Multimodal and Longitudinal Modeling for Predicting Alzheimer's Disease Progression

Inês Simões[1], Rita Mendes[1], Carlos Santiago[1], Maria Rita Verdelho[1], Catarina Barata[1]

Institute for Systems and Robotics (ISR/IST), LARSyS, Instituto Superior Técnico, Universidade de Lisboa, Portugal
inesmcsimoes@tecnico.ulisboa.pt

Abstract. Early and accurate prediction of Alzheimer's Disease (AD) progression remains challenging due to multimodal and inter-patient heterogeneity, missing modalities, irregular follow-ups, and complex temporal dynamics. We propose a longitudinal heterogeneous graph neural network that models patient trajectories by integrating cognitive assessments, PET and fMRI, and demographic data within a unified temporal graph, explicitly encoding non-linear temporal gaps. Heterogeneous interactions capture cross-modal dependencies and naturally handle missing modalities, while recurrent mechanisms on cognitive nodes encode disease progression. To address class imbalance and the ordinal structure of clinical stages, we introduce a loss combining weighted cross-entropy and Wasserstein distance. Integrated gradients enhance interpretability by identifying clinically relevant predictors. Experiments on ADNI and OASIS-3 show our approach consistently outperforms multimodal recurrent and transformer-based baselines, with notable gains in recall for Mild Cognitive Impairment (MCI) and AD, supporting robust, clinically meaningful disease staging.

Keywords: Alzheimer's disease · Multimodal learning · Graph neural networks · Temporal modeling · Disease progression forecasting

1 Introduction

Dementia affects over 55 million people globally, with a new case every three seconds [2]. AD is its main cause, responsible for nearly two-thirds of cases, progressively impairing memory, cognition, and independence [20]. While incurable, early diagnosis enables interventions that slow progression and preserve quality of life [23,21]. Clinically, AD diagnosis combines medical history, neuropsychological tests, and neuroimaging (fMRI, PET) [9]. Patients are classified as cognitively normal (CN), with MCI (a transitional stage often associated with diagnostic uncertainty) [6], or with AD [4]. In this context, deep learning (DL) has demonstrated strong predictive capability in AD by uncovering patterns not readily detectable by clinicians [17]. However, most existing methods frame AD diagnosis as a static classification problem [1,22], disregarding its dynamic and

heterogeneous progression [12]. Although effective for cross-sectional classification, this paradigm is insufficient for predicting future AD progression, which remains the central challenge addressed in this work [25].

Temporal modeling has partially addressed this limitation. Kim et al. [14] proposed an interpretable temporal GNN for 24-month prognosis, but limited to unimodal data and fixed intervals. Multimodal approaches include Kumar et al.'s [15] hierarchical multitask framework for short-term MCI–AD progression and LongForMAD [13], a Transformer-based model assuming monotonic transitions, which shows performance decay over long horizons. The TADPOLE challenge [18] benchmarked multimodal AD progression, with XGBoost [5] performing best, and subsequent studies explored patient-level longitudinal features and cross-cohort generalization [8,27]. However, these methods employ fixed-window forecasts and aggregate follow-up data using basic summary statistics, thereby neglecting irregular temporal dynamics, cross-modal interactions, and the heterogeneity of available modalities across patients.

In contrast, we propose a heterogeneous multimodal temporal graph that represents each patient's full longitudinal history, capturing interactions across available modalities and irregular temporal dependencies. The model combines recurrent and attention mechanisms, employs a hybrid order-sensitive loss for clinically consistent staging, and includes interpretability analyses to identify key drivers of disease progression. The main contributions of this work are:

1. A heterogeneous multimodal temporal graph for AD progression modeling under irregular follow-ups, missing modalities, and individual trajectories.
2. A systematic comparison against recurrent, attention-based, and tree-based baselines previously validated for longitudinal AD modeling.
3. Comprehensive evaluation on two large-scale neuroimaging cohorts, including joint training and cross-dataset transfer analyses to assess robustness.
4. A hybrid ordinal-aware loss combining cross-entropy and Wasserstein distance to improve stability and clinically consistent predictions.
5. Interpretability analyses using integrated gradients to identify key predictors.

