Journal of
Dementia and Alzheimer's Disease

MDPI

Article

Machine Learning-Based Risk Stratification for Survey-Derived Incident Dementia Among Older Korean Adults: A National Longitudinal Person-Period Study

Gyeong-Min Lee [1], Jiwon Hong [2], Seok Jin Yun [1,3] and Jae-Hyun Kim [1,3,*]

[1] Research Institute for Healthcare Policy, Dankook University, Cheonan 31116, Republic of Korea; lgm910212@yonsei.ac.kr (G.-M.L.)
[2] Department of Civil and Environmental Engineering, Seoul National University, Seoul 08826, Republic of Korea; wonidone@snu.ac.kr
[3] Department of Health Administration, College of Public Health Sciences, Dankook University, Cheonan 31116, Republic of Korea
* Correspondence: jaehyun@dankook.ac.kr

Abstract

Background/Objectives: Population-based risk stratification for dementia is increasingly important in rapidly ageing societies. Developing prediction models for incident dementia using longitudinal survey data remains challenging because new events are uncommon, observations are repeatedly collected from the same individuals, and predictor variables span multiple health and social domains. This study aimed to develop and compare machine learning models for predicting the two-year risk of newly meeting a survey-derived dementia classification among community-dwelling older adults using longitudinal data from the Korean Longitudinal Study of Aging. **Methods**: We analysed 12,025 person-period observations from 6380 participants across the 2018–2020 and 2020–2022 intervals after explicitly restricting the risk set to records with known baseline cognition, no baseline dementia classification, and observed follow-up cognition. Sixty baseline predictors were evaluated with eight algorithms using 5-fold participant-level grouped cross-validation and temporally separated validation trained on 2018–2020 and tested on 2020–2022. Performance was assessed using the area under the receiver operating characteristic curve (ROC-AUC), the area under the precision-recall curve (AUPRC), Brier score, calibration, participant-cluster bootstrap confidence intervals (CIs), and top 5%, 10%, and 20% risk-group yield. **Results**: The survey-derived dementia classification was newly met in 126 of 12,025 person-period observations (1.05%). In grouped cross-validation, CatBoost achieved ROC-AUC 0.890 (95% CI, 0.869–0.910) and AUPRC 0.119 (95% CI, 0.081–0.179); XGBoost achieved ROC-AUC 0.877 and AUPRC 0.113. In temporally separated validation, CatBoost achieved the highest ROC-AUC (0.874); AUPRC values were uniformly low (0.048–0.087) with widely overlapping confidence intervals, so no algorithm was superior for precision-recall performance. The top 10% CatBoost risk group contained 60.3% of incident events in grouped cross-validation and 56.4% in temporal validation, with approximately 6-fold and 5.6-fold enrichment, respectively. Age remained the dominant predictor. Baseline mild cognitive impairment (MCI) was uncommon (39 observations) but strongly enriched for subsequent classification (12 events; 30.8%). **Conclusions**: Routinely collected longitudinal survey variables supported useful population-level risk ranking for subsequent transition to a survey-derived dementia classification, but absolute case-finding performance remained modest in this rare-event setting. The results support prioritisation for follow-up rather than diagnosis. Sensitivity analyses showed that performance was essentially unchanged after removing baseline survey year or highly missing predictors,

Check for updates

Academic Editor: Hualou Liang

Received: 30 July 2026
Revised: 12 September 2026
Accepted: 23 October 2026
Published: 4 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

whereas excluding baseline MCI reduced precision-recall performance. External validation, recalibration, and prospective evaluation remain necessary before implementation.

Keywords: dementia; machine learning; longitudinal cohort; oral health; risk stratification; calibration; survey-derived classification; prediction model

1. Introduction

Dementia is one of the leading causes of disability and dependency among older adults and represents an increasing public health challenge in rapidly ageing populations [1,2]. As life expectancy continues to increase, strategies that enable earlier identification of individuals at elevated risk have become increasingly important for supporting preventive interventions, allocating community health resources, and improving the efficiency of dementia care [3]. In population-based settings, however, predicting incident dementia remains difficult because disease onset is relatively uncommon, risk accumulates gradually over time, and determinants of cognitive decline extend across demographic, socioeconomic, behavioural, functional, and health-related domains [4,5].

Machine learning has attracted growing interest as an approach for risk prediction because it can model complex nonlinear relationships and interactions among large numbers of predictors without requiring strong parametric assumptions [6]. Recent studies have reported promising discrimination for dementia prediction using machine learning algorithms, particularly when combining multidimensional health information [7,8]. This approach has recently been extended to nationally representative panel surveys using only self-reported social, behavioural, and functional information, with reported receiver operating characteristic area under the curve (ROC-AUC) values of approximately 0.82 in the National Health and Aging Trends Study and 0.75 for 10-year incident dementia in the Sydney Memory and Ageing Study, where regularised regression performed at least as well as more complex learners when events were few [9,10]. Nevertheless, discrimination alone provides an incomplete assessment of clinical or public health utility [11]. For rare outcomes such as incident dementia, models with high ROC-AUC may still identify relatively few true cases at commonly used probability thresholds [12]. Evaluation of prediction models should therefore incorporate complementary measures, including precision-recall performance, calibration, threshold-dependent sensitivity and specificity, temporal validation, and the ability to enrich individuals at genuinely elevated risk [13]. Recent evidence reinforces this requirement: a systematic review and meta-analysis of externally validated dementia risk models, and an external validation study in the EPIC-Norfolk cohort, both show that discrimination commonly declines and calibration deteriorates when models are applied outside the setting in which they were developed [14,15].

Another challenge is the selection of predictors that can be collected efficiently in community settings. Although neuroimaging, genetic information, and circulating biomarkers have improved prediction in specialised clinical cohorts, these data are expensive and are rarely available in large-scale population surveys [16]. Prediction models based on routinely collected demographic, socioeconomic, health, and functional information may therefore provide greater applicability for population-level screening. Oral-health-related quality of life represents one potentially informative but underexplored domain. The Geriatric Oral Health Assessment Index (GOHAI) reflects functional limitation, psychosocial well-being, and discomfort related to oral health, and may capture aspects of nutritional status, frailty, self-care, and social disadvantage that are not fully represented by conventional health indicators [17]. Recent longitudinal syntheses support a link between dental status and

later cognitive outcomes: a meta-analysis of longitudinal data reported that tooth loss was associated with subsequent dementia and mild cognitive impairment (MCI), with stronger associations among edentulous individuals and possible attenuation among denture users [18,19]. Measures of healthy ageing, including social participation and functional capacity, may likewise provide complementary information for identifying individuals at increased risk [20,21].

Evidence from Korean panel data is accumulating but remains limited in scope. An interpretable machine learning analysis of the Korean Longitudinal Study of Aging (KLoSA) identified age and education as the dominant non-modifiable contributors, and oral health, depressive symptoms, household income, quality of life, and instrumental activities of daily living as leading modifiable contributors to cognitive function [22], and joint-trajectory modelling of 14 years of KLoSA data showed that frailty and cognitive decline frequently progress together in this population [23]. These studies modelled cognitive level or cognitive trajectory rather than the two-year transition to a survey-derived dementia classification, and they did not report participant-level grouped validation, rare-event-appropriate metrics, calibration, or the case yield obtainable at a realistic screening capacity.

