

RESEARCH ARTICLE

Interpretable multivariate survival models: Improving predictions for conversion from mild cognitive impairment to Alzheimer's disease via data fusion and machine learning

Diana Sofia Rosales-Gurmendi¹, Gerardo Alejandro Fumagal-González^{1*}, Jorge Orozco¹, Joshua Farber², Victor Treviño³, Emmanuel Martinez-Ledesma^{3,4}, Antonio Martinez-Torteya⁵, Fabiola Rosales-Gurmendi⁶, Jose Tamez-Peña³, for the Alzheimer's Disease Neuroimaging Initiative[†]

1 School of Engineering and Sciences, Tecnológico de Monterrey, Monterrey, Mexico, **2** Qmetrics Technologies LLC, Rochester, New York, United States of America, **3** School of Medicine and Health Sciences, Tecnológico de Monterrey, Monterrey, Mexico, **4** The Institute for Obesity Research, Tecnológico de Monterrey, Monterrey, Mexico, **5** Engineering Department, Universidad de Monterrey, Monterrey, Mexico, **6** School of Medicine, Universidad Peruana Cayetano Heredia, Lima, Peru

[†] Membership of the Alzheimer's Disease Neuroimaging Initiative is listed in the Acknowledgments.

* gerardofumagalgz@gmail.com



OPEN ACCESS

Citation: Rosales-Gurmendi DS, Fumagal-González GA, Orozco J, Farber J, Treviño V, Martinez-Ledesma E, et al. (2026) Interpretable multivariate survival models: Improving predictions for conversion from mild cognitive impairment to Alzheimer's disease via data fusion and machine learning. *PLoS One* 21(4): e0321671. <https://doi.org/10.1371/journal.pone.0321671>

Editor: Moeko Noguchi-Shinohara, Kanazawa university, JAPAN

Received: March 8, 2025

Accepted: March 18, 2026

Published: April 30, 2026

Copyright: © 2026 Rosales-Gurmendi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All relevant data are within the paper and its [Supporting Information](#) files. Data used in this study were

Abstract

Accurately predicting which individuals with mild cognitive impairment (MCI) will progress to Alzheimer's disease (AD) can improve patient care. This study examines the role of quantitative MRI (qMRI), cognitive evaluations, apolipoprotein $\epsilon 4$ (*APOE* $\epsilon 4$), and cerebrospinal fluid (CSF) biomarkers in Cox survival models to predict progression from MCI to AD. Data from 564 participants in the ADNI study, who transitioned from MCI to AD, were analyzed. The data set included 330 features encompassing qMRI, cognitive assessments, CSF biomarkers, and *APOE* $\epsilon 4$ status. Advanced machine learning (ML) methods were applied to evaluate the importance of these data sources, select relevant features, and develop interpretable Cox survival models within a cross-validation framework. The top optimized model achieved a sensitivity of 0.69, 95% CI [0.63, 0.76], and a specificity of 0.87, 95% CI [0.83, 0.90], and used all data sources. The results demonstrated that combining qMRI features with cognitive assessments, CSF biomarkers, and *APOE* $\epsilon 4$ status, analyzed using the BSWiMS model, resulted in a substantial improvement in the ability to predict progression from MCI to AD, achieving 81% precision and 87% specificity. These results exceed those obtained with other models evaluated. Finally, biomarker analysis showed that cognitive scores are the most relevant features to predict conversion, followed by CSF and qMRI biomarkers. These findings highlight the value of integrating multiple data sources in highly interpretable Cox survival models for the early identification of individuals at risk for AD.

obtained from the TADPOLE challenge dataset (<https://tadpole.grand-challenge.org>). This dataset has been made publicly available and anonymized since 16 June 2017, and it was derived from the ADNI database (<adni.loni.usc.edu>). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. ADNI is a cross-sectional and longitudinal follow-up observational study. The main goal of ADNI has been to see how neuroimaging, cognitive tests, fluid, and genetic biomarkers can be used together to figure out how MCI and early AD will progress. For up-to-date information, see www.adni-info.org. Furthermore, we have set the baseline data and the reduced data we used in the study on the following GitHub repository: <https://github.com/joseTamezPena/SurvivalTadpole/tree/main> csv of the data can be found under the folder "Data".

