

Supplementary Information

Evaluating Patient Representation Strategies for Early Prediction of Alzheimer’s Disease and Related Dementias from Longitudinal EHRs

Shahla Farzana, Ash Arian, Moïse Desvarieux, Tatjana Rundek, Habibul Ahsan

Supplementary Methods

Table S1: OMOP standard concepts used to define ADRD cases.

Type	Concept IDs/ICD-10 code (Name)
Conditions	443605/ICD-10/F01 (Vascular dementia), 380701/ICD-10/G31.83 (Diffuse Lewy body disease), 4196433/ICD-10/G31.83 (Senile dementia of the Lewy body type), 43021816/ICD-10/G30.8, ICD-10/F00.2 (Mixed dementia), 37109056/ICD-10/F01.50 (Vascular dementia without behavioral disturbance), 4043378/ICD-10/G31.09 (Frontotemporal dementia), 44782763/ICD-10/G31.83 (Lewy body dementia with behavioral disturbance), 378419/ICD-10/G30.9 (Alzheimer’s disease), 37018688/ICD-10/F01.51 (Vascular dementia with behavioral disturbance).
Medications	715997 (Donepezil), 733523 (Rivastigmine), 757627 (Galantamine), 701322 (Memantine).

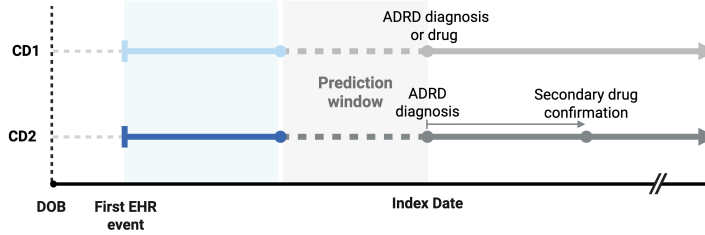


Fig. S1: Schematic illustration of the longitudinal relationship between EHR observation, prediction, and ADRD case ascertainment for the primary case definitions. For CD1, the index date was defined by the earliest qualifying ADRD diagnosis or anti-ADRDR medication exposure. For CD2, the index date corresponded to the qualifying ADRDR diagnosis, with a subsequent anti-ADRDR medication exposure serving as secondary confirmation. For each prediction horizon, the prediction window immediately preceded the index date, and only EHR information recorded before the start of that window was available for model input. DOB, date of birth; EHR, electronic health record; ADRDR, Alzheimer’s disease and related dementias.

Table S2: Index criteria distribution.

Index Criteria	Distribution (%)	
	AoU	UCM
Anti-ADRD medication	63.0%	52.2%
Alzheimer's disease	18.0%	20.3%
Vascular dementia	14.4%	24.9%
Frontotemporal dementia	2.5%	1.5%
Lewy body dementia	2.1%	1.2%
Mixed dementia	0.1%	-

Supplementary Results

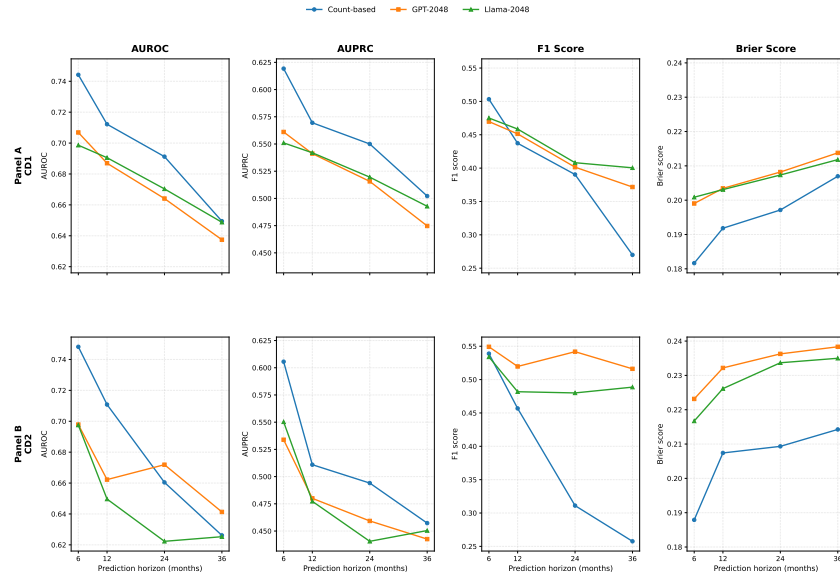


Fig. S2: Held-out test set results on UCM EHRs trained on UCM train set.

