Cross-dataset Evaluation of Dementia Longitudinal Progression Prediction Models

Chen Zhang[a], **B.T. Thomas Yeo**[a]

[a] *Centre for Sleep and Cognition (CSC) & Centre for Translational Magnetic Resonance Research (TMR), Yong Loo Lin School of Medicine, National University of Singapore, Singapore* zhang_chen@u.nus.edu

1. Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder with no cure, but early intervention can slow cognitive decline (1–3). Predicting AD progression is critical for timely treatment, caregiver planning, and optimizing clinical trial recruitment (4,5). Machine learning models leveraging multimodal biomarkers have shown promise in forecasting disease trajectory, but their generalizability across datasets remains an open question.

2. Substantial section

2.1 Related work

The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) challenge (6), involving 92 algorithms from 33 international teams, used multimodal biomarkers to predict clinical diagnosis, cognition, and ventricular volume up to five years into the future. The winning algorithm, FROG, utilized a Longitudinal-to-Cross-sectional (L2C) transformation, converting variable-length histories into fixed-length feature vectors by extracting key summary statistics (e.g., highest/lowest values, time since the highest/lowest values). This approach reformulates a longitudinal problem into a static prediction task, broadening the range of applicable machine learning models (e.g., XGBoost (7)) beyond conventional time-series approaches.

As far as we know, the L2C approach is relatively unique in the medical imaging community, where most methods fit models to entire longitudinal histories— either through parametric trajectory modeling, such as AD Course Map (AD_Map) (8), or dynamic state modeling, such as minimalRNN (9).

2.2 Method

While effective, FROG required separate models for different forecast windows and target variables, making it cumbersome. We propose L2C-FNN, a streamlined variant of FROG that replaces XGBoost with a single feedforward neural network (FNN), unifying prediction across all forecast windows and target variables.

TADPOLE evaluated models solely on the Alzheimer's Disease Neuroimaging Initiative (ADNI) (10,11) dataset, leaving cross-cohort generalizability untested. To address this, we trained models on the ADNI dataset and evaluated them on three independent cohorts—AIBL (N=402, Australia) (12), MACC (N=650, Singapore) (13), and OASIS (N=1260, North America) (14)—covering 2,312 participants and 13,200 timepoints across three continents.

ADNI participants were randomly divided into training, validation, and test sets (18:1:1) with 20 repetitions to ensure result stability (Figure 1). Care was taken to ensure non-overlapping test sets, covering the entirety of the ADNI cohort across the 20 data splits. Following TADPOLE convention, the first 50% of each participant's timepoints were used to predict the second 50%. Performance was evaluated through within- and cross-cohort comparisons. Model performance was evaluated using multiclass area under the operating curve (mAUC) (15) for clinical diagnosis, and Mean Absolute Error (MAE) for Mini-Mental State Examination (MMSE) and ventricle volume predictions.

Statistical significance was assessed using corrected resampled t-tests (16) for within-cohort evaluation and paired t-tests (17) for cross-cohort evaluation. Multiple comparisons were corrected with a false discovery rate of $q < 0.05$ (18). For more details, please refer to our preprint (19).

2.3 Results

2.3.1 Within-cohort comparisons

Figure 2 shows within-cohort clinical diagnosis prediction in ADNI. The three FROG variants: L2C-XGBw (original FROG), L2C-XGBnw, and L2C-FNN outperformed MinimalRNN, which in turn outperformed AD-Map. However, all models performed similarly for MMSE and ventricle volume prediction (Appendix Figure 1).

2.3.2 Cross-cohort comparisons

Figure 3 illustrates MMSE prediction errors across AIBL, MACC, and OASIS. L2C-FNN achieved the best overall performance, ranking first in AIBL and OASIS, while AD-Map performed slightly better in MACC. For ventricle volume prediction, overall L2C-FNN, L2C-XGBnw and AD-Map performed the best. The original FROG algorithm (L2C-XGBw) and MinimalRNN performed the worst (Appendix Figure 2). L2C-FNN also showed the best overall performance in clinical diagnosis prediction (Appendix Figure 3).