Funding: This research was supported by the Secretaría de Ciencia, Humanidades, Tecnología e Innovación (Secihti), with cloud computing resources provided through Microsoft's AI for Good Lab. In addition, the work was also partially supported by Secretaría de Educación Superior, Ciencia, Tecnología e Innovación, part of Gobierno de la República del Ecuador, and by the Strategic Research Group of Bioinformatics for Clinical Diagnosis from Tecnológico de Monterrey.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Alzheimer's disease (AD) is one of the most common cognitive disorders in old age [1]. The development of effective treatments or disease-modifying therapies is hampered by the complexity of aging and the lack of a clear understanding of the etiology and pathogenesis of AD [2]. The diagnosis of AD in the early stages of the disease is complex. Despite their mostly distinct pathophysiological features, these conditions are often misdiagnosed antemortem due to their overlapping cognitive dysfunction symptoms. Definitive diagnosis of clinical AD dementia is only possible through a postmortem neuropathological examination, given the current absence of accurate and reliable antemortem biomarkers [3–5]. The most accurate diagnostic test for AD requires a histopathological evaluation of brain tissue and can only be obtained through autopsy or biopsy. Without a biopsy, the diagnosis of a normal patient is defined as possible or probable AD according to patient reports, cognitive observation, and symptomatology [5]. Historically, AD was only diagnosed postmortem [4]; though research institutes capable of assessing amyloid and tau burden in living patients are challenging this historic paradigm [3,6].

Hence, understanding the process and each of the stages of AD is essential to developing effective treatments. It is noteworthy that mild cognitive impairment (MCI) is regarded as a transitional stage between normal aging and AD [7]. Therefore, in the context of the conversion of MCI to AD, some biomarkers have been identified in the AD literature. The Amyloid Beta ($A\beta$) and tau proteins from cerebrospinal fluid (CSF) [8,9], the scores in cognitive assessments [10], and the polymorphism of the fourth allele of the apolipoprotein E ($APOE \epsilon 4$) [11] have been recognized as validated risk factors for the conversion. On the other hand, some studies have shown that clinical information and imaging biomarkers can be used to predict patients who will undergo conversion from those who will not [4]. Imaging biomarkers related to AD have been found in positron-emitting tomography (PET) and magnetic resonance imaging (MRI) have a clear association with the evolution and presence of AD [12] and with the conversion of MCI to AD [13]. However, the impact of imaging biomarkers to determine the conversion rate between MCI to AD, as well as the relationship and the importance of the MRI features in survival models, has not been fully connected in studies.

Studies indicate that CSF biomarkers are increasingly being used to support the diagnosis of AD, especially to determine the difference between AD from non-AD dementia [14–16]. Different biomarkers can be extracted from CSF. The most established in the usual clinical practice include Amyloid Beta 1–42 ($A\beta$ 1–42), total tau (t-tau), phosphorylated tau in threonine 181 (p-tau) and in recent studies, phosphorylated tau in threonine 217 has been found to be a biomarker that outperforms p-tau 181 on the conversion diagnosis from MCI to AD. [14,17,18]. All of them have been previously described in AD pathogenesis. Despite their role in AD pathogenesis the cerebrospinal fluid biomarkers have an important utility for $A\beta$ -42/ β -40 and T-tau/ $A\beta$ 1–42 ratios, but limited specificity for distinguishing AD from Dementia with Lewy Bodies (DLB) and Progressive Nonfluent Aphasia (PNFA) [19,20].

Alternatively, the symptomatic AD process alters brain function; henceforth, cognitive assessments are usually the starting point for the brain health plan and are recommended for all patients who want to be screened for AD [21,22]. These assessments evaluate some important areas of brain function, memory and reasoning capacity, concentration, processing speed, and language. Depending on the country, language, or type of test, several experts have developed and validated standard tests such as those used in this work. Some authors have updated and proposed changes in the normal cognitive assessments to better classify or predict the conversion in patients, [23,24] and others proposed the combination of patterns of brain atrophy and cognitive assessment scores, the latter offering the highest predictive power for the conversion [25].

Other tests available for the understanding of the AD process are MRI and PET. These imaging modalities can visualize direct changes in brain structure, function, sugar metabolism, and AD build-up [26,27]. Therefore, AD-related imaging biomarkers with a clear association on the early development and presence of AD have been discovered in MRI and PET [26]. Furthermore, imaging biomarkers have been associated with the conversion from MCI to AD [28]. Consequently, MRI and PET are commonly used to monitor the progression of the disease and to detect the current stage of neuronal degeneration [29].

