




## Research Article

# Machine learning prediction of dementia conversion in mild cognitive impairment: A two- to six-year follow-up study

Valgeir Thorvaldsson<sup>1,2</sup> , Johan Svensson<sup>3,4</sup>, Emir Basic<sup>5</sup>, Michael Jonsson<sup>5,6</sup>, Petronella Kettunen<sup>5,6</sup> and Anders Wallin<sup>5,6</sup>

<sup>1</sup>Department of Psychology, University of Gothenburg, Gothenburg, Sweden, <sup>2</sup>Center for Ageing and Health, University of Gothenburg, Gothenburg, Sweden, <sup>3</sup>Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, <sup>4</sup>Region Västra Götaland, Skaraborg Central Hospital, Department of Internal Medicine, Skövde, Sweden, <sup>5</sup>R&D-Unit, Neuropsychiatry, Sahlgrenska University Hospital, Mölndal, Sweden and <sup>6</sup>Department of Psychiatry and Neurochemistry, Institute of Neuroscience of Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

### Abstract

**Objectives:** Mild cognitive impairment (MCI) involves measurable cognitive decline that does not yet significantly disrupt daily functioning but may signal increased risk of dementia. Reliable prediction of dementia conversion in MCI is essential for early intervention and optimized clinical trial design. This study aimed to evaluate the predictive performance of various machine learning (ML) classification algorithms using clinical and neuropsychological data. **Methods:** Data were drawn from the Gothenburg MCI Study and included 347 patients from a memory clinic, of whom 84 (24%) converted to dementia within two to six years. We applied 11 ML classification algorithms (logistic regression, linear discriminant analysis, naïve Bayes, k-nearest neighbors, LASSO, ridge regression, elastic net, decision tree, random forest, gradient boosting, and support vector machine (SVM)) to predict dementia conversion based on 54 clinical predictors (e.g., cerebrospinal fluid biomarkers, neuropsychological test scores, comorbidities, and demographics). In a second step, we included delta scores reflecting change in neuropsychological test performance from baseline to follow-up. **Results:** Without delta scores, LASSO, ridge, elastic net, random forest, and SVM performed best, achieving accuracy  $\geq 0.87$ , kappa = 0.64, and AUC-ROC  $\geq 0.90$ . These models demonstrated high specificity (0.94) but moderate sensitivity (0.68). Including delta scores improved performance, with ridge and elastic net achieving accuracy = 0.90, kappa = 0.73 and 0.72, AUC-ROC = 0.94, specificity = 0.96, and sensitivity = 0.73. The elastic net model yielded a positive predictive value of 0.85 and a negative predictive value of 0.92. **Conclusions:** ML models incorporating clinical and cognitive change data can accurately predict dementia conversion in MCI, supporting their utility in clinical decision-making.

**Keywords:** Mild cognitive impairment; dementia conversion; machine learning prediction; classification algorithms; neuropsychological test scores; cerebrospinal fluid biomarkers

(Received 8 August 2025; final revision 16 December 2025; accepted 16 December 2025; First Published online 13 January 2026)

### Statement of Research Significance

**Research Question(s) or Topic(s):** This study explored whether advanced computer-based methods can accurately predict which individuals with mild cognitive problems are at higher risk of developing dementia in the coming years. **Main Findings:** Using data from a memory clinic in Gothenburg, the study tested several computer-based prediction methods on clinical information from 347 patients. About one in four patients later developed dementia. The most accurate methods used a combination of medical test results, thinking and memory test scores, and measures of change in cognitive performance over time. The best-performing models correctly identified future dementia in about 9 out of 10 cases. **Study Contributions:** The findings show that prediction tools based on clinical and cognitive information, including how thinking skills change over time, can help identify individuals at higher risk for dementia. This has important implications for earlier diagnosis, timely intervention, and improved planning for research and treatment trials.

**Corresponding author:** Valgeir Thorvaldsson; Email: [valgeir.thorvaldsson@psy.gu.se](mailto:valgeir.thorvaldsson@psy.gu.se)

**Cite this article:** Thorvaldsson V., Svensson J., Basic E., Jonsson M., Kettunen P., & Wallin A. (2026) Machine learning prediction of dementia conversion in mild cognitive impairment: A two- to six-year follow-up study. *Journal of the International Neuropsychological Society*, 32: 11–23, <https://doi.org/10.1017/S135561772510177X>

© The Author(s), 2026. Published by Cambridge University Press on behalf of International Neuropsychological Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

## Introduction

Alzheimer's disease (AD) and other forms of dementias represent some of the most challenging global health issues, with profound societal, economic, and personal consequences (Wimo *et al.*, 2023). Mild cognitive impairment (MCI) is widely recognized as a putative stage of dementia and defined as a measurable decline in one or more cognitive domains that is greater than expected for an individual's age and education level without significantly interfering with daily functional abilities, thus not meeting the criteria for dementia (Petersen *et al.*, 1999; 2018). While MCI increases the risk for future cognitive decline, only about 30% to 50% of individuals with MCI will progress to dementia within a period of several years (Mitchell & Shiri-Feshki, 2009; Salemm *et al.*, 2025). Accurate prediction of which individuals with MCI are at highest risk for progression is essential for early intervention, targeted monitoring, and optimized clinical trial design.

A variety of machine learning (ML) algorithms have recently emerged as powerful tools for modeling complex and potentially nonlinear associations using high-dimensional data, offering significant promise in the context of individualized risk prediction. While previous studies have demonstrated the utility of ML algorithms in predicting dementia risk (e.g., Agostinho *et al.*, 2024; Bucholc *et al.*, 2023; Franciotti *et al.*, 2023; Valsdóttir *et al.*, 2024), few have conducted comprehensive comparative evaluation of multiple ML algorithms using a multimodal set of baseline predictors in a well-characterized MCI population. Notably, many of the existing studies (e.g., Cui *et al.*, 2011; Massetti *et al.*, 2022) have relied on data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), which provides a rich and reliable dataset. However, to strengthen and validate these findings, there is a need for additional studies based on independent cohorts with different recruitment settings and clinical characteristics (see e.g., Borchert *et al.*, 2023).

In the present study, we applied and compared 11 different ML algorithms to predict conversion from MCI to dementia, using baseline data collected during individuals' initial visits to a memory clinic. Our focus was on predicting dementia conversion occurring between two to six years following the baseline assessment. The predictive models incorporated a broad range of baseline variables reflecting multiple domains of relevance for dementia risk. These included, among others, data from a comprehensive neuropsychological assessment of cognitive function, cerebrospinal fluid (CSF) biomarkers, apolipoprotein E (*APOE*) allele status, clinical indicators of cerebrovascular and cardiovascular health, anthropometric measures, and basic demographic information.

In a second analytical step, we also evaluated whether the inclusion of cognitive decline over time would improve the prediction. Specifically, we added the differences score (T2-T1) from the cognitive assessments to our model, reflecting change in cognitive performance between the baseline and two-year follow-up assessment. This is based on prior evidence showing that cognitive decline is associated with elevated dementia risk (Amieva *et al.*, 2005; Thorvaldsson *et al.*, 2011).

By evaluating the predictive performance across a broad range of ML algorithms, we aimed to identify the approaches offering the most accurate and clinically informative predictions of dementia progression. Additionally, whenever possible, we explored the relative importance of each predictor across models, offering insight into cognitive, genetic and biological factors most relevant to dementia conversion prediction among individuals with MCI.

**Table 1.** Frequency of dementia diagnoses at each follow-up in the Gothenburg MCI study ( $n = 84$ )

Diagnoses	T2 (2 years)	T3 (4 years)	T4 (6 years)
Alzheimer's disease (AD)	29	7	2
Dementia of vascular type (VaD)	13	2	3
Mixed dementia (AD and VaD)	17	3	1
Lewy body dementia (LBD)	2		
Primary progressive aphasia (PPA)	1		
Non-specified uncertain dementia <sup>a</sup> (NUD)	2	1	1

Note. <sup>a</sup>Etiology could not be determined.

Whereas most previous studies have developed and validated ML models for predicting dementia conversion using large-scale research datasets such as ADNI, the present study used data derived from a memory clinic cohort. This distinction is important because memory clinic patients represent the population in which such predictive algorithms would ultimately be implanted in clinical practice. Compared with research database cohorts, memory clinic samples are more likely to be heterogeneous, encompassing a broader spectrum of cognitive impairment, comorbidities, educational background, and medication use (see Farias *et al.*, 2009; Gianattasio *et al.*, 2020; Snitz *et al.*, 2018). Consequently, they provide a more ecologically valid representation of real-world diagnostic populations. By evaluating the performance of multiple ML algorithms in this clinical context, the present study contributes to bridging the gap between research-oriented prediction models and their potential application in the routine diagnostic setting, offering insight into the generalizability and clinical utility of ML derived dementia prediction.

## Methods

### Participants

Data for the present study were drawn from the Gothenburg MCI Study (Wallin *et al.*, 2016), a longitudinal investigation of individuals with MCI. The sample consisted of 347 participants (201 females) who were identified with MCI at study entry. Participants had an average of 12.4 years of formal education ( $SD = 3.6$ ), and the mean age at baseline was 63.8 years ( $SD = 7.7$ ). Approximately 75% of the participants were referred to the memory clinic by general practitioners or medical specialists due to suspected MCI, whereas 25% were self-referred following concerns about cognitive decline. Baseline assessments began in the year 2000, with subsequent waves of data collection initiated in 2002, 2004, and 2006. Over the follow-up period of two to six years, 84 participants received a clinical diagnosis of dementia (see Table 1). Further details regarding the sample composition, including comparisons between individuals with stable MCI and those who converted to dementia, are presented in Table 2.

The study was approved by the Regional Ethical Review Board in Gothenburg, Sweden (diary numbers: L091-99 and T479-11), and was conducted in accordance with the principles of the Declaration of Helsinki (1975, as revised). Written informed consent was obtained from all participants prior to their inclusion in the study.

