

## Research and Applications

# HiMAL: Multimodal Hierarchical Multi-task Auxiliary Learning framework for predicting Alzheimer's disease progression

Sayantana Kumar , MS<sup>1,2,\*</sup>, Sean C. Yu , PhD<sup>2</sup>, Andrew Michelson, MD<sup>2,3</sup>, Thomas Kannampallil , PhD<sup>1,2,4</sup>, Philip R.O. Payne , PhD<sup>1,2</sup>,  
for the Alzheimer's Disease Neuroimaging Initiative

<sup>1</sup>Department of Computer Science and Engineering, McKelvey School of Engineering, Washington University in St. Louis, St. Louis, MO 63130, United States, <sup>2</sup>Institute for Informatics, Data Science and Biostatistics, Washington University School of Medicine, St. Louis, MO 63110, United States, <sup>3</sup>Division of Pulmonary and Critical Care, Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110, United States, <sup>4</sup>Department of Anesthesiology, Washington University School of Medicine, St. Louis, MO 63110, United States

\*Corresponding author: Sayantan Kumar, MS, Institute for Informatics, Data Science and Biostatistics, Washington University School of Medicine, 660 South Euclid Ave, Campus Box 8132, St. Louis, MO 63110, United States (sayantan.kumar@wustl.edu)

Thomas Kannampallil and Philip Payne have contributed equally and are considered co-senior authors of this work.

## Abstract

**Objective:** We aimed to develop and validate a novel multimodal framework **Hierarchical Multi-task Auxiliary Learning (HiMAL)** framework, for predicting cognitive composite functions as auxiliary tasks that estimate the longitudinal risk of transition from Mild Cognitive Impairment (MCI) to Alzheimer's Disease (AD).

**Materials and Methods:** HiMAL utilized multimodal longitudinal visit data including imaging features, cognitive assessment scores, and clinical variables from MCI patients in the Alzheimer's Disease Neuroimaging Initiative dataset, to predict at each visit if an MCI patient will progress to AD within the next 6 months. Performance of HiMAL was compared with state-of-the-art single-task and multitask baselines using area under the receiver operator curve (AUROC) and precision recall curve (AUPRC) metrics. An ablation study was performed to assess the impact of each input modality on model performance. Additionally, longitudinal explanations regarding risk of disease progression were provided to interpret the predicted cognitive decline.

**Results:** Out of 634 MCI patients (mean [IQR] age: 72.8 [67-78], 60% male), 209 (32%) progressed to AD. HiMAL showed better prediction performance compared to all state-of-the-art longitudinal single-modality single-task baselines (AUROC = 0.923 [0.915-0.937]; AUPRC = 0.623 [0.605-0.644]; all  $P < .05$ ). Ablation analysis highlighted that imaging and cognition scores with maximum contribution towards prediction of disease progression.

**Discussion:** Clinically informative model explanations anticipate cognitive decline 6 months in advance, aiding clinicians in future disease progression assessment. HiMAL relies on routinely collected electronic health records (EHR) variables for proximal (6 months) prediction of AD onset, indicating its translational potential for point-of-care monitoring and managing of high-risk patients.

## Lay Summary

Alzheimer's Disease (AD), neurodegenerative disorder, results in memory loss and cognitive decline, affecting millions of individuals globally. For patients with mild cognitive impairment (MCI), a prodromal phase of AD, it is important to monitor their risk of progressing to AD over time. Using longitudinal multimodal data on imaging, cognition and clinical data from a publicly available AD dataset, we developed and validated **HiMAL** a novel multimodal **Hierarchical Multi-task Auxiliary Learning** framework to estimate at every visit timepoint if a MCI patient will progress to AD within the next 6 months. HiMAL predicted forecasted neuropsychological composite function scores as auxiliary tasks at every timepoint and used the predicted forecasted scores to predict the main task of predicting progression. The hierarchical approach allows HiMAL to effectively use task dependencies, resulting in more informative features and better performance than state-of-the-art multimodal and multi-task methods. Additionally, HiMAL provided longitudinal explanations that can inform clinicians 6 months in advance the potential cognitive function decline that can lead to progression to AD in future. Built on routinely collected EHR data, HiMAL's flexibility in the selection of input modalities and auxiliary tasks suggest that it can be applied to various AD datasets and other clinical problems.

**Key words:** Alzheimer's disease progression; multimodal; multitask auxiliary learning; hierarchical; model explainability.

## Background and significance

Alzheimer Disease (AD) is a neurodegenerative disorder characterized by loss of memory and impaired cognition which affects millions of people worldwide.<sup>1-3</sup> AD-related brain

pathology, which includes the accumulation and deposition of amyloid- $\beta$  peptide and tau protein often occurs several years prior to the onset of clinical symptoms, making it challenging to detect AD early.<sup>4,5</sup> Mild cognitive impairment

Received: June 14, 2024; Revised: August 27, 2024; Editorial Decision: August 29, 2024; Accepted: August 30, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of the American Medical Informatics Association.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

(MCI) is considered a prodromal phase of AD that gradually progresses to AD over several years.<sup>6,7</sup> As only 10%-20% of MCI patients progress to an AD diagnosis, it is challenging to differentiate between stable MCI (sMCI) patients who do not progress to AD and progressive MCI (pMCI) patients who develop AD over time.<sup>8-10</sup>

Statistical techniques like linear and non-linear mixed effect models and survival analysis are common techniques in analyzing AD progression.<sup>11-14</sup> However, with the widespread applications of machine learning (ML) and deep learning (DL) models in clinical decision support systems, several studies have focused on developing ML and DL-based models for predicting the transition of patients from MCI to AD.<sup>15-21</sup> However, the existing research has three major limitations. First, much of the research has relied on single modality data such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), clinical data or genetic information.<sup>22-24</sup> However, the causal underpinnings of AD is multifactorial disease requiring multimodal data, with each data modality provides unique and complementary information, cumulatively enabling a more comprehensive understanding of disease pathophysiology and progression.<sup>9,19,25</sup> Second, most of the prior models were single-task models, predicting only a specific event such as sMCI versus pMCI or cognitive scores at a future timepoint.<sup>9,26,27</sup> However, at the point-of-care, clinicians have to determine the progression of the disease with concurrent prediction of multiple adverse events such as a future clinical diagnosis and forecasting the progression of cognitive performance over the natural course of the disease.<sup>22,28,29</sup> Finally, most existing research has adopted a cross-sectional prediction approach, relying on baseline data for predicting risk of progression.<sup>9,30</sup> However, the probability of progression to AD is dynamic,<sup>31</sup> and as such, models need to be flexible to make reliable, longitudinal predictions.

Our objective in this research is to utilize longitudinal data from multimodal EHR data sources such as imaging, cognition measures, and clinical variables to predict at every visit if a MCI patient will progress to AD within the next 6 months. Towards this end, we had research objectives: (1) develop a novel hierarchical multi-task DL model Hierarchical Multi-task Auxiliary Learning (HiMAL), that can predict forecasted neuropsychological composite functions as auxiliary tasks to estimate the main task of predicting progression risk and (2) provide longitudinal explanations about the potential factors behind disease progression. Compared to existing multimodal multitask studies on AD progression analysis where the tasks are predicted independently,<sup>20</sup> the hierarchical relation allows HiMAL to exploit task dependencies more effectively, leading to more informative feature representations and effective knowledge transfer between the tasks. To the best of our knowledge, HiMAL is the first hierarchical multitask multimodal DL framework that can jointly predict and explain longitudinal clinical outcomes within the same end-to-end framework.

## Methods

### Data sources

Data were sourced from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, established in 2003 through a public-private partnership.<sup>32</sup> ADNI aims to investigate the utility of combining various neuroimaging biomarkers and

neuropsychological assessments for tracking the progression of MCI and early Alzheimer's disease (AD). All participants in ADNI provided written informed consent, and study protocols were approved by respective institutional review boards at each site. Additional details about ADNI, including study protocols, eligibility criteria, and data access, can be found at <http://www.adni-info.org/>

### Cohort selection

We analyzed longitudinal data from ADNI1, ADNI2, ADNI GO, and ADNI3. Participants were selected if diagnosed with "MCI," "EMCI" (Early MCI), or "LMCI" (Late MCI) at baseline, with available T1-weighted MRI scans, cognitive test scores, and clinical data. MCI subjects reverting to Cognitively Normal (CN) during follow-up were excluded, due to the potential uncertainty of the clinical diagnosis, considering that AD is an irreversible form of dementia. MCI subjects progressing to AD within 5 years from baseline visit were labeled as pMCI; others as sMCI.<sup>33</sup> For sMCI, all available imaging and cognition visits within 5 years from baseline were included; for pMCI, visits until progression were selected. All cognition visits were aligned with imaging visits. Note that clinical variables were extracted only from baseline visits for both sMCI and pMCI samples. The cohort was split into training, validation, and test sets in a ratio of 75:15:10 for model training, hyperparameter tuning, and evaluation, respectively.