Using longitudinal data from KLoSA, this study developed and compared multiple machine learning algorithms for predicting the two-year risk of newly meeting a survey-derived dementia classification among community-dwelling older adults. This study emphasised participant-level grouped cross-validation, temporally separated validation, rare-event-appropriate metrics, calibration, and SHapley Additive exPlanations (SHAP)-based explainability. The objective was to evaluate the performance, incremental value, and potential public-health applicability of a multidimensional survey-based dementia risk-stratification framework, while explicitly distinguishing risk stratification from clinical diagnosis.

2. Materials and Methods

2.1. Study Design and Data Source

This longitudinal prediction study used KLoSA, a nationally representative panel survey of community-dwelling Korean adults. KLoSA collects biennial information on demographic characteristics, socioeconomic status, health conditions, functional status, employment, family structure, social participation, and quality of life. The source longitudinal file included 63,215 person-period observations from 11,174 participants across baseline waves from 2006 through 2020 and corresponding two-year follow-up waves through 2022.

For the present analysis, repeated waves were transformed into consecutive two-year person-period records. Predictors were measured at the beginning of each interval and follow-up cognitive classification was assessed at the subsequent wave. Analyses were restricted to the 2018–2020 and 2020–2022 intervals because the harmonised cognition-related definitions and oral-health measures required for this study were consistently available in these waves. From 13,428 records in these two intervals, 58 with missing baseline cognitive classification and 168 with prevalent baseline dementia classification were excluded. Among the remaining dementia-free records, 484 records with a recorded death and no follow-up cognition and 693 other records without follow-up cognition were excluded. The final analytic cohort therefore comprised 12,025 person-period observations from 6380 unique participants (Figure 1).

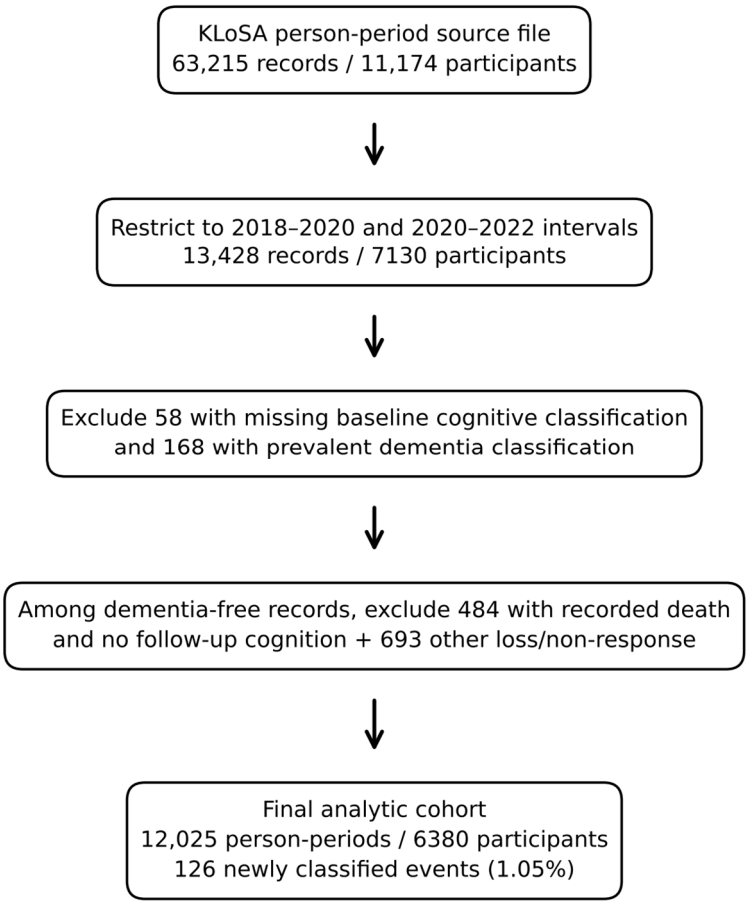


Figure 1. Study design and analytic cohort flow.

2.2. Study Population and Outcome Definition

The primary outcome was newly meeting the KLoSA survey-derived dementia classification at the subsequent two-year follow-up. The supplied KLoSA-derived cognition classification variable used codes corresponding to dementia classification (1), mild cognitive impairment (3), and neither category (0 or 5 in the harmonised analysis file). Because this endpoint is a survey-derived cognitive classification rather than an adjudicated clinical dementia diagnosis, it is referred to throughout as a 'survey-derived dementia classification', and participants are described as 'newly meeting the dementia classification' rather than as receiving a diagnosis. The risk set explicitly required a known baseline cognitive classification and absence of baseline dementia. Baseline MCI was allowed in the primary risk set because it is a clinically relevant high-risk state, but its influence was examined in a prespecified sensitivity analysis excluding baseline MCI.

2.3. Candidate Predictors

Sixty baseline predictors were selected a priori from routinely available KLoSA variables; analytic variable names and reader-friendly labels are listed in Supplementary Table S1. These covered age, sex, education, marital status, region and urbanicity, household composition, self-rated and subjective health, disability, chronic diseases and multimorbidity, falls, exercise, smoking, drinking, depressive symptoms, activities of daily living (ADL), instrumental activities of daily living (IADL), employment and retirement, household income and housing assets, quality of life, social activity, oral-health-related quality of life measured using GOHAI, denture use, and proxy indicators of active and successful ageing. Chronic disease variables included nine baseline conditions: hypertension, diabetes mellitus, cancer, chronic lung disease, liver disease, heart disease, cere-

brovascular disease, psychiatric disease, and arthritis/rheumatic disease. We additionally constructed a chronic disease count based on these nine conditions and indicators of multimorbidity (≥2 conditions) and severe multimorbidity (≥3 conditions). Additional composite indicators reflecting cardiometabolic, cardio-cerebrovascular, and neuropsychiatric disease burden were also included as predictors. The most influential constructs were defined as follows: GOHAI total score represented oral-health-related quality of life; self-rated/subjective health variables represented global perceived health; active- and successful-ageing proxy variables summarised functional and social dimensions generated in the harmonised analytic dataset. Baseline MCI and cognitive-impairment indicators were included as predictors, and their influence was evaluated by sensitivity analysis.

2.4. Data Preparation and Model Development

The analytic cohort and derived variables were prepared from the KLoSA source data using SAS 9.4 (SAS Institute Inc., Cary, NC, USA). The resulting analysis-ready dataset was analysed in a Python 3.12 workflow (Python Software Foundation, Wilmington, DE, USA). Data handling used pandas 2.2 and NumPy 2.1; model development used scikit-learn 1.6 (logistic regression, random forest, ExtraTrees, gradient boosting and AdaBoost), XGBoost 2.1, LightGBM 4.5, and CatBoost 1.2; model interpretation used the SHAP package 0.46; and figures were produced with Matplotlib 3.9. All random seeds were fixed at 42. Continuous and numerically coded predictors were imputed using the median estimated within each training split and standardised where required; categorical predictors were imputed using the most frequent category and encoded within the training data. All preprocessing steps were fitted separately within each training fold to avoid information leakage into validation data. Class weighting was applied as prespecified to logistic regression, random forest, ExtraTrees, and LightGBM, whereas no post hoc probability recalibration was applied to any model; the resulting differences in probability scale are taken into account when interpreting Brier scores across algorithms. Predictor-level missingness is reported in Supplementary Table S3. Three predictors had more than 40% missingness (retirement reason, retirement status, and hypertension in the harmonised file); a sensitivity analysis excluding predictors with >40% missingness produced similar CatBoost performance (ROC-AUC 0.888; AUPRC 0.114; Supplementary Table S5), reducing concern that these variables drove the main results.