2.3.3 Long-term prediction trends

Figure 4 presents a yearly breakdown of cross-cohort MMSE prediction from Figure 3 up to year 6, showing a decline in accuracy across all models as the prediction horizon increased. However, L2C-FNN consistently matched or outperformed other methods, except in MACC, where AD-Map was statistically superior in years 0–1 and 1–2 (Table 1). Similar trends were observed for ventricle volume and clinical diagnosis (Appendix Figures 4–5). Statistical comparisons (Appendix Table 1) confirm that L2C-FNN generally maintained strong performance across all time horizons and datasets.

3. Conclusion

Our benchmarking study showed that L2C-FNN achieved the best overall cross-cohort performance and demonstrated robust long-term prediction capabilities for dementia progression.

Cross-dataset Evaluation of Dementia Longitudinal Progression Prediction Models

Chen Zhang[a], **B.T. Thomas Yeo**[a]

[a] *Centre for Sleep and Cognition (CSC) & Centre for Translational Magnetic Resonance Research (TMR), Yong Loo Lin School of Medicine, National University of Singapore, Singapore* zhang_chen@u.nus.edu

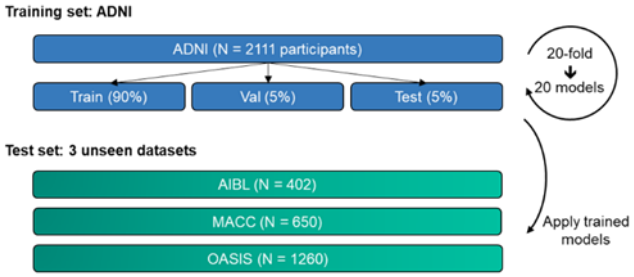
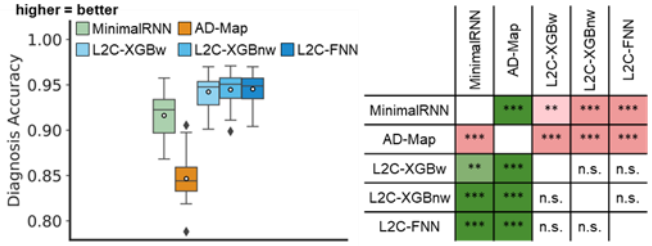
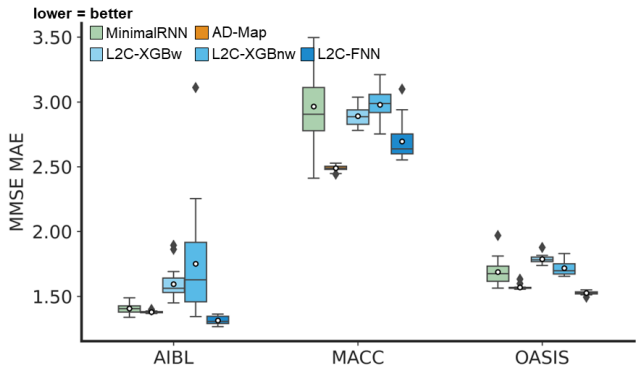


Fig. 1: Overview of model training on ADNI and evaluation on three external datasets.



	MinimalRNN	AD-Map	L2C-XGBw	L2C-XGBnw	L2C-FNN
MinimalRNN		***	**	***	***
AD-Map	***		***	***	***
L2C-XGBw	**	***		n.s.	n.s.
L2C-XGBnw	***	***	n.s.		n.s.
L2C-FNN	***	***	n.s.	n.s.	

Fig.2: Within-cohort (ADNI) prediction performance for clinical diagnosis prediction. Left: Boxplots show variability across 20 test sets. Right: Statistical differences between models. '***' $p < 0.00001$, '**' $p < 0.001$. 'n.s.' indicates non-significant results ($p > 0.05$) or those not surviving FDR correction. Each row compares a model against all others; green indicates better performance, red worse.