### Variables and data processing

#### Multimodal input feature selection

Imaging data used in our analysis included preprocessed MRI brain region volumes (gray matter volumes) obtained from T1-weighted MRI scans. We included 90 imaging features including volumes of 66 cortical regions (33 per hemisphere) and 24 subcortical regions (12 per hemisphere). Further details about how the regional brain volumes are extracted from MRI scans be found both online (<https://adni.loni.usc.edu/methods/mri-tool/mri-analysis/>)<sup>34</sup> and in the [Supplementary Material \(Section S1.1; Table S1\)](#). Following existing studies on AD progression, we included 12 neuropsychological cognitive assessments scores and 18 clinical features including demographics of participants (age, sex), vital signs and medical history (comorbidities). The full list of cognitive assessment scores and clinical variables can be found in the [Supplementary Material \(Section 1.2; Table S2 and Section 1.3; Table S3\)](#).

#### Composite cognitive functions as auxiliary tasks

We used 4 neuropsychological composite cognitive functions as our auxiliary tasks which have been validated to have associated with the risk of progressing from MCI to AD. These included composite scores for different domains of cognitive performance, namely memory (ADNI-MEM),<sup>35</sup> composite score for executive functioning (ADNI-EF),<sup>36</sup> composite score for language cognitive terms (ADNI-LAN) and composite score for visual-spatial (ADNI-VS).<sup>37</sup> Further, the use of a composite score increased measurement precision, helps avoid idiosyncratic features of a particular test that may capitalize on chance, and limits the number of statistical tests needed compared to analyzing each of the constituent parts separately.

### Data preprocessing

All imaging features of each patient were normalized by the intracranial volume (ICV) of that patient. Feature preprocessing of continuous time-series variables (imaging, cognition measures, and composite functions) included clipping the outlier values to the fifth and 95th percentile values and scaling between [0,1] using the MinMaxScalar package from sklearn. The training set was first scaled and the parameters from the scaled training set were used to standardize the validation and test set, respectively. All categorical features were one-hot encoded.

### Missing value imputation

Longitudinal data for each participant were collected at an interval of 6 months to 5 years from the baseline visit. However, not all participants had all modalities recorded at every visit. Missing values for all continuous time-series variables were imputed using the BRITS algorithm, a state-of-the-art imputation based on recurrent dynamics.<sup>38</sup> For more details about the BRITS algorithm and how missing values are imputed in our model, we encourage the reader to check the [Supplementary Material Section 2.2](#). Missing values for categorical variables were filled using the mode value of each variable. Following prevalent techniques in AD literature,<sup>39</sup> diagnosis information (MCI or AD) for each participant was forward filled from their previous recorded visit, if missing.

### Ground truth labels for clinical outcome and composite cognitive scores

For the primary task, we aimed to predict the risk of AD progression at each timepoint within the visit trajectory of an individual. Specifically, HiMAL predicted at each timepoint (visit) the probability (risk) of a subject progressing to AD by their next visit (within 6 months from current visit). The

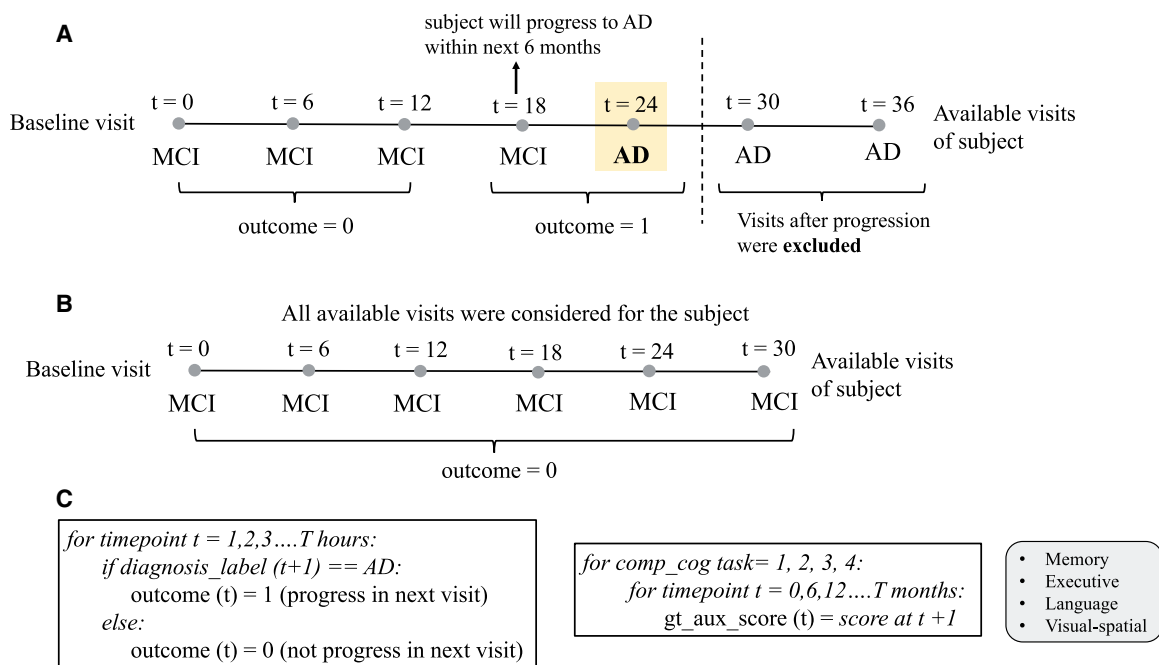
ground truth labels for the primary task were calculated as follows (Figure 1). At each timepoint, if the diagnosis label is AD for the next visit, we labelled that visit with outcome 1 or else 0. For example, if a pMCI subject progresses after 24 months from the baseline visit, the labels for bl (baseline), m06 (month 6), m12 (month 12) will be 0; while the labels for m18 (month 18), and m24 (month 24) will be 1 (Figure 1). Similar to the primary task, the predicted auxiliary tasks were also anticipated longitudinal predictions. Specifically, HiMAL predicted at each timepoint (visit) the forecasted composite scores at the next visit or after 6 months (Figure 1).

### Hierarchical multi-task auxiliary learning framework

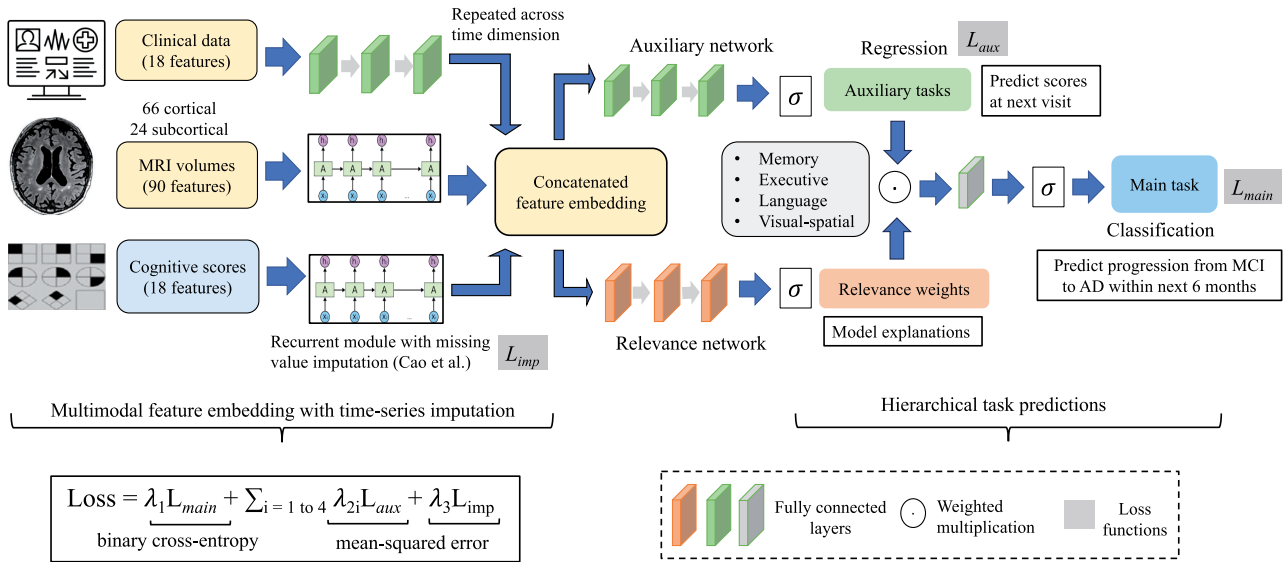
Our proposed HiMAL framework can be divided into 3 main components: (1) multimodal feature embedding with time-series embedding, (2) hierarchical task predictions, and (3) model explanations (Figure 2).

