MAGNET-AD: Multitask Spatiotemporal GNN for Interpretable Prediction of PACC and Conversion Time in Preclinical Alzheimer

Salma Hassan[1] ✉[0009−0006−5498−8319], Mostafa Salem[1,2][0000−0001−9915−8390], Vijay Papineni[3][0000−0002−6162−3290], Ayman Elsayed[3][0000−0003−3381−066X], and Mohammad Yaqub[1][0000−0001−6896−1105]

[1] Mohamed Bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE
{Salma.Hassan, Mostafa.Salem, Mohammad.Yaqub}@mbzuai.ac.ae
[2] Department of Computer Science, Faculty of Computers and Information, Assiut University, Assiut, Egypt
[3] Sheikh Shakhbout Medical City, Abu Dhabi, UAE
rkpvijay@gmail.com, ayaahmed@ssmc.ae

Abstract. Preclinical Alzheimer's Disease (AD) detection remains challenging due to the complex interplay of biological, structural, and temporal factors. Existing methods often struggle to integrate multimodal longitudinal data and predict key clinical outcomes. We propose MAGNET-AD, a novel multitask spatiotemporal graph neural network designed to predict the Preclinical Alzheimer's Cognitive Composite (PACC) score and time to AD conversion. MAGNET-AD offers three key contributions: (1) A dynamic heterogeneous graph architecture with weighted edges for hybrid fusion mechanisms, integrating static and dynamic multimodal data; (2) a temporal importance weighting loss function that adaptively learns critical time points while jointly optimizing time prediction and cognitive decline estimation; and (3) an interpretable attention framework that highlights key brain regions and genetic factors driving disease progression. MAGNET-AD achieves state-of-the-art performance with a concordance index of 0.858 for conversion time prediction and a mean square error of 1.983 for PACC prediction, outperforming existing deep learning approaches. These results underscore MAGNET-AD's potential for early AD risk assessment and monitoring, enabling broader clinical applications. The code is available at https://github.com/BioMedIA-MBZUAI/MAGNET-AD.

Keywords: Spatiotemporal Graph Neural Network · Multi-Task Learning · Dementia · Neuroimaging · Multimodal · Interpretability

1 Introduction

Alzheimer's Disease (AD) represents one of the most pressing challenges in global healthcare, with over 10 million new cases diagnosed worldwide each year—equivalent to one new case every 3.2 seconds [2]. Despite decades of research and drug discovery efforts, AD remains incurable, imposing an annual

economic burden of over $1 trillion globally [21]. Thus, early detection during the preclinical stage is critical, as interventions at this phase could alter the disease trajectory, slowing down the progression and improving patient outcomes.

Effective monitoring and prediction of AD progression require the capture of both cognitive decline and disease onset timing. The Preclinical Alzheimer Cognitive Composite (PACC) score quantifies subtle cognitive changes by combining four validated memory, recall, and cognitive assessments [6]. While PACC tracks cognitive function, predicting the time until AD development helps clinicians identify which patients need more immediate intervention and monitoring, which ultimately helps in allocating the limited medical resources effectively. These two prediction tasks have typically been approached separately despite their shared biological basis. Thus, MAGNET-AD aims to offer a unified multitask approach that leverages this connection to improve disease monitoring.

Previous research has made significant progress in utilizing machine learning and deep learning techniques for AD detection and progression prediction [1, 27]. While efforts to forecast the conversion from mild cognitive impairment (MCI) to AD using multimodal feature values have shown high accuracy for short-term predictions, they have struggled to address long-term disease progression effectively [17]. Moreover, studies applying deep learning to functional MRI data have demonstrated potential for early AD detection but face challenges in integrating multiple data modalities [25, 3, 7]. On the other hand, multimodal multitask deep learning models have improved predictive performance for AD progression using time-series data but offer limited interpretability of the underlying disease mechanisms [8]. Similarly, deep recurrent neural networks (RNN) have effectively captured temporal dependencies in longitudinal data but struggle to model complex multimodal interactions and provide insights into critical factors driving disease progression [18]. Graph Neural Networks (GNNs) have emerged as a promising approach for modeling complex data [10]. However, existing architectures struggle to combine time-invariant features with longitudinal spatiotemporal data and often lack effective mechanisms for data fusion, especially when handling inconsistent time points [28]. These limitations hinder their ability to capture interactions that drive disease progression. As a result, current approaches face challenges in temporal modeling, multimodal integration, and clinical interpretability, highlighting the need for more robust approaches.

