR	0.676 ± 0.014	0.685 ± 0.015	0.659 ± 0.014	0.642 ± 0.021	0.600 ± 0.017
RAVLT_i	nMSE	0.600 ± 0.020	0.581 ± 0.017	0.615 ± 0.022	0.611 ± 0.021	0.715 ± 0.021
	wR	0.637 ± 0.014	0.650 ± 0.012	0.634 ± 0.014	0.625 ± 0.018	0.568 ± 0.013
RAVLT_l	nMSE	0.752 ± 0.014	0.734 ± 0.015	0.790 ± 0.016	0.757 ± 0.016	0.910 ± 0.033
	wR	0.504 ± 0.014	0.520 ± 0.015	0.463 ± 0.018	0.496 ± 0.017	0.407 ± 0.019
RAVLT_f	nMSE	0.973 ± 0.009	0.962 ± 0.011	0.996 ± 0.009	0.977 ± 0.011	1.208 ± 0.030
	wR	0.176 ± 0.027	0.205 ± 0.026	0.104 ± 0.024	0.151 ± 0.028	0.072 ± 0.016
RAVLT_pf	nMSE	0.720 ± 0.031	0.701 ± 0.028	0.748 ± 0.024	0.706 ± 0.027	0.888 ± 0.033
	wR	0.532 ± 0.027	0.549 ± 0.026	0.497 ± 0.027	0.538 ± 0.029	0.443 ± 0.029
FAQ	nMSE	0.580 ± 0.031	0.566 ± 0.029	0.598 ± 0.017	0.601 ± 0.013	0.698 ± 0.022
	wR	0.655 ± 0.019	0.665 ± 0.018	0.648 ± 0.012	0.638 ± 0.011	0.587 ± 0.016
TRABSCOR	nMSE	0.694 ± 0.019	0.682 ± 0.022*	0.735 ± 0.018	0.723 ± 0.020	0.886 ± 0.052
	wR	0.557 ± 0.019	0.568 ± 0.020*	0.525 ± 0.011	0.536 ± 0.012	0.457 ± 0.016

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 generalisation in the case of a better average.

assessments. Additionally, we assessed overall performance across all tasks using normalized mean squared error (nMSE) and the weighted r-value (wR), both standard in MTL and medical research [14, 21, 24, 27]. Lower nMSE and rMSE values indicate better performance, while a higher wR reflects more favourable outcomes.

$$\text{nMSE}(Y, \hat{Y}) = \frac{\sum_{i=1}^t \|Y_i - \hat{Y}_i\|_2^2 / \sigma^2(Y_i)}{\sum_{i=1}^t n_i}, \quad (19)$$

$$\text{wR}(Y, \hat{Y}) = \frac{\sum_{i=1}^t \text{Corr}(Y_i, \hat{Y}_i) n_i}{\sum_{i=1}^t n_i}, \quad (20)$$

$$\text{rMSE}(y, \hat{y}) = \sqrt{\frac{\|y - \hat{y}\|_2^2}{n}}, \quad (21)$$

Table 7: Comparison of various methods for predicting the progression of longitudinal cognitive scores in AD (continued)

Data	Metric	FL-SGL	MinimalRNN	LSTM	AMCOT (Ours)
CDRSB	nMSE	0.579 ± 0.020	0.957 ± 0.086	0.956 ± 0.245	$0.519 \pm 0.020^*$
	wR	0.656 ± 0.016	0.215 ± 0.133	0.226 ± 0.122	$0.704 \pm 0.012^*$
ADAS11	nMSE	0.523 ± 0.019	0.482 ± 0.110	0.493 ± 0.139	$0.466 \pm 0.020^*$
	wR	0.695 ± 0.014	0.718 ± 0.084	0.716 ± 0.183	$0.735 \pm 0.012^*$
ADAS13	nMSE	0.492 ± 0.014	0.484 ± 0.092	0.490 ± 0.131	$0.451 \pm 0.018^*$
	wR	0.715 ± 0.010	0.722 ± 0.066	0.720 ± 0.183	$0.744 \pm 0.013^*$
MMSE	nMSE	0.600 ± 0.014	0.740 ± 0.126	0.708 ± 0.202	$0.521 \pm 0.025^*$
	wR	0.640 ± 0.010	0.508 ± 0.138	0.510 ± 0.158	$0.695 \pm 0.015^*$
RAVLT_i	nMSE	0.638 ± 0.014	0.616 ± 0.116	0.623 ± 0.176	$0.573 \pm 0.012^*$
	wR	0.604 ± 0.011	0.669 ± 0.074	0.676 ± 0.175	$0.657 \pm 0.007^*$
RAVLT_l	nMSE	0.751 ± 0.015	0.948 ± 0.105	0.959 ± 0.242	$0.707 \pm 0.015^*$
	wR	0.503 ± 0.015	0.176 ± 0.135	0.181 ± 0.117	$0.542 \pm 0.014^*$
RAVLT_f	nMSE	0.979 ± 0.007	1.032 ± 0.035	1.058 ± 0.273	$0.953 \pm 0.014^*$
	wR	0.138 ± 0.016	0.097 ± 0.086	0.071 ± 0.081	$0.220 \pm 0.025^*$
RAVLT_pf	nMSE	0.728 ± 0.026	0.523 ± 0.087	0.535 ± 0.139	$0.669 \pm 0.023^*$
	wR	0.517 ± 0.028	0.728 ± 0.058	0.711 ± 0.181	$0.581 \pm 0.020^*$
FAQ	nMSE	0.611 ± 0.020	0.582 ± 0.143	0.591 ± 0.175	$0.534 \pm 0.023^*$
	wR	0.628 ± 0.016	0.623 ± 0.106	0.615 ± 0.165	$0.685 \pm 0.016^*$
TRABSCOR	nMSE	0.736 ± 0.028	0.608 ± 0.100	0.583 ± 0.164	$0.689 \pm 0.019^*$
	wR	0.525 ± 0.023	0.695 ± 0.049	0.709 ± 0.181	$0.563 \pm 0.017^*$

Note: Table 7 is a continuation of Table 6.