2 Methodology

2.1 Problem Formulation

Let $\mathcal{P} = p_1, \ldots, p_N$ denote a cohort of patients followed longitudinally, with clinical data collected at irregular visits.For each visit, patient features comprise demographic and cognitive data, optionally complemented by one or more neuroimaging modalities. We aim to map the sequence of features up to time t to the diagnostic state at the next visit, which can be CN, MCI, or AD, enabling forecasting of future disease progression from each patient's longitudinal history. Patient trajectories are modeled as heterogeneous temporal graphs. Temporal irregularity is captured via time-difference features at node and edge levels, allowing non-uniform follow-up intervals without assuming fixed temporal decay. An overview of the framework is shown in Figure 1.

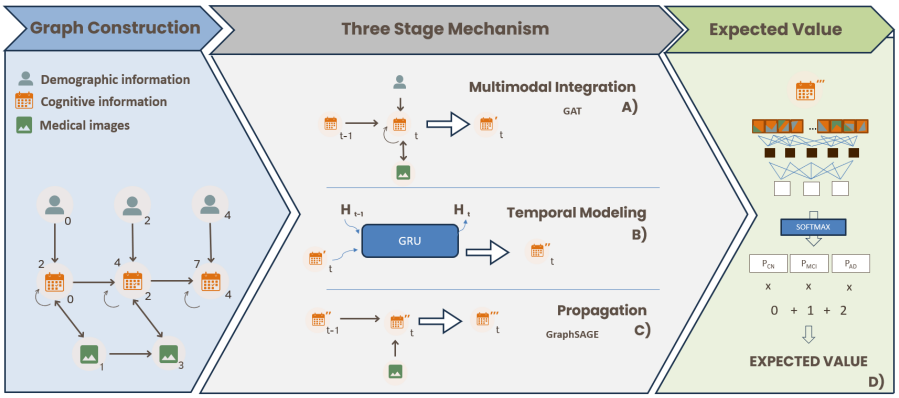

Fig. 1. Overview of the proposed heterogeneous multimodal temporal graph framework. Cognitive nodes contain temporal information (current visit at bottom right, predicted future visit at top left).

2.2 Heterogeneous Temporal Graph Construction

Each patient is represented as a heterogeneous graph $\mathcal{G}_p$ with nodes $\mathcal{V}_p$ (features) and edges $\mathcal{E}_p$ (relationships). Nodes are divided into three types: cognitive nodes encode clinical assessments and the observed diagnosis at the current visit, including global and domain-specific scores (e.g., memory, orientation, daily activities), plus temporal attributes (time since initial diagnosis and prediction horizon); demographic nodes encode static or slowly changing participant attributes; imaging nodes contain modality-specific embeddings from PET and fMRI scans using a pretrained MedicalNet backbone [3], with each modality modeled as an independent node. Edges capture temporal continuity and multimodal relationships: temporal edges link consecutive nodes of the same type, cross-modal edges connect demographic and imaging nodes to cognitive nodes. This representation naturally accommodates irregular visit intervals and missing modalities through absent nodes.

2.3 Model Architecture

Multimodal Integration (A): Given graph $\mathcal{G}_p$, node features are projected into a shared latent space via type-specific weights, followed by heterogeneous graph attention to model multimodal interactions and update each node v as:

$$\mathbf{h}_v^{(1)} = \sigma\Big(\sum_{u \in \mathcal{N}(v)} \alpha_{uv} \mathbf{W} \mathbf{h}_u^{(0)} \Big), \tag{1}$$

where α_{uv} are edge-type-specific attention coefficients and σ is a nonlinear activation. This allows the model to integrate information across modalities while emphasizing the most informative neighbors.

Temporal Modeling (B): To capture patient-specific trajectories, temporal dependencies are modeled with a recurrent unit on cognitive node embeddings. Cognitive nodes are chronologically ordered and processed via a GRU [7], including visit-interval features. Applying a GRU after graph-based feature extraction captures longitudinal progression on context-enriched representations, integrating past and current states while limiting oversmoothing and computational overhead compared to extra graph layers.

