Using Machine Learning to Forecast the Onset of Cognitive Decline

Andrew Gautier, Lingjian Zheng, Dr. Bernard Schreurs

Lane Department of Computer Science and Electrical Engineering & Rockerfeller Neuroscience Institute

West Virginia University

Abstract

Predicting the likelihood of patients progressing to Alzheimer's disease (AD) while they are presymptomatic is a challenging but important task for healthcare. To study early indicators of subjective cognitive decline (SCD), we selected a cohort from the National Alzheimer's Coordinating Center (**n = 4,407**) with four distinct cognitive trajectories. A gradient-boosted tree ensemble method was used to train a classification model for two binary classification tasks: progression to mild cognitive impairment (MCI) and progression to AD. Across all prediction windows, the MCI progression model AUROC was **85.6%,** and the AD progression model was **77%.** Lead time to diagnosis experiments showed promise that machine learning may provide utility in assisting healthcare professionals in assessing the risk of future cognitive decline.

Background

XGBoost [1] has gained popularity as a machine learning algorithm for its ability to perform well on sparse datasets and handle missing responses which are common in longitudinal clinical data.

Adelson et al. [2] trained and validated an XGBoost classifier to predict the progression of AD in a 48-month window. Our work will follow a similar approach but will evaluate a longer observation period and include the progression of normal cognition (CN) to MCI as well as MCI to AD.

Demographic Statistics

Table 1: **Demographic and Comorbidity Information**

Category	Subgroup	Stable CN *(n=2,981)*	CN to MCI *(n=212)*	Stable MCI *(n=359)*	MCI to AD *(n=855)*
Age (years)	51 60	344 (11.5%)	11 (5.2%)	41 (11.4%)	31 (3.6%)
	61 70	995 (33.4%)	55 (25.9%)	122 (34.0%)	173 (20.2%)
	71 80	979 (32.8%)	89 (42.0%)	139 (38.7%)	386 (45.1%)
	81 90	507 (17.0%)	48 (22.6%)	47 (13.1%)	224 (26.2%)
Sex	Male	972 (32.6%)	94 (44.3%)	186 (51.8%)	414 (48.4%)
	Female	2009 (67.4%)	118 (55.7%)	173 (48.2%)	441 (51.6%)
Race	White	2315 (77.7%)	165 (77.8%)	256 (71.3%)	713 (83.4%)
	Black	551 (18.5%)	38 (17.9%)	76 (21.2%)	109 (12.7%)
	Am. Indian	26 (0.9%)	1 (0.5%)	8 (2.2%)	7 (0.8%)
	Asian	73 (2.4%)	4 (1.9%)	9 (2.5%)	10 (1.2%)
	Other	15 (0.5%)	4 (1.9%)	10 (2.8%)	16 (1.9%)
Ethnicity	Hispanic	185 (6.2%)	3 (6.1%)	32 (8.9%)	70 (8.2%)
	Non-Hispanic	2796 (93.8%)	199 (93.9%)	327 (91.1%)	785 (91.8%)
Comorbidities	Alcohol abuse	14 (0.5%)	0 (0.0%)	4 (1.1%)	10 (1.2%)
	Heart attack	22 (0.7%)	2 (0.9%)	1 (0.3%)	11 (1.3%)
	Atrial fibrillation	171 (5.7%)	17 (8.0%)	12 (3.3%)	61 (7.1%)
	Diabetes	373 (12.5%)	38 (17.9%)	65 (18.1%)	125 (14.6%)
	Hypercholesterolemia	1477 (49.5%)	110 (51.9%)	195 (54.3%)	483 (56.5%)
	Hypertension	1597 (53.6%)	122 (57.5%)	206 (57.4%)	486 (56.8%)
	Vitamin B12 deficiency	82 (2.8%)	9 (4.2%)	13 (3.6%)	35 (4.1%)
	Depression	191 (6.4%)	30 (14.2%)	87 (24.2%)	201 (23.5%)
	Anxiety	104 (3.5%)	17 (8.0%)	33 (9.2%)	142 (16.6%)
	TBI	291 (9.8%)	24 (11.3%)	47 (13.1%)	76 (8.9%)