5.3. Comparison of Models

We conducted an extensive evaluation of our proposed method by comparing it against several leading models, including RMTL [44], TGL [21], cFSGl [26], FL_SGL [23], LSA [22], and AutoTR [24]. We varied parameter settings across a range $\lambda \in \{10^{-2}, 10^{-1}, 1, 10, 50, 100, 500, 1000\}$ to ensure robust testing. Our experimental setup involved analyzing ten distinct cognitive assessment score datasets, applied systematically to study the changes and progression in various stages of AD, as indicated by Eq. 3. Additionally, we compared the efficacy of our model against deep learning approaches, specifically MinimalRNN [45] and LSTM [46], to further validate its performance.

Our AMCOT method exhibited robust performance in predicting various cognitive scores, consistently outperforming established methods. AMCOT demonstrated exceptional predictive accuracy, achieving the lowest nMSE of 0.519 ± 0.020 and the highest

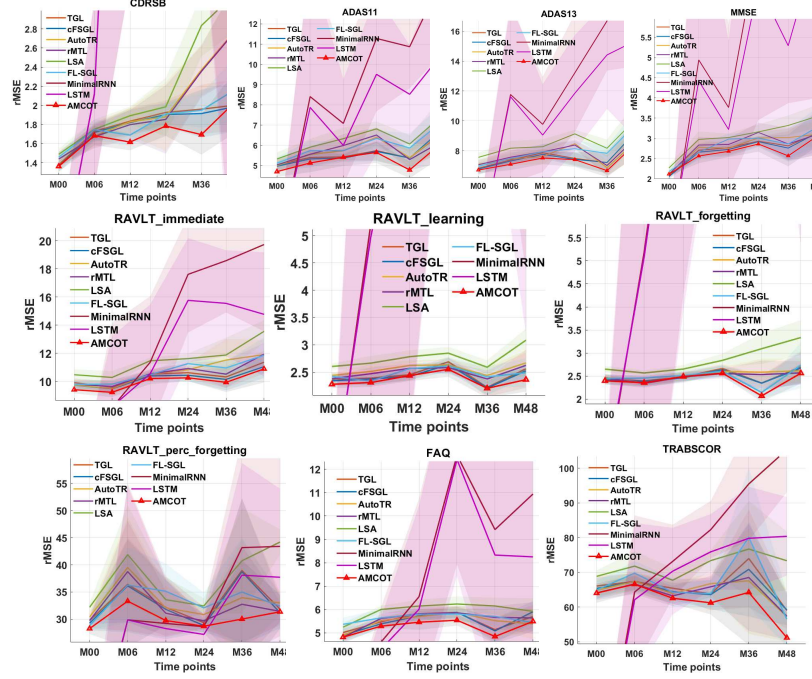


Fig 5: The average root Mean Squared Error (rMSE) compared to the baseline model is presented. Our method (AMCOT) involves multiple target cognitive scores. Therefore, when comparing single-target score tasks, we focus on optimizing the single-target scores.

wR of 0.704 ± 0.012 for CDRSB scores. It achieved the best performance among all compared methods in ADAS11 prediction, with an nMSE of 0.466 ± 0.020 and a wR of 0.735 ± 0.012 . For ADAS13, it recorded the best nMSE of 0.451 ± 0.018 and wR of 0.744 ± 0.013 . In MMSE evaluations, AMCOT showed superior performance with an nMSE of 0.521 ± 0.025 and a wR of 0.695 ± 0.015 . Although its wR was lower than LSTM's for RAVLT_i and RAVLT_pf scores, AMCOT maintained robust generalization. For RAVLT_pf scores, it achieved an nMSE of 0.669 ± 0.023 , slightly higher than MinimalRNN and LSTM models. In FAQ score predictions, AMCOT achieved the lowest nMSE of 0.534 ± 0.023 and a competitive wR of 0.685 ± 0.020 . Though it did not outperform cFSGL in TRABSCOR scores, its results remained highly competitive, surpassing LSTM and RNN in generalization performance.

Fig. 5 depicts the rMSE performance of various models across different time points. Due to their non-interpretable weight structures, deep learning models such as LSTM

Table 8: Ablation study of AMCOT on global cognitive scores.

Variant	CDRSB		ADAS11		ADAS13		MMSE	
	nMSE	wR	nMSE	wR	nMSE	wR	nMSE	wR
Temp. only	0.525 $\pm$ 0.022	0.694 $\pm$ 0.017	0.485 $\pm$ 0.034	0.727 $\pm$ 0.015	0.443 $\pm$ 0.011	0.749 $\pm$ 0.008	0.540 $\pm$ 0.013	0.681 $\pm$ 0.010
Tgt. Rel. only	0.581 $\pm$ 0.024	0.653 $\pm$ 0.020	0.540 $\pm$ 0.037	0.689 $\pm$ 0.016	0.494 $\pm$ 0.012	0.714 $\pm$ 0.010	0.594 $\pm$ 0.015	0.640 $\pm$ 0.011
Full AMCOT	0.510 $\pm$ 0.022	0.704 $\pm$ 0.016	0.475 $\pm$ 0.036	0.735 $\pm$ 0.016	0.431 $\pm$ 0.011	0.757 $\pm$ 0.008	0.528 $\pm$ 0.017	0.690 $\pm$ 0.012
Sig.	*** / †***	*** / †***	** / †***	*** / †***	*** / †***	*** / †***	** / †***	** / †***

Note: The “Sig.” row reports one-tailed significance for Full AMCOT against the two ablation variants over 10 folds. Plain superscripts denote comparisons with w/o target relationship, and † superscripts denote comparisons with w/o temporal smoothness. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, and ^{ns} denotes not significant. Exact p -values are reported in Table E.11 (Appendix E).