### Multimodal feature embedding with time-series imputation

HiMAL takes baseline clinical data, longitudinal MRI gray matter volumes (imaging data) and longitudinal cognitive assessment scores as the multimodal input. The time-invariant clinical features are passed through a set of fully connected layers to obtain the clinical embedding. Similarly, the multivariate time-series features (imaging and cognition data) are passed through a recurrent module to estimate the corresponding imaging and cognition embeddings, respectively. Note that the recurrent module consists of a series of Long Short-Term Memory<sup>40</sup> layers along with provisions for missing value imputation using the BRITS algorithm. BRITS is a data-driven imputation procedure where the missing values were imputed based on recurrent dynamics.<sup>38</sup> More details about missing value imputation can be found in the



**Figure 1.** Ground-truth label for clinical outcome (main task) and composite cognitive scores (auxiliary tasks) mortality. (A) Example pMCI subject who had available visits till 3 years from baseline ( $t=36$ ) and progressed at  $t=24$ . Outcome labels were 0 for  $t=0$  until  $t=12$  and 1 for  $t=18$  (progress at next visit) and  $t=24$ . Visits after progression ( $t=24$  onwards) were excluded. (B) Example sMCI who never progressed. All available visits till  $t=30$  were considered with all visits marked outcome label 0. (C) Pseudocode showing how ground truth labels are calculated for prediction of progression (left) and the 4 composite cognitive scores (right).



**Figure 2.** HiMAL: multimodal HiMAL framework for predicting risk of MCI to AD progression using forecasted cognitive composite scores. Time-series input features (imaging and cognition data) passed through recurrent modules with missing value imputation and time-invariant clinical data passed through a series of fully connected layers were concatenated into a multimodal feature embedding. The concatenated feature embedding was passed through 2 separate networks to predict the auxiliary tasks and the corresponding relevance weights at each timepoint. The main task was estimated by the weighted combination of the auxiliary scores and the relevance weights followed by a fully connected layer and sigmoid activation function.

Supplementary Material Section S2. Next, the embeddings from the individual modalities were concatenated to obtain the joint embedding space (Figure 2).

### Hierarchical task predictions

The concatenated multimodal feature embedding was subsequently passed through an auxiliary network of fully connected layers followed by a sigmoid activation function to predict the memory, executive functioning, language, and visual-spatial composite scores as the four auxiliary tasks (Figure 2). The main task of predicting MCI to AD progression was estimated by a weighted combination of the auxiliary tasks. Specifically, the predicted composite scores were multiplied with the relevance weights and passed through a fully connected layer, followed by a sigmoid activation function to predict the risk of progression. The relevance weights were learnt by passing the concatenated embedding through relevance network of fully connected layers followed by a sigmoid activation function. Similar to the main and auxiliary tasks, the relevance weights were also predicted for every visit timepoint of each participant in our cohort (Figure 2).

HiMAL differs from existing multitask learning methods where the primary and auxiliary tasks were predicted concurrently without any hierarchical dependency between them. The hierarchical nature of HiMAL allowed the model to exploit task dependencies more effectively, leading to more informative feature representations and effective knowledge transfer between the tasks.

### Model explanations

The relevance weights represented the contribution (importance) of each forecasted composite score towards the main task of predicting progression. Since progression prediction was calculated as a function of the forecasted composite scores at each timepoint, we aimed to analyze which of the

composite scores are important predictors of progression from MCI to AD. The relevance weights and the forecasted composite scores at each timepoint provided insights about how the contributions of different composites change over time from baseline visit to the onset of progression. For example, HiMAL can provide an explanation (at a particular timepoint) that the participant will probably progress to AD in the next 6 months due to possible decline in memory and executive functioning. Since the weights were jointly optimized along with the main and auxiliary tasks, the explanations were provided along with prediction within the same end-to-end framework, without relying on post hoc explanation techniques (Figure 2).

### Model training and evaluation

#### Loss function

The loss function of HiMAL can be written as:  $L = \lambda_1 L_{main} + \lambda_2 \sum_{c=1}^4 L_{aux} + \lambda_3 L_{imp}$  where  $L_{main}$  and  $L_{aux}$  are the binary cross-entropy loss for the main task classification and the mean-squared error for the auxiliary task regression, respectively. The imputation loss  $L_{imp}$  was also calculated using the mean-squared difference between the original and imputed values (for details, see<sup>38</sup> and Supplementary Material Section 2.2).  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  represent the coefficients of the different losses, respectively, and were set to  $\lambda_1 = 1$ ,  $\lambda_2 = 0.8$ , and  $\lambda_3 = 0.05$ , respectively, after hyperparameter tuning using grid search on the validation set.

#### Comparison with baselines

We compared the performance of HiMAL with validated state-of-the-art longitudinal benchmark models for AD progression prediction published in recent years. These models have been grouped into 4 categories as follows based on the number of modalities used (unimodal and multimodal) and the number of tasks optimized (single-task/multitask): (1) single-modality single task models (SMST),<sup>41</sup> (2) single-

modality multitask models (SMMT),<sup>42,43</sup> (3) multimodal single task models (MMST),<sup>25,44,45</sup> and (4) multimodal multitask models (MMMT).<sup>46–49</sup> The architectural designs of all these state-of-the-art longitudinal baseline models in the context of our data and tasks (both main and auxiliary) have been presented in the [Supplementary Material \(Figure S1\)](#). For models with a single modality (SMMT, SMST), the input can be either of the 3 modalities (imaging only, cognition only and clinical only). The models having multiple tasks (SMMT, MMT), predicted both the main and auxiliary tasks in a traditional multitask auxiliary learning approach where the main and auxiliary tasks are predicted independently in parallel ([Figure S1](#)). We also included traditional-sectional ML baselines like Support Vector Machines (SVM),<sup>50</sup> Random Forests (RF),<sup>51</sup> and XGBoost (XGB)<sup>52</sup> as baselines for comparison. We implemented SVM, RF, and XGB as cross-sectional single-modality models, such that they received all modalities and all timepoints in an aggregated way to predict progression from MCI to AD.

### Statistical analysis

All models (HiMAL and baselines) were trained with the same set of multimodal features and network parameters. Details of the parameter configurations are available in [Supplementary Material Table S4](#). We reported both the Area Under the Receiver Operating Characteristics (AUROC)<sup>53</sup> and the Area Under the Precision Recall Curve (AUPRC)<sup>54</sup> for the main task of progression prediction and Mean Squared Error (MSE) for the composite score regression. For all reported results, we computed the 95% confidence intervals (CIs) using the pivot bootstrap estimator<sup>55</sup> by sampling participants from the test dataset with replacement 200 times. We also performed the 2-sided Wilcoxon signed rank tests<sup>56</sup> to pairwise compare HiMAL and each of the baseline methods for the main task. The critical value  $P = .05$  was selected and the final  $P$  values were reported after false discovery rate (FDR) correction to adjust for multiple hypotheses.<sup>57</sup> We also performed an ablation study to understand the impact of each input modality on the performance of HiMAL. Specifically, the performance metrics on the main

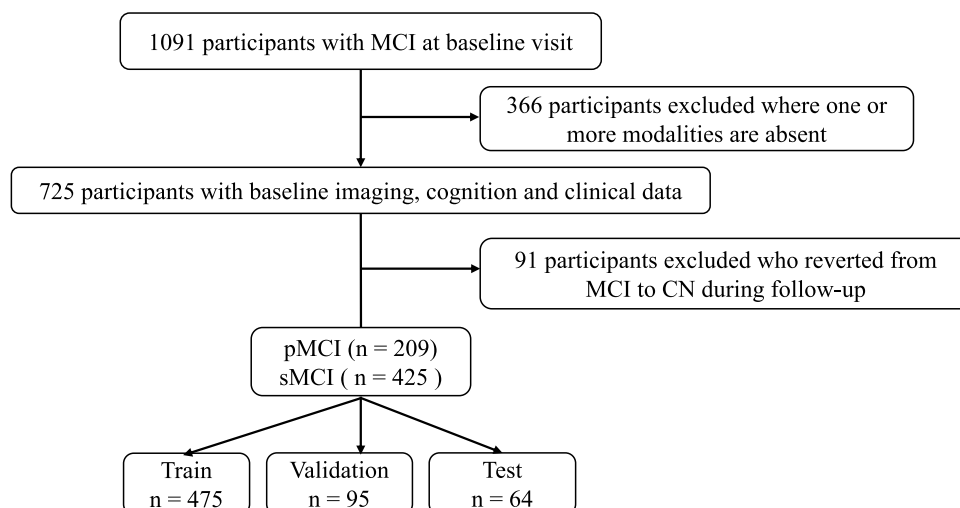
task (AUROC/AUPRC) and the auxiliary tasks (MSE) were compared for different combination of the 3 modalities.