Dataset Preparation

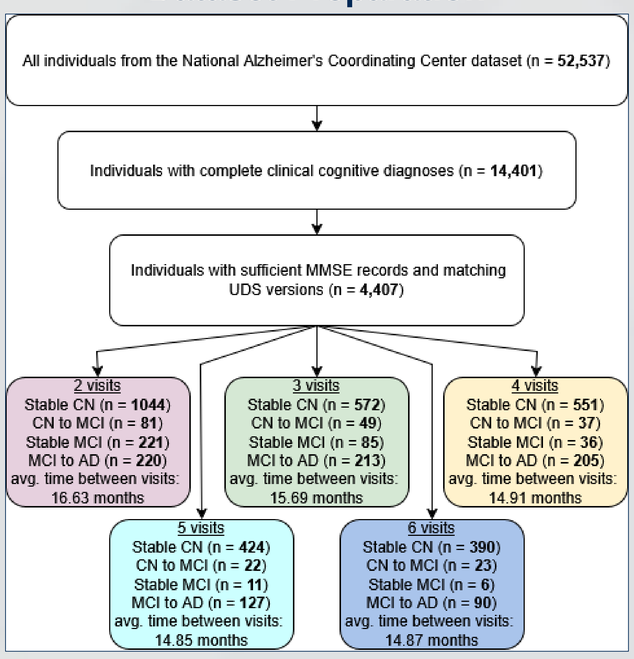


Fig 1: Cohort Filtering Criteria for Dataset

- Average time interval between visits: **15.67** months
- Training and testing set split between 80% and 20% randomly chosen.
  - Training set: (n = **3,526**)
  - Testing set: (n = **881**)
- Complete MMSE records were required for 2 and 3 visit cohorts, imputation of one missing record was allowed for 4 and 5, and 2 missing records were allowed for the 6-visit cohort.

Methodology

- **34** features were selected, combining static variables and longitudinal measurements.
- The target variable was cognitive status.

Table 2: **Features used to train ML algorithm across all time-points.**

Clinical History	Neuropsychiatric Assessments
Comorbidities	Mini-Mental State Examination (MMSE)
Years of Education	Functional Activities Questionnaire (FAQ)
Smoking History	Clinical Dementia Rating (CDR)
Body Mass Index	Geriatric Depression Scale (GDS)

Experiment 1: Patient Classification

XGBoost is an algorithm which can robustly handle complex datasets and missing values.

Parameters

- Estimators = **200**
- Maximum tree depth = **5**
- Learning rate = **0.1**
- Loss function = **Binary Logistic Regression**