Eight algorithms were evaluated: L2-regularised logistic regression, random forest, ExtraTrees, gradient boosting, AdaBoost, XGBoost, LightGBM, and CatBoost [24–30]. Hyperparameters were prespecified and applied consistently. Because multiple person-period observations could arise from the same participant, the primary internal validation used 5-fold participant-level grouped cross-validation, assigning all records from each participant to a single fold. All eight algorithms were fitted on the same analytic risk set, and small numerical differences among them are interpreted cautiously rather than as evidence of superiority.

2.5. Model Evaluation and Temporal Validation

Discrimination was evaluated using ROC-AUC and the area under the precision-recall curve (AUPRC), with AUPRC emphasised because it reflects positive-class performance under severe class imbalance [31]. Brier score was used for overall probabilistic accuracy. At the conventional 0.5 threshold, sensitivity, specificity, and F1 score were retained for comparability but were not considered the primary operational evaluation. To align evaluation with population risk stratification, we additionally examined the top 5%, 10%, and 20% of predicted risk and reported observed event rate, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and enrichment over the population

event rate. Calibration was summarised by decile plots, calibration intercept and slope, observed-to-expected (O:E) ratio, and event counts within risk bins [32]. Participant-cluster bootstrap resampling was used to obtain 95% confidence intervals (CIs) for ROC-AUC, AUPRC, and Brier score for the leading models.

Temporally separated validation trained each model on the 2018 baseline to 2020 follow-up interval and evaluated it in the 2020 baseline to 2022 follow-up interval. Because KLoSA is a continuing panel, this analysis was temporally separated but not participant-independent: 5645 of 5785 participants (97.6%) in the later test interval had also appeared in the earlier training interval. Only 140 test-period participants were participant-disjoint and only 3 events occurred in that subset; therefore, a stable participant-disjoint performance estimate was not considered defensible. This analysis is therefore described as temporally separated validation rather than as independent external validation.

2.6. Model Interpretation

SHAP values from the CatBoost model were used to characterise model behaviour [33]. Global importance was summarised by mean absolute SHAP value. SHAP values indicate contribution to model prediction and do not establish biological or causal importance. Because several predictors represented correlated domains, feature importance may be distributed or substituted across related variables. In particular, the GOHAI domain was additionally assessed by model ablation to determine whether its apparent SHAP prominence translated into unique incremental predictive value. Prespecified sensitivity analyses removed baseline survey year, removed the oral-health domain, excluded observations with baseline MCI, and excluded predictors with more than 40% missingness; simpler benchmark models (age alone; age and sex; age, sex, and education; and a parsimonious demographic/clinical model) were fitted under the same grouped cross-validation scheme.

2.7. Ethics and Reporting

This study used publicly available, de-identified secondary data and was exempt from institutional review board review in accordance with institutional policy. The original KLoSA survey obtained informed consent from all participants. This study was reported in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis + Artificial Intelligence (TRIPOD+AI) statement [13], with supporting reference to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement for the observational design elements. Completed checklists are provided as Supplementary Materials. Reporting emphasised transparent predictor definition, participant-level validation, calibration, temporal validation, and model explainability.

3. Results

3.1. Analytic Cohort and Baseline Characteristics

The analytic cohort comprised 12,025 person-period observations from 6380 participants. The survey-derived dementia classification was newly met in 126 observations (1.05%). The 2018–2020 interval contributed 6240 records and 71 events (1.14%), whereas the 2020–2022 interval contributed 5785 records and 55 events (0.95%) (Table 1). Mean age was 69.5 years (standard deviation [SD], 9.4), with a median of 69 years (interquartile range, 61–77) and a range of 55–101 years; the full age-group distribution is provided in Supplementary Table S9. Participants who newly met the dementia classification were substantially older (mean, 82.0 years; SD, 6.9), a difference of approximately 12.7 years. Women represented 58.1% of records and had a higher crude event rate than men (1.22% vs. 0.81%; crude odds ratio, 1.50); the sex association attenuated after age adjustment (odds ratio,

1.25; 95% CI, 0.85–1.83), and sex did not emerge among the leading SHAP predictors. Baseline MCI was present in 39 records, of which 12 (30.8%) subsequently met the dementia classification, compared with 114 of 11,986 (0.95%) among baseline non-MCI records. Baseline characteristics overall and by subsequent classification are shown in Table 2, and a comparison of included versus excluded target-interval records is provided in Supplementary Table S6.

Table 1. Analytic cohorts and event rates.

Prediction Interval	Analytic N	Event N	Event Rate (%)
2018–2020	6240	71	1.14
2020–2022	5785	55	0.95
Overall	12,025	126	1.05

Note. Each row represents a two-year prediction interval constructed from consecutive KLoSA waves. Analytic N denotes the number of eligible person-period observations at the interval baseline; Event N denotes the number of observations newly meeting the survey-derived dementia classification at the subsequent follow-up wave. KLoSA, Korean Longitudinal Study of Aging.

Table 2. Baseline characteristics overall and by subsequent survey-derived dementia classification.

Characteristic	Overall	No Event	Incident Classification
Age, years	69.48 (9.36)	69.34 (9.29)	82.03 (6.90)
Household income	3253.33 (3478.72)	3267.57 (3488.84)	1908.95 (1909.00)
GOHAI total score	50.90 (8.47)	50.97 (8.42)	44.98 (11.45)
Self-rated health (1 = very good to 5 = very poor)	3.68 (0.85)	3.67 (0.84)	4.47 (0.74)
Depressive symptom score	16.64 (4.39)	16.61 (4.35)	19.29 (6.65)
ADL limitation score	7.25 (2.05)	7.22 (1.90)	10.17 (7.00)
IADL limitation score	11.05 (4.62)	10.97 (4.38)	18.65 (12.97)
Number of chronic diseases	0.09 (0.31)	0.09 (0.31)	0.06 (0.24)
Number of social activities	0.98 (0.75)	0.98 (0.75)	0.40 (0.58)
Female	6986 (58.1%)	6901 (58.0%)	85 (67.5%)
Baseline MCI	39 (0.3%)	27 (0.2%)	12 (9.5%)
Urban residence	9172 (76.3%)	9088 (76.4%)	84 (66.7%)
Education: elementary school or less	4335 (36.0%)	4238 (35.6%)	97 (77.0%)
Education: middle school	2044 (17.0%)	2031 (17.1%)	13 (10.3%)
Education: high school	4074 (33.9%)	4058 (34.1%)	16 (12.7%)
Education: college or higher	1572 (13.1%)	1572 (13.2%)	0 (0.0%)
Region: metropolitan city	5057 (42.1%)	5020 (42.2%)	37 (29.4%)
Region: small or medium-sized city	4115 (34.2%)	4068 (34.2%)	47 (37.3%)
Region: rural area (eup/myeon)	2853 (23.7%)	2811 (23.6%)	42 (33.3%)

Note. Values are mean (SD) for continuous or numerically coded variables and n (%) for categorical variables. Education and region categories follow the KLoSA codebook; urban residence comprises metropolitan and small or medium-sized cities. MCI, mild cognitive impairment; GOHAI, Geriatric Oral Health Assessment Index; ADL, activities of daily living; IADL, instrumental activities of daily living.