Table S3: Demographic characteristics, healthcare utilization, and EHR history for the CD2 cohort. Number of encounters are reported as *mean (SD)*.

		AoU		UCM	
Variable	Sub-categories	Case (N=353)	Control (N=705)	Case (N=376)	Control (N=750)
Demographics					
Age	at index date, median	75.3	75.0	79.4	79.4
Gender	Male	51.0%	51.1%	38.8%	38.8%
	Female	47.3%	47.4%	61.2%	61.2%
	Unknown	1.7%	1.6%	-	-
Race	White	64.6%	64.7%	25.8%	25.9%
	Unknown	16.4%	16.5%	4.0%	4.0%
	Black or African American	13.3%	13.3%	68.9%	69.1%
	Others	3.4%	3.3%	0.5%	0.3%
Ethnicity	Asian	2.3%	2.3%	0.8%	0.8%
	NonHispanic	82.4%	82.7%	95.2%	93.7%
	Hispanic	13.9%	13.6%	3.7%	2.9%
	Unknown	3.7%	3.7%	1.1%	3.3%
Medical history length					
History years	median	15.4 ^a	11.4 ^a	8.0	7.5
Number of years, range	[4,6)	4.8% ^b	10.6% ^b	18.9%	22.8%
	[6,+)	95.2% ^b	89.4% ^b	81.1%	77.2%
Number of encounters	All types of encounters	378.6 (539.4)	187.8 (255.6)	69.5 (79.4)	52.9 (58.3)
	Ambulatory encounters	277.0 ^a (426.3)	123.1 ^a (205.6)	63.7 (75.2)	48.0 (54.7)
	Inpatient encounters	15.6 ^a (95.0)	5.6 ^a (21.2)	2.6 (5.3)	2.6 (6.1)
	Emergency encounters	7.2 ^a (33.3)	3.1 ^a (8.5)	3.1 ^a (4.9)	2.4 ^a (4.7)

^a Significant difference between case and control using Mann-Whitney U test.

^b Significant difference between case and control using Pearson chi-square test.

Table S4: Cross-validation performance of the best-performing feature variant within each representation strategy for CD1 across all prediction horizons.