Graph Propagation (C): After temporal encoding, multimodal information is reintegrated using heterogeneous GraphSAGE layers [10]:

$$\mathbf{h}_v^{(2)} = \sigma\left(\mathbf{W}_1 \mathbf{h}_v^{\mathrm{GRU}} + \mathbf{W}_2 \cdot \mathrm{AGG}(\{\mathbf{h}_u^{\mathrm{GRU}} : u \in \mathcal{N}(v)\})\right), \tag{2}$$

where $\mathbf{h}_v^{\mathrm{GRU}}$ is the GRU-updated embedding of cognitive nodes and AGG the neighborhood aggregation. This step consolidates temporal context, capturing patient history and multimodal relationships, while SAGEConv robustly aggregates neighboring information.

Prediction Layer (D): The final cognitive node embeddings are mapped to class probabilities via a fully connected layer and softmax. The expected cognitive trajectory is then computed as:

$$\text{expected value} = \sum_{c=0}^{2} p_c \cdot c, \tag{3}$$

where p_c is the predicted probability of class c. This expected value provides a continuous estimate of the cognitive state. The discrete prediction is obtained by selecting the class with the highest probability (`argmax`), while intermediate expected values (e.g., 0.7 or 1.3) reflect uncertainty and transitional stages.

2.4 Training Objective

AD diagnosis is an ordinal, multi-class task (CN < MCI < AD) with imbalanced classes. To account for both the ordinal structure and class imbalance while capturing transitional and uncertain cognitive states, we minimize a hybrid loss combining weighted cross-entropy and Wasserstein distance. To handle class imbalance, weighted cross-entropy assigns class-specific weights:

$$\mathcal{L}_{\mathrm{CE}}(y, \hat{y}) = -\sum_{i=1}^{C} w_i \, y_i \log(\hat{y}_i), \qquad w_i = \frac{N}{C \cdot N_i}, \tag{4}$$

where C is the number of classes, y_i the label, $\hat{y}_i$ the predicted probability, N the total number of samples, and N_i the number of samples in class i.

To respect the ordinal structure of AD stages, our loss leverages the one-dimensional closed-form of the Wasserstein-1 distance [11], computed from the cumulative distribution functions of the predicted and target labels, so that larger ordinal errors (e.g., CN instead of AD) are penalized more heavily.

$$\mathcal{L}_{W_1}(p,t) = \sum_{i=1}^{C} \left|\mathrm{CDF}_i(p) - \mathrm{CDF}_i(t)\right|, \quad \mathrm{CDF}_i(x) = \sum_{j=1}^{i} x_j. \tag{5}$$

The final objective combines the two losses, balancing standard classification with ordinal consistency, so that the model accounts both for class imbalance and for the severity of misclassifications along the AD progression spectrum.

$$\mathcal{L} = \alpha\, \mathcal{L}_{\mathrm{CE}} + (1-\alpha)\, \mathcal{L}_{W_1}, \quad \alpha \in [0,1]. \tag{6}$$

3 Experiments and Results

3.1 Dataset Statistics

The study uses two large-scale longitudinal cohorts, OASIS-3 [16] and ADNI [19], providing complementary information on the progression of AD. OASIS-3, after quality filtering (90% feature completeness, consistent diagnoses), includes 966 participants and 6,034 sessions (mean of 6.19 visits, baseline age of 74.39 ± 8.49 years, 59.8% women). Each visit included 11 cognitive, 22 demographic, and 512 imaging characteristics. Labels were imbalanced (5,189 CN, 418 MCI, 427 AD), with 150 progressing to MCI and 153 to AD; substantial reversals (136 MCI-CN, 25 AD-MCI) indicate heterogeneous trajectories. The mean interval between visits was 485 days. The ADNI includes 1,218 participants and 4,764 sessions (average of 3.91 visits, maximum of 19, baseline age of 72.05 ± 13.01 years, 52.6% male). The dimensions of the features are similar, except demographic features only contain 2 variables. The labels were unbalanced (3,442 CN, 326 MCI, 996 AD), with 125 progressing to MCI and 30 to AD; there were few reversals (5 AD-MCI, none MCI-CN), reflecting more stable trajectories. Follow-up intervals were irregular (median 410 days, mean 931 days) with an average forecast horizon of 469 days.