Table 9: Ablation study of AMCOT on RAVLT-related scores.

Variant	RAVLT _i		RAVLT _l		RAVLT _f		RAVLT _{pf}	
	nMSE	wR	nMSE	wR	nMSE	wR	nMSE	wR
Temp. only	0.588 $\pm$ 0.016	0.641 $\pm$ 0.012	0.728 $\pm$ 0.013	0.524 $\pm$ 0.015	0.965 $\pm$ 0.008	0.203 $\pm$ 0.019	0.660 $\pm$ 0.021	0.585 $\pm$ 0.017
Tgt. Rel. only	0.639 $\pm$ 0.014	0.600 $\pm$ 0.012	0.773 $\pm$ 0.013	0.479 $\pm$ 0.015	1.008 $\pm$ 0.011	0.121 $\pm$ 0.020	0.722 $\pm$ 0.026	0.531 $\pm$ 0.021
Full AMCOT	0.564 $\pm$ 0.017	0.660 $\pm$ 0.012	0.713 $\pm$ 0.015	0.538 $\pm$ 0.014	0.959 $\pm$ 0.012	0.224 $\pm$ 0.017	0.652 $\pm$ 0.022	0.593 $\pm$ 0.016
Sig.	*** / †***	*** / †***	** / †***	*** / †***	ns / †***	* / †***	** / †***	** / †***

Note: Same significance notation as in Table 8.

and MinimalRNN complicate feature selection based on interpretability. These models require initial cognitive scores as inputs, limiting the meaning of early predictions.

Our analysis for these models concentrates on later time points, using half of the longitudinal dataset for inputs and the remainder for outputs. These deep models showed optimal performance only during the first three time points of RAVLT_{pf} scores, with error rates for other scores increasing over time. Their reliance on complete datasets poses challenges for effective training, as they often overlook correlations between features and cognitive scores over time. In contrast, MTL approaches,

Table 10: Ablation study of AMCOT on functional scores.

Variant	FAQ		TRABSCOR	
	nMSE	wR	nMSE	wR
Temp. only	0.557 $\pm$ 0.022	0.668 $\pm$ 0.016	0.715 $\pm$ 0.019	0.544 $\pm$ 0.015
Tgt. Rel. only	0.613 $\pm$ 0.020	0.625 $\pm$ 0.016	0.753 $\pm$ 0.020	0.511 $\pm$ 0.015
Full AMCOT	0.542 $\pm$ 0.024	0.679 $\pm$ 0.017	0.707 $\pm$ 0.022	0.555 $\pm$ 0.017
Sig.	*** / †***	*** / †***	ns / †**	** / †***

Note: Same significance notation as in Table 8.

which analyze inter-task correlations, adapt more flexibly to available data. Notably, while cFSGL excelled in predicting TRABSCOR scores, it demonstrates that combining multiple objectives does not always yield optimal results for specific targets. The limited performance of AutoTR and LSA, which only analyzed six-time points, hinders their ability to fully leverage global temporal relationships. The selected task relevance level for LSA, $\alpha = 0.2$, may not have been optimal, affecting model performance. RMTL’s variable performance across tasks highlights the impact of anomalous tasks on overall results, emphasizing the need to address these to enhance model robustness and accuracy. These results highlight AMCOT’s effectiveness. It consistently outperforms other MTL methods by achieving lower nMSE and higher wR across cognitive scores. Its success is due to dynamic adjustments and effective integration of multi-temporal and cognitive data, enhancing predictive accuracy and model stability.

In this section, we conduct a comprehensive and in-depth ablation study to evaluate the contribution of each task relationship component within our model. To evaluate the individual contributions of the temporal smoothness constraint and the adaptive target relationship modeling in AMCOT, we conducted a comprehensive ablation study across all cognitive targets. As shown in Table 8, we compare three variants: (i) with temporal smoothness only, (ii) with target relationship modeling only, and (iii) with both components integrated. For each cognitive outcome, the results consistently demonstrate that jointly incorporating both the temporal smoothness constraint and the adaptive target relationship modeling leads to the best predictive performance, as evidenced by lower nMSE and higher wR values. Specifically, removing the target relationship modeling results in a significant performance decline for most targets (e.g., the nMSE for CDRSB increases from 0.510 to 0.525, and for ADAS13 from 0.431 to 0.443). Similarly, excluding the temporal smoothness constraint leads to even larger prediction errors (e.g., the nMSE for CDRSB increases from 0.510 to 0.581, and for ADAS13 from 0.431 to 0.494), underscoring its critical role in capturing longitudinal progression patterns. For all cognitive targets, introducing temporal smoothness regularization consistently improves both nMSE and wR compared to the baseline without target relationship modeling. This indicates that enforcing temporal consistency enables the model to better capture disease progression trends over time. Furthermore,

compared to using temporal smoothness alone, modeling the adaptive longitudinal interdependencies across multiple cognitive targets yields substantial performance gains, highlighting the importance of leveraging inter-target dependencies when predicting multidimensional cognitive trajectories. The combination of these two task perspectives not only improves the accuracy of task time point prediction, but also enhances temporal task consistency and cross-objective correlation learning, which is reflected in the consistent improvement of wR across all metrics. Notably, for cognitive assessments such as MMSE, RAVLT, and TRABSCOR, the combined multi-cognitive score task analysis achieved a significant reduction in nMSE, demonstrating the robustness of AMCOT across diverse cognitive domains. These findings validate the complementary effects of temporal smoothness and adaptive target dependency modeling in our framework and emphasize the necessity of simultaneously incorporating temporal and inter-task structural information when modeling longitudinal cognitive progression in AD.