3.2. Participant-Level Grouped Cross-Validation

Across the eight grouped cross-validation analyses, CatBoost had the numerically highest ROC-AUC (0.890), followed by XGBoost (0.877); their participant-cluster bootstrap 95% confidence intervals overlapped substantially. CatBoost ROC-AUC was 0.890 (95% CI, 0.869–0.910), AUPRC 0.119 (95% CI, 0.081–0.179), and Brier score 0.0098; XGBoost ROC-AUC was 0.877 (95% CI, 0.850–0.903), AUPRC 0.113 (95% CI, 0.074–0.180), and Brier score 0.0099 (Figure 2; Supplementary Table S8). These numerical rankings should not be interpreted as statistically meaningful superiority. At the conventional 0.5 threshold, tree-based models remained highly specific but insensitive (Table 3), reinforcing the need for capacity-based rather than fixed probability thresholds.

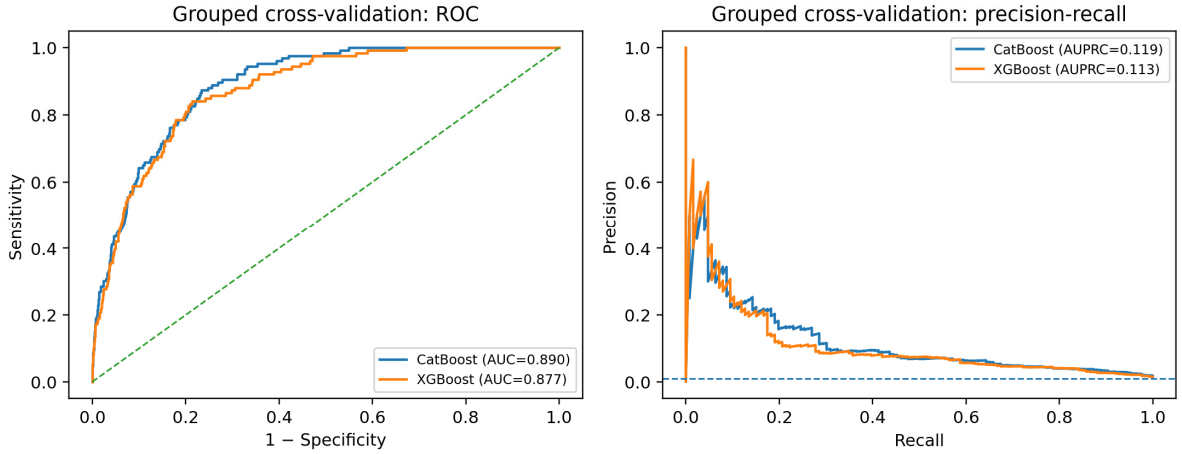


Figure 2. Participant-level grouped cross-validated ROC and precision-recall curves for the two leading models.

Table 3. Participant-level grouped cross-validated performance at the 0.5 threshold.

Model	ROC-AUC	AUPRC	Brier	Sensitivity	Specificity	F1
CatBoost	0.890	0.119	0.010	0.032	1.000	0.060
XGBoost	0.877	0.113	0.010	0.032	1.000	0.060
Gradient boosting	0.871	0.085	0.011	0.056	0.998	0.089
Random forest	0.870	0.073	0.014	0.032	0.998	0.051
AdaBoost	0.868	0.078	0.080	0.008	1.000	0.016
Extra trees	0.864	0.072	0.025	0.143	0.988	0.124
LightGBM	0.859	0.082	0.013	0.087	0.995	0.111
Logistic regression	0.856	0.099	0.126	0.754	0.815	0.079

Note. All records from the same participant were assigned to the same fold. Threshold-dependent metrics were computed at a probability cutoff of 0.5. The event-rate baseline for AUPRC was 0.010. ROC-AUC, area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve; F1, harmonic mean of precision and sensitivity.

3.3. Temporal Validation

In the temporally separated validation set (N = 5785; 55 events), CatBoost achieved the highest ROC-AUC (0.874; 95% CI, 0.836–0.911), with AUPRC 0.065 (95% CI, 0.041–0.108) and Brier score 0.0094; XGBoost achieved ROC-AUC 0.865 (95% CI, 0.823–0.901), AUPRC 0.067 (95% CI, 0.044–0.115), and Brier score 0.0093 (Table 4; Supplementary Table S8). AUPRC values were uniformly low and closely spaced across algorithms (0.048–0.087), and the numerically highest value was obtained by logistic regression (0.087) rather than by a boosting model; because the bootstrap intervals overlap widely, no algorithm can be regarded as superior for precision-recall performance in this interval. The low absolute AUPRCs emphasise that good rank discrimination did not translate into strong individual case identification. Because 97.6% of later-period participants had also contributed to the earlier interval, these results are interpreted as temporal transportability within the same panel rather than as participant-independent validation.

Table 4. Temporal validation performance: models trained on 2018–2020 and tested on 2020–2022.

Model	ROC-AUC	AUPRC	Brier	Sensitivity	Specificity	F1
CatBoost	0.874	0.065	0.009	0.000	1.000	0.000
AdaBoost	0.872	0.065	0.092	0.000	1.000	0.000
XGBoost	0.865	0.067	0.009	0.000	1.000	0.000

Table 4. *Cont.*

Model	ROC-AUC	AUPRC	Brier	Sensitivity	Specificity	F1
Gradient boosting	0.865	0.056	0.011	0.018	0.998	0.029
Random forest	0.864	0.062	0.015	0.055	0.995	0.069
Extra trees	0.831	0.048	0.027	0.127	0.986	0.097
Logistic regression	0.816	0.087	0.122	0.636	0.824	0.064
LightGBM	0.812	0.067	0.010	0.036	0.998	0.059

Note. Models were trained on the 2018–2020 interval and evaluated in the 2020–2022 interval. Threshold-dependent metrics were computed at a probability cutoff of 0.5. The event-rate baseline for AUPRC in the evaluation interval was 0.010. ROC-AUC, area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve; F1, harmonic mean of precision and sensitivity.

3.4. Calibration

Calibration of CatBoost was adequate for risk ranking but imperfect for absolute probabilities (Figure 3). In grouped cross-validation, the calibration intercept was −0.054, the slope 0.943, and the observed-to-expected (O:E) ratio 1.14 (126 observed vs. 110.2 expected events), indicating slight underprediction overall; the same direction was present in the highest decile (mean predicted risk 0.058 vs. observed event rate 0.063). In temporally separated validation, the intercept was −0.350, the slope 0.896, and the O:E ratio 1.03 (55 observed vs. 53.2 expected events). Calibration summaries are given in Supplementary Table S4 and event counts by decile of predicted risk in Supplementary Table S7; decile-level estimates should be interpreted cautiously because only 126 events overall and 55 events in the temporal interval were available. These findings support recalibration before absolute predicted risks are used operationally.