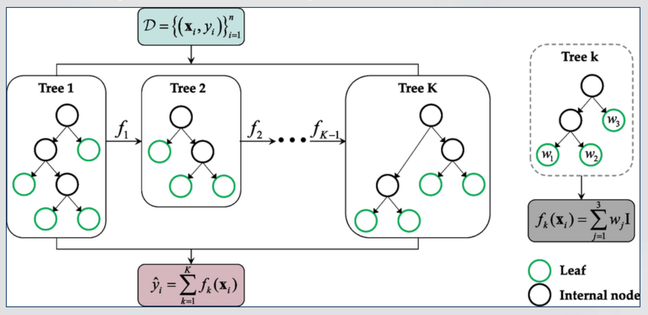


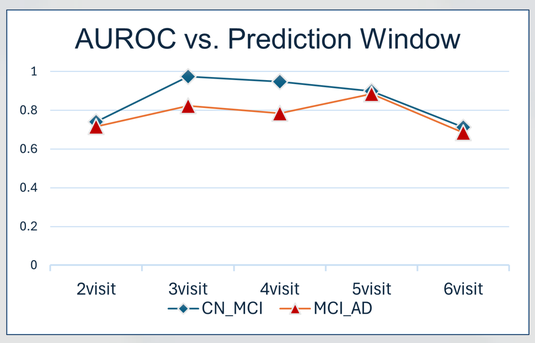
Figure 5: XGBoost Algorithm

Experiment 2: Forecasting Diagnosis Conversion

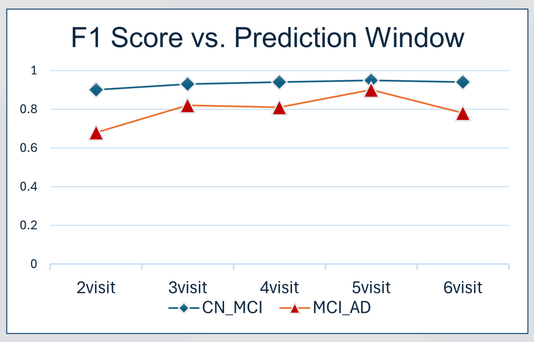
- The pretrained classifiers were retooled to predict class conversion probability interval with different thresholds for likelihood of conversion.
- Features were re engineered for each visit window using truncated time-series data.
- **Lead time Calculation**
  - Identified the **diagnosis visit**
  - Evaluated models retrospectively at earlier visits.
  - Recorded earliest visit where model probability ≥ threshold
  - **Lead time =** (Diagnosis visit time) – (Prediction visit time)

Results

(A)



(B)



(C)

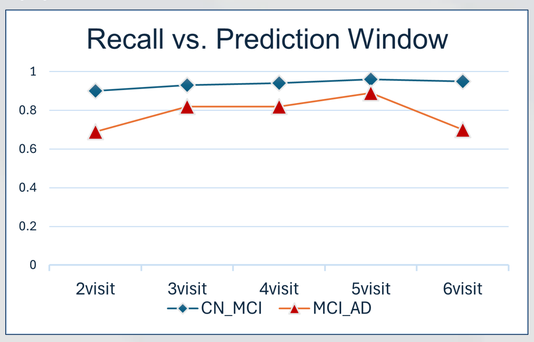


Fig 3: Performance metrics of XGBoost classifier evaluated across six cohorts.

Table 3: Model performance metrics across different visit intervals

	AUROC		Recall		F1 Score	
Visits	CN-MCI	MCI-AD	CN-MCI	MCI-AD	CN-MCI	MCI-AD
2 visit	0.74	0.716	0.90	0.69	0.90	0.68
3 visit	0.974	0.823	0.93	0.82	0.93	0.82
4 visit	0.947	0.785	0.94	0.82	0.94	0.81
5 visit	0.898	0.884	0.96	0.89	0.95	0.90
6 visit	0.712	0.684	0.95	0.70	0.94	0.78

Table 4: **Top 5 Most Important Features by Visit and Model Type**

Visit	CN-MCI Models	MCI-AD Models
2 visit	1. CDR_mean (47.08)	1. COR_max (10.47)
	2. CDR_V2 (43.07)	2. CDR_V2 (5.70)
	3. REMDATES_delta_V2-V1 (12.92)	3. REMDATES_V2 (2.12)
	4. CVMATT (9.88)	4. EVENTS_mean (1.99)
	5. MEALPREP_V1 (9.07)	5. PAVATTN_slope (1.97)
3 visit	1. CDR_max (34.62)	1. CDR_max (8.82)
	2. CDR_mean (31.26)	2. CDR_V2 (8.28)
	3. TAXES_delta_V2-V1 (15.48)	3. PAYATTN_delta_V2-V1 (2.89)
	4. HISPANIC (11.89)	4. CDR_V3 (2.80)
	5. SHOPPING_V3 (11.49)	5. STOVE_delta_V3-V1 (2.67)
4 visit	1. CDR_mean (54.86)	1. CDR_V4 (6.81)
	2. BILLS_V1 (22.92)	2. EVENTS_V4 (3.56)
	3. CDR_max (14.47)	3. EVENTS_V2 (1.75)
	4. BILLS_delta_V2-V1 (12.41)	4. PAYATTN_slope (1.72)
	5. MMSE_max (11.15)	5. MEALPREP_max (1.69)
5 visit	1. CDR_mean (24.20)	1. CDR_mean (2.54)
	2. CDR_max (23.00)	2. TRAVEL_mean (2.42)
	3. CDR_V4 (17.71)	3. MMSE_VS (2.34)
	4. TODAC30_V1 (17.70)	4. TRAVEL_slope (2.02)
	5. STOVE_slope (12.83)	5. REMDATES_VS (1.59)
6 visit	1. CDR_max (34.90)	1. REMDATES_V5 (1.29)
	2. TAXES_slope (32.35)	2. GDS_V2 (1.15)
	3. TAXES_delta_VG-V1 (30.38)	3. BML_slope (1.11)
	4. SHOPPING_V4 (25.79)	4. EVENTS_mean (1.03)
	5. GDS_delta_V3-V1 (23.97)	5. PAYATTN_mean (0.98)