In response to these challenges, we introduce **MAGNET-AD**, a multitask spatiotemporal graph neural network (STGNN) for preclinical AD. Our main contributions are threefold:

1. **STGNN with Hybrid Fusion:** We develop the first spatiotemporal graph neural network that jointly predicts PACC and AD progression outcomes while integrating multimodal data through a hybrid fusion mechanism, incorporating both static and dynamic features with informative edge weights.
2. **Temporally-Aware Loss Function:** We propose a novel loss function that strategically identifies critical time points in disease progression for accurate predictions by learning key temporal patterns.

3. **Interpretable Attention Framework:** We design an attention-based framework that provides actionable, interpretable clinical insights by identifying key brain and genetic biomarkers.

To our knowledge, this is the **first** implementation of STGNN for **time-to-event** prediction, which aims to enable earlier accurate AD risk assessment and provide interpretable insights for clinicians to improve patient outcomes.

2 Methodology

Pre-Processing. In neuroimaging, harmonizing multimodal datasets presents a significant challenge. Therefore, we developed a pre-processing pipeline to standardize and handle these complex data. T1 MRI scans underwent spatial normalization to MNI152 space, followed by brain extraction, bias field correction, and intensity normalization [12, 16]. Functional MRI data were co-registered to the processed T1 images, followed by slice timing correction, motion correction, spatial smoothing, and temporal filtering. SynthSeg [5], a robust deep learning-based segmentation tool trained on elderly and diseased brain scans, generated 32 anatomical segmentation masks. These masks were used to extract 107 region-specific radiomic features. Then, the AnatCL foundational model, trained on ADNI and OASIS, was used to extract meaningful feature embeddings for each structure [4]. For genetic data, 100 AD-associated genes were selected based

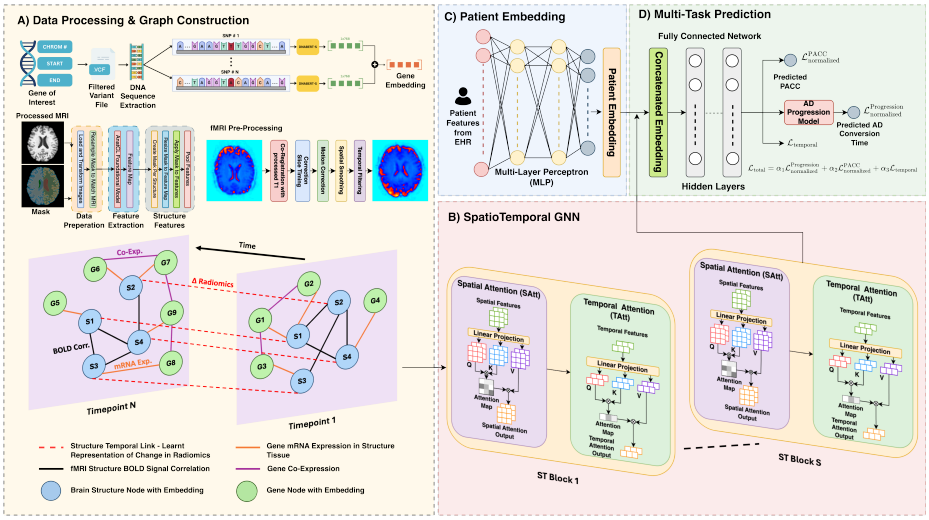


Fig. 1: Overview of the MAGNET-AD framework for AD progression prediction and patient classification. Starting with (A) data pre-processing and heterogeneous temporal graph construction, which is fed to (B) STGNN with attention mechanisms to process the graph, which is concatenated with patient embeddings (C) for multi-task predictions of PACC and time to AD conversion (D).

on their relevance to AD pathophysiology. DNA sequences with variants were extracted and processed using DNABERT-S to generate gene embeddings [29]. Gene co-expression data was obtained from GeneMania [24], while mRNA expression levels were sourced from the Human Protein Atlas [11].