It should be noted that deep learning approaches offer distinct advantages in data-rich settings. Models such as LSTM and MinimalRNN can capture highly nonlinear temporal dynamics and benefit from end-to-end feature learning when training data are abundant. As shown in Table 7, MinimalRNN achieves competitive performance on RAVLT_pf, suggesting that recurrent architectures may better exploit nonlinear co-progression patterns for certain cognitive scores when sufficient longitudinal observations are available. In the present clinical setting, however, sample sizes are limited ($n \leq 1,532$, reducing to 289 at M48), cognitive targets are heterogeneous, and biomarker interpretability is a primary requirement — conditions under which the structured inductive biases of MTL-AMCOT provide practical advantages over black-box deep models. From a deployment perspective, MTL-AMCOT additionally runs on a standard CPU, has a convex objective amenable to systematic cross-validation, and produces directly interpretable weight matrices — in contrast with recurrent baselines such as LSTM and MinimalRNN, which are typically trained on GPUs, require joint tuning of architecture and optimisation hyperparameters, and rely on black-box internal representations.

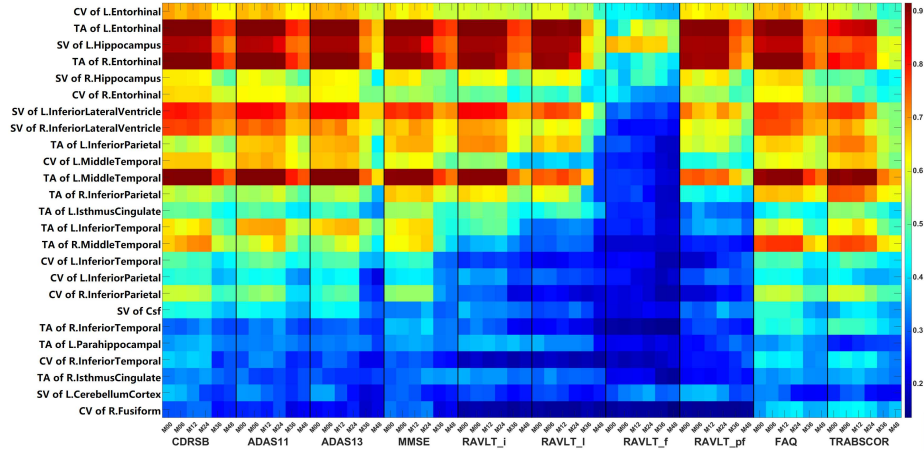


Fig 6: The stability vectors of the 25 stable MRI features generated by AMCOT. Longitudinal stability selection was performed by combining ten target scores. The larger the value, the more stable the feature.

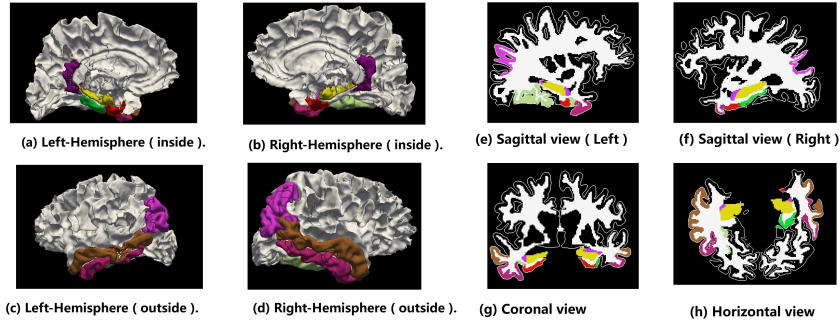


Fig 7: AMCOT uses stability selection to choose the brain maps with the highest ROIs. (a) - (d) are the selected cortical ROIs, and (e) - (g) are the selected subcortical ROIs.

5.4. Multiple Targets Stabilize Temporal Patterns of Biomarkers

The 25 most stable biomarkers identified by MTL-AMCOT (Fig. 6) can be divided into two categories based on existing clinical evidence.

Confirmed biomarkers. Several identified regions are well-established in the AD literature. The left hippocampus volume is a recognised stable biomarker across numerous studies [47, 48], with significant atrophy linked to short-term memory loss and particular vulnerability during the prodromal phase. Dysfunction in the entorhinal cortex [49] frequently manifests early in disease progression, while the inferotemporal cortex [50] is associated with memory performance. The inferior lateral ventricles [51]

are frequently enlarged in AD and serve as a sensitive metric for assessing MCI and AD progression. These findings are consistent with prior neuroimaging studies and validate the biological plausibility of the MTL-AMCOT framework.

Potentially novel findings. In addition, MTL-AMCOT highlights several regions that have received less systematic attention in longitudinal AD biomarker research, including the isthmus of the cingulate gyrus, inferior parietal lobule, and cerebellar cortex. The model’s consistent selection of cerebrospinal fluid (CSF) components aligns with recent findings on cognitive recovery [10, 52], though the specific CSF features identified here warrant further clinical validation. These findings should be interpreted as hypothesis-generating rather than confirmatory, and prospective studies with independent cohorts are needed to establish their prognostic relevance.

Fig. 7 showcases the stable MRI features identified by AMCOT in various brain views, highlighting regions with high stability based on our selection methodology. These mapped regions offer crucial insights into AD’s pathological changes, serving as key biomarkers for monitoring progression and evaluating treatment efficacy, thereby facilitating advancements in early diagnosis and targeted therapies

6. CONCLUSION

This study introduces MTL-AMCOT, a multi-task learning framework that jointly models longitudinal dynamics and cross-score dependencies across ten heterogeneous cognitive objectives for AD progression prediction. By structuring inter-task associations into an adaptive multi-cognitive target temporal matrix and incorporating sparse group lasso regularisation, MTL-AMCOT surpasses multiple baseline models on the ADNI dataset. A dynamic ADMM penalty strategy further reduces average iterations by 43.61% and execution time by 19.22% compared with static ADMM, enabling practical use on large-scale longitudinal cohorts. Stability selection across multiple targets simultaneously identifies established biomarkers (hippocampus, entorhinal cortex, inferiorotemporal cortex) and potentially novel ones (isthmus cingulate, cerebellar cortex), providing directly interpretable clinical insights.