Table 5: **MCI Lead time performance at different thresholds**

Thresholds	0.5	0.6	0.7	0.8	0.9
Early Predictions	16 (7.5%)	11 (5.2%)	10 (4.7%)	8 (3.8%)	7 (3.3%)
Mean Lead Time (months)	21.6	21.4	22.0	23.5	24.7

Table 6: **AD Lead time performance at different thresholds**

Thresholds	0.5	0.6	0.7	0.8	0.9
Early Predictions	160 (18.7%)	147 (17.2%)	126 (14.7%)	92 (10.8%)	66 (7.7%)
Mean Lead Time	20.1	20.3	20.8	21.2	20.5

Average observation period in years

- 2 visit = **2.78**
- 3 visit = **3.92**
- 4 visit = **4.97**
- 5 visit = **6.19**
- 6 visit = **7.44**

Discussion

- The MCI progression classifier performed better than the AD overall.
- High F1 score and Recall across the prediction window suggest robustness to imbalanced data and low false negatives.
- Small sample size likely impacted the 6-visit cohort.
- We have tested a longer observation period than comparative literature [2].
- Features which represented **Instrumental Activities of Daily Living** (IADLs) were consistently highly rated in feature importance of both progression models.

Future Work

Future projects will seek to incorporate both clinical and imaging data to advance the study of early Alzheimer's detection. Combining machine learning for clinical data and deep learning architectures for highly complex image analysis has shown promising results in biomarker discovery [3].The NACC has a dataset of MRI and PET imaging modalities, which could be used in conjunction with these progression models to further discover early indicators of AD.

Literature Cited

[1] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

[2] Adelson, R. P., Garikipati, A., Maharjan, J., Ciobanu, M., Barnes, G., Singh, N. P., Dinenno, F. A., Mao, Q., & Das, R. (2023). Machine Learning Approach for Improved Longitudinal Prediction of Progression from Mild Cognitive Impairment to Alzheimer's Disease. *Diagnostics (Basel, Switzerland)*, *14*(1), 13. https://doi.org/10.3390/diagnostics14010013

[3] Long, J. M., Coble, D. W., Xiong, C., Schindler, S. E., Perrin, R. J., Gordon, B. A., Benzinger, T. L. S., Grant, E., Fagan, A. M., Harari, O., Cruchaga, C., Holtzman, D. M., & Morris, J. C. (2022). Preclinical Alzheimer's disease biomarkers accurately predict cognitive and neuropathological outcomes. *Brain*, *145*(12), 4506–4518. https://doi.org/10.1093/brain/awac250

Acknowledgements

Thanks to my mentor Dr. Schreurs for his patience and support throughout this project, as well as the AGE-ADAR program for funding my research. Special thanks to Lingjian Zheng for his contributions.

GitHub URL: *https://github.com/Andrew-Gautier/AD-Early-Prediction*

West Virginia University

AGE-ADAR SCHOLARS PROGRAM