AIBL MMSE	Minimal RNN	AD-Map	L2C-XGBw	L2C-XGBnw	L2C-FNN
MinimalRNN		n.s.	**	***	*
AD-Map	n.s.		n.s.	**	n.s.
L2C-XGBw	**	n.s.		**	***
L2C-XGBnw	***	**	**		***
L2C-FNN	*	n.s.	***	***	

MACC MMSE	Minimal RNN	AD-Map	L2C-XGBw	L2C-XGBnw	L2C-FNN
MinimalRNN		***	n.s.	n.s.	**
AD-Map	***		***	***	**
L2C-XGBw	n.s.	***		n.s.	**
L2C-XGBnw	n.s.	***	n.s.		***
L2C-FNN	**	**	**	***	

OASIS MMSE	Minimal RNN	AD-Map	L2C-XGBw	L2C-XGBnw	L2C-FNN
MinimalRNN		**	*	n.s.	***
AD-Map	**		***	**	n.s.
L2C-XGBw	*	***		*	***
L2C-XGBnw	n.s.	**	*		***
L2C-FNN	***	n.s.	***	***	

Fig. 3: Cross-cohort MMSE prediction error (MAE) on three external datasets. Top: Boxplots show variability across 20 trained models. The x-axis denotes the test dataset. Bottom: Statistical differences between models. Each row compares a model against all others. Color and marker definitions follow Fig. 2.

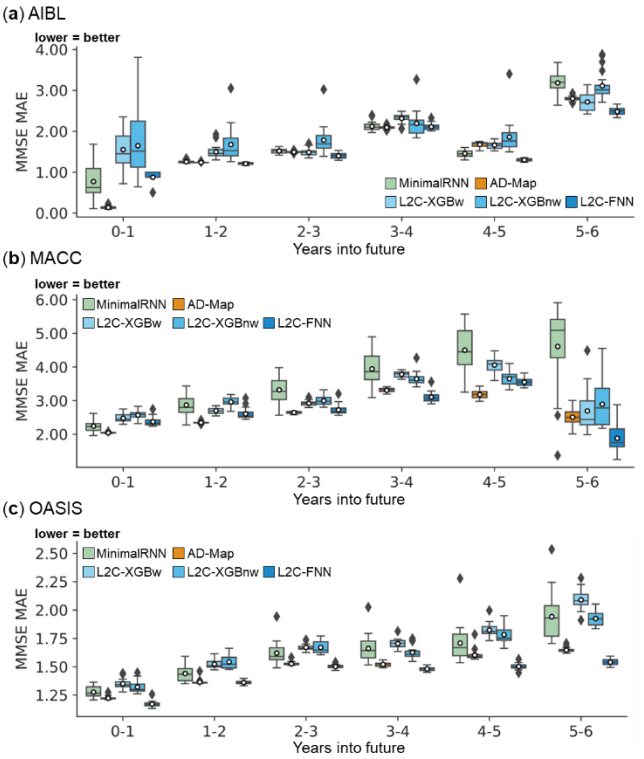


Fig. 4: Cross-cohort MMSE prediction performance (Fig. 3) broken down into yearly intervals up to 6 years.

	AIBL						MACC						OASIS					
MMSE	0-1	1-2	2-3	3-4	4-5	5-6	0-1	1-2	2-3	3-4	4-5	5-6	0-1	1-2	2-3	3-4	4-5	5-6
L2C-FNN vs MinimalRNN	ns	ns	ns	ns	ns	ns	ns	**	***	**	ns	ns	**	*	**	**	**	***
L2C-FNN vs AD-Map	ns	ns	ns	ns	*	ns	***	**	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
L2C-FNN vs L2C-XGBw	*	***	ns	ns	*	ns	ns	ns	ns	**	ns	ns	***	***	**	***	***	***
L2C-FNN vs L2C-XGBnw	ns	***	*	ns	**	ns	*	***	*	**	ns	ns	***	**	**	*	**	***

Table 1: Statistical significance between L2C-FNN and other approaches for cross-cohort MMSE prediction performance (Fig. 3) broken down into yearly intervals up to 6 years into the future. Color and marker definitions follow Fig. 2.