### Diagnosis of dementia disease

In patients who converted to dementia, the specific diagnoses were established using well-recognized clinical criteria. Alzheimer's

**Table 2.** Descriptive statistics for variables included in the machine learning prediction analysis of dementia conversion among individuals with mild cognitive impairment (MCI; excluding cognitive tests variables)

Variables	Total sample (N = 347)	MCI stable non-case (n = 263)	MCI converted cases (n = 84)	% missing
Females (%)	201 (58)	155 (59)	46 (55)	0
Age at baseline; M (SD)	63.79 (7.72)	66.45 (7.41)	62.94 (7.63)	0
Education (in years); M (SD)	12.39 (3.58)	11.32 (3.63)	12.73 (3.50)	0.86
Smoking/tobacco use (%)				0
No tobacco use	189 (54)	136 (52)	53 (63)	
Uses tobacco	124 (36)	100 (38)	24 (29)	
Former tobacco user	34 (10)	27 (10)	7 (8)	
Alcohol consumption				0.58
None	35 (10)	27 (10)	8 (10)	
Low or moderate	298 (86)	223 (85)	75 (89)	
High	12 (4)	11 (4)	1 (1)	
Global deterioration scale (%)				0
Stage 2. Subjective cognitive impairment	138 (40)	132 (50)	6 (7)	
Stage 3. Mild cognitive impairment	209 (60)	131 (50)	78 (93)	
Probable baseline etiology (%)				3.75
Non-vascular (primary)	227 (68)	176 (70)	51 (61)	
Vascular	35 (11)	21 (8)	14 (17)	
Mixed	72 (21)	54 (22)	18 (22)	
APOE genotype (%)				2.31
$\epsilon 2/\epsilon 3$	30 (9)	26 (10)	4 (5)	
$\epsilon 2/\epsilon 4$	13 (4)	11 (4)	2 (2)	
$\epsilon 3/\epsilon 3$	141 (42)	116 (45)	25 (30)	
$\epsilon 3/\epsilon 4$	119 (35)	85 (33)	34 (41)	
$\epsilon 4/\epsilon 4$	36 (11)	18 (7)	18 (22)	
CSF biomarkers				
A $\beta$ 1-42; M (SD)	602.22 (205.81)	644.49 (203.31)	474.92 (155.31)	2.88
Total tau; M (SD)	392.29 (278.58)	320.12 (166.97)	614.01 (408.20)	2.59
Phosphorylated tau181; M (SD)	58.5 (29.29); 52	53.69 (24.11)	75.68 (38.54)	10.37
Anthropometric measures				
Body mass index; M (SD)	24.98 (3.57)	25.12 (3.62)	24.56 (3.42)	2.30
Body circumference (cm)	87.18 (11.37)	87.72 (11.49)	85.63 (10.93)	7.78
Sagittal (abdominal) diameter (cm)	18.06 (2.91)	18.13 (2.97)	17.82 (2.72)	3.46
Hip circumference (cm)	100.3 (7.83)	100.61 (8.07)	99.36 (7.02)	5.48
Cerebro- and cardio-vascular health (%)				
Angina pectoris	6 (1.7)	5 (1.9)	1 (1.2)	0
Myocardial infarction	17 (4.9)	13 (4.9)	4 (4.8)	0
Hyperlipidemia	82 (23.8)	64 (24.5)	18 (21.7)	0.86
Transient ischemic attack	17 (4.9)	10 (3.8)	7 (8.3)	0.58
Stroke	15 (4.3)	8 (3.0)	7 (8.4)	0.29
Lacunar infarction	15 (4.3)	9 (3.4)	6 (7.2)	0.29
Diabetes	17 (4.8)	10 (3.8)	7 (8.3)	0
Peripheral vascular disease	9 (2.6)	5 (1.9)	4 (4.8)	0
Blood pressure systolic (mmHg); M (SD)	138.05 (18.88)	137.29 (19.25)	140.37 (17.6)	2.31
Blood pressure diastolic (mmHg); M (SD)	78.64 (9.27)	78.92 (9.11)	77.75 (9.74)	2.31
Pulse pressure mmHg; M (SD)	63.58 (9.34)	63.43 (9.39)	64.05 (9.21)	2.31
Cholesterol (total) mmol/L; M (SD)	5.54 (3.27)	5.84 (3.72)	5.86 (1.12)	2.56
Operation under narcosis (%)	206 (72)	59 (75)	147 (71)	17.58

disease (AD) was diagnosed according to the *Diagnostic and Statistical Manual of Mental Disorders* criteria for dementia, Third Edition, Revised (DSM-III-R; American Psychiatric Association, 1987), in conjunction with the NINCDS-ADRDA criteria for probable AD (McKhann et al., 1984). More specifically, for an AD diagnosis, the patient had to have a predominant parietotemporal lobe syndrome and no, or insignificant, cerebrovascular pathology on magnetic resonance imaging (MRI). Vascular dementia (VaD) was diagnosed based on the National Institute of Neurological Disorders and Stroke and Association Internationale pour la Recherche et l'Enseignement en Neurosciences (NINDS-AIREN) criteria (Román et al., 1993) and by the criteria for subcortical VaD proposed by Erkinjuntti et al. (2000). The latter emphasizes predominant impairments in executive functioning, processing speed, and gait, together with magnetic resonance imaging (MRI) evidence of cerebrovascular pathology judged to impose a significant burden on brain function. Mixed dementia (combined AD and VaD) was diagnosed in individuals who met the NINCDS-ADRDA

criteria for AD and also demonstrated cerebrovascular changes on MRI, though not of a severity deemed to cause major cerebral burden independently. Less common dementia subtypes were diagnosed using established consensus criteria: frontotemporal dementia (FTD) according to Neary et al. (1998), Lewy body dementia (LBD) according to McKeith et al. (1999), and primary progressive aphasia (PPA) according to Gorno-Tempini et al. (2011). A diagnosis of non ultra descriptum was assigned when the patient fulfilled the general criteria for clinical dementia, but did not meet the specific diagnostic criteria for any recognized dementia subtype. Importantly, information from cognitive assessments or other variables later used as predictors in the ML models was not incorporated into the clinical diagnostic decision-making process.

#### Identification of MCI

*Global deterioration scale.* Cognitive impairment was assessed using the Global Deterioration Scale (GDS; Auer & Reisberg,

1997), which rates the severity of cognitive decline across seven stages. In the present study, only stages 2 and 3 were included: stage 2 reflects subjective cognitive impairment, and stage 3 corresponds to MCI. Stage 1, indicating cognitively healthy individuals, and stages 4–7, reflecting dementia stages, were not represented (i.e., at baseline) in this sample. Detailed information on the GDS assessment procedure within the Gothenburg MCI study is provided in Wallin *et al.* (2016).

**Probable baseline etiology.** At baseline, all patients were assessed for probable etiological subtype based on vascular burden (i.e., the level of systemic and cerebrovascular risk factors) and classified into one of three categories: (1) vascular etiology, (2) mixed etiology, or (3) nonvascular etiology (i.e., primary degenerative etiology). This method differs from that of the assessment of the specific dementia diagnosis, which is based on cerebral manifestations without taking into account systemic manifestations. For further details, see Wallin *et al.* (2016).

### Neuropsychological assessments

The cognitive assessments followed the recommendations of the American Academy of Neurology, covering the domains of learning and memory, speed and attention, visuospatial ability, language, and executive functions. Multiple tests were included within each domain to obtain a comprehensive profile of cognitive functioning.

**Learning and memory** were measured using the Rey Auditory Verbal Learning Test (RAVLT; Geffen *et al.*, 1994), Wechsler Logical Memory Test (WLM; Wechsler, 1987), and Rey Complex Figure (RCF; Meyers & Meyers, 1995) delayed recall.

**Speed and attention** were assessed using the Digit Symbol Substitution and Digit Span subtests from the Wechsler Adult Intelligence Scale–Revised (WAIS-R; Wechsler, 1981), along with Trail Making Tests A and B (Reitan, 1985).

**Visuospatial ability** was evaluated with the Rey Complex Figure copy task (Meyers & Meyers, 1995), the Block Design subtest from the WAIS-R (Wechsler, 1981), and the Silhouettes subtest from the Visual Object and Space Perception Battery (VOSP; Binetti *et al.*, 1996).

**Language** function was assessed using the Token Test (Part V; Bandera *et al.*, 1985), the Boston Naming Test (Kaplan & Weintraub, 1983), Word Fluency (FAS; Crossley *et al.*, 1997), the Similarities subtest from the WAIS-R (Wechsler, 1981).

**Executive functions** were evaluated using a range of tasks including Parallel Serial Mental Operations (PaSMO; Nordlund *et al.*, 2005), Dual Task performance (Della Sala *et al.*, 1995), the computerized short version of the Wisconsin Card Sorting Test (WCST-CV64; Paolo *et al.*, 1996), and the Victoria version of the Stroop Test (Spreen & Strauss, 1998).

**General cognitive function** was also evaluated using the Mini-Mental State Examination (MMSE; Folstein *et al.*, 1975) and the Cognitive Estimation Test (CET; Shallice & Evans, 1978).

Testing was carried out across two sessions of approximately one to two hours each, alternating verbal and non-verbal tasks to minimize fatigue and test contamination. Special attention was given to avoid interference with memory assessments between immediate and delayed recall phases. Six trained psychologists administered the tests using a standardized procedure. About one-quarter of the sample received the Block Design, Similarities, and Digit Symbol Substitution subtests from the WAIS-III battery, while the remaining three-quarters completed the corresponding subtests from the older WAIS-R version. To integrate results across

test versions and avoid extensive missing data imputation, raw scores from these tests were z-score transformed and combined. Additional details regarding the cognitive tests and their application within the Gothenburg MCI Study are available in Nordlund *et al.* (2005).

### Other predictors

**Smoking and alcohol.** Smoking and alcohol consumption were assessed through interviews conducted by a memory clinic physician or nurse. Participants were asked about their use of tobacco products, smoking habits, and alcohol consumption, and the responses were recorded by the clinician. Tobacco use was categorized as no tobacco use/smoking, current tobacco use/smoking, or former tobacco user/current non-smoker. Alcohol consumption was classified into three categories: none, low or moderate consumption, and high consumption. No specific quantitative thresholds were predefined for these categories.

**Cerebrospinal fluid biomarkers.** CSF was obtained via lumbar puncture at the L3/L4 or L4/L5 interspace using a 22-gauge Quincke needle. Approximately 20 mL of CSF was collected into polypropylene tubes after discarding the initial fraction to avoid blood contamination. Samples were sent to the Clinical Neurochemistry Laboratory at Sahlgrenska University Hospital, Mölndal, where they were gently mixed, centrifuged at 2,000 g for 10 minutes at room temperature, and aliquoted into labeled polypropylene tubes. Aliquots designated for AD core biomarker analysis – amyloid- $\beta$ 1–42 (A $\beta$ 42), total tau (T-tau), and phosphorylated tau (P-tau181) – were stored at  $-20^{\circ}\text{C}$  for a maximum of one week prior to analysis. The remaining CSF was stored at  $-80^{\circ}\text{C}$ .