Lastly, the onset of AD requires a clear description of its risk factors. Age and gender are established risk factors for developing AD. Older people are at higher risk, while women are more prone than men to developing AD. *APOE* ϵ 4, the polymorphic allele of the apolipoprotein ϵ , is a known genetic risk factor for developing AD [29]. These known factors must be considered when developing screening or staging AD tests. Therefore, researchers have proposed the combination of risk factors, cognitive tests, MRI, and PET features with CSF biomarkers to improve the diagnostic accuracy of AD and to improve the predictive power of MCI to AD conversion [8].

The development of a good screening test requires the identification, characterization, and validation of biomarkers associated with AD and requires a good understanding of their evolution as well as their role in the disease process. Although there is plenty of research describing the association between biomarkers at the MCI stage and probable AD conversion, their discoveries have not effectively evaluated the time required from MCI to AD conversion [29–31]. This evaluation is important because some biomarkers may be associated with a slow conversion process (low-risk markers) while others with a swift conversion (high-risk markers) [32]. In this context, multivariate Cox regression models are a statistical tool that can be used to screen out low-risk markers vs. high-risk markers. Cox modeling incorporates the time to event in its fitting process and provides estimates of the hazard ratios (HR) of each potential biomarker. This feature of Cox modeling can be used to improve the understanding of biomarkers associated with the AD process. In addition, Cox proportional hazard has been widely used in survival studies [33].

The challenge of Cox modeling in biomarker discovery is that thousands of potential biomarkers can be associated with the risk of conversion. Traditional approaches use hypothesis-based feature selection; thus, findings have been limited to a small set of biomarkers [34]. Alternatives to the traditional approach are subset selection and regularization. Subset selection is based on computer algorithms that attempt to identify the best set of features associated with the disease process. Regularization methods use all the available features to estimate the total multivariate risk, and they solve the ill-posed problem by adding heuristic constraints. Statistical learning (SL) and machine learning (ML) strategies provide efficient and highly competitive regularization and subset selection methods. Embedded SL approaches like L1 regularization via LASSO or L2 regularization through RIDGE allow the exploration of multivariate models composed of hundreds of features [35]. L1 regularization via the LASSO also allows subset selection [36]. Bootstrap Step-Wise Model Selection (BSWiMS), golden section primal-dual active set (GPDAS), and sequential primal-dual active set (SPDAS) are other subset selection methods readily available to researchers [36,37]. Finally, researchers usually rely on traditional statistical methods controlled for false discovery rate (FDR) for feature selection (FS) [38,39].

The wide variety of methods available to researchers can make biomarker discovery a complex effort, especially when there is no clear choice of methodology for building/exploring survival models. To overcome this limitation, we propose a

unified approach for the study of Cox models in an ML setting. The approach is based on repeated cross-validating ML methods using the same training-testing sets across all the methods. The implementation evaluates LASSO, RIDGE, BSWIMS, GPDAS, SPDAS and FDR-controlled univariate filtering for building suitable survival models [40]. Thus, the result of the approach is a fair method comparison and a comprehensive evaluation of the role of each potential biomarker inside a Cox survival model. The primary goal of this paper is to provide a strict comparison of Cox models constructed from multivariate multi-data sources (imaging, risk factors, CSF, and cognitive assessments) that describe the MCI to AD conversion risk for each subject in a unified approach. Additionally, this study aims to highlight the relevance of the multiple biomarkers that are present in the conversion process of MCI to AD. Furthermore, a key objective is to provide a unique consensus-based Cox model that can be used to accurately predict the chances of conversion of MCI-diagnosed subjects. Subsequent sections present the data preparation, the utility of the unified approach for the comparison of ML models, and the role of the top biomarkers associated with MCI to AD conversion. A list of abbreviations and its definitions can be found at [S1 File](#) Supplementary Material.

Materials and methods

The ADNI/TADPOLE challenge datasets considered for this study were “D1—a comprehensive longitudinal data set for training,” and “D2—a comprehensive longitudinal data set on rollover subjects for forecasting”. The challenge included 1,737 individuals from the ADNI database with longitudinal observations. Each subject’s data included the diagnosis status, cognitive assessments, CSF biomarkers, qMRI, PET data and *APOE* ϵ 4 status. Detailed information regarding the rationale and the features contained in the TADPOLE challenge [39]. Github repository for data and experimentation can be found at

Ethics considerations

As mentioned earlier, data used in this study is publicly available and therefore no additional ethical approval was needed from an ethics committee; however, written informed consent for ADNI participants was obtained by the ADNI in accordance with the local legislation and ADNI requirements. ADNI studies follow Good Clinical Practices guidelines, the Declaration of Helsinki, and United States regulations (U.S. 21 CFR Part 50 and Part 56).