3.2 Implementation and Evaluation Details

All models were implemented in Python 3.10 with PyTorch 1.13.1 and PyTorch Geometric 2.6.1, running on Linux (Ubuntu 18.04) with CUDA 11.6 on an NVIDIA RTX 3090 GPU (24GB VRAM, cuDNN 8.3). Heterogeneous graphs included cognitive, demographic, and imaging nodes; missing modalities were omitted and missing values imputed with -99. Models were trained patient-wise with 80%/20% train/test splits. Hyperparameters were selected via stratified 3-fold cross-validation within the training split. The optimal configuration (64 hidden channels, learning rate 0.005, weight decay 10^{-3}, $\alpha = 0.3$) was trained up to 55 epochs with early stopping, accuracy varied by less than 1% across hyperparameter configurations. Final performance was averaged over 20 runs with the selected setup.

Table 1. Performance of GNN models under within-cohort evaluation and cross-cohort generalization performance (mean).

	OASIS → OASIS						ADNI → ADNI					
Model	Bal.Acc.	AUC	F1	Re_{CN}	Re_{MCI}	Re_{AD}	Bal.Acc.	AUC	F1	Re_{CN}	Re_{MCI}	Re_{AD}
GNN Cog	0.691	0.846	0.650	0.88	0.47	0.73	0.848	0.948	0.863	0.97	0.60	0.97
GNN Cog + Img	0.666	0.841	0.636	0.88	0.38	0.73	0.848	0.942	0.851	0.96	0.61	0.97
GNN Cog + Dem	0.714	0.865	0.655	0.86	0.51	0.77	**0.850**	**0.952**	**0.867**	**0.98**	0.60	0.97
GNN Multi	0.722	0.861	0.661	0.86	0.52	0.79	0.846	0.939	0.850	0.96	0.61	0.97
GNN Cog + GRU	0.717	0.861	0.659	0.86	0.54	0.75	0.849	0.948	0.855	0.96	0.61	0.97
GNN Multi+GRU	**0.742**	**0.865**	**0.666**	0.85	**0.58**	**0.79**	0.848	0.936	0.846	0.95	**0.63**	0.97
	OASIS → ADNI						ADNI → OASIS					
Model	Bal.Acc.	AUC	F1	Re_{CN}	Re_{MCI}	Re_{AD}	Bal.Acc.	AUC	F1	Re_{CN}	Re_{MCI}	Re_{AD}
GNN Multi+GRU	0.784	0.899	0.712	0.77	0.61	0.97	0.622	0.776	0.547	0.86	0.35	0.66

3.3 Impact of Multimodal Integration and Temporal Modeling

We evaluate six graph-based architectures of increasing complexity, from unimodal cognitive to full multimodal temporal models, under within-cohort and cross-cohort settings. Performance is measured using balanced accuracy, macro F1, macro one-vs-one ROC AUC, and class-wise recall, with statistical significance assessed via paired tests across runs (Table 1).

On OASIS, the multimodal temporal model (GNN Multi+GRU) achieves the best overall performance, significantly outperforming all ablations in BA and AUC ($p < 0.001$), driven by higher MCI and AD recall ($p < 0.001$) and a slight but significant CN recall reduction. Temporal modeling improves over static multimodal models ($p < 0.001$), while multimodality boosts unimodal temporal variants ($p \leq 0.001$). On ADNI, overall performance is higher and the simpler GNN Cog + Dem already achieves strong metrics (highest AUC, $p < 0.001$), leaving limited room for improvement; temporal modeling mainly improves MCI recall ($p = 0.015$). Cross-cohort evaluation with GNN Multi+GRU shows robust transfer from OASIS to ADNI (BA = 0.7835) but notable degradation in the opposite direction (BA = 0.6218), particularly for MCI and AD, likely due to OASIS's inclusion of inverse trajectories.

3.4 Computational Performance

In graph-based models, computational cost is a critical consideration, as graphs can be large and resource-intensive, leading to long training and inference times. To address this concern, we evaluate our models' runtime performance using the OASIS dataset, which contains relatively small patient graphs, averaging 16 nodes each.The empirical runtimes are reported in Table 2. For the GNN Multi+ GRU configuration, this corresponds to a mean inference time of 0.005ms per cognitive session, or approximately 0.03ms per patient, indicating that our method remains highly efficient even when additional data modalities and temporal modeling.

Table 2. Training and testing runtime per model (milliseconds).