The core CSF AD biomarker concentrations were measured using commercially available enzyme-linked immunosorbent assays (INNOTEST, Fujirebio, Ghent, Belgium) in accordance with established protocols (Andreasen *et al.*, 1999; Blennow *et al.*, 1995; Vanmechelen *et al.*, 2000). All analyses were conducted by board-certified laboratory technicians under a standardized laboratory quality control program to ensure consistency and reliability (Palmqvist *et al.*, 2014). To minimize between-assay variability, two or more internal control samples (aliquots of pooled CSF) were included in each analytical run. All procedures for lumbar puncture, pre-analytical handling, and biomarker analysis adhered to standardized protocols (Blennow *et al.*, 2010) and were conducted under accreditation by the Swedish Board for Accreditation and Conformity Assessment.

**Apolipoprotein E.** APOE genotyping in blood was performed at baseline using minisequencing, as previously described (Blennow *et al.*, 2000).

**Cerebrovascular, cardiovascular, and metabolic health markers.** History of angina pectoris, myocardial infarction, peripheral vascular disease, hyperlipidemia, transient ischemic attack, and stroke was recorded at the baseline clinical investigation. Diabetes was defined according to the 2006/2011 World Health Organization diagnostic criteria (Rydén *et al.*, 2013) as having a fasting plasma glucose level  $\geq 7.0$  mmol/L (126 mg/dL), a two-hour plasma glucose level  $\geq 11.1$  mmol/L (200 mg/dL) during an oral glucose tolerance test, or ongoing treatment with oral antidiabetic medication and/or insulin therapy.

**Anthropometric measures.** Body mass index was calculated as weight in kilograms (measured after an overnight fast and with light clothing) divided by height in meters squared ( $\text{kg}/\text{m}^2$ ). Waist circumference was measured at the midpoint between the lower

margin of the last palpable rib and the top of the iliac crest, using a non-stretchable measuring tape, with the participant standing and breathing out gently. Hip circumference was measured at the widest portion of the buttocks, with the tape held horizontally without compressing the skin. Sagittal abdominal diameter was measured in the supine position at the level of the iliac crest using a sliding-beam abdominal caliper, after a normal expiration. This measurement reflects abdominal height and serves as an indicator of visceral adiposity.

**Cholesterol.** Total blood cholesterol (mmol/L) was measured from fasting venous blood samples collected between 08:00 and 10:00 a.m. Analyses were conducted at the Clinical Chemistry Laboratory at Sahlgrenska University Hospital using standardized enzymatic methods.

**Pulse and blood pressure.** Blood pressure (BP) was measured in the sitting position after 10–15 minutes of rest. Two measurements were taken using an automated oscillometric device, and the mean of the two readings was used in the analyses. Systolic blood pressure (SBP) and diastolic blood pressure (DBP) were recorded in mmHg. Pulse pressure (PP) was calculated as the difference between systolic and diastolic blood pressure ( $PP = SBP - DBP$ ). Information about operation under narcosis was derived from medical records.

### Machine learning algorithms and model training

To evaluate and compare the performance of multiple ML algorithms in predicting dementia conversion, we used the caret (classification and regression training) package in R (Kuhn, 2008), which streamlines data preprocessing, model training, and performance evaluation. The following classification algorithms were implemented: logistic regression, linear discriminant analysis (LDA; Fisher, 1936), naïve Bayes (Hand & Yu, 2001), k-nearest neighbors (KNN; Cover & Heart, 1967), least absolute shrinkage and selection operator (LASSO; Tibshirani, 1996), ridge regression (Hoerl & Kennard, 1970), elastic net regression (Zou & Hastie, 2005), decision tree (Breiman et al., 1984), random forest (Breiman, 2001), gradient boosting (Friedman, 2001), and support vector machine (SVM; Cortes & Vapnik, 1995), following standard workflow (see e.g., James et al., 2021).

Model performance was evaluated using leave-one-out cross-validation (LOOCV), a resampling method well-suited for small to moderately sized datasets. In this approach, each observation serves once as the test set, while the remaining  $n-1$  cases form the training set. Within each training set, we performed internal resampling to tune hyperparameters, ensuring that the held-out test case was never used during model selection. This corresponds to a form of nested LOOCV, where the inner loop tunes the hyperparameters and the outer loop evaluates prediction error on each left-out subject. This approach prevents data leakage and yields unbiased performance estimates, reducing both bias and variance in model evaluation.

Data preprocessing, including standardization and missing data imputation, was conducted within the caret framework. Missing values accounted for 7.8% of the total data, primarily due to study design constraints, where certain cognitive tests were occasionally omitted for time-related reasons. Missing values were handled using k-nearest neighbor imputation ( $k = 5$ ), implemented via the caret package. Imputation was performed within each LOOCV training fold only, such that the imputation model was estimated using the training data and then applied to the held-out observation, thereby preventing data leakage. Where required,

predictors were centered and scaled prior to model fitting. For algorithms requiring hyperparameter tuning (KNN, LASSO, ridge, elastic net, decision tree, random forest, gradient boosting, and SVM), tuning grids were specified and optimized within the LOOCV procedure. Model performance was evaluated using multiple classification metrics, including accuracy, Cohen's Kappa, area under the receiver operating characteristic curve (AUC-ROC), F1-score, sensitivity, specificity, and the confusion matrix. Model calibration was additionally assessed using the calibration intercept and slope, obtained by fitting a logistic regression of the observed outcome on the logit of the model predicted probabilities, as well as Brier scores, which were computed as the mean squared difference between the predicted probabilities and the observed outcomes.

## Results

### Dementia diagnosis across follow-up visits

The frequency of dementia diagnoses across the three follow-up visits at years two, four, and six is presented in Table 1. Out of the 347 MCI patients at baseline, 84 (24%) received a dementia diagnosis during the follow-up period. Among these, 38 individuals (45%) were diagnosed with AD, 18 (21%) with VaD, 21 (25%) with mixed dementia (combined AD and VaD), 2 (2.4%) with LBD, 1 (1.2%) PPA, and 4 (4.8%) received a diagnosis of dementia of unknown etiology. None of the participants received a diagnosis of FTD. For the ML prediction models, the dementia outcome variable was coded as binary. Individuals who received any dementia diagnosis were coded as one, and those who remained stable were coded as zero. We further reiterate that none of the baseline predictors used in the ML classification was incorporated into the diagnostic decision-making.

### Descriptives of predictors

Tables 2 and 3 present descriptive statistics for the predictor variables used in the ML models. Table 2 includes non-cognitive variables such as demographic characteristics (sex, age, and education), lifestyle factors (tobacco and alcohol use), clinical evaluations (GDS), APOE genotype frequency, baseline clinical etiology evaluation, CSF biomarkers, anthropometric data, cardiovascular and cerebrovascular health indicators, and information regarding past surgeries involving anesthesia.

Table 3 summarizes cognitive test performance, organized by the following domains: learning and memory; speed and attention; visuospatial ability; language; and executive function. Information about global cognitive performance measures, including the MMSE and the CET are also presented. Cohen's  $d$  effect sizes comparing converters to non-converters suggest substantial baseline group differences. For example, the memory domain tests showed large effect sizes ( $d > 1.2$ ), with similarly large effects observed in mental speed ( $d = 1.07$  to  $1.4$ ), and visuospatial ability ( $d = 0.76$  to  $0.99$ ). Among non-cognitive variables, high effect sizes were noted for CSF biomarkers: A $\beta$ 42 ( $d = 0.88$ , 95% CI [0.62, 1.14]), T-tau ( $d = 0.79$ , 95% CI [0.51, 1.06]), and P-tau181 ( $d = 1.18$ , 95% CI [0.92, 1.45]). The GDS also demonstrated large group differences at baseline.

### Machine learning prediction model performance

Performance metrics for each ML algorithm are summarized in Table 4. Across models, classification accuracy ranged from 0.82 to 0.88, Cohen's kappa from 0.52 to 0.64, AUC from 0.82 to 0.92, and

**Table 3.** Descriptive statistics for the neuropsychological cognitive test variables included in the machine learning prediction analyses of dementia conversion among individuals with mild cognitive impairment (MCI)

	Total sample ( <i>N</i> = 347)		MCI stable non- cases ( <i>n</i> = 263)		MCI converted cases ( <i>n</i> = 84)		Effect size <sup>f</sup>		% missing
	M	SD	M	SD	M	SD	<i>d</i>	95% CI	
<b>Learning and memory</b>									
RAVLT <sup>a</sup> total	39.75	11.97	42.45	10.69	30.10	8.92	1.20	[0.93, 1.47]	8.07
RAVLT recall	7.26	3.75	8.31	3.35	4.12	3.07	1.27	[1.00, 1.54]	6.92
RAVLT delayed recall (30 min.)	7.11	3.95	8.27	3.42	3.63	3.36	1.36	[1.09, 1.64]	6.92
RAVLT recognition	14.02	2.13	14.37	1.78	12.92	2.71	0.71	[0.44, 0.98]	10.09
RAVLT maximum score	10.34	2.73	11.11	2.46	8.15	2.26	1.23	[0.95, 1.50]	12.39
WLM <sup>b</sup> immediate recall	22.56	7.19	24.13	6.35	15.82	6.71	1.29	[0.97, 1.62]	23.39
WLM delayed recall (30 min.)	17.73	9.02	19.82	7.75	8.90	8.74	1.37	[1.04, 1.70]	24.50
RCF <sup>c</sup> immediate recall	14.31	7.43	16.46	6.59	7.81	5.91	1.35	[1.07, 1.62]	7.20
RCF delayed recall (30 min.)	14.01	7.62	16.22	6.75	7.19	5.97	1.38	[1.10, 1.65]	8.07
<b>Speed and attention</b>									
Digit symbol substitution <sup>d</sup>	0.00	1.00	0.27	0.92	-0.82	0.75	1.24	[0.97, 1.51]	7.20
Trail making test A	42.88	16.16	38.94	12.81	54.70	19.23	1.07	[0.81, 1.34]	6.63
Trail making test B	107.15	54.28	90.98	36.96	156.06	67.67	1.40	[1.12, 1.67]	7.20
Digit span forward	6.05	1.11	6.14	1.12	5.76	1.02	0.35	[0.09, 0.61]	10.66
Digit span backward	4.69	2.35	4.90	2.61	4.08	1.10	0.35	[0.10, 0.61]	10.66
<b>Visuospatial ability</b>									
RCF copy	31.55	4.46	32.53	3.32	28.55	5.94	0.97	[0.70, 1.23]	6.92
Block design	0.00	1.00	0.22	0.95	-0.69	0.82	0.99	[0.72, 1.25]	7.49
VSOP Silhouettes	20.26	4.45	21.06	4.09	17.84	4.65	0.76	[0.50, 1.02]	7.20
<b>Language</b>									
Token test	19.43	2.54	19.91	2.05	17.94	3.26	0.82	[0.56, 1.09]	7.49
Boston naming Test	52.17	6.09	53.21	4.92	48.88	8.03	0.74	[0.48, 1.01]	9.80
Word fluency (FAS)	39.62	13.17	42.25	12.80	31.62	10.88	0.86	[0.60, 1.12]	6.92
Similarities	0.00	1.00	0.16	0.90	-0.48	1.13	0.66	[0.41, 0.92]	6.63
<b>Executive functions</b>									
Parallel serial mental operations	81.01	32.63	76.75	29.79	97.20	37.80	0.65	[0.36, 0.93]	15.56
Dual task	49.26	10.62	50.87	10.08	44.50	10.84	0.62	[0.30, 0.94]	40.63
Wisconsin card sorting test	39.29	14.07	40.35	13.88	33.91	14.01	0.46	[0.09, 0.84]	42.07
Stroop (Victoria version)	31.07	11.49	29.01	10.10	39.99	12.93	1.03	[0.69, 1.37]	32.28
<b>Global screening tests</b>									
MMSE <sup>e</sup>	28.47	1.46	28.71	1.24	27.70	1.80	0.72	[0.47, 0.97]	0
Cognitive estimation test	4.35	2.12	4.14	2.09	5.26	2.04	0.54	[0.20, 0.87]	33.43