## Results

[Figure 3](#) shows a flowchart of cohort selection including the inclusion and exclusion criteria. We first identified 1091 MCI subjects out of which 725 subjects were selected with simultaneous availability of T1-weighted MRI scan, neuropsychological cognitive test scores, and clinical data at the baseline visit. 91 subjects, who were clinically diagnosed as MCI at baseline visit but reverted to CN during follow-up were excluded from the study. Our final cohort consisted of 425 sMCI samples and 209 pMCI samples. The training, validation and test set included 475, 95, and 64 subjects, respectively.

### Sample characteristics

Descriptive statistics corresponding to the baseline visits for pMCI and sMCI cohorts, respectively, are shown in [Table 1](#). Our cohort consisted of more MCI individuals who remained stable in their trajectory (sMCI;  $n = 425$ ) compared to those who progressed (pMCI;  $n = 209$ ). The sMCI cohort had relatively younger individuals (mean  $\pm$  SD:  $[72.5 \pm 5.4]$  vs  $[73.7 \pm 5.26]$ ;  $P = .037$ ) and higher proportion of males ( $252/425 = 59\%$  vs  $115/209 = 55\%$ ;  $P = .007$ ) compared to the pMCI cohort. The level of education (in years) was slightly higher for the pMCI cohort (mean  $\pm$  SD:  $[16.2 \pm 2.55]$  vs  $[15.4 \pm 2.27]$ ), though the difference was not statistically significant. The pMCI cohort had less AD pathology than sMCI indicated by the generally high MMSE scores (mean  $\pm$  SD:  $[28.7 \pm 1.7]$  vs  $[24.8 \pm 2.1]$ ;  $P = .0035$ ). In terms of the composite cognitive scores, pMCI individuals exhibited more impairment (lower values) in memory (mean  $\pm$  SD:  $[-0.98 \pm 0.68]$  vs  $[0.52 \pm 0.81]$ ;  $P < .001$ ), executive functioning (mean  $\pm$  SD:  $[1.1 \pm 0.85]$  vs  $[0.53 \pm 0.76]$ ;  $P < .001$ ), language (mean  $\pm$  SD:  $[-1.01 \pm 1.1]$  vs  $[0.48 \pm 0.82]$ ;  $P = .008$ ), and visual-spatial skills (mean  $\pm$  SD:  $[-0.76 \pm 1.07]$  vs  $[0.07 \pm 0.74]$ ;  $P < .001$ ) compared to the sMCI cohort.



**Figure 3.** Flowchart of study participants. Abbreviations: CN = cognitively normal; MCI = mild cognitive impairment; pMCI = progressive MCI; sMCI = stable MCI.

## Model performance

HiMAL outperformed all baseline methods including traditional ML classifiers, single task and unimodal baselines in terms of both AUROC and AUPRC (Table 2). The non-DL models SVM, RF, and XGBoost were cross-sectional ignoring the temporal aspect of the data, and hence exhibited a general trend of lower prediction performance compared to their DL counterparts. Overall, we observed that using multiple modalities is beneficial compared to using only a single modality as demonstrated by the superior performance of MMST and MMT over SMST and SMMT. For a particular modality configuration (eg, SMST vs SMMT or MMST vs MMT), multitask models performed better than their single-task counterparts. The statistically performance improvement of HiMAL over MMT (Table 2;  $P = .037$ ) indicates that the hierarchical connection between tasks allowed HiMAL to model task dependencies more effectively, leading to more informative feature representations and effective knowledge transfer between the tasks.

We observed similar trends of HiMAL performing better (lower MSE) than the unimodal and single-task baselines for the auxiliary task regression (Table 3). HiMAL had the

**Table 1.** Descriptive statistics of the pMCI and sMCI cohorts, respectively.

Variables	Total cohort	pMCI	sMCI
Number of samples, $n$ (%)	634 (100%)	209 (33%)	425 (67%)
Age (years), mean $\pm$ SD	72.8 $\pm$ 4.75	73.7 $\pm$ 5.26	72.5 $\pm$ 5.4
Male: Female	375:259	115:94	252:173
Education (years), mean $\pm$ SD	15.9 $\pm$ 2.76	15.4 $\pm$ 2.27	16.2 $\pm$ 2.55
<sup>a</sup> MMSE, mean $\pm$ SD	27.5 $\pm$ 1.9	24.8 $\pm$ 2.1	28.7 $\pm$ 1.7
<sup>b</sup> Memory, mean $\pm$ SD	0.13 $\pm$ 0.93	-0.98 $\pm$ 0.68	0.52 $\pm$ 0.81
<sup>b</sup> Executive, mean $\pm$ SD	0.125 $\pm$ 1.07	-1.1 $\pm$ 0.85	0.53 $\pm$ 0.76
<sup>b</sup> Language, mean $\pm$ SD	0.07 $\pm$ 0.91	-1.01 $\pm$ 1.1	0.48 $\pm$ 0.82
<sup>c</sup> Visual-spatial, mean $\pm$ SD	-0.12 $\pm$ 0.85	-0.76 $\pm$ 1.07	0.07 $\pm$ 0.74

All variables corresponded to the baseline visit.

<sup>a</sup> Range from [0,30], lower scores indicating more impairment.

<sup>b</sup> Range from [-3,3] with lower values indicating more impairment.

<sup>c</sup> Range from [-3,1] with lower values indicating more impairment.

Abbreviations: MMSE = Mini-Mental State Examination; SD = standard deviation.

lowest MSE for memory and language and were second best for executive functioning and visual-spatial. Overall, multimodal models exhibited better performance for all the composite cognitive scores compared to their unimodal counterparts as demonstrated by the superior performance of MMST and MMT over SMST and SMMT, respectively (Table 3). We observed a general trend of multitask models performing better than single-task models, which was consistent across all the four composite scores.

## Model explanations

The forecasted composite scores and the corresponding relevance scores at each timepoint provided insights (explanations) about how the contributions of different composites change over time from baseline visit to the onset of progression. Here we visualized the longitudinal explanations (every 6 months) for a particular pMCI individual who progressed 48 months after the first baseline visit (Figure 4). At each timestep (visit), the outcome (progression) labels from  $t=0$  to  $t=36$  were 0 (not progress) and 1 (progress) for  $t=42$  and  $t=48$ . As the predicted probability of progression rises from  $t=36$ , the model was shown to pay more relevance to memory and language composites (Figure 4). In other words, HiMAL predicted that the MCI patient will progress to AD within the next 6 months because due to forecasted decline cognitive skills related to memory and language (Figure 4). Low relevance scores assigned to executive functioning and visual-spatial can be attributed to the fact that these scores were relatively constant with negligible changes throughout the trajectory (Figure 4). The cognitive trajectory of a sMCI patient who did not progress indicate no significant decline in any of the cognitive functions and hence were assigned very low relevance scores (Supplementary Material; Figure S2).

## Ablation studies on the number of modalities (HiMAL)

Using all 3 modalities (imaging + cognition + clinical) as input resulted in the best prediction performance of HiMAL for both AD progression prediction (AUROC, AUPRC) and composite cognitive score regression (MSE) (Figure 5).

**Table 2.** Performance on the main task (AD progression prediction) in terms of AUROC (higher = better) and AUPRC (higher = better) for HiMAL and the baselines.