Inclusivity in global research

Additional information regarding the ethical, cultural, and scientific considerations specific to inclusivity in global research is included in the Supporting Information [S2 Checklist](#).

Subjects

The general eligibility, inclusion, and exclusion criteria used in this study are summarized in [Fig 1](#). From a total of 1,737 individuals with a baseline diagnosis of MCI, who were common among the TADPOLE individuals ($\gg 30\%$), 622 of these subjects were further excluded due to not having complete observation baseline data; in that sense, only 1,115 subjects met criteria for inclusion as part of the MCI group recruited for ADNI. Likewise, 551 of these subjects were excluded for having the condition of non-convert (NC) and Alzheimer’s diagnosis. Finally, 564 MCI subjects were included, with 191 subjects converted and 373 non-converted to assess the difference between groups in the study and to evaluate the importance of features.

Clinical data

We considered the main features as potential predictors of MCI to dementia conversion in our analysis; therefore, we divided the features into three major groups. The features detailing levels of three different proteins, $A\beta$ 1–42, t-tau, and p-tau, were included and studied in a group labeled as CSF features. It should be noted that in the ADNI database version used for this study, the phosphorylated tau biomarker that is recorded is pTau181, measured in CSF and, in

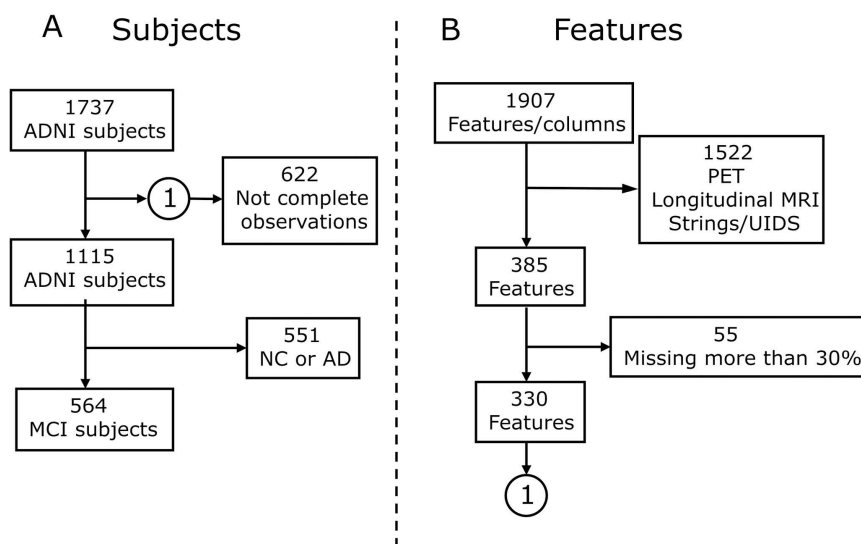


Fig 1. Data inclusion and exclusion diagram. (A) Subject selection, a baseline of 1737 subjects from ADNI; 622 subjects with missing data or few observations were excluded and 564 subjects with MCI and complete observations were included. (B) Feature selection, a total of 1907 features were identified from ADNI-TADPOLE. Of these, 1522 PET and longitudinal analysis features with FreeSurfer were excluded, along with features containing missing data, leaving 330 features per patient.