Graph Construction. We propose a new temporal multimodal graph framework that introduces a hybrid data fusion mechanism, seamlessly integrating dynamic neuroimaging patterns with time-invariant genetic markers to disentangle the intricate progression of AD, illustrated in Fig. 1A. The graph architecture integrates four distinct types of relationships: functional connectivity between brain regions, which is derived from BOLD signal correlations, gene co-expression patterns, gene-structure relationships through mRNA expression, and learned temporal relationships of structural changes through radiomics. Each node in the graph represents either a brain structure (blue nodes) or a gene (green nodes) enriched with their respective embeddings. The temporal dimension connects multiple time points to model disease progression. The framework learns the temporal evolution of brain structural changes while simultaneously incorporating static genetic factors that have been strategically categorized as early, mid, and late-stage biomarkers of AD. This hybrid approach allows us to capture both the dynamic and static nature of disease progression.

MAGNET-AD. Our approach implements a heterogeneous STGNN through dual attention mechanisms: spatial attention captures relationships between brain structures and genetic factors within each timepoint, while temporal attention models the evolution of structural changes across visits, as shown in Fig. 1B. The spatial attention (SAtt) mechanism operates through cross-modal graph attention layers and multi-head attention, enabling the model to learn complex interactions between brain regions and their associated genetic factors. Both attention mechanisms are computed through a series of learned transformations, where the attention score A is calculated as the product of a trainable matrix V_a and a non-linear transformation of the input features X_i. These features undergo multiple linear transformations with the dimensions depending on whether we are computing SAtt (using node count N) or TAtt (using temporal length T_{r-1}). The final attention weights are normalized using a softmax function. TAtt is incorporated through learned edge weights derived from radiomics features, quantifying structural changes between consecutive time points. This heterogeneous graph structure processes both dynamic (temporal MRI) and static (genetic) features through hierarchical processing: first modeling gene-gene interactions through stage-specific (early, mid, late) graph layer, then capturing gene-structure relationships via cross-modal attention, and finally processing spatial-temporal connections through these dual attention mechanisms. This architecture enables comprehensive disease progression, which is modeled by capturing both immediate structural changes and underlying genetic influences.

Multi-Task Learning. MAGNET-AD employs a multi-task learning strategy

illustrated in Fig. 1D. The learned graph embeddings, which capture spatiotemporal disease progression patterns, are concatenated with patient-level feature embeddings extracted from electronic health records (EHR) shown in Fig. 1C, forming a unified representation. This concatenated representation is fed to two specialized prediction heads: an AD progression head and a PACC regression head. The model simultaneously optimizes both tasks through a hybrid loss function illustrated in Equation 1, where α_1, α_2, and α_3 are learnable weights that balance the contribution of each loss component. The $\mathcal{L}^{\text{Progression}}_{\text{normalized}}$ term measures how accurately our model predicts both conversion time and relative patient risk rankings [15]. The $\mathcal{L}^{\text{PACC}}_{\text{normalized}}$ component represents the normalized mean squared error loss for PACC score prediction. The temporal regularization term $\mathcal{L}_{\text{temporal}}$ introduces a novel adaptive weighting mechanism that emphasizes critical time points in disease progression, illustrated in Equation 2 where $w_t(i)$ represents learned temporal importance weights for timepoint i. The temporal decay function $\beta(\Delta t)$ incorporates a distance-aware penalty that decreases with larger time intervals (Δt) between consecutive visits, controlled by the hyperparameter γ, which ensures that rapid changes in predictions between closely spaced timepoints are penalized more heavily and allows the model to adapt its predictions based on varying time intervals between patient visits.

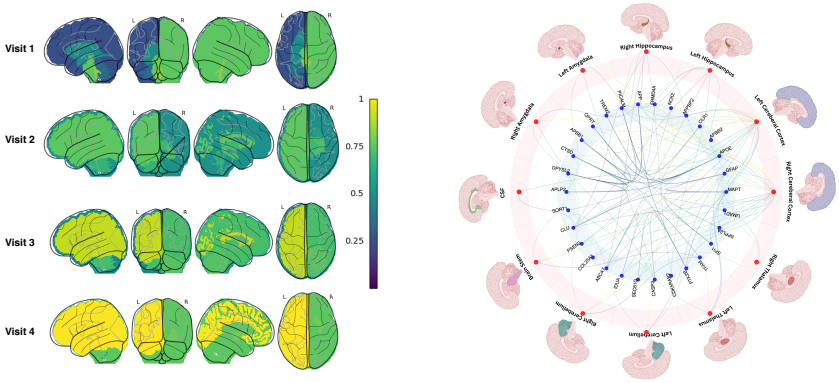
$$\mathcal{L}_{\text{total}} = \alpha_1 \mathcal{L}^{\text{Progression}}_{\text{normalized}} + \alpha_2 \mathcal{L}^{\text{PACC}}_{\text{normalized}} + \alpha_3 \mathcal{L}_{\text{temporal}}, \tag{1}$$