Notes. <sup>a</sup>Rey Auditory Verbal Learning Test. <sup>b</sup>Wechsler Logical Memory. <sup>c</sup>Rey Complex Figure. <sup>d</sup>Digit Symbol Substitution. The Block Design, Digit Symbol Substitution, and Similarities were transformed to a z-score scale. <sup>e</sup>Mini Mental State Examination. <sup>f</sup>Cohen's *d* effect size using pooled standard deviations.

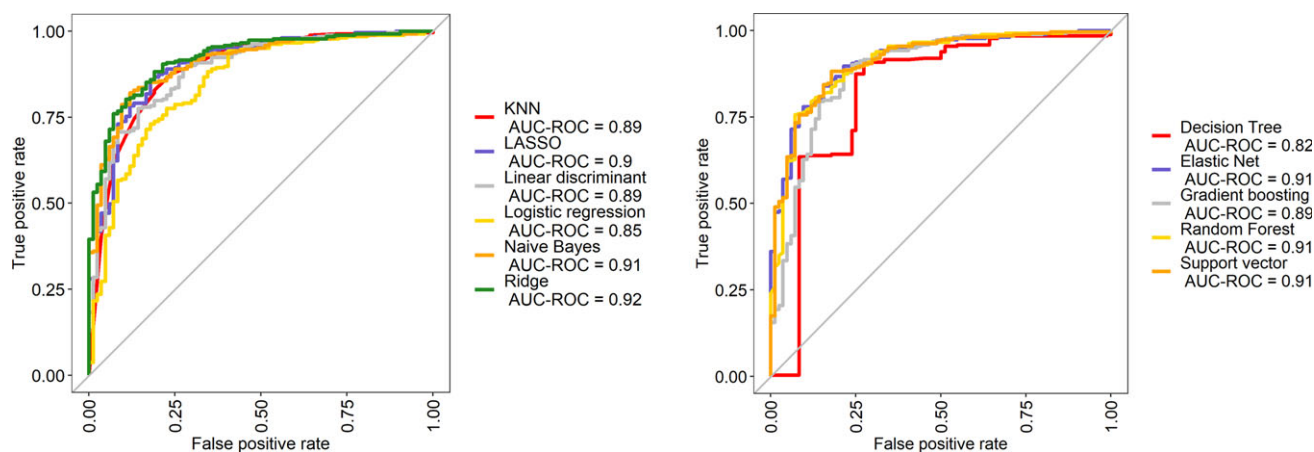
**Table 4.** Performance of machine learning algorithms in predicting dementia diagnosis two to six years after baseline among individuals with mild cognitive impairment (MCI; *N* = 347; prevalence = 84 cases, 24%)

Algorithm	Accuracy <sup>a</sup>	95% CI	Kappa	AUC-ROC	95% CI	F1-score	Sensitivity (recall)	Specificity	TP	TN	FP	FN
Logistic regression	0.82	[0.77, 0.86]	0.52	0.85	[0.81, 0.89]	0.64	0.67	0.86	56	228	35	28
Linear discriminant analysis	0.85	[0.81, 0.89]	0.60	0.89	[0.86, 0.92]	0.69	0.68	0.91	57	239	24	27
Naïve Bayes	0.86	[0.82, 0.89]	0.62	0.91	[0.88, 0.94]	0.71	0.69	0.91	58	241	22	26
KNN	0.86	[0.81, 0.89]	0.56	0.89	[0.86, 0.92]	0.64	0.54	0.96	45	252	11	39
LASSO	0.87	[0.83, 0.91]	0.64	0.90	[0.87, 0.93]	0.72	0.67	0.94	56	247	16	28
Ridge	0.88	[0.83, 0.91]	0.64	0.92	[0.89, 0.95]	0.72	0.67	0.94	56	248	15	28
Elastic net	0.87	[0.83, 0.91]	0.64	0.91	[0.88, 0.94]	0.72	0.68	0.94	57	246	17	27
Decision tree	0.86	[0.81, 0.89]	0.60	0.82	[0.78, 0.86]	0.68	0.67	0.92	56	241	22	28
Random forest	0.87	[0.83, 0.90]	0.62	0.91	[0.88, 0.94]	0.70	0.62	0.95	52	251	12	32
Gradient boosting	0.87	[0.83, 0.91]	0.64	0.89	[0.86, 0.92]	0.72	0.67	0.94	56	247	16	28
Support vector machine	0.87	[0.83, 0.90]	0.63	0.91	[0.88, 0.94]	0.71	0.67	0.94	56	246	17	28

Notes. <sup>a</sup>The no-information rate was 0.76 (i.e., the accuracy obtained by always predicting the most common outcome, or no subsequent dementia diagnosis). All models were evaluated using leave-one-out cross validation (LOOCV). Missing data was handled using *k*-nearest neighbors (KNN) imputation with *k* = 5. The confusion matrix includes true positive (TP), true negative (TN), false positive (FP), and false negative (FN); LASSO: least absolute shrinkage and selection operator. Calibration statistics for these models are provided in Supplementary Table S1.

F1 from 0.64 to 0.72. Most models demonstrated high specificity (0.86 to 0.96) but relatively lower sensitivity (i.e., recall; 0.54 to 0.68). A summary of the calibration statistics is provided in Supplementary Table S1.

The regularized regression models (LASSO, ridge, and elastic net) performed well, each with accuracy of 0.87 or more, AUC values between 0.90 and 0.92, and Brier scores of 0.09 and 0.10. The elastic net model had the highest balanced accuracy ((0.68 + 0.94)/



**Figure 1.** Receiver operating characteristic (ROC) curves for the machine learning algorithms predicting dementia conversion two to six years after baseline among individuals with mild cognitive impairment (MCI).

2 = 0.81). These models apply penalties to shrink or eliminate coefficients to avoid overfitting. The LASSO can reduce some coefficients to zero, effectively performing variable selection, the ridge shrinks coefficients toward zero without eliminating them, and elastic net combines both approaches. Given the observed prevalence of dementia (24%) in the MCI sample, the elastic net model achieved a positive predictive value (PPV) of approximately 0.78 and a negative predictive value (NPV) of approximately 0.90. This indicates that, when applied to a similar population, 78% of individuals predicted by the model to convert to dementia are expected to receive a future dementia diagnosis, while 90% of those predicted not to convert are expected to remain free of a dementia diagnosis. These values reflect the model's practical utility in estimating an individual's likelihood of future diagnostic outcomes based on the predicted conversion status.

Tree-based models also performed strongly. Random forest had an accuracy of 0.87 and AUC of 0.91 but showed lower sensitivity (0.62) but high specificity (0.95). The gradient boosting model had an AUC of 0.89, while the decision tree had higher sensitivity (0.67) but lower specificity (0.92). The SVM also performed well, with an accuracy of 0.87 and AUC of 0.91. Some models, like KNN, showed lower kappa (0.56) and poor sensitivity (0.54) but high specificity (0.96). Logistic regression and LDA had the poorest overall performance. The Naïve Bayes model yielded the highest sensitivity, though at the cost of somewhat lower specificity. The ROC for all of the algorithms are shown in Figure 1.

### Variable importance

Figure 2 shows the standardized variable importance (relative to the top-ranked variable set at 100) for the ridge regression and random forest models. While random forest importance rankings will vary across runs due to its inherent randomness (bagging and feature sampling), consistent top predictors emerged. These included P-tau181, recall memory scores (RCF, RAVLT, WLM), mental speed (Digit Symbol Substitution, TMT-A, TMT-B), and executive function (Stroop Test). A $\beta$ 42 generally also ranked among the most important predictors. Overall, the strongest predictors of dementia conversion were cognitive measures of recall memory, speed, and executive function, along with CSF biomarkers, particularly P-tau181 and A $\beta$ 42. Comparable plots are provided in the Supplementary Material (Figures S1 to S3) for

those algorithms for which variable importance ranking was feasible.