<https://doi.org/10.1371/journal.pone.0321671.g001>

recent phases, also in plasma. The biomarker pTau217 is not available because its validation as a sensitive indicator is more recent, and its inclusion has not yet been adopted in public cohorts such as ADNI. These measures were obtained during baseline evaluation at the University of Pennsylvania Medical Center. On the other hand, we included information from several neuropsychological tests, labeling the entire group of features as cognitive assessment features. Due to the nature of the disease, an ADNI examiner interviewed the patient to determine all the cognitive assessment scores. A complete description of the assessment acquisition is found in the ADNI manual. Likewise, five assessments were used: Clinical Dementia Rating Sum of Boxes (CDRSB) [40], Alzheimer’s Disease Assessment Scale (ADAS) [41], Mini-Mental State Examination (MMSE) [42], Rey’s Auditory Verbal Learning Test (RAVLT) [43], and Functional Assessment Questionnaire (FAQ) [44]. The 346 qMRI measurements provided by the University of California San Francisco (UCSF) were included and labeled as Radiomic features. Each MRI dataset was post-processed using FreeSurfer v4.3, a processing program with the function of (a) automated model-based reconstruction of the brain’s cortical surface and subcortical structures and (b) morphometric analysis. Finally, variables such as gender and the *APOE* ϵ 4 biomarker status were also included; for the latter, it was considered that it had at least one allele ϵ 4. [Table 1](#) shows the descriptive characteristics of the main features of the dataset. A full table of names and definitions of the TADPOLE MCI features can be found at

Based on the previously mentioned data, the methodology process can be outlined in [Fig 2](#). In which it begins with the inclusion and exclusion of data from the ADNI-TADPOLE set. This is followed by a rigorous preprocessing phase focusing on individuals with MCI, aimed at harmonizing the diverse data sources. Subsequently, the process transitions into a comprehensive evaluation of machine learning models, comparing their predictive performance to determine the most accurate configurations for identify MCI to AD conversion. To complement these analyses, six Cox regression models, adjusted for survival analyses, are compared to evaluate their capacity for predicting progression over time. Each stage of the workflow incorporates specific procedures, such as cross-validation, to minimize inter-cohort bias and ensure the validity and reliability of the results. This integrative approach not only strengthens the reproducibility of findings but also emphasizes their clinical relevance, providing actionable insights into the progression of AD.

Table 1. Demographic data with features sectioned by non-converter and converter groups.

Feature	Units/Scale	No Converter (373)	Converter (191)
AGE (years)	years	71.93 (7.65)	73.53 (7.07)**
Gender (M/F)	Male / Female	217 — 156	119 — 72
APOE ε4	categoric	218 — 155	67 — 124***
MMSE score	0 - 30	28.11 (1.69)	26.96 (1.79)***
FAQ	0 - 30	1.99 (3.19)	5.24 (4.68)***
ADAS11 (cognitive score)	0 - 70	8.52 (3.82)	12.93 (4.64)***
ADAS13 (cognitive score)	0 - 85	13.70 (5.79)	20.87 (6.22)***
RAVLT immediate	# words	37.69 (10.71)	28.59 (7.20)***
RAVLT % forgetting	# words (%)	52.82 (30.46)	76.12 (29.11)***
RAVLT learning	# words	4.80 (2.53)	3.07 (2.31)***
RAVLT forgetting	# words	5.08 (2.43)	5.08 (2.97)
ABETA	(pg/mL)	1154.23 (580.01)	724.01 (325.90)***
TAU	(pg/mL)	256.44 (117.93)	342.46 (133.00)***
PTAU	(pg/mL)	24.27 (13.15)	34.33 (15.10)***
ICV	mm	115.15 (3.95)	115.65 (3.48)
Whole Brain	mm	884.40 (21.25)	868.01 (20.44)***
Ventricles	mm	280.03 (49.43)	296.33 (46.40)***

* p<0.05, ** p<0.01, *** p<0.001.

<https://doi.org/10.1371/journal.pone.0321671.t001>

Data conditioning and pre-processing

We computed the time to conversion from the provided data. The event time for subjects that converted consisted of the difference in days between the date of their first AD diagnosis and the baseline date. The event time for stable MCI subjects consisted of the difference in days between the date of the baseline and the date of the last recorded follow-up visit. MCI-stable subjects were labeled as censored. We pre-processed the radiomic features as follows: We compute the cubic root of all qMRI volumes and the square root of all qMRI areas. All qMRI data was normalized by dividing each one by intracranial volume. After that, the qMRI measures of the left and right sides of the brain were described by the mean, absolute, and relative differences. Hence, these last features can be used to check dementia issues due to brain asymmetry.