$$\mathcal{L}_{\text{temporal}} = \sum_i w_t(i) \cdot ((\hat{y}_{i+1} - \hat{y}_i)^2 \cdot \beta(\Delta t)), \text{ where } \beta(\Delta t) = \frac{2}{1 + \exp(\gamma \cdot |\Delta t|)}. \tag{2}$$

3 Experimental Setup

Dataset. The Anti-Amyloid Treatment in Asymptomatic Alzheimer's Disease (A4) [22] study dataset consists of longitudinal data from over 1,787 clinically normal older patients with elevated amyloid levels. Each participant underwent baseline screening and a varying number of follow-up visits. The data includes T1w MRI, fMRI, genetic data, and EHR data. Unlike commonly used datasets such as ADNI [19] or OASIS [13], A4 stands out as the **only** dataset specifically focused on preclinical AD, providing complete multimodal data.

Implementation Details. For reproducibility, the data were split into five stratified cross-validation folds, ensuring an equal percentage of censored data across folds. PyTorch Geometric library [9] was used to train the model for 150 epochs with an early stopping patience of 30, batch size of 128, and random seed of 42. We implemented our model with the following hyperparameters: 16 spatio-temporal (ST) blocks with 4 multi-head attention layers. Complete hyperparameters are outlined in the code. The concordance index (C-index) was



(a) Change of brain structure importance based on attention score across time.

(b) Chord diagram for top 30 gene-to-gene and gene-to-structure connection.

Fig. 2: Interpretability plots for temporal analysis of structure importance and gene connections.

used to evaluate the model's prognostic performance, as it measures the ability to correctly rank patients while accounting for censoring.

Ablation Studies. We conducted comprehensive ablation studies to validate the design choices. First, we optimized hyperparameters to determine optimal learning parameters. We then systematically evaluated the contribution of MRI data alone versus with genetic data along with varying edge types between gene-to-gene, gene-to-structure, and adding the change in radiomics features as the temporal edge weight, as outlined in Table 2, followed by an analysis of BOLD signal thresholds for functional connectivity. Finally, we validated our architectural choices by comparing MAGNET-AD against temporal deep learning approaches (Table 1) and established prognosis models.

Table 1: Performance comparison of temporal models; results are reported as the mean ± standard deviation; the best is bolded, and the second best is underlined.

Baseline Models	**C-Index ↑**	**MSE ↓**
LSTM [20]	0.5301 ± 0.017	3.1744 ± 0.031
Temporal Convolution Network [14]	0.6231 ± 0.013	2.8342 ± 0.027
Temporal Multimodal Transformer [26]	0.7112 ± 0.016	2.7302 ± 0.017
GCN STGNN [10]	0.8041 ± 0.013	2.1322 ± 0.027
MAGNET-AD w/o $\mathcal{L}_{\text{temporal}}$	$\underline{0.8224} \pm \underline{0.013}$	$\underline{2.0891} \pm \underline{0.029}$
MAGNET-AD (Ours)	$\mathbf{0.8582 \pm 0.012}$	$\mathbf{1.9831 \pm 0.028}$

Table 2: MAGNET-AD performance comparison with different edge weight configurations and data types. The filled circle (•) indicates inclusion, and the empty circle (○) indicates exclusion.