Model	Training Time (ms)	Testing Time (ms)
GNN Cog	210	1.4
GNN Multi	840	4.0
GNN Multi + GRU	1650	6.6

3.5 Combined Loss

For the GNN Multi+GRU model, using cross-entropy alone favored CN (recall 0.84) but underperformed for MCI (0.56), while Wasserstein loss alone overemphasized MCI (0.97) at the cost of CN and AD performance. In contrast, the combined loss ($\alpha = 0.3$) achieved the best overall balance, yielding the highest balanced accuracy (0.742) and consistently strong recall across CN (0.85), MCI (0.58), and AD (0.79).

3.6 Qualitative Analysis of Longitudinal Predictions

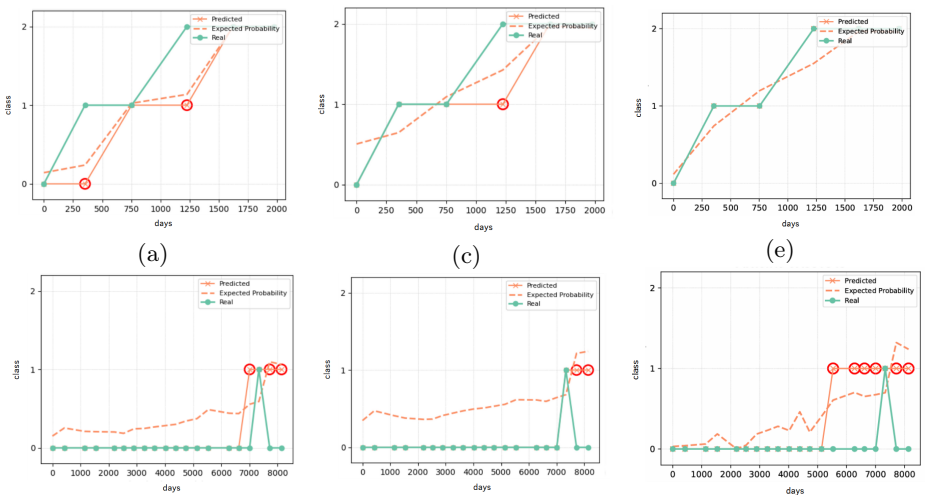

Fig. 2. Longitudinal prediction of the model. Examples of patients showing the actual trajectory (green) and model predictions (dashed orange) over time (CN=0, MCI=1, AD=2). The first column shows results for GNN Cog, the second column for GNN Cog + GRU, and the third column for our proposed GNN Multi + GRU model. (a) Difficulty in consistently monitoring clinical progression. (b) Early prediction of conversion to MCI. (c) Trajectory better aligned with reality. (d) Consistent prediction of MCI conversion. (e) Trajectory closely follows the real course. (f) Earlier prediction of conversion with the overall trend correctly captured.

To provide a qualitative insight, Figure 2 presents examples of longitudinal predictions at the individual level. Using only cognitive information (GNN cog), the model often struggles to consistently monitor disease progression, sometimes showing a delayed alignment with reality (a), or anticipating conversion to MCI earlier than observed (b). The incorporation of temporal dynamics through GRU leads to more stable trajectories: GNN cog+GRU is able to follow the clinical course more faithfully (c), producing consistent conversion predictions (d). When multimodal information is also included (GNN multi+GRU), predictions are more aligned with observed trajectories (e), and even in more complex cases (f), the general trend is correctly captured in advance. All models showed difficulty in reclassifying subjects as CN after reaching a more severe stage. In fact, the model rarely predicts a direct reversion to CN immediately after indicating MCI or AD. This inertia reflects its temporal memory and the persistence of clinical features associated with the more severe state. Once a conversion is predicted, subsequent visits typically exhibit patterns consistent with that diagnosis, making an immediate improvement unlikely. In some cases, the model does eventually predict a return to CN, but only after several inconsistent sessions, suggesting that sufficient longitudinal evidence is required before it revises its internal representation of the patient's state.

3.7 Alternative strategies for longitudinal modeling:

To evaluate the impact of relational graph structure in longitudinal disease modeling, we compared the proposed graph-based architecture against sequential and non-sequential baselines on the OASIS dataset, which includes inverse diagnostic trajectories and irregular follow-up intervals. The evaluated baselines consisted of a GRU [26], a time-aware Transformer [24], and a patient-level XGBoost model [27]. All models were trained using identical patient-level splits and comparable optimization settings to ensure a fair comparison.