Prediction horizon	Representation strategy	Selected variant	Classifier	AUROC	AUPRC	F1 score	Brier score
6	Count-based	Count+Smoking	LightGBM	0.806 $\pm$ 0.000	0.690 $\pm$ 0.009	0.644 $\pm$ 0.011	0.162 $\pm$ 0.002
	CLMBR-T	Emb+Smoking	LR	0.748 $\pm$ 0.016	0.609 $\pm$ 0.023	0.592 $\pm$ 0.017	0.190 $\pm$ 0.004
	GPT-512	Emb+Smoking	LR	0.769 $\pm$ 0.009	0.637 $\pm$ 0.017	0.613 $\pm$ 0.011	0.183 $\pm$ 0.003
	GPT-2048	Emb+Smoking	LR	0.765 $\pm$ 0.015	0.631 $\pm$ 0.028	0.603 $\pm$ 0.018	0.184 $\pm$ 0.005
	Llama-512	Emb+Smoking	LR	0.736 $\pm$ 0.012	0.585 $\pm$ 0.016	0.583 $\pm$ 0.012	0.195 $\pm$ 0.003
	Llama-2048	Emb+Smoking	LR	0.759 $\pm$ 0.013	0.620 $\pm$ 0.018	0.602 $\pm$ 0.010	0.187 $\pm$ 0.004
	Mamba-16K	Emb+Smoking	LR	0.742 $\pm$ 0.014	0.603 $\pm$ 0.027	0.580 $\pm$ 0.004	0.192 $\pm$ 0.004
12	Count-based	Count	LightGBM	0.772 $\pm$ 0.003	0.640 $\pm$ 0.010	0.608 $\pm$ 0.012	0.175 $\pm$ 0.002
	CLMBR-T	Emb+Lifestyle	LR	0.723 $\pm$ 0.012	0.568 $\pm$ 0.021	0.561 $\pm$ 0.029	0.198 $\pm$ 0.004
	GPT-512	Emb+Lifestyle	LR	0.737 $\pm$ 0.012	0.587 $\pm$ 0.022	0.581 $\pm$ 0.009	0.194 $\pm$ 0.004
	GPT-2048	Emb+Lifestyle	LR	0.738 $\pm$ 0.013	0.590 $\pm$ 0.016	0.578 $\pm$ 0.028	0.193 $\pm$ 0.004
	Llama-512	Emb+Smoking	LR	0.717 $\pm$ 0.015	0.555 $\pm$ 0.016	0.567 $\pm$ 0.010	0.200 $\pm$ 0.005
	Llama-2048	Emb+Smoking	LR	0.740 $\pm$ 0.013	0.588 $\pm$ 0.015	0.591 $\pm$ 0.015	0.193 $\pm$ 0.004
	Mamba-16K	Emb+Smoking	LR	0.723 $\pm$ 0.007	0.577 $\pm$ 0.017	0.571 $\pm$ 0.004	0.197 $\pm$ 0.002
24	Count-based	Count+Alcohol	LightGBM	0.738 $\pm$ 0.003	0.597 $\pm$ 0.005	0.582 $\pm$ 0.002	0.186 $\pm$ 0.000
	CLMBR-T	Emb+Lifestyle	LR	0.682 $\pm$ 0.015	0.520 $\pm$ 0.019	0.530 $\pm$ 0.014	0.226 $\pm$ 0.004
	GPT-512	Emb+Lifestyle	LR	0.702 $\pm$ 0.017	0.539 $\pm$ 0.023	0.553 $\pm$ 0.018	0.221 $\pm$ 0.004
	GPT-2048	Emb+Smoking	LR	0.711 $\pm$ 0.015	0.558 $\pm$ 0.021	0.559 $\pm$ 0.016	0.200 $\pm$ 0.003
	Llama-512	Emb+Lifestyle	LR	0.691 $\pm$ 0.010	0.522 $\pm$ 0.009	0.549 $\pm$ 0.008	0.206 $\pm$ 0.003
	Llama-2048	Emb+Smoking	LR	0.713 $\pm$ 0.014	0.552 $\pm$ 0.017	0.559 $\pm$ 0.016	0.201 $\pm$ 0.004
	Mamba-16K	Emb+Lifestyle	LR	0.692 $\pm$ 0.015	0.528 $\pm$ 0.020	0.549 $\pm$ 0.009	0.206 $\pm$ 0.003
36	Count-based	Count+Smoking	LightGBM	0.710 $\pm$ 0.004	0.557 $\pm$ 0.010	0.546 $\pm$ 0.022	0.194 $\pm$ 0.002
	CLMBR-T	Emb+Lifestyle	LR	0.663 $\pm$ 0.010	0.498 $\pm$ 0.016	0.504 $\pm$ 0.043	0.212 $\pm$ 0.002
	GPT-512	Emb+Lifestyle	LR	0.692 $\pm$ 0.013	0.521 $\pm$ 0.021	0.545 $\pm$ 0.012	0.206 $\pm$ 0.003
	GPT-2048	Emb+Smoking	LR	0.696 $\pm$ 0.009	0.531 $\pm$ 0.006	0.536 $\pm$ 0.007	0.222 $\pm$ 0.001
	Llama-512	Emb+Smoking	LR	0.677 $\pm$ 0.006	0.509 $\pm$ 0.012	0.522 $\pm$ 0.045	0.209 $\pm$ 0.002
	Llama-2048	Emb+Lifestyle	LR	0.694 $\pm$ 0.008	0.530 $\pm$ 0.011	0.548 $\pm$ 0.006	0.205 $\pm$ 0.002
	Mamba-16K	Emb+Lifestyle	LR	0.669 $\pm$ 0.014	0.498 $\pm$ 0.021	0.529 $\pm$ 0.017	0.211 $\pm$ 0.002

Values are reported as mean $\pm$ standard deviation across three cross-validation folds. Within prediction horizon, and model family, only the feature variant selected according to the prespecified cross-validation model-selection procedure is shown. Lower Brier scores indicate better probabilistic accuracy.