Several limitations remain. Inter-task constraints can induce negative transfer when

objectives are weakly coupled; the ablation study shows that RAVLT_f and TRABSCOR yield non-significant improvements from the target-relationship component ($p = 0.096$ and $p = 0.064$ respectively; Tables 9–10), consistent with their low inter-task correlations in Fig. 2. This is partly attributable to the use of Pearson correlation, which captures only linear dependencies; nonlinear measures such as distance correlation or kernel-based metrics are a natural extension. Furthermore, mean imputation may bias parameter estimates under informative missingness at later time points, where attrition is most pronounced (n declines from 1,532 at baseline to 289 at M48; Table 4). Robust MTL formulations, inverse-probability weighting, and missingness-aware MTL are identified as priorities for future work. Recent work has begun to address informative missingness explicitly — for instance, adaptive latent disease-state learning for biomarker detection under missing modalities [13] — and integrating such mechanisms into the AMCOT framework is a concrete direction for future work.

Beyond AD, AMCOT offers a general mechanism for capturing temporal correlations between heterogeneous targets. Future investigations will assess its utility in other neurodegenerative diseases such as Parkinson’s disease and non-neurological disorders such as depression, to confirm its effectiveness and adaptability across different disease contexts.

Acknowledgments

The authors thank the Alzheimer’s Disease Neuroimaging Initiative for providing data for this paper. This research was supported by the National Natural Science Foundation of China (No.62301452). Data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu).

Declarations

This manuscript uses generative AI tools primarily to improve readability and language (grammar, spelling, and style). All content has been reviewed and approved by the authors, who bears full responsibility for it.

Prior publication. This manuscript extends our conference paper presented at the 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) [53]. Compared with the conference version, we (i) introduce an ADMM-accelerated optimization scheme that improves training stability, and (ii) evaluate a broader set of cognitive scores to provide more comprehensive progression assessment.

References

- [1] P. Scheltens, B. De Strooper, M. Kivipelto, H. Holstege, G. Ch  telat, C. E. Teunissen, J. Cummings, W. M. van der Flier, Alzheimer’s disease, *The Lancet* 397 (10284) (2021) 1577–1590.
- [2] S. Gauthier, C. Webster, S. Servaes, et al., *World Alzheimer Report 2022: Life after diagnosis: Navigating treatment, care and support*, Alzheimer’s Disease International, London, England, 2022.
- [3] C. P. Hughes, L. Berg, W. Danziger, L. A. Coben, R. L. Martin, A new clinical scale for the staging of dementia, *The British journal of psychiatry* 140 (6) (1982) 566–572.
- [4] M. F. Folstein, L. N. Robins, J. E. Helzer, The mini-mental state examination, *Archives of general psychiatry* 40 (7) (1983) 812–812.
- [5] W. G. Rosen, R. C. Mohs, K. L. Davis, A new rating scale for alzheimer’s disease., *The American journal of psychiatry* 141 (11) (1984) 1356–1364.
- [6] M. Schmidt, *Rey auditory verbal learning test*, Western Psychological Services Los Angeles, 1996.
- [7] R. I. Pfeffer, T. T. Kurosaki, C. Harrah Jr, J. M. Chance, S. Filos, Measurement of functional activities in older adults in the community, *Journal of gerontology* 37 (3) (1982) 323–329.
- [8] J. D. Corrigan, N. S. Hinkeldey, Relationships between parts a and b of the trail making test, *Journal of clinical psychology* 43 (4) (1987) 402–409.

- [9] A. Khan, S. Zubair, Development of a three tiered cognitive hybrid machine learning algorithm for effective diagnosis of alzheimer's disease, *Journal of King Saud University-Computer and Information Sciences* 34 (10) (2022) 8000–8018.
- [10] L. M. Shaw, H. Vanderstichele, M. Knapik-Czajka, C. M. Clark, P. S. Aisen, R. C. Petersen, K. Blennow, H. Soares, A. Simon, P. Lewczuk, et al., Cerebrospinal fluid biomarker signature in alzheimer's disease neuroimaging initiative subjects, *Annals of neurology* 65 (4) (2009) 403–413.
- [11] P. Vemuri, C. R. Jack Jr, Role of structural mri in alzheimer's disease, *Alzheimer's research & therapy* 2 (4) (2010) 23.
- [12] J. Xu, C. Yuan, X. Ma, H. Shang, X. Shi, X. Zhu, Interpretable medical deep framework by logits-constraint attention guiding graph-based multi-scale fusion for alzheimer's disease analysis, *Pattern Recognition* 152 (2024) 110450.
- [13] Z. Chen, F. Zhang, Y. Zhang, J. Zhu, Q. Li, Y. Liu, Adaptive latent disease state learning for multimodal alzheimer's disease biomarker detection with missing modalities, *Pattern Recognition* (2025) 112389.
- [14] J. Zhou, J. Liu, V. A. Narayan, J. Ye, A. D. N. Initiative, et al., Modeling disease progression via multi-task learning, *NeuroImage* 78 (2013) 233–248.
- [15] P. Cao, X. Shan, D. Zhao, M. Huang, O. Zaiane, Sparse shared structure based multi-task learning for mri based cognitive performance prediction of alzheimer's disease, *Pattern Recognition* 72 (2017) 219–235.
- [16] Y. Zhang, Q. Yang, A survey on multi-task learning, *IEEE transactions on knowledge and data engineering* 34 (12) (2021) 5586–5609.
- [17] R. Nishihara, L. Lessard, B. Recht, A. Packard, M. Jordan, A general analysis of the convergence of admm, in: *International conference on machine learning*, PMLR, 2015, pp. 343–352.
- [18] B. He, X. Yuan, On the $o(1/n)$ convergence rate of the douglas–rachford alternating direction method, *SIAM Journal on Numerical Analysis* 50 (2) (2012) 700–709.