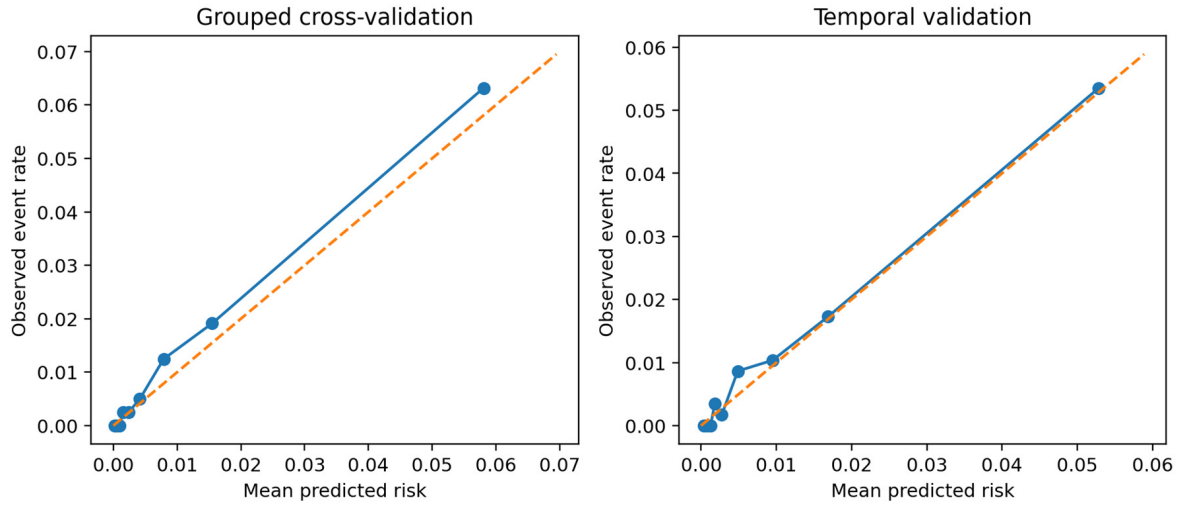


Figure 3. CatBoost calibration in participant-level grouped cross-validation and temporally separated validation.

3.5. Capacity-Based Risk Stratification

Because a probability cutoff of 0.5 is uninformative at this event rate, the models were additionally evaluated by the yield obtained within the highest-risk 5%, 10%, and 20% of predicted risk. In grouped cross-validation, the top 5% of CatBoost-predicted risk contained 53 of 126 events (sensitivity, 42.1%) with an observed event rate of 8.8%, corresponding to approximately 8.4-fold enrichment over the population event rate; the top 10% contained 76 events (60.3%) with an observed event rate of 6.3% (approximately 6-fold enrichment); and the top 20% contained 99 events (78.6%) with an observed event rate of

4.1% (3.9-fold enrichment). In temporally separated validation, the top 5% contained 23 of 55 events (41.8%) with an observed event rate of 7.9% (8.3-fold enrichment), and the top 10% contained 31 events (56.4%) with an observed event rate of 5.4% (5.6-fold enrichment). XGBoost produced closely similar yields. Full positive predictive value, sensitivity, and specificity results are given in Table 5.

Table 5. Capacity-based risk-stratification performance for CatBoost and XGBoost.

Validation	Model	Risk Group	N Selected	Events	PPV	Sensitivity	Specificity	Enrichment
Grouped CV	CatBoost	Top 5%	602	53	0.088	0.421	0.954	8.4×
Grouped CV	CatBoost	Top 10%	1203	76	0.063	0.603	0.905	6.0×
Grouped CV	CatBoost	Top 20%	2405	99	0.041	0.786	0.806	3.9×
Grouped CV	XGBoost	Top 5%	602	49	0.081	0.389	0.954	7.8×
Grouped CV	XGBoost	Top 10%	1203	74	0.062	0.587	0.905	5.9×
Grouped CV	XGBoost	Top 20%	2405	99	0.041	0.786	0.806	3.9×
Temporal validation	CatBoost	Top 5%	290	23	0.079	0.418	0.953	8.3×
Temporal validation	CatBoost	Top 10%	579	31	0.054	0.564	0.904	5.6×
Temporal validation	CatBoost	Top 20%	1157	41	0.035	0.745	0.805	3.7×
Temporal validation	XGBoost	Top 5%	290	22	0.076	0.400	0.953	8.0×
Temporal validation	XGBoost	Top 10%	579	31	0.054	0.564	0.904	5.6×
Temporal validation	XGBoost	Top 20%	1157	40	0.035	0.727	0.805	3.6×

Note. Risk groups were defined by ranking predicted probabilities within each validation setting. Enrichment is the observed event rate within the risk group divided by the overall event rate (1.05% in grouped cross-validation and 0.95% in the temporal validation interval). CV, cross-validation; PPV, positive predictive value.

3.6. SHAP-Based Model Interpretation

Age was the dominant contributor to CatBoost predictions, followed by household income, educational attainment, GOHAI total score, self-rated health, subjective health status, and the number of social activities (Table 6; Figure 4). This pattern indicates that the model drew on information spanning biological ageing, socioeconomic position, perceived health, oral-health-related quality of life, and social participation. Directional interpretation of coded categorical predictors requires reference to the original KLoSA codebook and was therefore not reduced to linear risk statements. Model ablation showed that removing the GOHAI and denture-use variables did not reduce CatBoost discrimination (ROC-AUC 0.891 vs. 0.890 in the full model; AUPRC 0.124 vs. 0.119; Supplementary Table S5). A high SHAP ranking for GOHAI should therefore be interpreted as predictive information that is partly redundant with correlated vulnerability domains rather than as evidence that oral health is an independent causal driver of the outcome.

Table 6. Top 15 predictors in the CatBoost model by mean absolute SHAP value.

Rank	Predictor	Mean Absolute SHAP Value
1	Age	0.617
2	Household income	0.160
3	Educational attainment	0.151
4	GOHAI total score	0.125
5	Self-rated health	0.097
6	Subjective health status	0.096
7	Number of social activities	0.090
8	Successful ageing proxy	0.089
9	Residential region	0.081
10	Active ageing proxy	0.053
11	Number of living children	0.051
12	Depressive symptom score	0.050

Table 6. *Cont.*

Rank	Predictor	Mean Absolute SHAP Value
13	Urbanicity	0.047
14	Current drinking	0.041
15	Quality of life score	0.040

Note. Values are mean absolute SHAP values obtained from the CatBoost model. Larger values indicate a greater average contribution to model output and do not imply causal effects. Directional interpretation of coded categorical predictors requires reference to the original KLoSA codebook. SHAP, SHapley Additive exPlanations; GOHAI, Geriatric Oral Health Assessment Index; KLoSA, Korean Longitudinal Study of Aging.