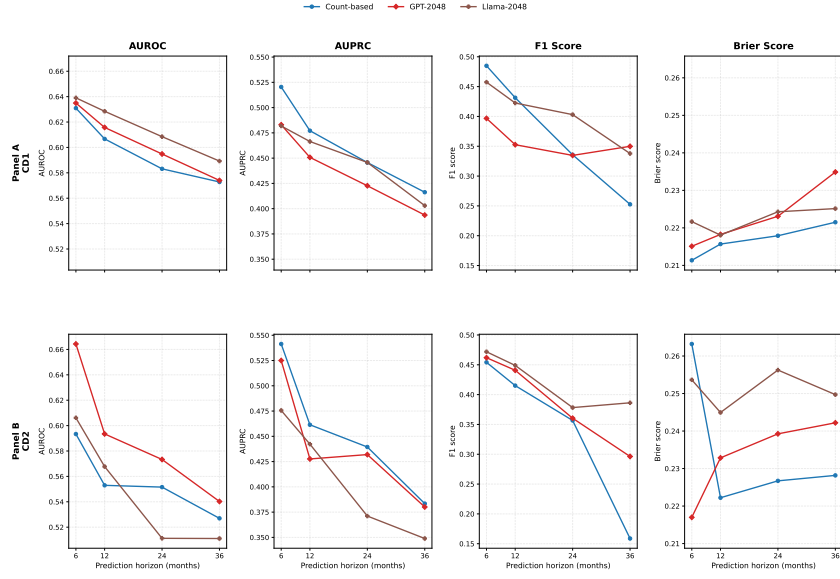


Fig. S3: Held-out test set results on UCM EHRs with zero-shot transfer from AoU train set.

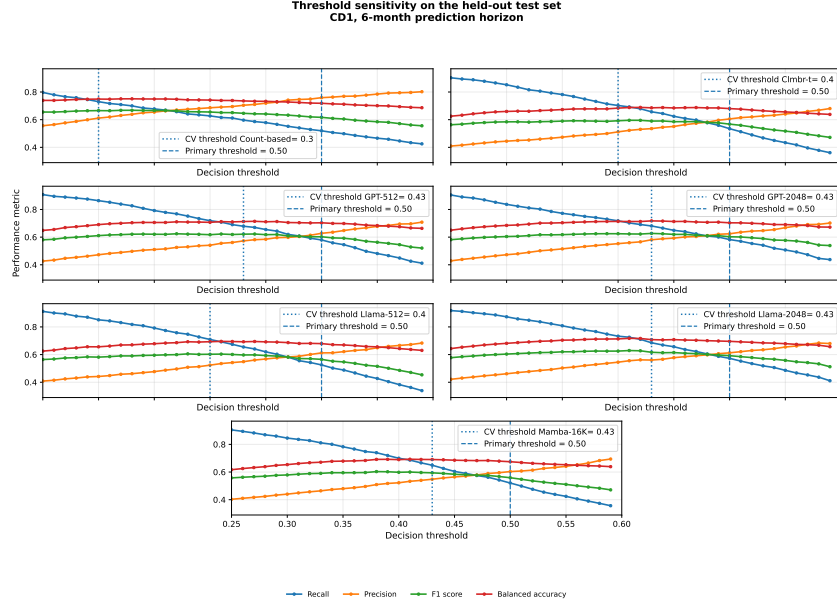


Fig. S4: Threshold sensitivity analysis as identified from the held-out test-set results of AoU dataset, in the CD1 cohort at the 6-month prediction horizon. Recall, precision, F1 score, and balanced accuracy were evaluated across decision thresholds from 0.25 to 0.60 for the best-performing representative models from the count-based and CLMBR model families. Dashed vertical lines indicate the fixed threshold of 0.50 used in the primary test-set analysis, whereas dotted vertical lines indicate the model-specific thresholds selected during cross-validation. Increasing the decision threshold improved precision but reduced recall, while F1 score and balanced accuracy remained comparatively stable across intermediate thresholds.