- [19] C. Song, S. Yoon, V. Pavlovic, Fast admm algorithm for distributed optimization with adaptive penalty, in: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), AAAI Press, Phoenix, Arizona, USA, 2016, pp. 753–759. [doi:10.1609/aaai.v30i1.10069](https://doi.org/10.1609/aaai.v30i1.10069).
- [20] C. R. Jack Jr, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, et al., The alzheimer’s disease neuroimaging initiative (adni): Mri methods, *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 27 (4) (2008) 685–691.
- [21] J. Zhou, L. Yuan, J. Liu, J. Ye, A multi-task learning formulation for predicting disease progression, in: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 2011, pp. 814–822.
- [22] M. Zhou, Y. Zhang, T. Liu, Y. Yang, P. Yang, Multi-task learning with adaptive global temporal structure for predicting alzheimer’s disease progression, in: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, 2022, pp. 2743–2752.
- [23] X. Liu, P. Cao, A. R. Gonçalves, D. Zhao, A. Banerjee, Modeling alzheimer’s disease progression with fused laplacian sparse group lasso, *ACM Transactions on Knowledge Discovery from Data (TKDD)* 12 (6) (2018) 1–35.
- [24] M. Zhou, P. Yang, Automatic temporal relation in multi-task learning, in: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023, pp. 3570–3580.
- [25] J. Skinner, J. O. Carvalho, G. G. Potter, A. Thames, E. Zelinski, P. K. Crane, L. E. Gibbons, A. D. N. Initiative, The alzheimer’s disease assessment scale-cognitive-plus (adas-cog-plus): an expansion of the adas-cog to improve responsiveness in mci, *Brain imaging and behavior* 6 (4) (2012) 489–501.
- [26] J. Zhou, J. Liu, V. A. Narayan, J. Ye, Modeling disease progression via fused

- sparse group lasso, in: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 2012, pp. 1095–1103.
- [27] N. Simon, J. Friedman, T. Hastie, R. Tibshirani, A sparse-group lasso, *Journal of computational and graphical statistics* 22 (2) (2013) 231–245.
 - [28] W. Liang, K. Zhang, P. Cao, X. Liu, J. Yang, O. R. Zaiane, Exploiting task relationships for alzheimer’s disease cognitive score prediction via multi-task learning, *Computers in Biology and Medicine* 152 (2023) 106367.
 - [29] F. Zheng, H. Wang, T. Liu, F. Feng, K. Wu, C. Liang, A. D. N. Initiative, et al., Structured knowledge-inspired two-stage knowledge alignment framework for alzheimer’s disease diagnosis, *Pattern Recognition* (2025) 112234.
 - [30] B. Lei, Y. Liang, J. Xie, Y. Wu, E. Liang, Y. Liu, P. Yang, T. Wang, C. Liu, J. Du, et al., Hybrid federated learning with brain-region attention network for multi-center alzheimer’s disease detection, *Pattern Recognition* 153 (2024) 110423.
 - [31] X.-A. Bi, W. Shen, D. Chen, W. Chen, L. Xu, X. Luo, L. Zhang, Agent propagation generative adversarial networks with large foundation models for alzheimer’s disease risk prediction, *Pattern Recognition* (2026) 113927.
 - [32] T. Wang, Q. Dai, H. Lu, Ce-ah: A contrast-enhanced attention hierarchical network for alzheimer’s disease diagnosis based on structural mri, *Pattern Recognition* 169 (2026) 111986.
 - [33] S. Mohsen, Alzheimer’s disease detection using deep learning and machine learning: a review, *Artificial Intelligence Review* 58 (9) (2025) 262.
 - [34] M. R. Shaikh, A. Jeyabose, R. V. Arjunan, Deep learning for alzheimer’s disease: advances in classification, segmentation, subtyping, and explainability, *BioMedical Engineering OnLine* 24 (1) (2025) 150.
 - [35] L. Barnard, H. Botha, N. Corriveau-Lecavalier, J. Graff-Radford, E. Dicks, V. Gogineni, G. Zhang, B. J. Burkett, D. R. Johnson, S. J. Huls, et al., An fdg-pet-based machine learning framework to support neurologic decision-making in alzheimer disease and related disorders, *Neurology* 105 (2) (2025) e213831.