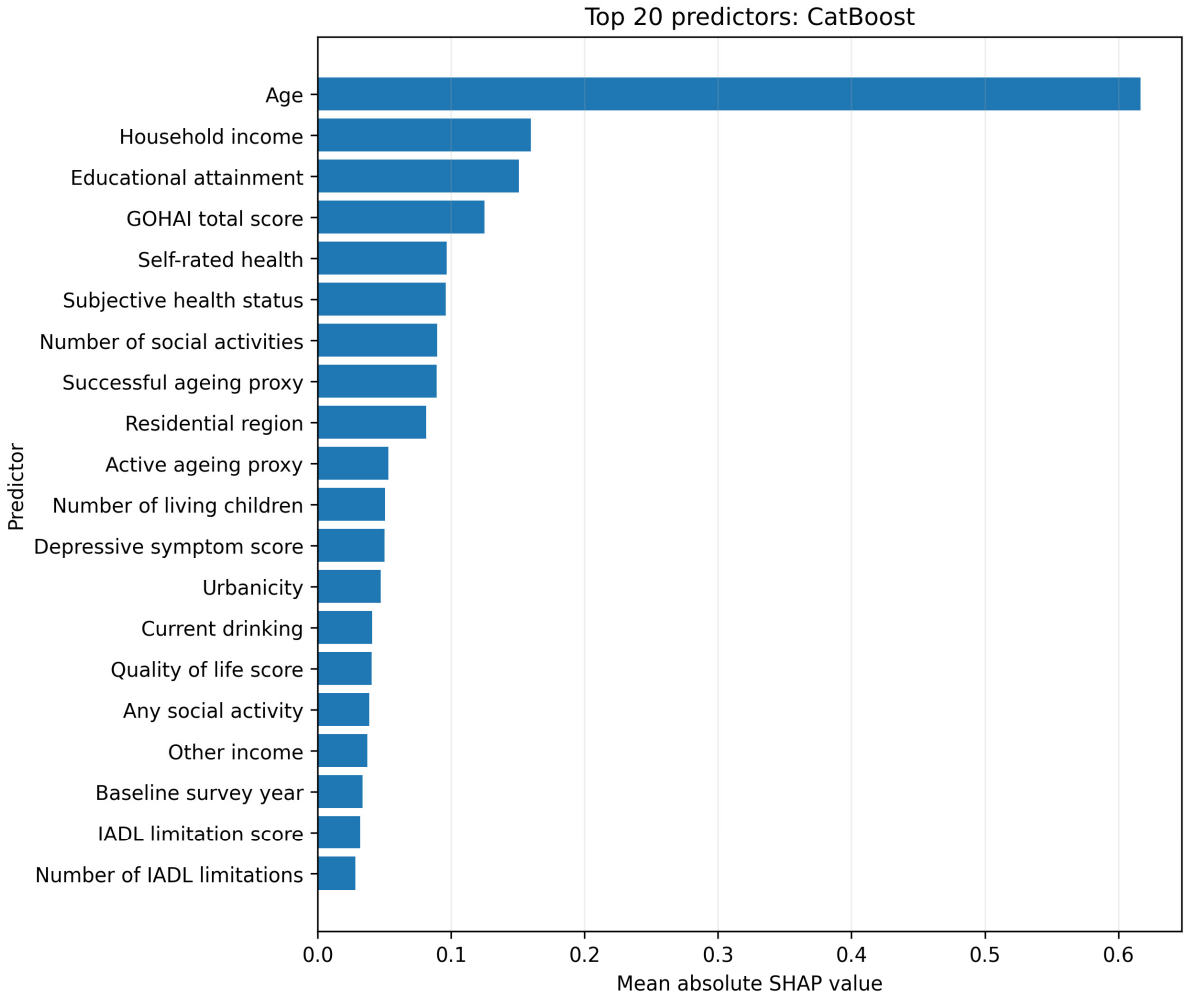


Figure 4. Top 20 predictors by mean absolute SHAP value.

4. Discussion

This study developed and evaluated machine learning models for two-year transition to a survey-derived dementia classification using nationally representative longitudinal KLoSA person-period data. The analytic risk set, restricted to records with known baseline cognition, no baseline dementia classification, and observed follow-up cognition, comprised 12,025 person-period observations and 126 events. The leading models retained good discrimination, but their overlapping uncertainty intervals and modest AUPRCs argue against emphasising small algorithmic rank differences [13,31,32]. The principal finding is therefore not that one boosting algorithm is uniquely superior, but that routinely collected

multidomain survey information can meaningfully enrich near-term dementia-classification risk in a very low-prevalence setting [8,34].

The practical value of the models is clearer when evaluation is organised around screening capacity. In grouped cross-validation, the highest-risk 10% of participants captured 60.3% of events at an observed event rate of 6.3%, and in temporally separated validation, the corresponding values were 56.4% and 5.4% (Table 5). These yields are more informative for public-health prioritisation than performance at an arbitrary 0.5 probability threshold [12].

The temporally separated analysis should be interpreted cautiously. Most participants in the later interval also contributed to the earlier interval, so the analysis evaluates stability across calendar periods within the same panel rather than participant-independent transportability. A participant-disjoint temporal subset could be identified, but it included only 140 participants and 3 events, which was too sparse for a meaningful performance estimate. Future external validation should therefore use a genuinely independent cohort [13,35]. Mortality is another important source of informative attrition: among dementia-free records in the two target intervals, 484 had a recorded death and no follow-up cognitive classification, while 693 had other loss or non-response. The present analysis is conditional on observed follow-up cognition and does not model death as a formal competing risk; this may favour healthier survivors and should be considered when interpreting population applicability.

Age was by far the strongest predictive signal. Participants who developed the survey-derived dementia classification were, on average, approximately 12.7 years older than non-events. Although women had a higher crude event rate, the association with sex attenuated after age adjustment, suggesting that age composition explained much of the crude difference. Baseline MCI, in contrast, defined a small but very high-risk subgroup: 30.8% of baseline-MCI observations transitioned to the dementia classification. Excluding baseline MCI reduced CatBoost ROC-AUC from 0.890 to 0.866 and AUPRC from 0.119 to 0.059. Thus, the primary model partly serves the clinically relevant task of identifying progression among participants already showing cognitive vulnerability; the sensitivity analysis clarifies performance among those without baseline MCI.

Comparison with simpler models further contextualised incremental value. An age-only logistic model achieved ROC-AUC 0.853 and AUPRC 0.057; adding sex produced essentially no improvement, and age plus sex and education achieved ROC-AUC 0.859. A parsimonious demographic/clinical model that included age, sex, education, income, baseline MCI, self-rated health, depressive symptoms, ADL/IADL, and chronic disease burden achieved ROC-AUC 0.885 and AUPRC 0.101. The full CatBoost model therefore provided only modest incremental discrimination over a strong simple model, although it improved precision-recall performance. This finding reinforces the need to balance predictive gain against complexity when considering implementation [11].

The GOHAI finding warrants similarly cautious interpretation. GOHAI ranked highly in SHAP importance, but removing GOHAI and denture-use variables did not impair predictive performance. GOHAI may therefore act as a marker of correlated constructs such as frailty, nutrition, socioeconomic disadvantage, access to care, social isolation, or early functional decline rather than provide unique incremental information [17,36]. Population-level evidence continues to link dental status to subsequent cognitive outcomes [18,19], but the present ablation indicates that this information is largely redundant once a broad set of socioeconomic, functional, and health variables is available. Likewise, removing baseline survey year left CatBoost performance essentially unchanged (ROC-AUC 0.890; AUPRC 0.127; Supplementary Table S5), arguing against material reliance on period-specific

artifacts. These sensitivity analyses shift the interpretation from isolated feature importance toward the broader multidomain vulnerability profile captured by the model [37].