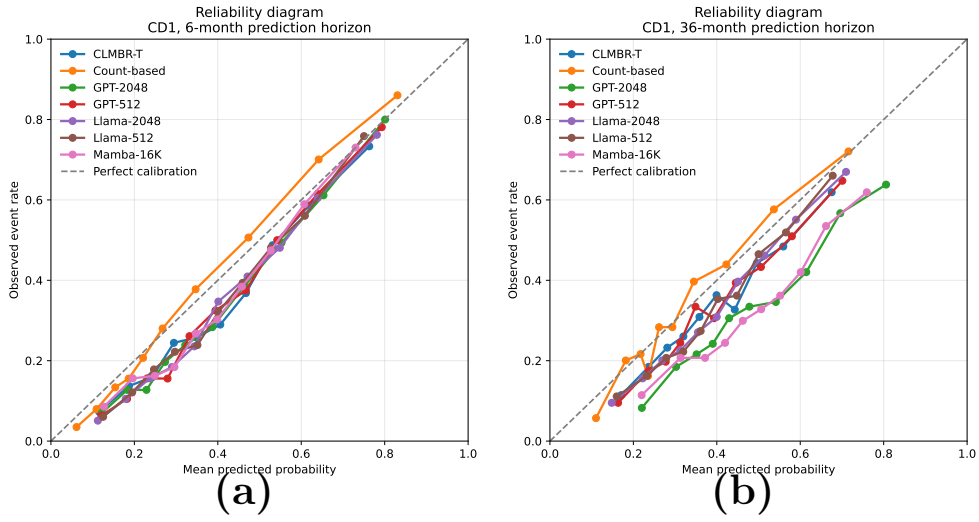


Fig. S5: Reliability curve as identified from held-out test set of AoU dataset for CD1 cohort at (a) 6-month prediction horizon (b) 36-month prediction horizon.

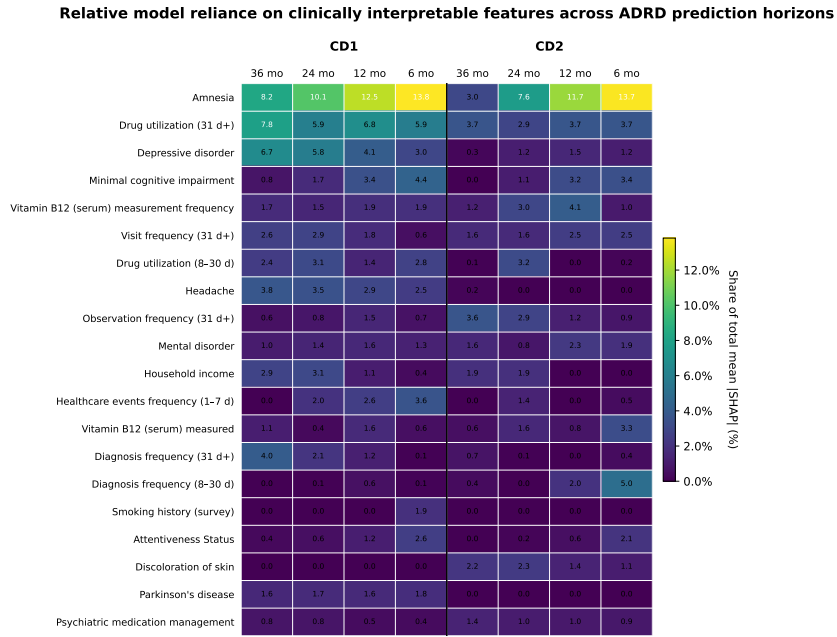


Fig. S6: Cross-horizon feature importance for the top 20 predictors.