- [36] G. Mårtensson, D. Ferreira, T. Granberg, L. Cavallin, K. Oppedal, A. Padovani, I. Rektorova, L. Bonanni, M. Pardini, M. G. Kramberger, et al., The reliability of a deep learning model in clinical out-of-distribution mri data: a multicohort study, *Medical Image Analysis* 66 (2020) 101714.
- [37] H. Xu, K. M. J. Shuttleworth, Medical artificial intelligence and the black box problem: a view based on the ethical principle of “do no harm”, *Intelligent Medicine* 4 (1) (2024) 52–57.
- [38] T. Alam, J. Suchi, A. Saha, F. Pathan, A. Das, M. M. Kabir, M. Mridha, Advances in alzheimer’s disease diagnosis with machine learning and deep learning techniques: a comprehensive review, *Artificial Intelligence Review* (2026).
- [39] M. Moritz, E. Topol, P. Rajpurkar, Coordinated ai agents for advancing healthcare, *Nature Biomedical Engineering* 9 (4) (2025) 432–438.
- [40] O. Freyer, S. Jayabalan, J. N. Kather, S. Gilbert, Overcoming regulatory barriers to the implementation of ai agents in healthcare, *Nature Medicine* 31 (10) (2025) 3239–3243.
- [41] S. G. Kwak, Lasso regression analysis: applications in dyslipidemia and cardiovascular disease research, *Journal of Lipid and Atherosclerosis* 14 (3) (2025) 289.
- [42] J. Zhang, H. Li, G. Fang, Z. Zhu, X. Li, Y. Chen, Z. Hou, W. Zhuang, Y. Liu, J. Wang, et al., Development of a lasso dynamic prediction system for interbody cage subsidence following olif surgery, *npj Digital Medicine* 8 (1) (2025) 656.
- [43] B. Wohlberg, Admm penalty parameter selection by residual balancing, *arXiv preprint arXiv:1704.06209* (2017).
- [44] P. Gong, J. Ye, C. Zhang, Robust multi-task feature learning, in: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012, pp. 895–903.
- [45] M. Nguyen, T. He, L. An, D. C. Alexander, J. Feng, B. T. Yeo, A. D. N. Initiative, et al., Predicting alzheimer’s disease progression using deep recurrent neural networks, *NeuroImage* 222 (2020) 117203.

- [46] M. Nguyen, N. Sun, D. C. Alexander, J. Feng, B. T. Yeo, Modeling alzheimer's disease progression using deep recurrent neural networks, in: 2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI), IEEE, 2018, pp. 1–4.
- [47] Y. Mu, F. H. Gage, Adult hippocampal neurogenesis and its role in alzheimer's disease, *Molecular neurodegeneration* 6 (1) (2011) 85.
- [48] B. T. Hyman, G. W. Van Hoesen, A. R. Damasio, C. L. Barnes, Alzheimer's disease: cell-specific pathology isolates the hippocampal formation, *Science* 225 (4667) (1984) 1168–1170.
- [49] K. M. Igarashi, Entorhinal cortex dysfunction in alzheimer's disease, *Trends in neurosciences* 46 (2) (2023) 124–136.
- [50] P. Bentley, J. Driver, R. J. Dolan, Modulation of fusiform cortex activity by cholinesterase inhibition predicts effects on subsequent memory, *Brain* 132 (9) (2009) 2356–2371.
- [51] S. M. Nestor, R. Rupsingh, M. Borrie, M. Smith, V. Accomazzi, J. L. Wells, J. Fogarty, R. Bartha, A. D. N. Initiative, Ventricular enlargement as a possible measure of alzheimer's disease progression validated using the alzheimer's disease neuroimaging initiative database, *Brain* 131 (9) (2008) 2443–2454.
- [52] T. Iram, F. Kern, A. Kaur, S. Myneni, A. R. Morningstar, H. Shin, M. A. Garcia, L. Yerra, R. Palovics, A. C. Yang, et al., Young csf restores oligodendrogenesis and memory in aged mice via fgf17, *Nature* 605 (7910) (2022) 509–515.
- [53] X. Fan, M. Zhou, Y. Zhang, J. Qi, Y. Yang, P. Yang, Adaptive multi-cognitive objective temporal task approach for predicting ad progression, in: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2024, pp. 1933–1938. [doi:10.1109/BIBM62325.2024.10822758](https://doi.org/10.1109/BIBM62325.2024.10822758).

Appendix A. Cholesky Decomposition

For (10), the Cholesky decomposition technique is employed due to its numerical robustness in solving linear systems, especially for positive definite matrices. Within the multi-task learning framework, this decomposition efficiently addresses regularization challenges related to coupling terms by enabling fast matrix factorization and inversion. The solution is formulated as:

$$X_i^T X_i w_i + \rho w_i + \rho W f_i = X_i^T y_i - u_i + \rho \theta_i - V g_i + \rho \mathfrak{R} g_i, \quad (\text{A.1})$$

where X_i is the feature matrix for the i -th task, w_i denotes the parameter vector, y_i is the response vector, and ρ is the regularization parameter.

For simplification, define:

$$H_i w_i = q_i, \quad \forall i \in \{1, \dots, t\}, \quad (\text{A.2})$$

where:

$$H_i = X_i^T X_i + \rho(1 + F_{ii})I_{p \times p}, \quad (\text{A.3})$$

$$q_i = X_i^T y_i - u_i + \rho \theta_i - \rho V g_i + \rho \mathfrak{R} g_i - \rho \sum_{k=1, k \neq i}^t w_k F_{ki}. \quad (\text{A.4})$$

Appendix B. Derivation of the AMCOT Penalty Term $\mathfrak{R}$

The combined temporal–target error decomposes as follows. Assuming minimal differences between adjacent time points ($y_{n,\tau+1}^i \approx y_{n,\tau+1}^j$) and that highly correlated tasks share similar scores ($y_n^i \approx c(i, j) \cdot y_n^j$):

$$\begin{aligned} \mathfrak{R} &= \sum_{i=1}^O \mathfrak{R}_i \\ &= \sum_{\substack{i=1 \\ i \neq j}}^O \left(\sum_{\tau=1}^T (y_{n,\tau}^i - c^{(i,j)} \cdot y_{n,\tau}^j) + \sum_{\tau=1}^{T-1} (y_{n,\tau+1}^i - y_{n,\tau}^i) \right) \\ &= \sum_{\substack{i=1 \\ i \neq j}}^O X^i \left(\sum_{\tau=1}^T (W^\tau - M_\tau^{(i,j)} \cdot W^j) + \sum_{\tau=1}^{T-1} (W_{\tau+1} - W_\tau) \right), \end{aligned} \quad (\text{B.1})$$

which is exactly the AMCOT penalty $\|\mathfrak{R}\|$ in Eq. (6).