Acknowledgments

We would like to thank Christina Rabe, and Paul Manser from Team FROG for sharing their code with us, which significantly facilitated the current study. This research is supported by the NUS Yong Loo Lin School of Medicine (NUHSRO/2020/124/TMR/LOA), the Singapore National Medical Research Council (NMRC) LCG (OFLCG19May-0035), NMRC CTG-IIT (CTGIIT23jan-0001), NMRC STaR (STaR20nov-0003), NMRC OF-IRG (OFIRG24jan-0030), Singapore Ministry of Health (MOH) Centre Grant (CG21APR1009), the Temasek Foundation (TF2223-IMH-01), and the United States National Institutes of Health

Cross-dataset Evaluation of Dementia Longitudinal Progression Prediction Models

Chen Zhang[a], **B.T. Thomas Yeo**[a]

[a] *Centre for Sleep and Cognition (CSC) & Centre for Translational Magnetic Resonance Research (TMR), Yong Loo Lin School of Medicine, National University of Singapore, Singapore* zhang_chen@u.nus.edu

(R01MH120080 & R01MH133334). Our computational work was partially performed on resources of the National Supercomputing Centre, Singapore (https://www.nscc.sg). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of the Singapore NMRC, MOH, Temasek Foundation or USA NIH.
Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (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. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (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.
Data were provided in part by OASIS-3: Longitudinal Multimodal Neuroimaging: Principal Investigators: T. Benzinger, D. Marcus, J. Morris; NIH P30 AG066444, P50 AG00561, P30 NS09857781, P01 AG026276, P01 AG003991, R01 AG043434, UL1 TR000448, R01 EB009352. AV-45 doses were provided by Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly.

References

1. Jack CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimers Dement J Alzheimers Assoc. 2018 Apr;14(4):535–62.

2. Hampel H, Hardy J, Blennow K, Chen C, Perry G, Kim SH, et al. The Amyloid-β Pathway in Alzheimer's Disease. Mol Psychiatry. 2021 Oct;26(10):5481–503.

3. Van Dyck CH, Swanson CJ, Aisen P, Bateman RJ, Chen C, Gee M, et al. Lecanemab in Early Alzheimer's Disease. N Engl J Med. 2023 Jan 5;388(1):9–21.

4. Zhang R, Simon G, Yu F. Advancing Alzheimer's Research: A Review of Big Data Promises. Int J Med Inf. 2017 Oct;106:48–56.

5. Oxtoby NP, Shand C, Cash DM, Alexander DC, Barkhof F. Targeted Screening for Alzheimer's Disease Clinical Trials Using Data-Driven Disease Progression Models. Front Artif Intell. 2022;5:660581.

6. Marinescu RV, Oxtoby NP, Young AL, Bron EE, Toga AW, Weiner MW, et al. The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) Challenge: Results after 1 Year Follow-up. Mach Learn Biomed Imaging. 2021 Dec 31;1(December 2021 issue):1–60.

7. 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 [Internet]. New York, NY, USA: Association for Computing Machinery; 2016 [cited 2024 Jul 20]. p. 785–94. (KDD '16). Available from: https://doi.org/10.1145/2939672.2939785

8. Maheux E, Koval I, Ortholand J, Birkenbihl C, Archetti D, Bouteloup V, et al. Forecasting individual progression trajectories in Alzheimer's disease. Nat Commun. 2023 Feb 10;14(1):761.

9. Nguyen M, He T, An L, Alexander DC, Feng J, Yeo BTT. Predicting Alzheimer's disease progression using deep recurrent neural networks. NeuroImage. 2020 Nov 15;222:117203.

10. Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J Magn Reson Imaging JMRI. 2008 Apr;27(4):685–91.

11. Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. Neurology. 2010 Jan;74(3):201–9.

Cross-dataset Evaluation of Dementia Longitudinal Progression Prediction Models

Chen Zhang[a], **B.T. Thomas Yeo**[a]

[a] *Centre for Sleep and Cognition (CSC) & Centre for Translational Magnetic Resonance Research (TMR), Yong Loo Lin School of Medicine, National University of Singapore, Singapore* zhang_chen@u.nus.edu

12. Fowler C, Rainey-Smith SR, Bird S, Bomke J, Bourgeat P, Brown BM, et al. Fifteen Years of the Australian Imaging, Biomarkers and Lifestyle (AIBL) Study: Progress and Observations from 2,359 Older Adults Spanning the Spectrum from Cognitive Normality to Alzheimer's Disease. J Alzheimers Dis Rep. 2021;5(1):443–68.

13. Hilal S, Tan CS, van Veluw SJ, Xu X, Vrooman H, Tan BY, et al. Cortical cerebral microinfarcts predict cognitive decline in memory clinic patients. J Cereb Blood Flow Metab Off J Int Soc Cereb Blood Flow Metab. 2020 Jan;40(1):44–53.

14. LaMontagne PJ, Benzinger TLs, Morris JC, Keefe S, Hornbeck R, Xiong C, et al. OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease [Internet]. 2019 [cited 2024 Jun 1]. Available from: http://medrxiv.org/lookup/doi/10.1101/2019.12.13.19014902

15. Hand DJ, Till RJ. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Mach Learn. 2001 Nov 1;45(2):171–86.

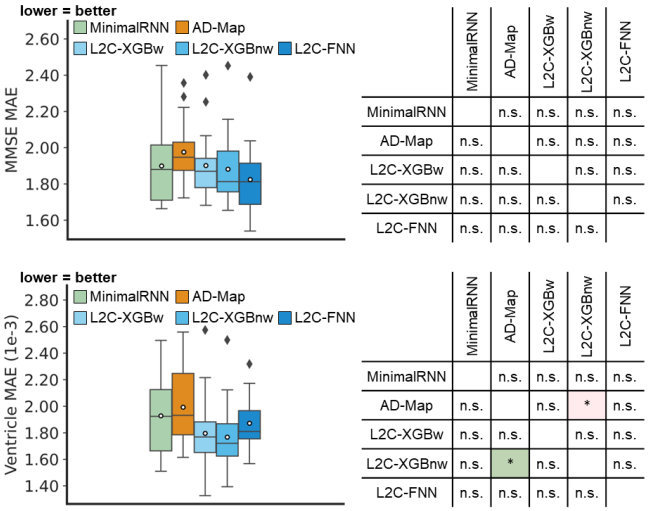
16. Bouckaert RR, Frank E. Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms. In: Dai H, Srikant R, Zhang C, editors. Berlin, Heidelberg: Springer Berlin Heidelberg; 2004 [cited 2023 Nov 7]. p. 3–12. (Kanade T, Kittler J, Kleinberg JM, Mattern F, Mitchell JC, Nierstrasz O, et al., editors. Lecture Notes in Computer Science; vol. 3056). Available from: http://link.springer.com/10.1007/978-3-540-24775-3_3

17. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. New York: Routledge; 1988. 567 p.

18. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol. 1995;57(1):289–300.

19. Zhang C, An L, Wulan N, Nguyen KN, Orban C, Chen P, et al. Cross-dataset Evaluation of Dementia Longitudinal Progression Prediction Models [Internet]. medRxiv; 2024 [cited 2024 Nov 27]. p. 2024.11.18.24317513. Available from: https://www.medrxiv.org/content/10.1101/2024.11.18.24317513v1