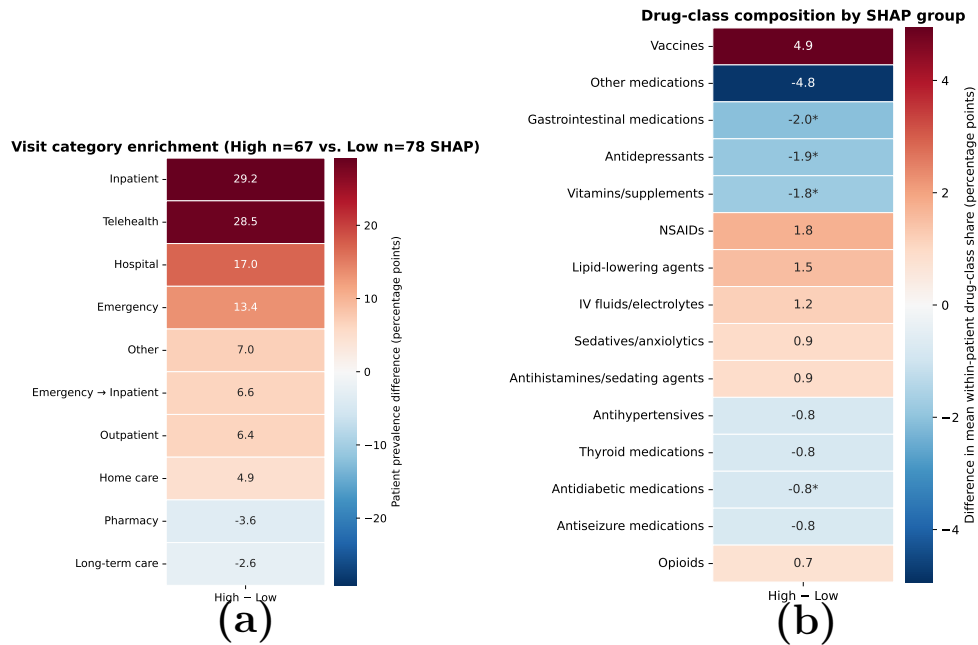


Fig. S7: Decomposition of the major healthcare utilization features identified by SHAP. (a) Difference in the prevalence of encounter types between patients with high and low SHAP values for the visit-frequency feature. Values represent absolute percentage-point differences in the proportion of patients with at least one encounter of each visit type before the prediction cutoff. (b) Difference in normalized therapeutic-class composition between high- and low-SHAP groups (high $n = 78$ vs. low $n = 43$) for the drug-utilization feature. Values represent percentage-point differences in the mean within-patient proportion of medication exposures belonging to each therapeutic class. Asterisks denote Benjamini–Hochberg-adjusted $q < 0.05$.

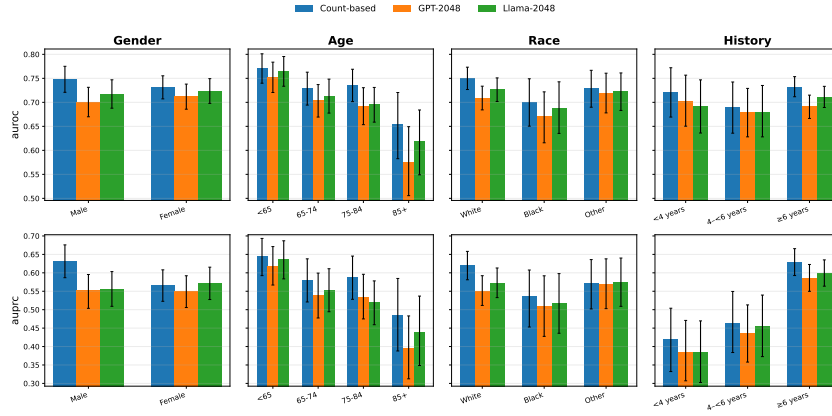


Fig. S8: Performance across clinically relevant patient subgroups of AoU dataset. AUROC (top) and AUPRC (bottom) of the representative Count-based, GPT-2048, and LLaMA-2048 models on the held-out test set for the CD1 cohort at the 36-month prediction horizon, stratified by sex, age, race, and pre-prediction clinical history length. Error bars denote 95% bootstrap percentile intervals estimated within each subgroup.