Data	Temporal Weights	Learned Radiomics	Gene-Struct	Gene-Gene	C-Index ↑	MSE ↓
MRI	○	○	○	○	0.5799 ± 0.011	2.5072 ± 0.021
	●	○	○	○	0.6377 ± 0.011	2.4224 ± 0.034
	●	●	○	○	0.6703 ± 0.013	2.3675 ± 0.021
MRI + Gen	○	○	●	○	0.7227 ± 0.010	2.2757 ± 0.051
	●	○	●	○	0.7386 ± 0.010	2.2115 ± 0.023
	●	●	●	○	0.7929 ± 0.013	2.1521 ± 0.008
MRI + Gen	○	○	●	●	0.8127 ± 0.009	2.1168 ± 0.015
	●	○	●	●	0.8241 ± 0.008	2.0694 ± 0.034
	●	●	●	●	**0.8582 ± 0.012**	**1.9831 ± 0.028**

Table 3: Comparison of C-Index and MSE across different models assessed at various timepoints.

# Visits	LSTM [20]		TCN [14]		Transformer [26]		GCN [10]		MAGNET-AD (Ours)	
	C-Index ↑	MSE ↓	C-Index ↑	MSE ↓	C-Index ↑	MSE ↓	C-Index ↑	MSE ↓	C-Index ↑	MSE ↓
2	0.4483	3.7384	0.5138	3.1468	0.5467	3.0765	0.6141	2.5314	**0.6823**	**2.4224**
3	0.4603	3.6013	0.5461	3.1285	0.5789	2.9952	0.6609	2.3108	**0.7568**	**2.2156**
4	0.4831	3.5412	0.5586	3.1175	0.6239	2.9072	0.6916	2.2819	**0.7832**	**2.1932**
5	0.5125	3.3214	0.5875	2.9153	0.6605	2.8654	0.7741	2.2564	**0.8102**	**2.1427**

4 Results and Discussion

Graph Data Fusion Structure. Table 2 reveals critical insights into the relative importance and interplay of different edge types in our architecture. In the MRI-only configuration, adding temporal weights significantly improves performance, improving the C-Index from 0.5799 to 0.6377 and reducing MSE from 2.5072 to 2.4224. This improvement suggests that temporal weights are particularly crucial for progression prediction and have an impact on cognitive assessment. The integration of learned radiomics features further enhances the model's performance, yielding a 5.1% improvement in C-Index and a 2.3% reduction in MSE. The introduction of genetic information marks a pivotal enhancement in the model's capabilities. The addition of gene-structure edges yields the most substantial single improvement observed, with the C-Index jumping dramatically from 0.6703 to 0.7929 and MSE reducing from 2.3675 to 2.1521. This marked improvement suggests that gene-structure relationships capture fundamental disease mechanisms that neither temporal nor radiomics features alone can detect. The subsequent integration of gene-gene interactions further refines the model's predictive power, improving the C-Index to 0.8582 and reducing MSE to 1.9831. The RMSE value indicates that our predictions deviate by 1.41 points from the actual PACC scores. Perhaps most notably, the impact of temporal weights remains significant even in the presence of genetic information, indicating that these components capture complementary rather than redundant information

about disease progression. The final model's performance demonstrates super-additive effects, with improvements exceeding what would be expected from the simple sum of individual components, as demonstrated by the results of the baseline models.

Model Architecture Comparison. The comparison with baseline temporal models (Table 1) reveals MAGNET-AD's superior performance in both tasks. While LSTM [20] shows basic temporal modeling capability, transformers [26] demonstrate improved performance. MAGNET-AD's significant improvement over these baselines (C-Index improvement of 20.7% and MSE reduction of 27.4% compared to Transformer) validates our graph-based spatiotemporal approach. This superior performance can be attributed to MAGNET-AD's hybrid data fusion mechanism, which offers three key advantages. First, unlike traditional sequential models that treat biomarkers as independent features, MAGNET-AD explicitly models complex interactions through both functional and structural connections, which better captures subtle disease-related changes. Second, while transformers rely solely on self-attention, our architecture combines attention mechanisms with edge-weighted temporal connections derived from radiomics features, providing more robust temporal dependency modeling. Third, our hybrid fusion approach allows for distinct treatment of static genetic markers and dynamic imaging features, with the gene-structure edges serving as a bridge, enabling the model to better leverage both time-invariant risk factors and dynamic disease progression patterns. This architectural design effectively captures the complex spatiotemporal patterns characteristic of preclinical AD, with the temporal loss function improving performance over the base MAGNET-AD model.