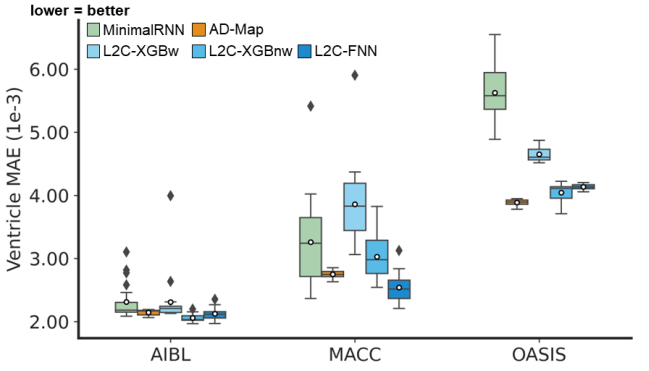
Appendix. Supplementary figures and tables



	MinimalRNN	AD-Map	L2C-XGBw	L2C-XGBnw	L2C-FNN
MinimalRNN		n.s.	n.s.	n.s.	n.s.
AD-Map	n.s.		n.s.	n.s.	n.s.
L2C-XGBw	n.s.	n.s.		n.s.	n.s.
L2C-XGBnw	n.s.	n.s.	n.s.		n.s.
L2C-FNN	n.s.	n.s.	n.s.	n.s.	

	MinimalRNN	AD-Map	L2C-XGBw	L2C-XGBnw	L2C-FNN
MinimalRNN		n.s.	n.s.	n.s.	n.s.
AD-Map	n.s.		n.s.	*	n.s.
L2C-XGBw	n.s.	n.s.		n.s.	n.s.
L2C-XGBnw	n.s.	*	n.s.		n.s.
L2C-FNN	n.s.	n.s.	n.s.	n.s.	

Appendix Fig. 1: Within-cohort (ADNI) prediction performance. Left: Boxplots show variability across 20 test sets for MMSE (top) and ventricle volume (bottom) predictions. Right: Statistical differences between models for MMSE (top) and ventricle volume (bottom). Color and marker definitions follow Fig. 2.



AIBL Ventricle	Minimal RNN	AD-Map	L2C-XGBw	L2C-XGBnw	L2C-FNN
MinimalRNN		n.s.	n.s.	*	*
AD-Map	n.s.		n.s.	n.s.	n.s.
L2C-XGBw	n.s.	n.s.		*	n.s.
L2C-XGBnw	*	n.s.	*		n.s.
L2C-FNN	*	n.s.	n.s.	n.s.	

MACC Ventricle	Minimal RNN	AD-Map	L2C-XGBw	L2C-XGBnw	L2C-FNN
MinimalRNN		**	***	*	***
AD-Map	**		***	n.s.	n.s.
L2C-XGBw	***	***		***	***
L2C-XGBnw	*	n.s.	***		***
L2C-FNN	***	n.s.	***	***	

OASIS Ventricle	Minimal RNN	AD-Map	L2C-XGBw	L2C-XGBnw	L2C-FNN
MinimalRNN		***	*	***	***
AD-Map	***		***	n.s.	n.s.
L2C-XGBw	*	***		***	***
L2C-XGBnw	***	n.s.	***		n.s.
L2C-FNN	***	n.s.	***	n.s.	

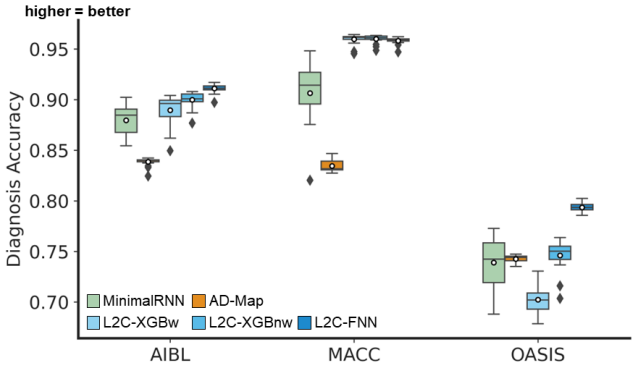
Appendix Fig. 2: Cross-cohort ventricle volume prediction error (MAE) on three external datasets. Top: Boxplots show variability across 20 trained models. The x-axis denotes the

Cross-dataset Evaluation of Dementia Longitudinal Progression Prediction Models

Chen Zhang[a], **B.T. Thomas Yeo**[a]

[a] *Centre for Sleep and Cognition (CSC) & Centre for Translational Magnetic Resonance Research (TMR), Yong Loo Lin School of Medicine, National University of Singapore, Singapore* zhang_chen@u.nus.edu

test dataset. Bottom: Statistical differences between models. Each row compares a model against all others. Color and marker definitions follow Fig. 2.