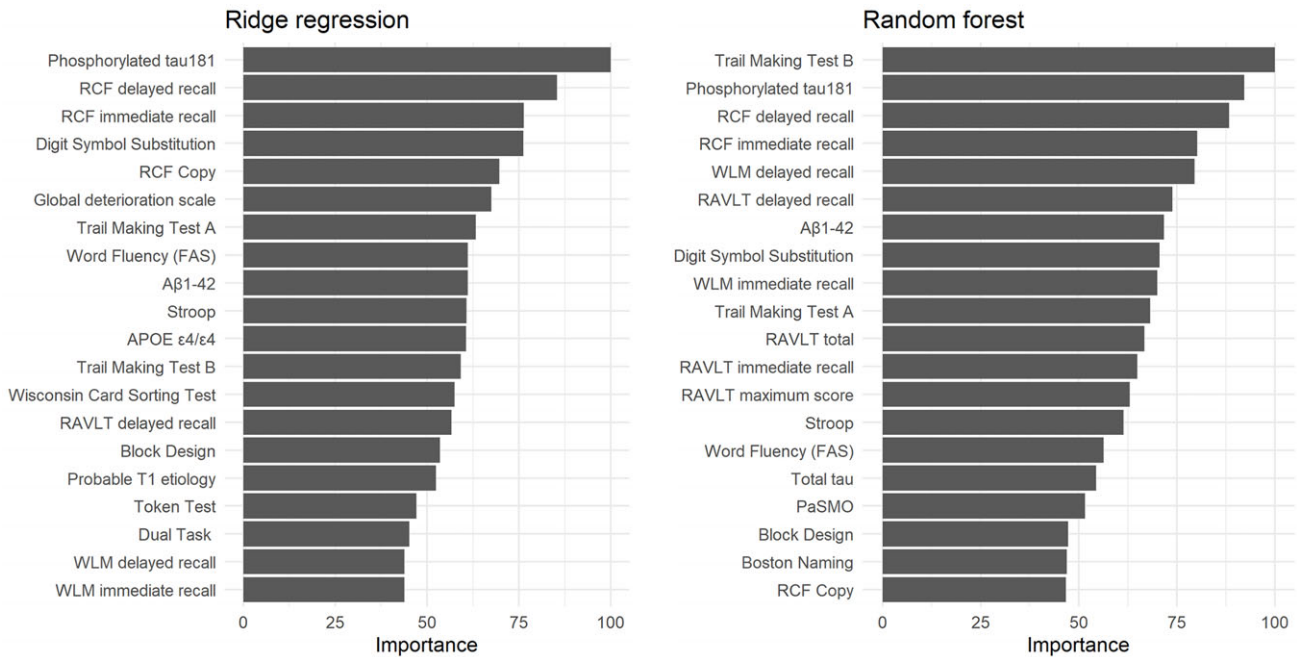
### Inclusion of cognitive delta scores

The findings presented in Table 4 indicate that most of the evaluated algorithms achieved relatively high specificity (>0.90) but lower sensitivity (<0.70). In an effort to improve sensitivity, we extended the models by incorporating measures of change in cognitive performance, using delta scores calculated as the difference between follow-up (T2) and baseline (T1) cognitive test results. The performance of these enhanced models is summarized in Table 5 (see also calibration statistics in Supplementary Table S1).

Among these models, ridge and elastic net regression yielded the highest classification performance, both achieving overall accuracy above 0.90, AUC values exceeding 0.94, and Brier score of 0.08. Specifically, the elastic net model achieved a specificity of 0.96 and a sensitivity of 0.73, representing an improvement from 0.68 in the elastic net model without delta scores. Similarly, the ridge model showed a sensitivity of 0.73, up from 0.67 in its non-delta counterpart. The ridge model slightly outperformed the elastic net by correctly classifying one additional non-converter. The gradient boosting model also performed strongly, with an accuracy of 0.90, AUC of 0.93, specificity of 0.95, sensitivity of 0.73, and Brier score of 0.09. It identified the same number of true converters as the ridge and elastic net models but misclassified one additional non-converter compared to the elastic net model. Figure 3 displays the ROC curves for the ML algorithms incorporating cognitive delta information. These results illustrate substantial gains in discriminative ability when change scores are considered in the modeling process.

### Variable importance when using delta scores

Variable importance rankings for the elastic net and gradient boosting models are shown in Figure 4. Consistent with earlier models (those not incorporating delta scores), the most influential predictors included cognitive measures of memory, processing speed, and executive functioning, as well as CSF biomarkers such as P-tau181. Notably, several delta scores contributed meaningfully to predictive performance, for example changes in MMSE and the Similarities test. Comparable plots are provided in the



**Figure 2.** The 20 most important variables for predicting subsequent dementia conversion two to six years after baseline among individuals with mild cognitive impairment (MCI) in the ridge regression and random forest models. Variable importance was standardized relative to the most important variable. Comparable plots for the other models are presented in the Supplementary Material (Figures S1–S3).

**Table 5.** Performance of machine learning algorithms in predicting dementia diagnosis two to six years after baseline among individuals with mild cognitive impairment (MCI), including cognitive delta scores ( $N = 347$ ; prevalence = 84 cases, 24%)

Algorithm	Accuracy <sup>a</sup>	95% CI	Kappa	AUC-ROC	95% CI	F1-score	Sensitivity (recall)	Specificity	TP	TN	FP	FN
Logistic regression <sup>b</sup>	0.81	[0.77, 0.85]	0.51	0.79	[0.74, 0.84]	0.64	0.68	0.86	57	225	38	27
Linear discriminant analysis	0.88	[0.84, 0.9]	0.68	0.92	[0.89, 0.95]	0.76	0.74	0.93	62	245	18	22
Naïve Bayes	0.84	[0.79, 0.87]	0.52	0.87	[0.83, 0.91]	0.62	0.56	0.92	47	243	20	37
KNN	0.86	[0.82, 0.90]	0.54	0.91	[0.88, 0.94]	0.61	0.45	0.99	38	261	2	46
LASSO	0.89	[0.85, 0.92]	0.67	0.94	[0.92, 0.96]	0.74	0.63	0.97	53	256	7	31
Ridge	0.90	[0.87, 0.93]	0.73	0.94	[0.92, 0.96]	0.79	0.73	0.96	61	253	10	23
Elastic net	0.90	[0.87, 0.93]	0.72	0.94	[0.92, 0.96]	0.78	0.73	0.96	61	252	11	23
Decision tree	0.85	[0.81, 0.89]	0.59	0.77	[0.72, 0.82]	0.68	0.66	0.92	55	241	22	29
Random forest	0.89	[0.85, 0.92]	0.66	0.92	[0.89, 0.96]	0.72	0.61	0.98	51	257	6	33
Gradient boosting	0.90	[0.86, 0.93]	0.71	0.93	[0.90, 0.96]	0.78	0.73	0.95	61	251	12	23
Support vector machine	0.89	[0.85, 0.92]	0.70	0.93	[0.90, 0.96]	0.76	0.71	0.95	60	249	14	24

Notes. <sup>a</sup>The no-information rate was 0.76 (i.e., the accuracy obtained by always predicting the most common outcome, or no subsequent dementia diagnosis). <sup>b</sup>The logistic regression model did not converge when delta scores were included. All models were evaluated using leave-one-out cross validation (LOOCV). Missing data was handled using k-nearest neighbors (KNN) imputation with  $k = 5$ . The confusion matrix includes true positive (TP), true negative (TN), false positive (FP), and false negative (FN); LASSO: least absolute shrinkage and selection operator. Calibration statistics for these models are provided in Supplementary Table S1.

Supplementary Material (Figures S4 to S6) for the other algorithms where such ranking could be derived.

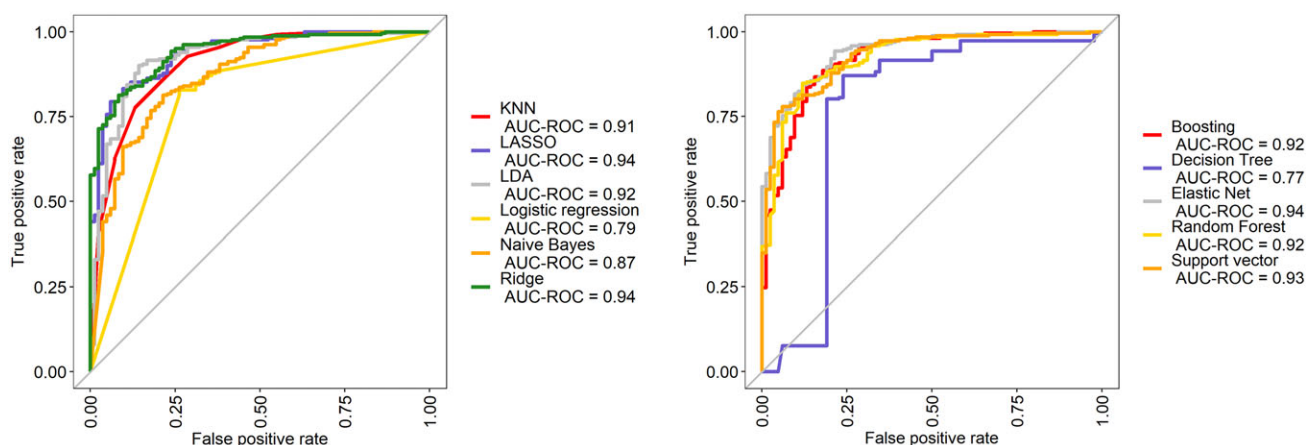
Finally, using the observed dementia prevalence of 24% in the MCI sample, and the elastic net model's specificity (0.96) and sensitivity (0.73), we calculated a positive predictive value (PPV) of 0.85 and a negative predictive value (NPV) of 0.92. This indicates that, in a population with similar prevalence, individuals predicted to convert to dementia have an 85% probability of receiving a future dementia diagnosis, while those predicted not to convert have a 92% probability of remaining dementia-free.

## Discussion

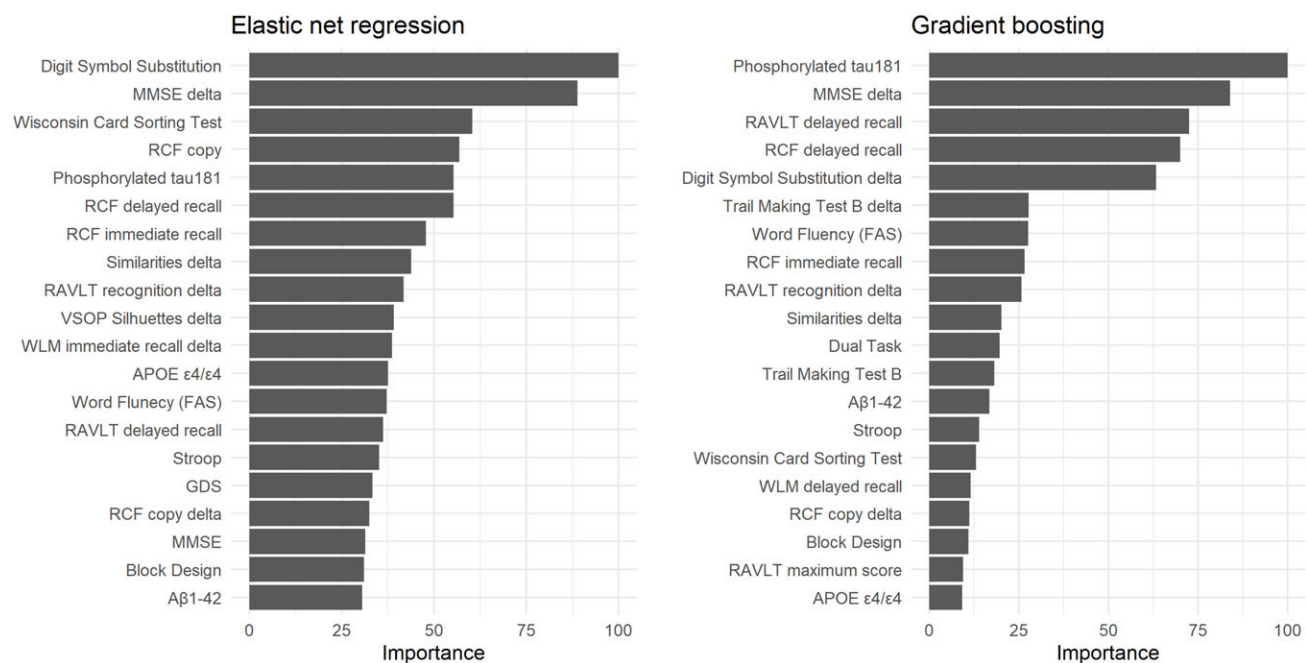
In this study, we aimed to evaluate the accuracy of ML algorithms in predicting future dementia conversion among individuals

diagnosed with MCI, using clinically available baseline data and cognitive follow-up information. Our findings demonstrate that accurate and meaningful prediction of dementia conversion is feasible. Specifically, several ML algorithms, particularly regularized regression models, such as elastic net and ridge regression, achieved high levels of accuracy (up to 90%) and excellent discrimination ( $AUC-ROC > 0.94$ ), with sensitivity and specificity values suggesting practical utility in identifying both converters and non-converters.