Cox modeling via machine learning subset selection

To explore the subset of features and its association with the MCI to AD conversion, we used machine learning to train Cox regression models. Cox models explore the relationship between the time of the event and the possible explanatory variables. The model estimates the hazard, λ_i , of the subject i given the observed feature vector $X_i = \{X_{i1}, \dots, X_{ip}\}$, and the unknown baseline hazard $\lambda(t_0)$, i.e., given by:

$$\lambda_i(t|X_i) = \lambda(t_0)e^{X_i \cdot \beta} \tag{1}$$

where $\beta = \{\beta_1, \dots, \beta_p\}$ is the vector of coefficients. The fitting is commonly performed by a maximum likelihood estimation method providing the β values and relative hazard ratios. Thus, the Cox model provides an estimate of the total hazard or risk of conversion, given the observed features for an individual. Due to the large set of possible qMRI features to be

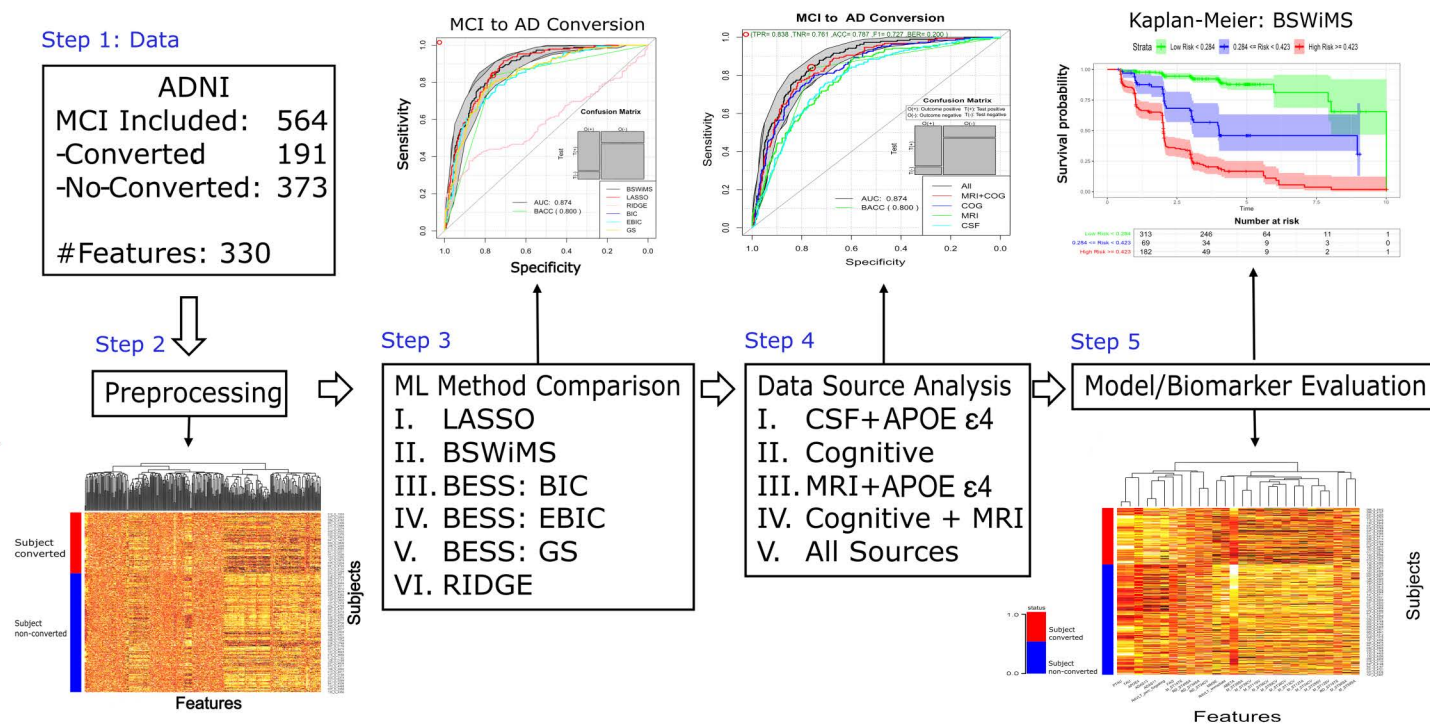


Fig 2. Diagram on the process of analysis for the discovered biomarkers and data fusion (CSF: cerebrospinal fluid, MRI: magnetic resonance imaging, Cog: cognitive, APOE ϵ 4 (apolipoprotein E4)) by cross-validation and Machine Learning methods comparison evaluating the performance of 6 Cox models.

<https://doi.org/10.1371/journal.pone.0321671.g002>

considered in some of the Cox models, ML methods were used to find an optimal set of features and their corresponding coefficients that mimicked the observed rate of conversion.