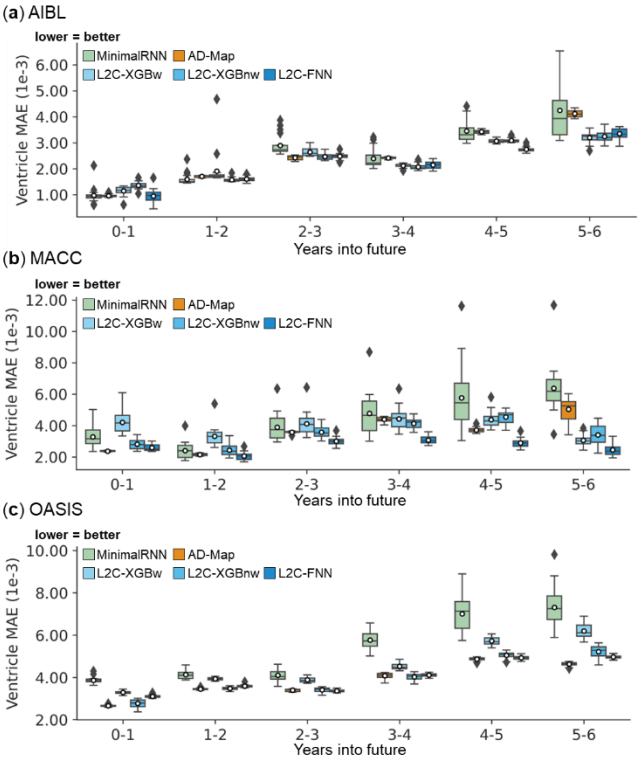


AIBL Diagnosis	Minimal RNN	AD-Map	L2C-XGBw	L2C-XGBnw	L2C-FNN
MinimalRNN		n.s.	n.s.	n.s.	n.s.
AD-Map	n.s.		*	*	**
L2C-XGBw	n.s.	*		n.s.	n.s.
L2C-XGBnw	n.s.	*	n.s.		n.s.
L2C-FNN	n.s.	**	n.s.	n.s.	

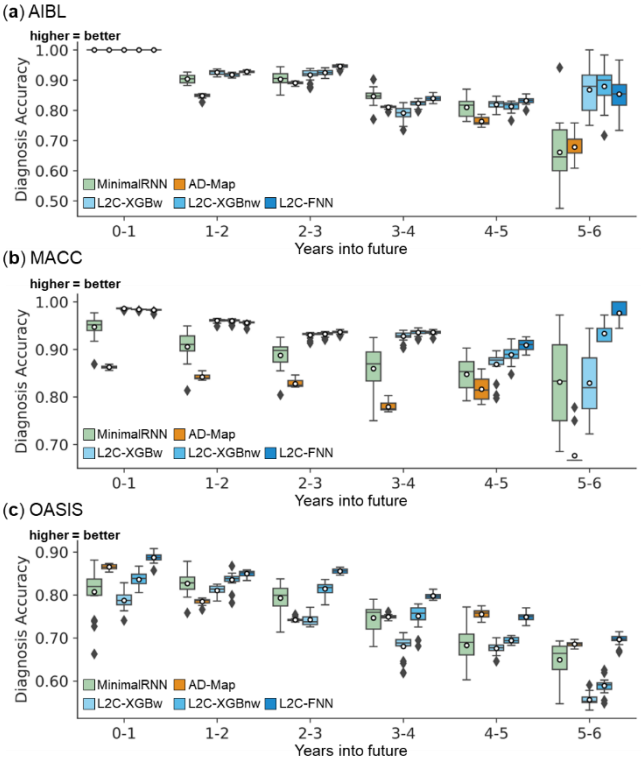
MACC Diagnosis	Minimal RNN	AD-Map	L2C-XGBw	L2C-XGBnw	L2C-FNN
MinimalRNN		**	*	*	*
AD-Map	**		**	**	**
L2C-XGBw	*	**		n.s.	n.s.
L2C-XGBnw	*	**	n.s.		n.s.
L2C-FNN	*	**	n.s.	n.s.	