One key finding across models was the relatively high specificity (often  $> 0.90$ ) compared to lower sensitivity (typically  $< 0.70$ ), a pattern observed in previous studies using ML for dementia prediction in MCI populations (e.g., Bucholc *et al.*, 2023; Massetti *et al.*, 2022). This discrepancy is partly driven by class imbalance, with converters comprising only 24% of our sample. As



**Figure 3.** Receiver operating characteristic (ROC) curves for the machine learning algorithms predicting dementia conversion two to six years after baseline among individuals with mild cognitive impairment (MCI), based on models including cognitive delta (i.e., difference) scores. The logistic regression model did not converge when delta scores were included.



**Figure 4.** The 20 most important variables for predicting subsequent dementia conversion two to six years after baseline among individuals with mild cognitive impairment (MCI) in the elastic net regression and gradient boosting models that included the cognitive delta (i.e., difference) scores as predictors. Variable importance was standardized relative to the most important variable. Comparable plots for the other models are presented in the Supplementary Material (Figures S4–S6).

a result, models tend to favor the majority, non-converting class, which inflates specificity. Nevertheless, the best-performing models, particularly the elastic net and ridge regression when enhanced with cognitive change scores, demonstrated improved sensitivity (0.73), highlighting the value of incorporating longitudinal data into predictive modeling.

Comparing our model performance to previous studies, our results appear relatively strong and, in some cases, comparable to or exceeding earlier findings. For instance, the ridge regression model (excluding cognitive delta scores) achieved an AUC-ROC of 0.92, an accuracy of 0.88, and a Cohen’s kappa of 0.64, indicating high discrimination and substantial agreement beyond chance, despite the modest number of converters ( $n = 84$ ). These values

align closely with those reported by Guo et al. (2024), who found AUC of 0.92 for two-year prediction and 0.89 for six-year prediction of dementia conversion based on clinical prediction scores. In contrast, Chang & Brainerd (2022), using latent neuropsychological and MRI factors, in addition to CSF, reported an AUC of 0.86 for a two-year prediction window, highlighting the strength of our neuropsychological and CSF-based model even without imaging. Furthermore, the inclusion of cognitive delta scores substantially improved performance across models. For example, in the elastic net model, sensitivity increased from 0.68 to 0.73, with accuracy rising to 0.90 and AUC reaching 0.94. These results underscore the predictive utility of changes in cognitive performance, which captures dynamic processes of decline that

precede a formal dementia diagnosis (Hall et al., 2000; Wilson et al., 1999).

Model comparisons further illuminate the mechanisms underlying performance differences. Regularized regression methods (LASSO, ridge, and elastic net) consistently outperformed traditional statistical approaches such as logistic regression and LDA. This reflects their capacity to handle high-dimensional and multicollinear data, mitigate overfitting, and capture complex relationships through penalization strategies. In contrast, logistic regression and LDA, both linear models without built-in regularization, are more sensitive to multicollinearity and model bias, which limits their utility in heterogeneous clinical datasets.

Tree-based models showed variable performance depending on model complexity. While the simple decision tree performed modestly, the random forest and gradient boosting models yielded excellent AUC values (0.91 and 0.89, respectively) and high specificity, although sensitivity remained modest. This pattern reflects the known strengths and weaknesses of tree-based methods: Random forests reduce variance through bagging but may struggle with capturing rare outcomes, while boosting improves bias reduction but may overfit without careful tuning. The relatively poorer performance of the simple decision tree reflects a lack of pruning and regularization, causing it to fit noise rather than meaningful patterns.

Our study extends prior ML work on forecasting dementia conversion from MCI by partly replicating earlier findings while addressing key gaps. Most previous evaluations have relied on large, population-based cohort, whereas we analyzed a clinical sample. This offers a more direct view of generalizability and clinical transportability given the case-mix, comorbidities, and referral patterns typical of practice. A further contribution is our extensive neuropsychological test battery, which included multiple cognitive domains to a greater extent than is usual and provides evidence that such a battery can improve prediction relative to minimalist cognitive screens. Importantly, we also incorporated information about individual differences in change over time (i.e., delta scores), moving beyond the predominantly cross-sectional information used in many prior studies.

### Variable importance

In terms of variable importance, the most influential predictors were consistent across models. Measures of delayed recall memory (e.g., RCF, RAVLT, and WLM), processing speed (e.g., Digit Symbol, TMT-A/B), and executive function (e.g., Stroop) were highly ranked, aligning with prior findings that these domains decline early in the dementia trajectory (Howieson et al., 2008; Thorvaldsson et al., 2011). Among the CSF biomarkers, P-tau181 emerged as the most critical predictor, followed by A $\beta$ 42. These markers are well-established indicators of Alzheimer's disease pathology and have been shown to correlate with cognitive impairment severity and progression (see e.g., Nelson et al., 2012; Thorvaldsson et al., 2010). Interestingly, APOE  $\epsilon$ 4 carrier status contributed relatively little to the prediction, possibly because its effect is more pronounced earlier in the disease course or because its predictive utility is captured indirectly through CSF and cognitive variables.

Importantly, delta scores in cognition, especially changes in recall memory, MMSE and Similarities, substantially improved prediction. The finding that change in Similarities, a measure of crystallized intelligence, contributed meaningfully to prediction

supports work suggesting that decline in crystallized abilities may signal neurodegeneration, even when static (baseline) performance remains within normal limits (Thorvaldsson et al., 2011). This reinforces the importance of repeated cognitive measurements in clinical follow-up, particularly for capturing subtle intraindividual decline that may not be evident from a single time point assessment.

Across algorithms, a relatively small subset of the 54 candidate predictors consistently drove performance, appearing at the top of the importance rankings. These were primarily CSF biomarkers (i.e., P-tau181 and A $\beta$ 42) and neuropsychological measures (i.e., episodic memory, processing speed, and executive functioning). Many other variables were repeatedly down ranked and contributed little to increment value, suggesting that a more parsimonious predictor panel is both feasible and desirable for clinical use. Importantly, most of these core measures are already collected at many memory clinics, and emerging blood-based assays are likely to further enhance the practicality.

### Limitations

The principal limitation is the small sample size ( $N = 347$ ; converters  $n = 84$ ), which increases the risk of overfitting and high variance, potentially yielding unstable coefficients and feature importance and may lead to inflated cross-validation estimates. A direct consequence of the small sample, especially the low counts within each diagnostic subtype, is that we were unable to reliably evaluate subtype-specific prediction. Replication in larger, preferable multi-site, cohorts is warranted to obtain more precise and generalizable performance estimates. A second limitation is the absence of brain imaging data (e.g., structural MRI, fluorodeoxyglucose positron emission tomography, amyloid/tau PET), which could capture neurodegeneration and pathology more precisely. Third, we lack ambulatory assessments that track short-term fluctuation and change. Finally, although the diagnostic evaluations were conducted by approximately the same team of clinicians throughout the study period, we cannot entirely exclude the possibility of minor variations in diagnostic procedures, which could introduce subtle inconsistencies in case classification.

### Future research

While our models performed well, improving sensitivity remains a key challenge for future research. Enhancing sensitivity is critical for clinical applications, where the cost of false negatives (i.e., missed converters) can be substantial. Future efforts might explore the inclusion of additional longitudinal data (e.g., frequent ambulatory cognitive assessments via smartphone apps) to better capture early signs of decline (Nicosia et al., 2023; Polk et al., 2025; Öhman et al., 2021). High-resolution, temporally dense data may help to characterize short-term variability and early drift, which traditional assessments miss (Nesselroade, 1991). Moreover, including neuroimaging biomarkers (Ewers et al., 2011; Mohammed & Malhotra, 2025), genetic risk scores (Najar et al., 2023; de Rojas et al., 2021), and digital behavioral phenotyping (Knox et al., 2025; Kourtis et al., 2019) may further improve individualized prediction.

It is also worth noting that the CSF biomarker P-tau181 provided the strongest signal among fluid markers, perhaps because of its closer relationship to neurofibrillary tangle pathology and cognitive symptoms than A $\beta$ 42. In parallel, there is rising interest in blood-based biomarkers (Hansson et al., 2023; Palmqvist et al., 2021) which sensitively index tau accumulation and align well with established CSF and imaging measures. Because

blood sampling is substantially easier and less invasive than CSF via lumbar puncture and imaging, these assays are scalable for large longitudinal cohorts and may become especially valuable features for ML models predicting conversion from MCI to dementia. Furthermore, the dominant role of delayed recall memory tests underscores the centrality of episodic memory in the early manifestation of AD and its strong predictive utility in identifying at-risk individuals.

Lastly, our findings underscore the importance of calculating predictive values (PPV and NPV) in the context of the actual prevalence. For instance, based on our sample's dementia conversion rate (24%), the elastic net model performance translates to a PPV of 0.85 and NPV of 0.92. These values suggest that the model would be clinically informative. When it predicts future dementia, it is likely correct in 85% of cases, and when it predicts stability, it is correct in 92% of cases. However, these estimates will vary with base rates, and recalibration may be needed for different MCI populations.

## Conclusion

Taken together, our results demonstrate that ML models, especially those incorporating regularization and cognitive change scores, can accurately predict dementia conversion in MCI patients using clinical data. Key predictors include CSF P-tau181, delayed recall memory, and cognitive processing speed. Enhancing prediction, particularly sensitivity, will likely depend on the integration of repeated measures and novel data streams, moving us closer to reliable, individualized dementia risk assessment in clinical practice.

**Supplementary material.** The Supplementary Material for this article can be found at <https://doi.org/10.1017/S135561772510177X>.