There are several strategies for building Cox models [35,45–48]. In this paper, we evaluated three open-source ML packages. These packages provided us with different strategies for building Cox models. The first ML method was BSWiMS, part of the FRESA.CAD R package [35], is a supervised model-selection method aimed at generating a unique statistical model that predicts a user-specified outcome, in this case, a survival outcome. The statistical model is constructed by bagging a set of Cox models, where each Cox model is composed of a set of model-wise statistically significant features [46]. The second evaluated package was the glmnet R package that implements the Penalized Cox Regression (CoxNet) algorithm. CoxNet fits the Cox model regularized by L1 or L2 or a mixture of them using the elastic net penalty [49]. We executed CoxNet with their provided 10-fold cross-validation function (cv.glmnet) to determine the optimal weight of the regularization constraint. Furthermore, we ran cv.glmnet with two alpha values: alpha = 1 (feature selection with LASSO regularization) and alpha = 0 (ridge or L2 regularization). The third evaluated package was the BeSS R package. This package implements the Golden Section Primal-Dual Active Set algorithm (GSPDAS), aiming for the selection of the best set of features of the Cox model. Like CoxNet, BeSS can use different strategies for subset selection. The default configuration uses the GSPDAS algorithm. The second BeSS option is SPDAS based on the Bayesian information criterion (BIC). The third option runs the SPDAS algorithm with the EBIC. In summary, we evaluated six different Cox models: BSWiMS, LASSO, RIDGE, BESS:BIC, BESS:EBIC, and BESS:GS.

Cox model validation and evaluation

The main aim of this paper was the comprehensive evaluation of the Cox models for the prediction of the MCI to AD conversion. To ensure a fair comparison, we employed a repeated holdout cross-validation (RHOCV) approach across all machine learning (ML) strategies [50]. This consistent evaluation framework was applied to each set of features provided by the TADPOLE challenge (Fig 2). The test results of the RHOCV were used to compare and explore the performance of the ML alternatives. The RHOCV strategy is part of the FRESA.CAD R package. The RHOCV method creates multiple sets of training and testing sets. At each interaction, the input data is randomly divided into a training and testing set. The training set is used for the model or feature selection, while the holdout set is predicted by the trained method. Once all the holdout predictions are generated, the test results are evaluated and compared between ML strategies. The RHOCV implementation uses the Survival R package to calculate the final Cox predictions of each model.

The estimated coefficients β^j on each training set T_j were used to get the subject i linear predictions f_i^j of the holdout set T_j^c at each repetition:

$$f_i^j = X_i \cdot \beta^j, \quad \forall i \in T_j^c \text{ and } \forall X_i \in A_j \tag{2}$$

Once the linear predictions were obtained for each repetition, the test results were ensemble by computing the median prediction for each subject as $\tilde{f}_i = \text{median}(\{f_i^1, \dots, f_i^R\})$. The ensemble prediction was used to divide the subjects into high-risk (HR: $\tilde{f} \geq t_o$) and low-risk (LR: $\tilde{f} < t_o$) groups, where t_o is the decision threshold. We assumed that censored subjects belong to the low-risk group, while true MCI to AD conversions are in the high-risk group. The receiver operating characteristic (ROC) plots and their area under the curve (AUC) with their corresponding 95% confidence intervals (CI) were computed for the risk prediction using the pROC package [51]. Accuracy (ACC), sensitivity (SEN), and specificity (SPE), describing the ability of the Cox models to predict censored vs. uncensored subjects, were computed based on the number of true positives (TP) and true negatives (TN) given by:

$$TP = |(\tilde{f} \geq t_o) \cap \text{uncensored}| \tag{3}$$

$$TN = |(\tilde{f} < t_o) \cap \text{censored}| \tag{4}$$

$$ACC = \frac{TP + TN}{|\text{uncensored} + \text{censored}|} \tag{5}$$

$$SEN = \frac{TP}{|\text{uncensored}|} \tag{6}$$

$$SPE = \frac{TN}{|\text{censored}|} \tag{7}$$

The survival performance was evaluated using the concordance index (c-index). The c-index measures the fraction of all order pairs of subjects ϵ_{ij} whose predicted survival times are correctly ordered among all subjects that can be ordered. It can be written as:

$$c = \frac{1}{|\epsilon|} \sum_{t_i \in \text{uncensored}} \sum_{t_j > t_i} 1_{\tilde{t}_i < \tilde{t}_j} \tag{8}$$

where the indicator function $1_{a < b} = 1$ if $a < b$, and 0 otherwise, $|e|$ is the number of ordered pairs. \tilde{t}_i is the median of the predicted survival time and t_i is the actual observed time of the uncensored subject i . The values of the c-index range from 0 to 1, where 1 implies a perfect concordance between observed and predicted times.