Model	AUROC	*P-value (AUROC)	AUPRC	*P-value (AUPRC)
SVM	0.714 [0.705-0.728]	<.001 <sup>***</sup>	0.416 [0.405-0.431]	<.001 <sup>***</sup>
RF	0.688 [0.67-0.695]	<.001 <sup>***</sup>	0.392 [0.377-0.408]	<.001 <sup>***</sup>
XGBoost	0.725 [0.708-0.738]	<.001 <sup>***</sup>	0.436 [0.422-0.451]	<.001 <sup>***</sup>
<b>SMST</b>				
• Imaging only	0.747 [0.725-0.758]	<.001 <sup>***</sup>	0.432 [0.415-0.452]	<.001 <sup>***</sup>
• Cognition only	0.715 [0.726-0.735]	<.001 <sup>***</sup>	0.429 [0.421-0.437]	<.001 <sup>***</sup>
• Clinical only	0.637 [0.615-0.656]	<.001 <sup>***</sup>	0.365 [0.296-0.421]	<.001 <sup>***</sup>
<b>SMMT</b>				
• Imaging only	0.894 [0.865-0.918]	.018 <sup>*</sup>	0.542 [0.517-0.568]	<.001 <sup>***</sup>
• Cognition only	0.825 [0.805-0.847]	<.001 <sup>***</sup>	0.472 [0.452-0.481]	<.001 <sup>***</sup>
• Clinical only	0.723 [0.708-0.734]	<.001 <sup>***</sup>	0.443 [0.422-0.461]	<.001 <sup>***</sup>
<b>MMST</b>	0.875 [0.862-0.898]	<.001 <sup>***</sup>	0.574 [0.507-0.595]	.0027 <sup>**</sup>
<b>MMMT</b>	0.902 [0.875-0.933]	.037 <sup>*</sup>	0.595 [0.537-0.625]	.0085 <sup>**</sup>
<b>HiMAL</b>	0.923 [0.915-0.937]	—	0.623 [0.605-0.644]	—

The numbers in [] indicate 95% confidence intervals using the pivot bootstrap estimator by sampling participants from the test dataset with replacement 200 times. The  $P$ -values for AUROC and AUPRC were calculated by pairwise comparison between the proposed (last row) and each of the baselines using the Wilcoxon signed rank tests adjusted for FDR correction.

Significant  $P$ -values are highlighted in bold with <sup>\*</sup> $0.01 < P < .05$ , <sup>\*\*</sup> $0.005 < P < .01$ , <sup>\*\*\*</sup> $P < .001$ .

Abbreviations: MMT = multimodal multitask; MMST = multimodal single task; RF = Random Forest; SMMT = single modality multitask; SMST = Single-modality single task; SVM = Support Vector Machines.

**Table 3.** Performance on the auxiliary task regression (memory, executive functioning, language, and visual-spatial) in terms of MSE (lower = better) for HiMAL and the baselines.

Model	Regression (MSE)			
	Memory	Executive	Language	Visual-spatial
<u>SMST</u>				
• Imaging only	0.352 [0.332-0.367]	0.376 [0.362-0.385]	0.406 [0.376-0.428]	0.416 [0.435-0.468]
• Cognition only	0.395 [0.365-0.412]	0.358 [0.334-0.375]	0.352 [0.341-0.368]	0.416 [0.405-0.44]
• Clinical only	0.526 [0.508-0.551]	0.516 [0.507-0.528]	0.62 [0.606-0.651]	0.575 [0.542-0.594]
<u>SMMT</u>				
• Imaging only	0.234 [0.195-0.252]	0.224 [0.212-0.245]	0.27 [0.256-0.296]	0.295 [0.276-0.307]
• Cognition only	0.345 [0.292-0.363]	0.326 [0.312-0.34]	0.334 [0.315-0.347]	0.305 [0.286-0.337]
• Clinical only	0.422 [0.391-0.456]	0.435 [0.423-0.45]	0.387 [0.362-0.403]	0.406 [0.392-0.418]
<u>MMST</u>	0.185 [0.158-0.207]	0.206 [0.188-0.221]	0.24 [0.23-0.256]	0.245 [0.228-0.267]
<u>MMMT</u>	0.21 [0.19-0.232]	<b>0.175 [0.152-0.188]</b>	0.252 [0.232-0.265]	<b>0.216 [0.195-0.238]</b>
<u>HiMAL</u>	<b>0.182 [0.162-0.195]</b>	0.195 [0.17-0.225]	<b>0.232 [0.213-0.256]</b>	0.225 [0.218-0.247]

The numbers in [] indicate 95% confidence intervals using the pivot bootstrap estimator by sampling participants from the test dataset with replacement 200 times. Bold values indicate the model with the best performance (lowest MSE) for each cognitive function

Abbreviations: MMTT = multimodal multitask; MMST = multimodal single task; SMMT = single modality multitask; SMST = single-modality single task.

Comparing across modalities, using only imaging data or cognition data as input led to a competitive performance across the different tasks. Combining both imaging and cognition data as bimodal input often resulted in improved performance compared to using either modality as standalone input. This was observed particularly for AUROC (Figure 5A), memory (Figure 5C), and visual-spatial (Figure 5F). However, using only clinical data showed a loss in prediction performance, as demonstrated by the relatively lower AUROC, AUPRC and relatively higher MSE for the composite scores. Combining clinical data as an additional modality to either imaging or cognition features did not result in a significant performance improvement compared to using only imaging or only cognition as unimodal input.

## Discussion

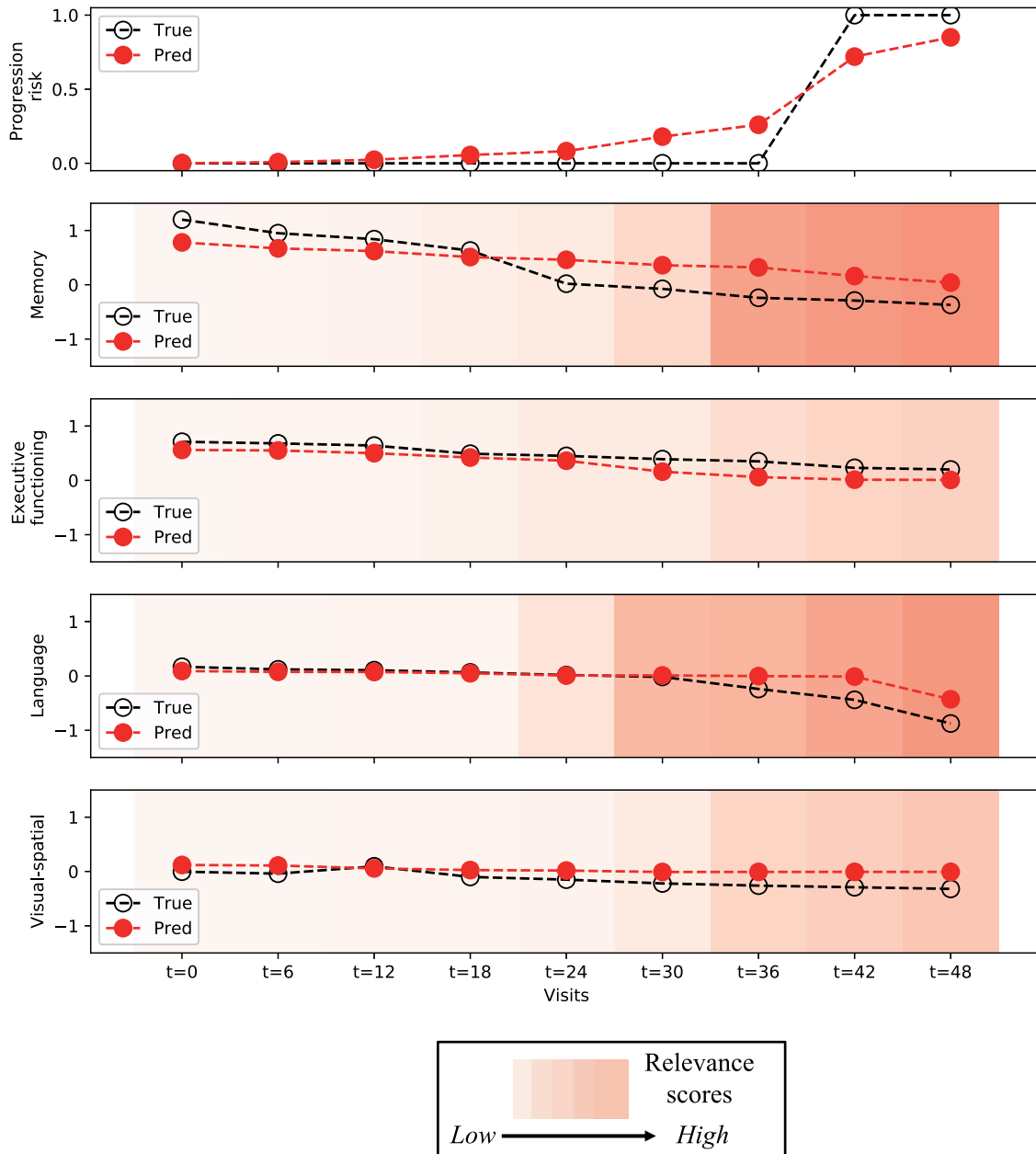
In this study, we used longitudinal multimodal data on imaging, self-reported cognition measures, and clinical data for predicting the longitudinal risk of MCI to AD progression within the next 6 months. We developed and validated HiMAL: a multimodal hierarchical multitask auxiliary learning framework to predict every 6 months, a set of neuropsychological composite cognitive function scores. Next, we used the forecasted composite function scores at every timepoint to predict the future risk of progression from MCI to AD. Experiments were conducted on the publicly available ADNI dataset, and the findings showed that HiMAL performed better than state-of-the-art unimodal and single-task baselines for predicting both the risk of AD progression (Table 2) and the composite scores (Table 3). Further, HiMAL also provided explanations about the potential factors behind disease progression by analyzing the contributions of each auxiliary task (Figure 4). Ablation analysis revealed imaging and cognition data as the most influential modalities towards the outcome.