Appendix C. Algorithm Pseudocode

Algorithm 1: Construction of the R Matrix for Temporal–Target Task Associations

```

1: Input:  $e_1, e_2, e_3, T$  (temporal and target relevance between tasks)
2: Output:  $R$ 
3: Initialize  $H_1, \dots, H_o \leftarrow 0 \in \mathbb{R}^{T \times (T-1)}$ 
4: for  $i = 1 \rightarrow T - 1$  do
5:   Set  $H_1, H_2, \dots, H_o$  to  $[e_1, -e_1]$ 
6: end for
7: Initialize matrix  $M$  as a zero matrix of size  $|O| \times |O|$ 
8: for  $\tau = 1 \rightarrow |T|$  do
9:    $M_\tau^{i,j} = c_\tau^{i,j}$ 
10: end for
11: for  $i = 1 \rightarrow |O|$  do
12:   for  $j = 1 \rightarrow |O|$  do
13:     if  $i \neq j$  then
14:        $R_\tau^{i,j} = M_\tau^{i,j} \times H_\tau$ 
15:     else
16:        $R_\tau^{i,j} = |O| \times M_\tau^{i,j} \times H_\tau$ 
17:     end if
18:   end for
19: end for

```

Algorithm 2: ADMM-Based Algorithm with Dynamic Adjustment of Quadratic Penalty Coefficient

```

1: Input:  $X, Y, \lambda_1, \lambda_2, \lambda_3$ 
2: Output:  $W$ 
3: Initialize  $W_0, \Theta, \mathfrak{R}, U, V \leftarrow 0, \tau \leftarrow 1, \rho_0 \leftarrow 0.01$ 
4: repeat
5:   if  $iter > 1$  then
6:     Compute  $r^{k+1}, s^{k+1}$  via (16),(17)
7:     Update  $\rho^{k+1}$  via (18)
8:   end if
9:   Update  $W^{k+1}, \Theta^{k+1}, \mathfrak{R}^{k+1}, U^{k+1}, V^{k+1}$  via (10)–(15)
10:  if  $L_k > L_{k-1}$  then
11:    Backtrack to  $(k - 1)$ th, stop adjusting  $\rho^k$ , set  $\tau \leftarrow 1.618$ 
12:  end if

```

13: **until** convergence

Appendix D. Derivation of the W -Update Step

Starting from Eq. (10), the augmented Lagrangian sub-problem for W can be written in scaling form as:

$$W^{k+1} = \arg \min_W \frac{1}{2} \|XW - Y\|_F^2 + \frac{\rho}{2} \|W - (\Theta + \frac{U}{\rho})\|_F^2 + \frac{\rho}{2} \|WR - (\mathfrak{R} + \frac{V}{\rho})\|_F^2. \quad (\text{D.1})$$

Setting the gradient with respect to W to zero yields the linear system:

$$X^T XW + \rho W + \rho WF = X^T Y - U + \rho \Theta - VG + \rho \mathfrak{R}G, \quad (\text{D.2})$$

where $F = HH^T$ and $G = H^T$. This system is solved efficiently per task using Cholesky decomposition, exploiting the positive-definiteness of the coefficient matrix $X_i^T X_i + \rho(1 + F_{ii})I_{p \times p}$.

Appendix E. Full Statistical Significance Results for Ablation Study

Table E.11 reports the exact one-tailed p -values for all comparisons in the ablation study (Table 8). Tests were conducted across the 10 cross-validation folds using a paired t -test when per-fold differences satisfied Shapiro–Wilk normality ($p \geq 0.05$), and the Wilcoxon signed-rank test otherwise.

Table E.11: Exact one-tailed p -values for Full AMCOT vs. each ablation variant.

Score	w/o Tgt. Rel.		w/o Temp. Sm.	
	nMSE	wR	nMSE	wR
CDRSB	$5.60 \times 10^{-4}***$	$7.30 \times 10^{-4}***$	$1.19 \times 10^{-7}***$	$9.77 \times 10^{-4}***$
ADAS11	$2.83 \times 10^{-3}**$	$6.61 \times 10^{-4}***$	$1.40 \times 10^{-7}***$	$3.06 \times 10^{-8}***$
ADAS13	$2.18 \times 10^{-4}***$	$2.96 \times 10^{-4}***$	$1.53 \times 10^{-8}***$	$5.15 \times 10^{-8}***$
MMSE	$8.21 \times 10^{-3}**$	$2.60 \times 10^{-3}**$	$3.42 \times 10^{-6}***$	$8.48 \times 10^{-7}***$
RAVLT _i	$9.35 \times 10^{-7}***$	$2.56 \times 10^{-7}***$	$9.77 \times 10^{-4}***$	$6.51 \times 10^{-11}***$
RAVLT _l	$1.65 \times 10^{-3}**$	$4.56 \times 10^{-4}***$	$8.16 \times 10^{-7}***$	$7.93 \times 10^{-8}***$
RAVLT _f	$9.64 \times 10^{-2}^\dagger$	$1.06 \times 10^{-2}^*$	$2.65 \times 10^{-8}***$	$2.11 \times 10^{-8}***$
RAVLT _{pf}	$8.62 \times 10^{-3}**$	$1.27 \times 10^{-3}**$	$1.22 \times 10^{-7}***$	$9.77 \times 10^{-4}***$
FAQ	$7.98 \times 10^{-4}***$	$3.39 \times 10^{-4}***$	$2.02 \times 10^{-8}***$	$1.34 \times 10^{-8}***$
TRABSCOR	$6.40 \times 10^{-2}^\dagger$	$6.23 \times 10^{-3}**$	$4.18 \times 10^{-5}***$	$3.98 \times 10^{-6}***$

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, and $^\dagger p \geq 0.05$.

Paired t -test was used when per-fold differences passed Shapiro–Wilk normality; otherwise, the Wilcoxon signed-rank test was used. All tests were one-tailed.