**Funding statement.** This work was supported by grants from the Riksbankens Jubileumsfond (P21-0112), Konung Gustaf V:s och Drottning Victorias Frimurarestiftelse, Demensfonden, Anna-Lisa och Bror Björnssons stiftelse, Axel Linders Stiftelse, Stiftelsen Handlanden Hjalmar Svenssons forskningsfond, Stiftelsen familjen Nils Winbergs fond, Maud & Birger Gustavssons Stiftelser, Agneta Prytz-Folkes och Gösta Folkes stiftelse, Gun och Bertil Stohnes Stiftelse, Gunvor och Josef Anér's stiftelse, Stiftelsen Långmanska kulturfonden, Kungliga Vetenskaps- och Vitterhetssamhället i Göteborg, Stiftelsen Psykiatriska Forskningsfonden, Magnus Bergvalls Stiftelse, Stiftelsen för Gamla Tjänarinnor, Syskonen Svenssons Fond För Medicinsk Forskning, Greta och Einar Askers stiftelse, Wilhelm & Martina Lundgrens Vetenskapsfond, Åhlénstiftelsen, and by the Swedish state: ALFGBG-965744, ALFGBG-724331, and ALFGBG-720661.

**Competing interests.** The authors report no conflicts of interest.

## References

- Agostinho, D., Simões, M., & Castelo-Branco, M. (2024). Predicting conversion from mild cognitive impairment to Alzheimer's disease: a multimodal approach. *Brain Communications*, 6(4), 1–20.
- American Psychiatric Association. (1987). *Diagnostic and Statistical Manual of Mental Disorders*. (3rd ed., rev.). American Psychiatric Association.
- Amieva, H., Jacqmin-Gadda, H., Orgogozo, J.-M., Carret, N. L., Helmer, C., Letenneur, L., Barberger-Gateau, P., Fabrigoule, C., & Dartigues, J.-F. (2005). The 9 year cognitive decline before dementia of the Alzheimer's type: A prospective population-based study. *Brain*, 128, 1093–1101.
- Andreasen, N., Hesse, C., Davidsson, P., Minthon, L., Wallin, A., & Winblad, B. (1999). Cerebrospinal fluid beta-amyloid(1-42) in Alzheimer disease: Differences between early- and late-onset Alzheimer disease and stability during the course of disease. *Archives of Neurology*, 56, 673–680.
- Auer, S., & Reisberg, B. (1997). The GDS/FAST staging system. *International Psychogeriatrics*, 9, 167–171.
- Bandera, R., Capitani, E., & Della Sala, S. (1985). Discrimination between senile dementia Alzheimer type patients and education-matched normal controls by means of a 6-test set. *Italian Journal of Neurological Sciences*, 6, 339–344.
- Binetti, G., Cappa, S. F., Magni, E., Padovani, A., Bianchetti, A., & Trabucchi, M. (1996). Disorders of visual and spatial perception in the early stage of Alzheimer's disease. *Annals of the New York Academy of Sciences*, 777, 221–225.
- Blennow, K., Hampel, H., Weiner, M., & Zetterberg, H. (2010). Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nature Reviews Neurology*, 6, 131–144.
- Blennow, K., Ricksten, A., Prince, J. A., Brookes, A. J., Emahazion, T., Wasslavik, C., Bogdanovic, N., Andreasen, N., Båtsman, S., Marcusson, J., Nägga, K., Wallin, A., Regland, B., Olofsson, H., Hesse, C., Davidsson, P., Minthon, L., Jansson, A., Palmqvist, L., & Rymo, L. (2000). No association between the alpha2-macroglobulin (A2M) deletion and Alzheimer's disease, and no change in A2M mRNA, protein, or protein expression. *Journal of Neural Transmission*, 107, 1065–1079.
- Blennow, K., Wallin, A., Agren, H., Spenger, C., Siegfried, J., & Vanmechelen, E. (1995). Tau protein in cerebrospinal fluid: A biochemical marker for axonal degeneration in Alzheimer disease? *Molecular and Chemical Neuropathology*, 26, 231–245.
- Borchert, R. J., Azevedo, T., Badhwar, A. P., Bernal, J., Betts, M., Bruffaerts, R., Burkhart, M. C., Dewachter, I., Gellersen, H. M., Low, A., Lourida, I., Machado, L., Madan, C. R., Malpetti, M., Mejia, J., Michopoulou, S., Muñoz-Neira, C., Pepys, J., Peres, M., . . . Rittman, T. (2023). Artificial intelligence for diagnostic and prognostic neuroimaging in dementia: A systematic review. *Alzheimer's & Dementia*, 19, 5885–5904.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Chapman & Hall.
- Bucholtz, M., Titarenko, S., Ding, X., Canavan, C., & Chen, T. (2023). A hybrid machine learning approach for prediction of conversion from mild cognitive impairment to dementia. *Expert Systems with Applications*, 217, 119541.
- Chang, M., & Brainerd, C. J. (2022). Predicting conversion from mild cognitive impairment to Alzheimer's disease with multimodal latent factors. *Journal of Clinical & Experimental Neuropsychology*, 44, 316–335.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273–297.
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13, 21–27.
- Crossley, M., D'Arcy, C., & Rawson, N. S. (1997). Letter and category fluency in community-dwelling Canadian seniors: A comparison of normal participants to those with dementia of the Alzheimer or vascular type. *Journal of Clinical and Experimental Neuropsychology*, 19, 52–62.
- Cui, Y., Liu, B., Luo, S., Zhen, X., Fan, M., Liu, T., . . . Jin, J. S. (2011). Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. *PLoS ONE*, 6(7), 1–10.
- de Rojas, I., Moreno-Grau, S., Tesi, N., Grenier-Boley, B., Andrade, V., Jansen, I. E., Pedersen, N. L., Stringa, N., Zettergren, A., Hernández, I., Montreal, L., Antúnez, C., Antonell, A., Tankard, R. M., Bis, J. C., Sims, R., Bellenguez, C., Quintela, I., González-Perez, A., . . . Ruiz, A. (2021). Common variants in Alzheimer's disease and risk stratification by polygenic risk scores. *Nature Communications*, 12, 1–16.
- Della Sala, S., Baddeley, A., Papagno, C., & Spinnler, H. (1995). Dual-Task paradigm: A means to examine the central executive. In J. Grafman, K. J. Holyoak, & F. Boller (Eds.), *Structure and Functions of the Human Prefrontal Cortex* (pp. 161–171). New York Academy of Sciences.
- Erkinjuntti, T., Inzitari, D., Pantoni, L., Wallin, A., Scheltens, P., Rockwood, K., Roman, G. C., Chui, H., & Desmond, D. W. (2000). Research criteria for subcortical vascular dementia in clinical trials. *Journal of Neural Transmission Supplement*, 59, 23–30.
- Ewers, M., Sperling, R. A., Klunk, W. E., Weiner, M. W., & Hampel, H. (2011). Neuroimaging markers for the prediction and early diagnosis of Alzheimer's disease dementia. *Trends in Neurosciences*, 8, 430–442.
- Farias, S. T., Mungas, D., Reed, B. R., Harvey, D., & DeCarli, C. (2009). Progression of mild cognitive impairment to dementia in clinic- vs community-based cohorts. *Archives of Neurology*, 66, 1151–1157.

- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, 179–188.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.
- Franciotti, R., Nardini, D., Russo, M., Onofri, M., & Sensi, S. L. (2023). Comparison of machine learning-based approaches to predict conversion to Alzheimer’s disease from mild cognitive impairment. *Neuroscience*, 514, 143–152.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29, 1189–1232.
- Geffen, G. M., Butterworth, P., & Geffen, L. B. (1994). Test–retest reliability of a new form of the auditory verbal learning test (AVLT). *Archives of Clinical Neuropsychology*, 9, 303–316.
- Gianattasio, K. Z., Bennet, E. E., Wei, J., Mehrotra, M., Mosley, T., Gottesman, R. F., & Power, M. C. (2020). Generalizability of findings from a clinical sample to a community-based sample: A comparison of ADNI and ARIC. *Alzheimers & Dementia*, 17, 1265–1276.
- Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., Ogar, J. M., Rohrer, J. D., Black, S., Boeve, B. F., Manes, F., Dronkers, N. F., Vandenberghe, R., Rascofsky, K., Patterson, K., Miller, B. L., Knopman, D. S., Hodges, J. R., Mesulam, M. M., & Grossman, M. (2011). Classification of primary progressive aphasia and its variants. *Neurology*, 76, 1006–1014.
- Guo, H., Sun, S., Yang, Y., Ma, R., Wang, C., Zheng, S., Wang, X., Li, G., & for the Alzheimer’s Disease Neuroimaging Initiative the Alzheimer’s Disease Metabolomics Consortium. (2024). A novel score to predict individual risk for future Alzheimer’s disease: A longitudinal study of the ADNI cohort. *Journal of Alzheimer’s Disease*, 101, 923–936.
- Hall, C. B., Lipton, R. B., Sliwinski, M., & Stewart, W. F. (2000). A change point model for estimating the onset of cognitive decline in preclinical Alzheimer disease. *Statistics in Medicine*, 19, 1555–1566.
- Hand, D. J., & Yu, K. (2001). Idiot’s Bayes: Not so stupid after all? *International Statistical Review*, 69, 385–398.
- Hansson, O., Blennow, K., Zetterberg, H., & Dage, J. (2023). Blood biomarkers for Alzheimer 301;s disease in clinical practice and trials. *Nature Aging*, 3, 506–519.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12, 55–67.
- Howieson, D. B., Carlson, N. E., Moore, M. M., Wasserman, D., Abendroth, C. D., Payne-Murphy, J., & Kaye, J. A. (2008). The trajectory of mild cognitive impairment onset. *Journal of the International Neuropsychological Society*, 14, 192–198.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning with application in R*. (2nd ed.). Springer.
- Kaplan, E. G., & Weintraub, S. (1983). *The Boston naming test*. (2nd ed.). Lea and Febiger.
- Knox, S., Aghamoosa, S., Heider, P. M., Cutty, M., Wright, A., Scherbakov, D., & Obeid, J. S. (2025). AI approaches for phenotyping Alzheimer’s disease and related dementias using electronic health records. *Alzheimer’s & Dementia*, 11, 1–9.
- Kourtis, L., Regele, O. B., Wright, J. M., & Jones, G. B. (2019). Digital biomarkers for Alzheimer’s disease: The mobile/wearable devices opportunity. *NPI Digital Medicine*, 2, 1–9.
- Kuhn, M. (2008). Caret: Classification and regression training (R package version 7.0.1) [Computer software]. <https://CRAN.R-project.org/package=caret>.
- Massetti, N., Russo, M., Franciotti, R., Nardini, D., Mandolini, G. M., Granzotto, A., Bomba, M., Delli Pizzi, S., Mosca, A., Scherer, R., Onofri, M., Sensi, S. L., & For the Alzheimer’s Disease Neuroimaging Initiative (ADNI) the Alzheimer’s Disease Metabolomics Consortium (ADMC). (2022). A machine learning-based holistic approach to predict the clinical course of patients within the Alzheimer’s disease spectrum. *Journal of Alzheimer’s Disease*, 85, 1639–1655.
- McKeith, I. G., Perry, E. K., & Perry, R. H. (1999). Report of the second dementia with Lewy body international workshop: Diagnosis and treatment. Consortium on Dementia with Lewy Bodies. *Neurology*, 53, 902–905.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer’s disease: Report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer’s disease. *Neurology*, 34, 939–944.
- Meyers, J. E., & Meyers, K. R. (1995). *Rey Complex Figure Test and Recognition Trial*. Psychological Assessment Resources, Inc.
- Mitchell, A. J., & Shiri-Feshki, M. (2009). Rate of progression of mild cognitive impairment to dementia: Meta-analysis of 41 robust inception cohort studies. *Acta Psychiatrica Scandinavica*, 119, 252–265.
- Mohammed, S., & Malhotra, N. (2025). Predicting Alzheimer’s disease onset: A machine learning framework for early diagnosis using biomarkers data. *Computer Methods and Programs in Biomedicine Update*, 8, 1–14.
- Najar, J., Thorvaldsson, V., Kern, S., Skoog, J., Waern, M., Zetterberg, H., Blennow, K., Skoog, I., & Zettergren, A. (2023). Polygenic risk scores for Alzheimer’s disease in relation to cognitive change: A representative sample from the general population followed over 16-years. *Neurobiology of Disease*, 189, 1–7.
- Neary, D., Snowden, J. S., Gustafson, L., Passant, U., Stuss, D., Black, S., Freedman, M., Kertesz, A., Robert, P. H., Albert, M., Boone, K., Miller, B. L., Cummings, J., & Benson, D. F. (1998). Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. *Neurology*, 51, 1546–1554.
- Nelson, P. T., Alafuzoff, I., Bigio, E. H., Bouras, C., Braak, H., Cairns, N. J., Castellani, R. J., Crain, B. J., Davies, P., Tredici, K. D., Duyckaerts, C., Frosch, M. P., Haroutunian, V., Hof, P. R., Hulette, C. M., Hyman, B. T., Iwatsubo, T., Jellinger, K. A., Jicha, G. A., . . . Beach, T. G. (2012). Correlation of Alzheimer disease neuropathologic changes with cognitive status: A review of the literature. *Journal of Neuropathology & Experimental Neurology*, 71, 362–381.
- Nesselroade, J. R. (1991). The warp and the woof of the developmental fabric. In R. M. Downs, L. S. Liben, & D. S. Palermo (Eds.), *Visions of aesthetics, the environment, and development: The legacy of Joachim F. Wohlwill*. (pp. 213–240). Lawrence Erlbaum Associates.
- Nicosia, J., Aschenbrenner, A. J., Balota, D. A., Sliwinski, M. J., Tahan, M., Adams, S., Stout, S. S., Wilks, H., Gordon, B. A., Benzinger, T. L. S., Fagan, A. M., Xiong, C., Bateman, R. J., Morris, J. C., & Hassenstab, J. (2023). Unsupervised high-frequency smartphone-based cognitive assessments are reliable, valid, and feasible in older adults at risk for Alzheimer’s disease. *Journal of the International Neuropsychological Society*, 29, 459–471.
- Nordlund, A., Rolstad, S., Hellström, P., Sjögren, M., Hansen, S., & Wallin, A. (2005). The Goteborg MCI study: Mild cognitive impairment is a heterogenous condition. *Journal of Neurology and Neurosurgery Psychiatry*, 76, 1485–1490.
- Öhman, F., Hassenstab, J., Berron, D., Schöll, M., & Papp, K. V. (2021). Current advances in digital cognitive assessment for preclinical Alzheimer’s disease. *Alzheimer’s & Dementia*, 13, e12217.
- Palmqvist, S., Tideman, P., Cullen, N., Zetterberg, H., Blennow, K., Stomrud, E., Janelidze, S., Mattsson-Carlgen, N., Hansson, O., & The Alzheimer’s Disease Neuroimaging Initiative. (2021). Prediction of future Alzheimer’s disease dementia using plasma phospho-tau combined with other accessible measures. *Nature Medicine*, 27, 1034–1042.
- Palmqvist, S., Zetterberg, H., Blennow, K., Vestberg, S., Andreasson, U., Brooks, D. J., & Hansson, O. (2014). Accuracy of brain amyloid detection in clinical practice using cerebrospinal fluid beta-amyloid 42: A cross-validation study against amyloid positron emission tomography. *JAMA Neurology*, 71, 1282–1289.
- Paolo, A. M., Axelrod, B. N., Tröster, A. I., Blackwell, K. T., & Koller, W. C. (1996). Utility of a Wisconsin card sorting test short form in persons with Alzheimer’s and Parkinson’s disease. *Journal of Clinical and Experimental Neuropsychology*, 18, 892–897.
- Petersen, R. C., Lopez, O., Armstrong, M. J., Getchius, T. S. D., Ganguli, M., Gloss, D., & Rae-Grant, A. (2018). Practice guideline update summary: Mild cognitive impairment: Report of the guideline development, dissemination, and implementation subcommittee of the American academy of neurology. *Neurology*, 90, 126–135.
- Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology*, 56, 303–308.
- Polk, S. E., Öhman, F., Hassenstab, J., König, A., Papp, K. V., Schöll, M., & Berron, D. (2025). A scoping review of remote and unsupervised digital

- cognitive assessments in preclinical alzheimer's disease. *NPJ Digital Medicine*, 8, 1–14.
- Reitan, R. M., & Wolfson, D. (1985). *The Halstead-Reitan neuropsychological test battery*. Neuropsychology Press.
- Román, G. C., Tatemichi, T. K., Erkinjuntti, T., Cummings, J. L., Masdeu, J. C., Garcia, J. H., & Scheinberg, P. (1993). Vascular dementia: Diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology*, 43, 250–260.
- Rydén, L., Grant, P. J., Berne, C., Cosentino, F., Danchin, N., Deaton, C., Javier, E., Hans-Peter, H., Heikki, H., Michel, M., Nikolaus, M., Linda, M., Jan, O., Carlo, P., Petar, S., Miguel, S. U., Marja-Riita, T., Michal, T., . . . Xuereb, R. G. (2013). ESC guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD: The task force on diabetes, pre-diabetes, and cardiovascular diseases of the European Society of Cardiology (ESC) and developed in collaboration with the European Association for the Study of Diabetes (EASD). *European Heart Journal*, 34, 3035–3087.
- Salemme, S., Lombardo, F. L., Lacorte, E., Sciancalepore, F., Remoli, G., Bacigalupo, I., Piscopo, P., Zamboni, G., Rossini, P. M., Cappa, S. F., Perani, D., Spadin, P., Tagliavini, F., Vanacore, N., & Ancidoni, A. (2025). The prognosis of mild cognitive impairment: A systematic review and meta-analysis. *Alzheimer's & Dementia*, 17, 1–15.
- Shallice, T., & Evans, M. E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex*, 14, 294–303.
- Snitz, B. E., Wang, T., Cloonan, Y. K., Jacobsen, E., Chang, C.-C. H., Huges, T. F., Kamboh, M. I., & Ganguli, M. (2018). Risk of progression from subjective cognitive decline to mild cognitive impairment. The role of study setting. *Alzheimer's & Dementia*, 14, 734–742.
- Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests*. (2nd ed.). Oxford University Press.
- Thorvaldsson, V., MacDonald, S. W. S., Fratiglioni, L., Winblad, B., Kivipelto, M., Laukka, E. J., Skoog, I., Sacuiu, S., Guo, X., Östling, S., Börjesson-Hanson, A., Gustafson, D., Johansson, B., & Bäckman, L. (2011). Onset and rate of cognitive change before dementia diagnosis: Findings from two Swedish population-based longitudinal studies. *Journal of the International Neuropsychological Society*, 17, 154–162.
- Thorvaldsson, V., Nordlund, A., Reinvang, I., Blennow, K., Zetterberg, H., Wallin, A., & Johansson, B. (2010). Memory in individuals with mild cognitive impairment in relation to APOE and CSF A $\beta$ <sub>42</sub>. *International Psychogeriatrics*, 22, 598–606.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 58, 267–288.
- Valsdóttir, V., Jónsdóttir, M. K., Magnúsdóttir, B. B., Chang, M., Hu, Y.-H., Gudnason, V., Launer, L. J., & Stefánsson, H. (2024). Comparative study of machine learning methods for modeling associations between risk factors and future dementia cases. *GeroScience*, 46, 737–750.
- Vanmechelen, E., Vanderstichele, H., Davidsson, P., Van Kerschaver, E., Van Der Perre, B., Sjögren, M., Andreasen, N., & Blennow, K. (2000). Quantification of tau phosphorylated at threonine 181 in human cerebrospinal fluid: A sandwich ELISA with a synthetic phosphopeptide for standardization. *Neuroscience Letters*, 285, 49–52.
- Wallin, A., Nordlund, A., Jonsson, M., Lind, K., Edman, Å., Göthlin, M., Stålhammar, J., Eckerström, M., Kern, S., Börjesson-Hanson, A., Carlsson, M., Olsson, E., Zetterberg, H., Blennow, K., Svensson, J., Öhrfelt, A., Bjerke, M., Rolstad, S., Eckerström, C. (2016). The Gothenburg MCI study: Design and distribution of Alzheimer's disease and subcortical vascular disease diagnoses from baseline to 6-year follow-up. *Journal of Cerebral Blood Flow Metabolism*, 36, 114–131.
- Wechsler, D. (1981). *WAIS-R manual. The Psychological Corporation*.
- Wechsler, D. (1987). *Wechsler memory scale-revised manual*. The Psychological Corporation.
- Wilson, R. S., Beckett, L. A., Bennett, D. A., Albert, M. S., & Evans, D. A. (1999). Change in cognitive function in older persons from a community population: Relation to age and Alzheimer disease. *Archives of Neurology*, 56, 1274–1279.
- Wimo, A., Seeher, K., Cataldi, R., Cyhlarova, E., Dielemann, J. L., Frisell, O., Guerchet, M., Jönsson, L., Malaha, A. K., Nichols, E., Pedroza, P., Prince, M., Knapp, M., & Dua, T. (2023). The worldwide costs of dementia in 2019. *Alzheimer's & Dementia*, 19, 2865–2873.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67, 301–320.