Table 3 summarizes the performance of all evaluated approaches. The proposed GNN-based model achieved the highest Balanced Accuracy (0.742) and ROC-AUC (0.866), while also improving recall for the clinically challenging MCI class. In particular, the increase in MCI recall suggests improved sensitivity to transitional disease stages, which are often difficult to distinguish in longitudinal clinical data.

Statistical significance was assessed using Welch's t-test across cross-validation folds for Balanced Accuracy, F1-score, and ROC-AUC. The proposed model significantly outperformed both the GRU and Transformer baselines across all metrics ($p < 0.001$). Compared to XGBoost, the GNN achieved significantly higher Balanced Accuracy and ROC-AUC ($p < 0.001$), while differences in F1-score were not statistically significant ($p = 0.324$). Overall, these findings indicate that incorporating relational graph structure enhances longitudinal modeling performance and improves robustness under class imbalance, particularly for intermediate disease stages such as MCI.

Table 3. Performance of baseline compared to the proposed model.

Model	Bal.Acc.	AUC	F1	Re_{CN}	Re_{MCI}	Re_{AD}
GRU [26]	0.704	0.840	0.654	0.885	0.525	0.703
Transformer [24]	0.705	0.824	0.627	0.855	0.541	0.720
XGBoost [27]	0.712	0.858	0.667	**0.914**	0.431	0.793
GNN Multi+GRU	**0.742**	**0.866**	**0.665**	0.851	**0.583**	**0.793**

3.8 Comparison with State-of-the-Art Methods

Comparisons to external methods are provided for contextual purposes only, as differences in prediction horizons, cohort selection, and evaluation protocols prevent fully controlled reproduction. All models, including ours, utilize the ADNI dataset, which provides a common basis for approximate benchmarking. As shown in Table 4, our model achieves the highest reported AUROC (0.936), demonstrating competitive performance relative to prior approaches. Unlike prior methods, our framework models asynchronous multimodal visits and irregular follow-ups via a heterogeneous graph, allowing flexible handling of real-world longitudinal data.

Table 4. Comparison of AUROC scores with state-of-the-art approaches on the ADNI dataset.

Metric	HiMAL[15]	LongForMAD[13]	L2C-TabPFN [8]	FROG [8]	GNN MULTI+GRU
AUROC	0.923	0.920	0.913	0.925	**0.936**

3.9 Relevance of Cognitive Attributes for Prediction

To understand which cognitive node features are most relevant for prediction, we used the Integrated Gradients method, implemented through the Explainer module of the PyTorch Geometric library. This approach allows us to interpret the model's behavior, quantifying the contribution of each individual attribute to decision-making. The analysis (Figure 3) revealed that *category* and *days_to_future* had the greatest relative importance. This is expected, as they encode the patient's current diagnostic state and the temporal distance to the target session, respectively. Following these two, memory and judgment emerged as the most influential cognitive attributes, consistent with their well-established role in the clinical diagnosis of Alzheimer's disease. In contrast, features related to daily functioning, such as personal care (*perscare*) and communication abilities (*commun*), had a reduced impact on the model's decisions.