Temporal Robustness. Table 3 demonstrates MAGNET-AD's robust performance across varying numbers of patient visits, showcasing its effectiveness in real-world clinical scenarios where complete longitudinal data may be limited. With just two visits, MAGNET-AD achieves a C-Index of 0.6823 and MSE of 2.4224, significantly outperforming other baseline models. The model's performance steadily improves with additional visits and maintains a consistent performance advantage over baseline models [14]. This demonstrates MAGNET-AD's ability to effectively leverage even limited temporal data for reliable prediction.

Clinical Interpretability. The attention patterns revealed in Fig. 2 demonstrate strong alignment with established Braak staging of AD pathology [23]. The temporal progression of structure node importance mirrors the characteristic spread of tau pathology, beginning with high attention weights in medial temporal regions (Visits 1-2), progressing through limbic regions (Visits 2-3), and ultimately showing widespread neocortical involvement (Visit 4). This chord diagram validates this pattern and illustrates how genetic factors modulate this progression through complex gene-to-structure interactions. Such correspondence between MAGNET-AD's learned attention patterns and well-established neuropathological staging provides biological validation to our model.

5 Conclusion

We introduced MAGNET-AD, a novel multitask STGNN that achieves state-of-the-art performance in preclinical AD progression prediction. Through its innovative hybrid fusion architecture, adaptive temporal weighting, and interpretable attention mechanisms, MAGNET-AD demonstrates superior capability in modeling the complex interplay between genetic factors and longitudinal brain changes. Beyond its immediate clinical applications in early AD risk assessment, MAGNET-AD's framework establishes a new paradigm for progression prediction in preclinical AD. The success of our approach in modeling spatiotemporal disease dynamics while maintaining interpretability suggests promising applications across broader medical domains where early detection and progression monitoring are critical. Future directions include extending the model to integrate blood-based biomarkers and exploring its potential for personalized treatment optimization through patient-specific progression patterns.

Acknowledgments. This publication was supported by the Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI) Graduate Student and Postdoctoral Researcher Conference Travel Grant Initiative.

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

References

1. Abuhmed, T., El-Sappagh, S., Alonso, J.M.: Robust hybrid deep learning models for Alzheimer's progression detection. Knowledge-Based Systems **213**, 106688 (2021)
2. Alzheimer's Disease International: Dementia facts & figures: Dementia statistics (2024), https://www.alzint.org/about/dementia-facts-figures/dementia-statistics
3. Baltrušaitis, T., Ahuja, C., Morency, L.P.: Multimodal machine learning: A survey and taxonomy. IEEE transactions on pattern analysis and machine intelligence **41**(2), 423–443 (2018)
4. Barbano, C.A., Brunello, M., Dufumier, B., Grangetto, M.: Anatomical foundation models for brain MRIs. arXiv preprint arXiv:2408.07079 (2024)
5. Billot, B., Greve, D.N., Puonti, O., Thielscher, A., Van Leemput, K., Fischl, B., Dalca, A.V., Iglesias, J.E., et al.: SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. Medical image analysis **86**, 102789 (2023)
6. Donohue, M.C., Sperling, R.A., Salmon, D.P., Rentz, D.M., Raman, R., Thomas, R.G., Weiner, M., Aisen, P.S., et al.: The Preclinical Alzheimer Cognitive Composite: Measuring Amyloid-related Decline. JAMA neurology **71**(8), 961–970 (2014)
7. Du, L., Liu, K., Zhu, L., Yao, X., Risacher, S.L., Guo, L., Saykin, A.J., Shen, L., Initiative, A.D.N.: Identifying progressive imaging genetic patterns via multi-task sparse canonical correlation analysis: a longitudinal study of the ADNI cohort. Bioinformatics **35**(14), i474–i483 (2019)