These results can be placed alongside recent work. Machine learning models developed in other national panel surveys have reported ROC-AUC values of approximately 0.75 to 0.82 for incident dementia using self-reported information [9,10]; the higher discrimination observed here probably reflects the short two-year horizon and the inclusion of baseline cognitive-status indicators, as demonstrated by the MCI-exclusion sensitivity analysis. Recent syntheses of externally validated dementia risk models, and external validation in the EPIC-Norfolk cohort, show that discrimination frequently declines and calibration deteriorates outside the development setting [14,15], which reinforces our decision to describe the present temporal analysis as transportability within a single panel rather than as external validation. Within Korean panel data, our leading predictors agree with the interpretable machine learning analysis of KLoSA [22], while the present study adds a prospective two-year transition endpoint, participant-level grouped validation, rare-event metrics, calibration, and capacity-based yield; the joint progression of frailty and cognitive decline reported in the same panel [23] is consistent with our finding that a multidomain vulnerability profile, rather than any single domain, carries the predictive signal. Finally, the modest absolute yield observed at realistic screening capacity parallels the finding that a high-risk screening strategy would fail to identify a substantial proportion of future dementia cases [38], and supports framing the model as a prioritisation tool rather than a case-finding instrument.

Limitations

First, the event count remained small (126 overall; 55 in the later interval), limiting precision for model comparison and calibration; accordingly, cluster-bootstrap confidence intervals and event counts by calibration bin are reported. Second, the endpoint is a KLoSA survey-derived cognitive classification rather than clinically adjudicated dementia and should not be interpreted as a clinical diagnosis. Third, baseline MCI was included in the primary risk set; although this is relevant to real-world risk stratification, it means the primary model partly captures MCI-to-dementia-classification transition, as demonstrated by the MCI-exclusion sensitivity analysis. Fourth, temporal validation was not participant-independent because of substantial panel overlap. Fifth, death and non-response created informative attrition, and the analysis did not use formal competing-risk methods; because the survey-derived cognition classification is observed only among participants who survived and responded at follow-up, competing-risk and inverse-probability-of-censoring estimators were not identifiable within this data structure, and linkage to mortality records is a priority for future work. Sixth, predictor missingness was substantial for several variables; although imputation was restricted to training folds and excluding predictors with >40% missingness produced similar performance, missing-not-at-random bias remains possible. Seventh, several predictors were correlated or represented alternative codings of related constructs, so SHAP importance cannot establish independent or causal effects. Finally, external validation in an independent cohort and prospective impact evaluation are required before implementation.

5. Conclusions

Machine learning models using routinely collected longitudinal survey information provided useful risk enrichment for two-year transition to a survey-derived dementia classification among older Korean adults. CatBoost showed good discrimination, but the central contribution was actionable ranking rather than algorithmic superiority: the top 10% risk group captured approximately 60% of events in grouped cross-validation and

56% in temporally separated validation. Performance remained stable after removing survey year and highly missing predictors, while baseline MCI accounted for an important component of predictive performance. The model should be viewed as a population-level prioritisation tool, not a diagnostic classifier, and requires independent external validation, recalibration, and prospective evaluation before implementation.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jdad3040053/s1, Table S1: Predictor labels and analytic variable names; Table S2: Prespecified model settings; Table S3: Predictor-level missingness in the analytic cohort; Table S4: Calibration summary; Table S5: Sensitivity analyses and simpler benchmark models; Table S6: Baseline comparison of included versus excluded target-interval records; Table S7: CatBoost calibration deciles with event counts; Table S8: Participant-cluster bootstrap confidence intervals for the leading models; Table S9: Age distribution by subsequent survey-derived dementia classification; Figure S1: Temporal-validation receiver operating characteristic and precision-recall curves for all models; TRIPOD+AI and STROBE checklists.

Author Contributions: Conceptualization, J.-H.K.; methodology, G.-M.L., J.H. and S.J.Y.; software, S.J.Y.; formal analysis, G.-M.L., J.H. and S.J.Y.; investigation, G.-M.L., J.H. and S.J.Y.; data curation, S.J.Y.; writing—original draft preparation, G.-M.L. and J.H.; writing—review and editing, S.J.Y.; visualization, S.J.Y.; supervision, J.-H.K.; project administration, J.-H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki. Ethical review and approval were waived for this study because it consisted of a secondary analysis of publicly available, fully de-identified data from the Korean Longitudinal Study of Aging (KLoSA), in which individual participants are not identifiable to the investigators. Research using such data is exempt from institutional review board deliberation under Article 15 of the Enforcement Rule of the Bioethics and Safety Act of the Republic of Korea.

Informed Consent Statement: Informed consent was obtained from all subjects at the time of enrolment in the original Korean Longitudinal Study of Aging survey by the data-collecting agency. Because the present study analysed de-identified secondary data, additional informed consent was not required.

Data Availability Statement: The data presented in this study are available in the Korean Longitudinal Study of Aging (KLoSA) data repository at https://survey.keis.or.kr (accessed on 22 October 2026). These data were derived from the following resources available in the public domain: Korean Longitudinal Study of Aging, Korea Employment Information Service, https://survey.keis.or.kr. The analytic code generated during this study is available from the corresponding author upon reasonable request, subject to the terms of the data provider.

Acknowledgments: The authors thank the Korea Employment Information Service for granting access to the Korean Longitudinal Study of Aging data.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. World Health Organization. *Dementia Fact Sheet*; World Health Organization: Geneva, Switzerland, 2025.
2. He, Z.; Zhang, H.; Hu, G.; Qiao, Y.; Yin, C.; Li, J.; Lin, H.; Wu, A.; Qin, D.; Law, B.Y.-K. The current status, trends, and challenges of Alzheimer's disease and other dementias in Asia (1990–2036). *Front. Public Health* **2025**, *13*, 1583339. [CrossRef] [PubMed]
3. Livingston, G.; Huntley, J.; Liu, K.Y.; Costafreda, S.G.; Selbæk, G.; Alladi, S.; Ames, D.; Banerjee, S.; Burns, A.; Brayne, C.; et al. Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. *Lancet* **2024**, *404*, 572–628. [CrossRef] [PubMed]