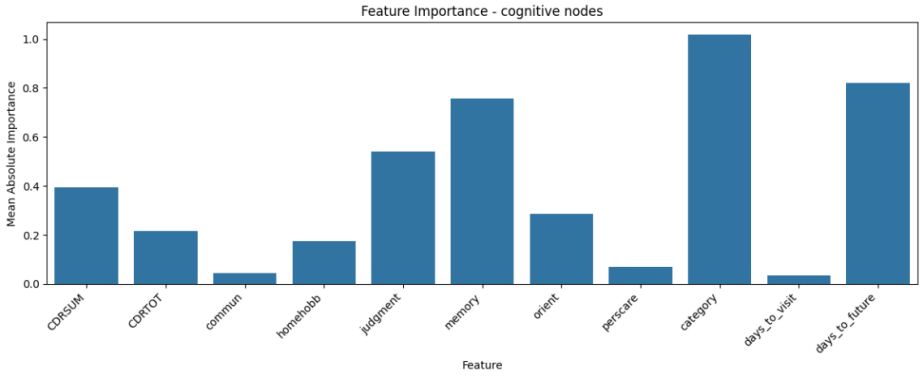

Fig. 3. Feature importance of cognitive nodes obtained with the IG method

3.10 Analysis of Temporal and Demographic Factors in Prediction Errors

To understand model behavior, we analyzed how accuracy varies with participant age and prediction horizon (*days_to_future*). Correctly predicted participants were younger (mean: 72, median: 72 years) than incorrectly predicted ones (mean: 77, median: 79 years), suggesting that advanced age increases forecasting difficulty due to higher heterogeneity or accelerated disease progression. Furthermore, incorrect predictions were associated with longer intervals between visits (mean: 607 days vs. 472 days for correct ones), confirming that longer prediction horizons increase error likelihood.

Notably, in 156 initially "incorrect" cases, the participant's subsequent diagnosis later matched the model's prediction, showing its sensitivity to subtle changes before formal clinical confirmation. Conversely, the model struggled with short-term fluctuations, predicting immediate return to CN in only 2 out of 64 MCI-to-CN reversions; in most cases, transient MCI predictions persisted. Finally, while neuropsychiatric or medical comorbidities (e.g., hypothyroidism, mood disorders, alcoholism) increased cognitive variability and misclassifications, they did not systematically cause errors, as the model successfully predicted progression for several such individuals.

4 Conclusion

This work presents a heterogeneous multimodal temporal graph framework for forecasting AD progression. By integrating cognitive, demographic, and neuroimaging data while handling missing modalities and irregular visits, the model captures complex longitudinal patterns. The hybrid ordinal-aware loss improves prediction stability and clinical consistency, and interpretability analyses highlight meaningful cognitive indicators. Results demonstrate strong predictive performance, particularly for the clinically challenging MCI stage, highlighting the

framework's potential for personalized disease monitoring and early risk identification. A partial anonymized implementation is available at `https://anonymous.4open.science/r/MLM_PAP-EA56/README.md`

Acknowledgments This work was supported by LARSyS FCT funding (DOI: 10.54499/LA/P/0083/2020, UID/50009/2025), partially funded with grant 2023.02043.BDANA from Fundação para a Ciência e Tecnologia, and the FCT project MMIST [2024.06929.RESTART].

Competing Interests The authors declare that they have no competing interests relevant to the content of this article.

References

1. Almubark, I., Alsegehy, S., Jiang, X., Chang, L.C.: Classification of alzheimer's disease, mild cognitive impairment, and normal controls with multilayer perceptron neural network and neuropsychological test data. In: IJCCI. pp. 439–446 (2020)
2. Alzheimer's Disease International: Dementia statistics (2025), accessed: 2025-01-10
3. Alzubaidi, L., Santamaría, J., Manoufali, M., Mohammed, B., Fadhel, M.A., Zhang, J., Al-Timemy, A.H., Al-Shamma, O., Duan, Y.: Mednet: pre-trained convolutional neural network model for the medical imaging tasks. arXiv preprint arXiv:2110.06512 (2021)
4. Burns, B., Xue, Y., Scharre, D.W., Ning, X.: Flexible multimodal neuroimaging fusion for alzheimer's disease progression prediction. In: International Workshop on Applications of Medical AI. pp. 235–245. Springer (2025)
5. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. pp. 785–794 (2016)
6. Chiari-Correia, R.D., Tumas, V., Santos, A.C., Salmon, C.E.G.: Structural and functional differences in the brains of patients with mci with and without depressive symptoms and their relations with alzheimer's disease: an mri study. Psychoradiology **3**, kkad008 (2023)
7. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555 (2014)
8. Ding, Y., Ren, J., Lu, J., Kwak, G.H., Iraji, A., Fedorov, A.: Longitudinal progression prediction of alzheimer's disease with tabular foundation model. arXiv preprint arXiv:2508.17649 (2025)
9. Hakami, F., Madkhali, M.A., Saleh, E., Ayoub, R., Moafa, S., Moafa, A., Alnami, B., Maashi, B., Khubrani, S., Busayli, W., et al.: Awareness and perception toward alzheimer's disease among residents living in the jazan province, saudi arabia: a cross-sectional study. Cureus **15**(9) (2023)
10. Hamilton, W.L., Ying, R., Leskovec, J.: Inductive representation learning on large graphs (2018)
11. Hou, L., Yu, C.P., Samaras, D.: Squared earth mover's distance-based loss for training deep neural networks. arXiv preprint arXiv:1611.05916 (2016)
12. Jiang, Y., Ding, H., Chen, H., Lan, J., Teng, X., Cheng, G.W., Li, Z., Xie, H., Yoo, J.S., Cai, J.: M3 ad: Multi-task multi-gate mixture of experts for alzheimer's disease diagnosis with conversion pattern modeling. arXiv preprint arXiv:2508.01819 (2025)