8. El-Sappagh, S., Abuhmed, T., Islam, S.R., Kwak, K.S.: Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data. Neurocomputing **412**, 197–215 (2020)
9. Fey, M., Lenssen, J.E.: Fast graph representation learning with PyTorch Geometric. arXiv preprint arXiv:1903.02428 (2019)
10. He, S., Luo, Q., Du, R., Zhao, L., He, G., Fu, H., Li, H.: STGC-GNNs: A GNN-based traffic prediction framework with a spatial–temporal granger causality graph. Physica A: Statistical Mechanics and its Applications **623**, 128913 (2023)
11. Human Protein Atlas: The Human Protein Atlas. https://www.proteinatlas.org/ (2024)
12. Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H.P., Heiland, S., Wick, W., et al.: Automated brain extraction of multisequence MRI using artificial neural networks. Human brain mapping **40**(17), 4952–4964 (2019)
13. LaMontagne, P.J., Benzinger, T.L., Morris, J.C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A.G., et al.: OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. medrxiv pp. 2019–12 (2019)
14. Lea, C., Flynn, M.D., Vidal, R., Reiter, A., Hager, G.D.: Temporal convolutional networks for action segmentation and detection. In: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 156–165 (2017)
15. Lee, C., Zame, W., Yoon, J., Van Der Schaar, M.: Deephit: A deep learning approach to survival analysis with competing risks. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
16. Lowekamp, B.C., Chen, D.T., Ibáñez, L., Blezek, D.: The design of SimpleITK. Frontiers in neuroinformatics **7**, 45 (2013)
17. Minhas, S., Khanum, A., Alvi, A., Riaz, F., Khan, S.A., Alsolami, F., A Khan, M.: Early MCI-to-AD conversion prediction using future value forecasting of multimodal features. Computational intelligence and neuroscience **2021**(1), 6628036 (2021)
18. Nguyen, M., He, T., An, L., Alexander, D.C., Feng, J., Yeo, B.T., Initiative, A.D.N., et al.: Predicting alzheimer's disease progression using deep recurrent neural networks. NeuroImage **222**, 117203 (2020)
19. Petersen, R.C., Aisen, P.S., Beckett, L.A., Donohue, M.C., Gamst, A.C., Harvey, D.J., Jack Jr, C., Jagust, W.J., Shaw, L.M., Toga, A.W., et al.: Alzheimer's disease neuroimaging initiative (ADNI) clinical characterization. Neurology **74**(3), 201–209 (2010)
20. Schmidhuber, J., Hochreiter, S., et al.: Long short-term memory. Neural Comput **9**(8), 1735–1780 (1997)
21. Skaria, A.P.: The economic and societal burden of Alzheimer disease: managed care considerations. The American journal of managed care **28**(10 Suppl), S188–S196 (2022)
22. Sperling, R.A., Aisen, P.S.: Anti-amyloid treatment in asymptomatic alzheimer's disease (A4) and longitudinal evaluation of amyloid risk and neurodegeneration (LEARN) study longitudinal results. In: Alzheimer's Association International Conference. ALZ (2023)
23. Therriault, J., Zimmer, E.R., Benedet, A.L., Pascoal, T.A., Gauthier, S., Rosa-Neto, P.: Staging of alzheimer's disease: past, present, and future perspectives. Trends in molecular medicine **28**(9), 726–741 (2022)

24. Warde-Farley, D., Donaldson, I., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C.T., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G.B., Bader, G.D., Morris, Q.: The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Research **38**(suppl_2), W214–W220 (2010). https://doi.org/10.1093/nar/gkq537
25. Warren, S.L., Moustafa, A.A.: Functional magnetic resonance imaging, deep learning, and alzheimer's disease: A systematic review. Journal of Neuroimaging **33**(1), 5–18 (2023)
26. Xu, P., Zhu, X., Clifton, D.A.: Multimodal learning with transformers: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(10), 12113–12132 (2023)
27. Yang, L., Wang, X., Guo, Q., Gladstein, S., Wooten, D., Li, T., Robieson, W.Z., Sun, Y., Huang, X., Initiative, A.D.N.: Deep learning based multimodal progression modeling for Alzheimer's disease. Statistics in Biopharmaceutical Research **13**(3), 337–343 (2021)
28. Zheng, S., Zhu, Z., Liu, Z., Guo, Z., Liu, Y., Yang, Y., Zhao, Y.: Multi-modal graph learning for disease prediction. IEEE Transactions on Medical Imaging **41**(9), 2207–2216 (2022)
29. Zhou, Z., Wu, W., Ho, H., Wang, J., Shi, L., Davuluri, R.V., Wang, Z., Liu, H.: Dnabert-s: Learning species-aware DNA embedding with genome foundation models. ArXiv (2024)