4. Stephan, B.C.; Brain, J.; Buchanan, T.; Burley, C.V.; Burton, E.; Dunne, J.; Myers, B.; Sabatini, S.; Stephan, W.; Tang, E.Y.H. External Validation of Dementia Risk Prediction Models: Towards evidence-based implementation. *Alzheimers Dement.* **2025**, *21*, e101969. [CrossRef]
5. Hou, X.-H.; Feng, L.; Zhang, C.; Cao, X.-P.; Tan, L.; Yu, J.-T. Models for predicting risk of dementia: A systematic review. *J. Neurol. Neurosurg. Psychiatry* **2019**, *90*, 373–379. [CrossRef] [PubMed]
6. Rajkomar, A.; Dean, J.; Kohane, I. Machine learning in medicine. *N. Engl. J. Med.* **2019**, *380*, 1347–1358. [CrossRef] [PubMed]
7. Wang, Y.; Liu, S.; Spiteri, A.G.; Huynh, A.L.H.; Chu, C.; Masters, C.L.; Goudey, B.; Pan, Y.; Jin, L. Understanding machine learning applications in dementia research and clinical practice: A review for biomedical scientists and clinicians. *Alzheimers Res. Ther.* **2024**, *16*, 175. [CrossRef] [PubMed]
8. Javeed, A.; Dallora, A.L.; Berglund, J.S.; Ali, A.; Ali, L.; Anderberg, P. Machine learning for dementia prediction: A systematic review and future research directions. *J. Med. Syst.* **2023**, *47*, 17, Erratum in *J. Med. Syst.* **2023**, *48*, 87. [CrossRef] [PubMed]
9. Alam, A.; Rabbani, M.G.; Prybutok, V.R. Artificial Intelligence-Enabled Dementia Risk Prediction for Smart and Sustainable Healthcare: An Interpretable Machine Learning Study Using NHATS. *Appl. Sci.* **2026**, *16*, 2180. [CrossRef]
10. Chalmers, R.A.; Cervin, M.; Choo, C.; Numbers, K.; Mather, K.A.; Brodaty, H.; Kochan, N.A.; Sachdev, P.S.; Medvedev, O.N. Machine learning models for dementia risk prediction: Evidence from the Sydney Memory and Ageing Study. *npj Dement.* **2026**, *2*, 27. [CrossRef]
11. Vickers, A.J.; Elkin, E.B. Decision curve analysis: A novel method for evaluating prediction models. *Med. Decis. Mak.* **2006**, *26*, 565–574. [CrossRef] [PubMed]
12. Minus, E.; Coley, R.Y.; Shortreed, S.M.; Williamson, B.D. Behavior of prediction performance metrics with rare events. *J. Clin. Epidemiol.* **2025**, *189*, 112046. [CrossRef] [PubMed]
13. Collins, G.S.; Moons, K.G.M.; Dhiman, P.; Riley, R.D.; Beam, A.L.; Van Calster, B.; Ghassemi, M.; Liu, X.; Reitsma, J.B.; van Smeden, M.; et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **2024**, *385*, e078378, Erratum in *BMJ* **2024**, *385*, q902. [CrossRef] [PubMed]
14. Stephan, B.C.M.; Brain, J.; Anstey, K.J.; Buchanan, T.; Burley, C.V.; Burton, E.; Dunne, J.; Errington, L.; Gorringe, M.; Guan, Z.; et al. Discriminative performance of externally validated dementia risk prediction models: A systematic review and meta-analysis. *BMC Med.* **2026**, *24*, 132. [CrossRef] [PubMed]
15. Brain, J.; Tang, E.Y.H.; Burley, C.V.; Dunne, J.; Figueredo, G.; Greene, L.; Siervo, M.; Tully, P.J.; Turnbull, D.; Guan, Z.; et al. External validation of dementia risk prediction models in the EPIC-Norfolk cohort: A UK-based cohort study. *npj Dement.* **2026**, *2*, 44. [CrossRef] [PubMed]
16. Camacho, M.; Atehortúa, A.; Wilkinson, T.; Gkontra, P.; Lekadir, K. Low-cost predictive models of dementia risk using machine learning and exposome predictors. *Health Technol.* **2025**, *15*, 355–365. [CrossRef]
17. Motoishi, Y.; Yamanashi, H.; Kitamura, M.; Hayashida, H.; Kawasaki, K.; Nobusue, K.; Nonaka, F.; Honda, Y.; Shimizu, Y.; Kawashiri, S.-Y.; et al. Oral health-related quality of life is associated with physical frailty: A cross-sectional study of Japanese community-dwelling older adults. *J. Gen. Fam. Med.* **2021**, *22*, 271–277. [CrossRef] [PubMed]
18. Chen, H.; Ji, L.; Wang, Y.; Klineberg, I.; Chen, H. Dentition-Cognition Relationship in Aging Populations: A Meta-Analysis of Longitudinal Data. *J. Oral Rehabil.* **2026**, *53*, 924–938. [CrossRef] [PubMed]
19. Yeung, C.A. Is tooth loss associated with subsequent cognitive decline and dementia in ageing populations? *Evid.-Based Dent.* **2026**, *27*, 1–2. [CrossRef] [PubMed]
20. Sommerlad, A.; Kivimaki, M.; Larson, E.B.; Singh-Manoux, A.; Röhr, S.; Shirai, K.; Livingston, G. Social participation and risk of developing dementia. *Nat. Aging* **2023**, *3*, 532–545. [CrossRef] [PubMed]
21. Blotenberg, I.; Boekholt, M.; Michalowsky, B.; Platen, M.; Rodriguez, F.S.; Teipel, S.; Hoffmann, W.; Thyrian, J.R. Predictors of Life Expectancy in Community-Dwelling People with Dementia–The Role of Social Support. *Alzheimers Dement.* **2024**, *20*, e084646. [CrossRef]
22. Sohn, M.; Yang, J.; Lee, J.H.; Choi, D. Predictive factors for dementia among older adults in South Korea: An interpretable machine learning analysis. *Alzheimers Res. Ther.* **2025**, *17*, 246. [CrossRef] [PubMed]
23. Jung, Y.; Lyu, J. Joint Trajectories of Frailty and Cognitive Function in Older Koreans Based on 14-Year Period Data From Korean Longitudinal Study of Aging (KLoSA). *Geriatr. Gerontol. Int.* **2026**, *26*, e70672. [CrossRef] [PubMed]
24. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
25. 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, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
26. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [CrossRef]
27. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
28. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]

29. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.
30. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 6638–6648.
31. Saito, T.; Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE* **2015**, *10*, e0118432. [CrossRef] [PubMed]
32. Van Calster, B.; McLernon, D.J.; van Smeden, M.; Wynants, L.; Steyerberg, E.W. Calibration: The Achilles heel of predictive analytics. *BMC Med.* **2019**, *17*, 230. [CrossRef] [PubMed]
33. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.
34. Mohanannair Geethadevi, G.; Quinn, T.J.; George, J.; Anstey, K.J.; Bell, J.S.; Sarwar, M.R.; Cross, A.J. Multi-domain prognostic models used in middle-aged adults without known cognitive impairment for predicting subsequent dementia. *Cochrane Database Syst. Rev.* **2023**, *6*, CD014885. [CrossRef] [PubMed]
35. Collins, G.S.; de Groot, J.A.; Dutton, S.; Omar, O.; Shanyinde, M.; Tajar, A.; Voysey, M.; Wharton, R.; Yu, L.-M.; Moons, K.G.; et al. External validation of multivariable prediction models: A systematic review of methodological conduct and reporting. *BMC Med. Res. Methodol.* **2014**, *14*, 40. [CrossRef] [PubMed]
36. Zhu, Z.; Xu, J.; Lin, Y.; Chai, K.; Zhou, Y.; Jia, R.; Ni, H.; Wu, M.; Wen, C.; Qiao, Y.; et al. Correlation between nutritional status and oral health quality of life, self-efficacy of older inpatients and the influencing factors. *BMC Geriatr.* **2022**, *22*, 280. [CrossRef] [PubMed]
37. Walsh, S.; Klee, M.; Hui, E.K.; Saeed, U.; Waters, S.; Kuhn, I.; Siette, J.; Orgeta, V.; Stafford, J.; Smith, L.J.; et al. Social determinants of dementia: A scoping review. *Alzheimers Dement.* **2025**, *21*, e70524. [CrossRef] [PubMed]
38. Walsh, S.; Wallace, L.; Merrick, R.; Hayat, S.; Luben, R.; Mytton, O.; Lafortune, L.; Brayne, C. How many future dementia cases would be missed by a high-risk screening program? A retrospective cohort study in a population-based cohort. *Alzheimers Dement.* **2024**, *20*, 6278–6286. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.