13. Karaman, B.K., Sabuncu, M.R.: Assessing the significance of longitudinal data in alzheimer's disease forecasting. In: International Conference on AI in Healthcare. pp. 3–16. Springer (2024)
14. Kim, M., et al.: Interpretable temporal graph neural network for prognostic prediction of alzheimer's disease using longitudinal neuroimaging data. In: IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (2021)
15. Kumar, S., Yu, S.C., Michelson, A., Kannampallil, T., Payne, P.R.: Himal: Multimodal hi erarchical m ulti-task a uxiliary l earning framework for predicting alzheimer's disease progression. JAMIA open **7**(3), ooae087 (2024)
16. LaMontagne, P.J., Benzinger, e.a.: Oasis-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer's disease. medRxiv (2019). https://doi.org/10.1101/2019.12.13.19014902
17. Liu, C., Huang, F., Qiu, A.: Monte carlo ensemble neural network for the diagnosis of alzheimer's disease. Neural Networks **159**, 14–24 (2023). https://doi.org/https://doi.org/10.1016/j.neunet.2022.10.032
18. Marinescu, R.V., Oxtoby, N.P., Young, A.L., Bron, E.E., Toga, A.W., Weiner, M.W., Barkhof, F., Fox, N.C., Klein, S., Alexander, D.C., et al.: Tadpole challenge: prediction of longitudinal evolution in alzheimer's disease. arXiv preprint arXiv:1805.03909 (2018)
19. Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L.: The alzheimer's disease neuroimaging initiative. Neuroimaging Clinics **15**(4), 869–877 (2005)
20. Nippak, P., Begum, H.: Recognizing early signs of dementia: Subtle symptoms often mistaken for normal aging. Journal of Dementia **8**(5), 1–3 (2024)
21. Pais, M., Martinez, L., Ribeiro, O., Loureiro, J., Fernandez, R., Valiengo, L., Canineu, P., Stella, F., Talib, L., Radanovic, M., et al.: Early diagnosis and treatment of alzheimer's disease: new definitions and challenges. Brazilian Journal of Psychiatry **42**, 431–441 (2020)
22. Praneeth, B., Reddy, P., Bhuvaneshwar, P., ., M., Moharir, M., R, A.: Deep learning algorithms for early alzheimer's diagnosis: From data to prognosis. pp. 366–371 (10 2023). https://doi.org/10.1109/ICSSAS57918.2023.10331657
23. Rasmussen, J., Langerman, H.: Alzheimer's disease–why we need early diagnosis. Degenerative neurological and neuromuscular disease pp. 123–130 (2019)
24. Susetzky, T., Qiu, H., Braren, R., Rueckert, D.: A holistic time-aware classification model for multimodal longitudinal patient data. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 24–33. Springer (2025)
25. van der Veere, P.J., Hoogland, J., Visser, L.N., Van Harten, A.C., Rhodius-Meester, H.F., Sikkes, S.A., Venkatraghavan, V., Barkhof, F., Teunissen, C.E., van de Giessen, E., et al.: Predicting cognitive decline in amyloid-positive patients with mild cognitive impairment or mild dementia. Neurology **103**(3), e209605 (2024)
26. Yeo, B.T.: Cross-dataset evaluation of dementia longitudinal progression prediction models
27. Zhang, C., An, L., Wulan, N., Nguyen, K.N., Orban, C., Chen, P., Chen, C., Zhou, J.H., Liu, K., Yeo, B.T., et al.: Cross-dataset evaluation of dementia longitudinal progression prediction models. Human Brain Mapping **46**(11), e70280 (2025)