For the primary task of predicting risk of MCI to AD progression, multitask models performed better than their single-task counterparts (eg, SMST vs SMMT or MMST vs MMTT; Table 2). Additionally, optimizing only a particular composite score did not lead to a drop in performance (MSE) for that score compared to when all the scores are optimized together (MMST vs MMTT vs proposed; Table 3). Both

these observations highlighted the benefits of jointly optimizing multiple tasks compared to learning a single task. In terms of architectural design, HiMAL was closest to MMTT where the main and auxiliary task were predicted concurrently without any hierarchical dependence between the tasks. The superior prediction performance of HiMAL on both primary and auxiliary tasks (Tables 2 and 3) validated our design choice of having a hierarchical relation between the tasks. The superior performance of HiMAL can be attributed to the fact that the hierarchical relationship allowed the model to exploit task dependencies more effectively, leading to more informative feature representations and effective knowledge transfer between the tasks.

There exists a plethora of studies predicting the risk of MCI to AD progression at a single endpoint.<sup>8,9,48</sup> However, the probability of progression from MCI to AD is dynamic and should be assessed using a longitudinal lens. Further, at the point-of-care, clinicians must determine the progression of the disease with concurrent prediction of multiple adverse events such as a future clinical diagnosis and forecasting the progression of cognitive performance over the natural course of the disease. HiMAL provided a more holistic approach for monitoring the risk of progression over a 6-month period, with simultaneous prediction of the functional decline in different domains of human cognition. Further, HiMAL is built on routinely collected EHR variables from a publicly available multimodal AD dataset, which encourages the translational ability of our work on other AD datasets. We believe that HiMAL can be a useful tool for early identification of patients in the prodromal phase of AD, who are at risk of progressing in future. This could potentially lead to therapeutic interventions to delay the disease progression over time and could be helpful for tailoring disease management and planning future care.

In addition to forecasting risk of disease progression, HiMAL also provided a complementary approach of providing explanatory support to the predictions. As black-box DL models become more prevalent in healthcare, clinicians seek not only superior prediction performance but also explanations for model decisions to enhance interpretability and trustworthiness.<sup>49,58,59</sup> Recent research on AD progression only identified the important features for a prediction, providing a snapshot of model explanations at a fixed



**Figure 4.** Model explanations of an pMCI individual who progressed 48 months (4 years) after the baseline visit. The topmost plot represents progression risk while the following points represent the composite functions. The x-axis represents visits in months and the plotted values represent either the ground truth label (black) or the predicted value (red). The colormap within each row indicates the weight/importance (relevance score) assigned to the forecasted composite scores where dark hues corresponded to higher relevance given to a particular composite score at a specific timepoint. Relevance scores were scaled between 0 and 1 for easier interpretation. Composite scores, predicted between [0,1] after sigmoid activation function (Figure 2), were rescaled to Z-scores with lower values indicating greater impairment.

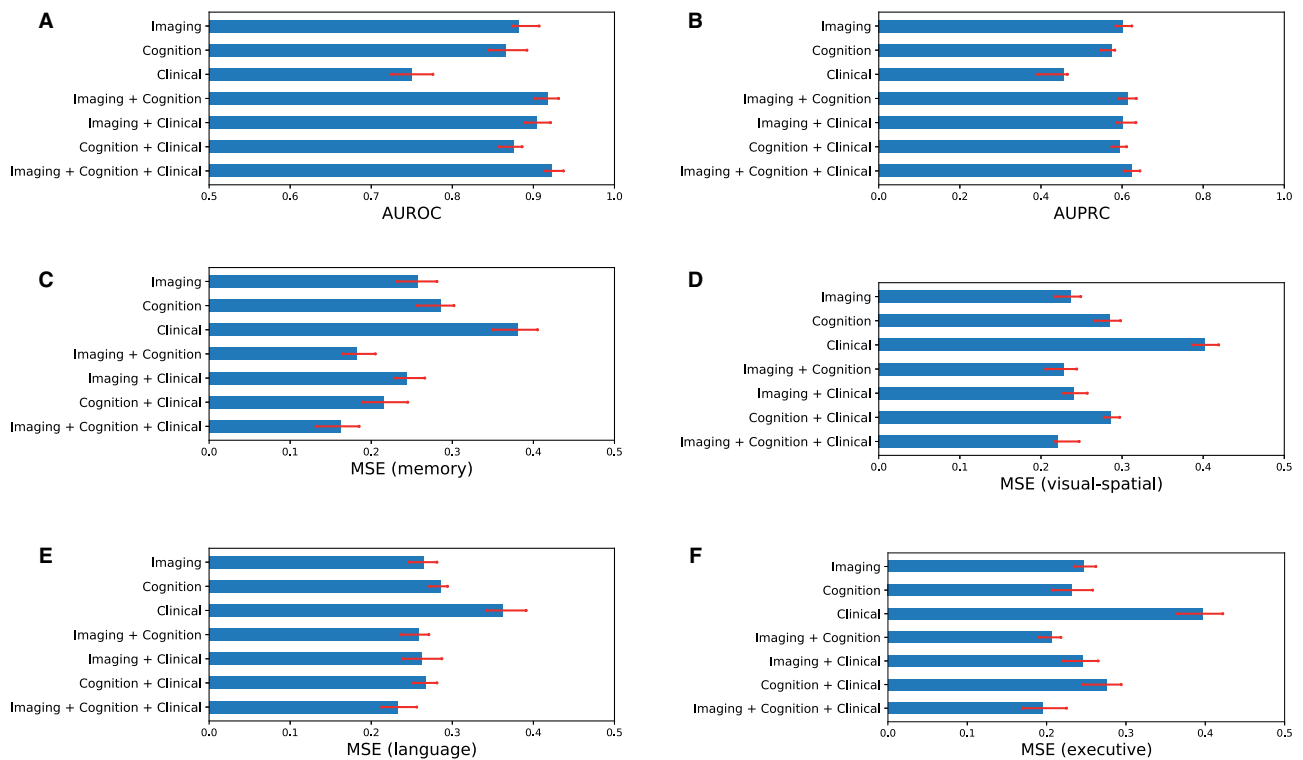
timepoint.<sup>58,60</sup> On the contrary, HiMAL provided longitudinal explanations about the potential reasons behind high/low risk prediction throughout the visit trajectory. The longitudinal explanations are clinically informative in the sense that they can inform clinicians 6 months in advance the potential decline in those specific cognitive areas that can lead to progression to AD in future (Figure 4).

Longitudinal studies have demonstrated that changes in cognitive functions such as memory, executive function, language, and visual-spatial skills are significantly correlated with the progression from MCI to AD.<sup>35-37</sup> However, it is important to note that the trajectory illustrated in Figure 4

represents a single patient's progression and may not be indicative of the entire dataset. Each patient, whether a progressor or non-progressor, has a unique trajectory, reflecting potential declines in specific cognitive function scores (Figure 4, Figure S2). This patient-specific approach aligns with the broader objective of DL-driven precision diagnostics, enabling the development of individualized treatment plans based on personalized cognitive trajectories.