OASIS Diagnosis	Minimal RNN	AD-Map	L2C-XGBw	L2C-XGBnw	L2C-FNN
MinimalRNN		n.s.	**	n.s.	n.s.
AD-Map	n.s.		n.s.	n.s.	*
L2C-XGBw	**	n.s.		**	**
L2C-XGBnw	n.s.	n.s.	**		**
L2C-FNN	n.s.	*	**	**	

Appendix Fig. 3: Cross-cohort ventricle clinical diagnosis prediction accuracy (mAUC) on three external datasets. Top: Boxplots show variability across 20 trained models. The x-axis denotes the test dataset. Bottom: Statistical differences between models. Each row compares a model against all others. Color and marker definitions follow Fig. 2.



Appendix Fig. 4: Cross-cohort ventricle volume prediction performance (Appendix Fig. 2) broken down into yearly intervals up to 6 years.



Appendix Fig. 5: Cross-cohort clinical diagnosis prediction performance (Appendix Fig. 3) broken down into yearly intervals up to 6 years.

Cross-dataset Evaluation of Dementia Longitudinal Progression Prediction Models

Chen Zhang[a], **B.T. Thomas Yeo**[a]

[a] *Centre for Sleep and Cognition (CSC) & Centre for Translational Magnetic Resonance Research (TMR), Yong Loo Lin School of Medicine, National University of Singapore, Singapore* zhang_chen@u.nus.edu

	AIBL						MACC						OASIS					
Ventricle	0-1	1-2	2-3	3-4	4-5	5-6	0-1	1-2	2-3	3-4	4-5	5-6	0-1	1-2	2-3	3-4	4-5	5-6
L2C-FNN vs MinimalRNN	ns	ns	*	ns	***	ns	**	**	**	**	*	ns	*	ns	ns	*	ns	**
L2C-FNN vs AD-Map	ns	ns	ns	ns	*	ns	ns	ns	ns	*	ns	ns	ns	ns	ns	ns	ns	ns
L2C-FNN vs L2C-XGBw	ns	**	ns	ns	ns	ns	***	***	***	***	*	ns	ns	ns	*	ns	ns	**
L2C-FNN vs L2C-XGBnw	*	ns	ns	ns	ns	ns	ns	***	**	***	*	ns	ns	ns	ns	ns	ns	ns
Diagnosis	0-1	1-2	2-3	3-4	4-5	5-6	0-1	1-2	2-3	3-4	4-5	5-6	0-1	1-2	2-3	3-4	4-5	5-6
L2C-FNN vs MinimalRNN	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
L2C-FNN vs AD-Map	ns	**	ns	ns	ns	ns	**	**	**	**	ns	ns	ns	ns	**	ns	ns	ns
L2C-FNN vs L2C-XGBw	ns	ns	**	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	*	**	ns	**
L2C-FNN vs L2C-XGBnw	ns	ns	*	ns	ns	ns	ns	ns	ns	ns	ns	ns	**	ns	ns	ns	ns	*

Appendix Table 1: Statistical significance between L2C-FNN and other approaches for cross-cohort ventricle volume (Appendix Fig. 2) and clinical diagnosis (Appendix Fig. 3) prediction performance broken down into yearly intervals up to 6 years into the future. Color and marker definitions follow Fig. 2.