HiMAL is also flexible with the selection of modalities and auxiliary tasks, encouraging reproducibility on other clinical problems besides AD. First, our framework can easily accommodate additional modalities with modality-specific modules





**Figure 5.** Performance of HiMAL for different combinations of input modalities in terms of (A) AUROC, (B) AUPRC, (C) MSE (memory), (D) MSE (executive), (E) MSE (language), and (F) visual-spatial. The red bars represent the 95% confidence interval using the pivot bootstrap estimator by sampling participants from the test dataset with replacement 200 times.

to generate the feature embedding as shown in Figure 2. Finally, the composite cognitive functions as auxiliary tasks were calculated from individual cognitive assessment scores which are readily available in most AD datasets. The rationale behind the design choice of selecting composite cognitive functions as auxiliary tasks was that the aggregated summary scores represented a particular cognitive domain as a whole and facilitated clinically intuitive units of explanation for analysis of disease progression. However, we would like to emphasize that HiMAL does not depend on the availability of these composite cognitive functions. Other scores like MMSE, ADAS, and CDR-SB can serve as suitable alternatives for auxiliary tasks if necessary.

### Limitations and scope for future work

Our study has several limitations and opportunities for future research. First, our findings are based on a single AD dataset from North America, limiting generalizability. Future studies should validate our approach on diverse cohorts such as OASIS-3,<sup>61</sup> AIBL,<sup>62</sup> and MIRIAD.<sup>63</sup> Additionally, HiMAL has been trained and tested on populations from the same dataset, for example, ADNI. Evaluating the generalizability of HiMAL on unseen independent datasets remains an important avenue for future work. Second, HiMAL only incorporates imaging, cognition, and clinical data, excluding other modalities like PET imaging and genetic data commonly used in AD research.<sup>64</sup> However, HiMAL can accommodate additional modalities, provided there is sufficient longitudinal data and sample size. Third, HiMAL was developed retrospectively, necessitating evaluation in a prospective setting where patient outcomes are unknown. Finally, we

emphasize that this are prototype models to demonstrate methods. In order to translate these models into deployment, more rigorous evaluation would be needed including prospective validation and detailed case review.

### Conclusions

In this study, we utilized longitudinal data from MRI, cognition, and clinical sources to predict the risk of MCI to AD progression. Our hierarchical multitask deep framework HiMAL predicted neuropsychological composite functions as auxiliary tasks at each timepoint to estimate progression risk. Results on an AD dataset showed superior performance compared to unimodal and single-task approaches. Ablation analysis revealed imaging and cognition data as the most influential modalities towards the outcome. Clinically informative model explanations anticipate cognitive decline 6 months in advance, aiding clinicians in future AD progression assessment. Flexibility in input modalities and auxiliary task selection indicated the translational impact of HiMAL extends to diverse clinical prediction problems.

### Acknowledgments

Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the

Laboratory for Neuro Imaging at the University of Southern California.

## Author contributions

All authors contributed to the design of the methodology and the experiments. Sayantan Kumar implemented the data pre-processing, modeling, and data analysis, and manuscript writing. Sean C. Yu provided guidance on data analysis, model building, and validation. Andrew Michelson, Thomas Kannappallil, and Philip Payne provided conceptualization, clinical interpretation, and manuscript writing. All authors participated in manuscript revision and editing and approved the final version for submission

## Supplementary material

Supplementary material is available at *JAMIA Open* online.

## Funding

The preparation of this report was supported by the Centene Corporation contract (P19-00559) for the Washington University-Centene ARCH Personalized Medicine Initiative. Data collection and sharing for this project was funded by the ADNI (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada.

## Conflicts of interest

The authors do not have any competing interests to disclose.

## Data availability

ADNI data used in this study are publicly available and can be requested following ADNI Data Sharing and Publications Committee guidelines: <https://adni.loni.usc.edu/data-samples/access-data/>. Implementation code for this project will be made public upon manuscript acceptance.

## References

1. Kumar S, Oh I, Schindler S, Lai AM, Payne PR, Gupta A. Machine learning for modeling the progression of Alzheimer disease dementia using clinical data: a systematic literature review. *JAMIA Open*. 2021;4(3):ooab052.
2. Richards M, Brayne C. What do we mean by Alzheimer's disease? *BMJ*. 2010;341:c4670.
3. Scheltens P, Blennow K, Breteler MMB, et al. Alzheimer's disease. *Lancet*. 2016;388(10043):505-517.
4. Iddi S, Li D, Aisen PS, Rafii MS, Thompson WK, Donohue MC, Alzheimer's Disease Neuroimaging Initiative. Predicting the course of Alzheimer's progression. *Brain Inf*. 2019;6(1):6-18.
5. Murphy MP, LeVine IH. Alzheimer's disease and the amyloid- $\beta$  peptide. *J Alzheimers Dis*. 2010;19(1):311-323.
6. Petersen RC. Mild cognitive impairment. *Continuum (Minneapolis)*. 2016;22(2 Dementia):404-418.
7. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the pre-clinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):280-292.
8. Lee G, Nho K, Kang B, Sohn K-A, Kim D; for Alzheimer's Disease Neuroimaging Initiative. Predicting Alzheimer's disease progression using multi-modal deep learning approach. *Sci Rep*. 2019;9(1):1952.
9. Qiu S, Chang GH, Panagia M, Gopal DM, Au R, Kolachalama VB. Fusion of deep learning models of MRI scans, mini-mental state examination, and logical memory test enhances diagnosis of mild cognitive impairment. *Alzheimers Dement (Amst)*. 2018;10:737-749.
10. Li H, Habes M, Wolk DA, Fan Y, Initiative ADN; Alzheimer's Disease Neuroimaging Initiative and the Australian Imaging Biomarkers and Lifestyle Study of Aging. A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. *Alzheimers Dement*. 2019;15(8):1059-1070.
11. Mirabnahrzham G, Ma D, Beaulac C, et al. Predicting time-to-conversion for dementia of Alzheimer's type using multi-modal deep survival analysis. *Neurobiol Aging*. 2023;121:139-156. <https://doi.org/10.1016/j.neurobiolaging.2022.10.005>
12. Berezuk C, Khan M, Callahan BL, et al. Sex differences in risk factors that predict progression from mild cognitive impairment to Alzheimer's dementia. *J Int Neuropsychol Soc*. 2023;29(4):360-368. <https://doi.org/10.1017/S1355617722000297>
13. Mattsson-Carlgnen N, Salvadó G, Ashton NJ, et al. Prediction of longitudinal cognitive decline in preclinical Alzheimer disease using plasma biomarkers. *JAMA Neurol*. 2023;80(4):360-369. <https://doi.org/10.1001/jamaneurol.2022.5272>
14. Maheux E, Koval I, Ortholand J, et al. Forecasting individual progression trajectories in Alzheimer's disease. *Nat Commun*. 2023;14(1):761. <https://doi.org/10.1038/s41467-022-35712-5>
15. Zhang D, Wang Y, Zhou L, et al. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage*. 2011;55(3):856-867.
16. Cheng B, Liu M, Zhang D, Munsell BC, Shen D. Domain transfer learning for MCI conversion prediction. *IEEE Trans Biomed Eng*. 2015;62(7):1805-1817.
17. Wee C-Y, Yap P-T, Shen D; and for the Alzheimer's Disease Neuroimaging Initiative. Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. *Hum Brain Mapp*. 2013;34(12):3411-3425.
18. Moore PJ, Lyons TJ, Gallacher J, et al. Random forest prediction of Alzheimer's disease using pairwise selection from time series data. *PLoS One*. 2019;14(2):e0211558.
19. Moradi E, Pepe A, Gaser C, et al. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage*. 2015;104:398-412.
20. Cheng H, Yuan S, Li W, et al. De-accumulated error collaborative learning framework for predicting Alzheimer's disease progression. *Biomed Signal Process Control*. 2024;89:105767. <https://doi.org/10.1016/j.bspc.2023.105767>

21. Maruszak A, Silajdžić E, Lee H, et al. Predicting progression to Alzheimer's disease with human hippocampal progenitors exposed to serum. *Brain*. 2023;146(5):2045-2058. <https://doi.org/10.1093/brain/awac472>
22. Nie L, Zhang L, Meng L, Song X, Chang X, Li X. Modeling disease progression via multisource multitask learners: a case study with Alzheimer's disease. *IEEE Trans Neural Netw Learn Syst*. 2016;28(7):1508-1519.
23. Kumar S, Payne PR, Sotiras A. Normative modeling using multimodal variational autoencoders to identify abnormal brain volume deviations in Alzheimer's disease. *SPIE Medical Imaging 2023. Computer-Aided Diagnosis*. Vol. 12465. SPIE; 2023:1246503.
24. Kumar S, Payne P, Sotiras A. Improving normative modeling for multi-modal neuroimaging data using mixture-of-product-of-experts variational autoencoders. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)* (pp. 1-5). IEEE; 2024.
25. Golovanevsky M, Eickhoff C, Singh R. Multimodal attention-based deep learning for Alzheimer's disease diagnosis. *J Am Med Inform Assoc*. 2022;29(12):2014-2022.
26. Weiner MW, Veitch DP, Aisen PS, et al. Recent publications from the Alzheimer's disease neuroimaging initiative: reviewing progress toward improved AD clinical trials. *Alzheimer's. Dement*. 2017;13(4):e1-e85.
27. Ito K, Corrigan B, Zhao Q, et al. Disease progression model for cognitive deterioration from Alzheimer's disease neuroimaging initiative database. *Alzheimers Dement*. 2011;7(2):151-160.
28. Zhang D, Shen D, Initiative ADN; Alzheimer's Disease Neuroimaging Initiative. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *NeuroImage*. 2012;59(2):895-907.
29. Liao Q, Ding Y, Jiang ZL, Wang X, Zhang C, Zhang Q. Multi-task deep convolutional neural network for cancer diagnosis. *Neurocomputing*. 2019;348:66-73.
30. Ding X, Bucholtz M, Wang H, et al. A hybrid computational approach for efficient Alzheimer's disease classification based on heterogeneous data. *Sci Rep*. 2018;8(1):9774.
31. Van der Mussele S, Franssen E, Struyfs H, et al. Depression in mild cognitive impairment is associated with progression to Alzheimer's disease: a longitudinal study. *J Alzheimers Dis*. 2014;42(4):1239-1250.
32. Mueller SG, Weiner MW, Thal LJ, et al. The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clin N Am*. 2005;15(4):869-877.
33. Wang C, Li Y, Tsuboshita Y, et al. A high-generalizability machine learning framework for predicting the progression of Alzheimer's disease using limited data. *NPJ Digital Medicine*. 2022;5(1):43.
34. Jack CR, Bernstein MA, Fox NC, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J Magn Reson Imaging*. 2008;27(4):685-691.
35. Crane PK, Carle A, Gibbons LE, et al. Development and assessment of a composite score for memory in the Alzheimer's disease neuroimaging initiative (ADNI). *Brain Imaging Behav*. 2012;6(4):502-516.
36. Gibbons LE, Carle AC, Mackin RS, et al. A composite score for executive functioning, validated in Alzheimer's disease neuroimaging initiative (ADNI) participants with baseline mild cognitive impairment. *Brain Imaging Behav*. 2012;6(4):517-527.
37. Choi S-E, Mukherjee S, Gibbons LE, et al. Development and validation of language and visuospatial composite scores in ADNI. *Alzheimer's & Dementia. Transl Res Clin Interv*. 2020;6:e12072.
38. Cao W, Wang D, Li J, Zhou H, Li L, Li Y. Brits: Bidirectional recurrent imputation for time series. *Adv Neural Inf Process Syst*. 2018:31.
39. Runtti H, Mattila J, Van Gils M, et al. Quantitative evaluation of disease progression in a longitudinal mild cognitive impairment cohort. *J Alzheimers Dis*. 2014;39(1):49-61.
40. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735-1780.
41. Jung W, Kim SE, Kim JP, et al. Deep learning model for individualized trajectory prediction of clinical outcomes in mild cognitive impairment. *Front Aging Neurosci*. 2024;16:1356745. <https://doi.org/10.3389/fnagi.2024.1356745>
42. Zhou J, Liu J, Narayan VA, Ye J, Initiative ADN, Alzheimer's Disease Neuroimaging Initiative. Modeling disease progression via multi-task learning. *NeuroImage*. 2013;78:233-248.
43. Liang W, Zhang K, Cao P, Liu X, Yang J, Zaiane OR. Exploiting task relationships for Alzheimer's disease cognitive score prediction via multi-task learning. *Comput Biol Med*. 2023;152:106367. <https://doi.org/10.1016/j.compbiomed.2022.106367>
44. Zhang J, He X, Liu Y, Cai Q, Chen H, Qing L. Multi-modal cross-attention network for Alzheimer's disease diagnosis with multi-modality data. *Comput Biol Med*. 2023;162:107050. <https://doi.org/10.1016/j.compbiomed.2023.107050>
45. Zhang R, Yuan J, Sun Y, et al. Machine learning on longitudinal multi-modal data enables the understanding and prognosis of Alzheimer's disease progression. *iScience*. 2023;27(7):110263. <https://doi.org/10.2139/ssrn.4411420>
46. Liu X, Li W, Miao S, Liu F, Han K, Bezabih TT. HAMMF: hierarchical attention-based multi-task and multi-modal fusion model for computer-aided diagnosis of Alzheimer's disease. *Comput Biol Med*. 2024;176:108564. <https://doi.org/10.1016/j.compbiomed.2024.108564>
47. Zhang Z-C, Zhao X, Dong G, Zhao X-M. Improving Alzheimer's disease diagnosis with Multi-Modal PET embedding features by a 3D Multi-Task MLP-Mixer neural network. *IEEE J Biomed Health Inform*. 2023;27(8):4040-4051. <https://doi.org/10.1109/JBHI.2023.3280823>
48. El-Sappagh S, Abuhmed T, Islam SR, Kwak KS. Multimodal multi-task deep learning model for Alzheimer's disease progression detection based on time series data. *Neurocomputing*. 2020;412:197-215.
49. El-Sappagh S, Alonso JM, Islam SR, Sultan AM, Kwak KS. A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease. *Sci Rep*. 2021;11(1):2660.
50. Rabeh AB, Benzarti F, Amiri H. Diagnosis of Alzheimer diseases in early step using SVM (support vector machine). *2016 13th International conference on Computer Graphics, Imaging and Visualization (CGiV)*, IEEE; 2016:364-367.
51. Velazquez M, Lee Y; Alzheimer's Disease Neuroimaging Initiative. Random Forest model for feature-based Alzheimer's disease conversion prediction from early mild cognitive impairment subjects. *PLoS One*. 2021;16(4):e0244773.
52. Nguyen K, Nguyen M, Dang K, et al. Early Alzheimer's disease diagnosis using an XG-Boost model applied to MRI images. *Biomed Res Ther*. 2023;10(9):5896-5911.
53. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit*. 1997;30(7):1145-1159.
54. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. *Proceedings of the 23rd ACM International Conference on Machine Learning*. 2006:233-240.
55. Tibshirani RJ, Efron B. An introduction to the bootstrap. *Monogr Statist Appl Probab*. 1993;57:1-436.
56. Demšar J. Statistical comparisons of classifiers over multiple data sets. *J Machine Learn Res*. 2006;7:1-30.
57. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc: Ser B (Methodological)*. 1995;57(1):289-300.
58. Bogdanovic B, Eftimov T, Simjanoska M. In-depth insights into Alzheimer's disease by using explainable machine learning approach. *Sci Rep*. 2022;12(1):6508.
59. Yu L, Xiang W, Fang J, Chen Y-PP, Zhu R. A novel explainable neural network for Alzheimer's disease diagnosis. *Pattern Recognit*. 2022;131:108876.
60. Sudar KM, Nagaraj P, Nithisaa S, Aishwarya R, Aakash M, Lakshmi SI. Alzheimer's Disease Analysis using Explainable Artificial Intelligence (XAI). *2022 International Conference on*

- Sustainable Computing and Data Communication Systems (ICSCDS)*, IEEE; 2022:419-423.
61. LaMontagne PJ, Benzinger TL, Morris JC, et al. OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. *Alzheimers Dement.* 2018;14(7S\_Part\_20): P1097-P1097.
  62. Ellis KA, Bush AI, Darby D, et al. The Australian imaging, biomarkers and lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *Int Psychogeriatr.* 2009;21(4):672-687.
  63. Malone IB, Cash D, Ridgway GR, et al. MIRIAD—public release of a multiple time point Alzheimer's MR imaging dataset. *Neuroimage.* 2013;70:33-36.
  64. Veitch DP, Weiner MW, Aisen PS, et al. Understanding disease progression and improving Alzheimer's disease clinical trials: Recent highlights from the Alzheimer's disease neuroimaging initiative. *Alzheimers Dement.* 2019;15(1):106-152.