We evaluated the prediction benefit using decision curve analysis (DCA). The DCA curves indicated a net benefit of the predicted probability of the models. The analysis of the predicted probability was used to estimate the high-risk threshold at 90% specificity and the middle-risk threshold at 80% specificity. The 90% threshold was then used to evaluate the classification performance of the models.

The visualization of the predicted survival groups, high-risk vs. mid-risk and low-risk, was done using Kaplan-Meier (KM) plots of the `survminer` R package [52]. The statistical significance of the difference between the survival groups was evaluated by the log-rank test [53] given by:

$$Z = \frac{\sum_{i=1}^N (O_{HRI} - E_{HRI})}{\sqrt{\sum_{i=1}^N V_i}} \xrightarrow{d} N(0, 1), \tag{9}$$

and

$$V_i = \frac{O_i \left(\frac{N_{HRI}}{N_i} \right) \left(1 - \frac{N_{HRI}}{N_i} \right) (N_i - O_i)}{N_i - 1}, \tag{10}$$

where O_i is the actual number of events, N_i is the number of subjects below rank, and N_{HRI} is the number of subjects at high risk.

Feature source and Cox-models

To evaluate the predictive performance of the Cox models, we conducted a series of experiments using the Recursive Hyperparameter Optimization with Cross-Validation (RHOCV) procedure, repeated 50 times for robustness, similar to the work by [54]. Each run allocated 70% of the subjects for training and 30% for testing, and the models were constructed by selecting features dynamically in each iteration. This approach enabled the identification of the most frequently selected features across experiments, providing insight into their relative importance. The experiments were designed to assess the impact of different data sources on the models. Five independent experiments were performed: Experiment 1 used CSF and *APOE* ϵ 4 features. Experiment 2 studied six Cognitive Assessment features. Experiment 3 constructed models from the MRI and *APOE* ϵ 4. Experiment 4 used cognitive assessment and MRI. Experiment 5 used all the features from CSF, cognitive assessment, MRI, and *APOE* ϵ 4 as part of the feature set. Each experiment consisted of the independent execution of the RHOCV procedure 50 times. Each run used 70% of the subjects for training while 30% was used for testing. The procedure builds models selecting features in each run; hence the analysis plan included the exploration of the common top features discovered in each experiment.

Biomarker evaluation via model feature analysis

The study evaluates the performance of Cox models, which notably do not require harmonized data, with detailed results presented in Table 2. In a subsequent experiment, outlined in the subsection “Biomarker evaluation via model feature analysis,” we explored the hazard ratios of the top Cox model features associated with the conversion from MCI to AD. This experiment involved splitting the data into training (70%) and testing (30%) sets, followed by z-standardization of all features using the training set to estimate the mean and variance of continuous features, with z-scores subsequently applied to the testing set. To ensure robust results and minimize the effect of outliers, the

Table 2. ML subset selection method comparison.

Method	ROC AUC	c-Index	Accuracy	Sensitivity	Specificity
BSWiMS	0.87 [0.85,0.90]	0.85 [0.82,0.87]	0.81 [0.78,0.84]	0.69 [0.63,0.76]	0.87 [0.83,0.90]
LASSO	0.88 [0.85,0.90]	0.85 [0.82,0.87]	0.79 [0.75,0.82]	0.64 [0.57,0.71]	0.87 [0.83,0.90]
BESS:BIC	0.83 [0.80,0.87]	0.82 [0.79,0.84]	0.79 [0.75,0.82]	0.63 [0.56,0.70]	0.86 [0.83,0.90]
BESS:EBIC	0.83 [0.80,0.87]	0.82 [0.79,0.84]	0.79 [0.75,0.82]	0.63 [0.56,0.70]	0.86 [0.83,0.90]
BESS:GS	0.84 [0.80,0.87]	0.82 [0.79,0.85]	0.79 [0.76,0.82]	0.63 [0.56,0.70]	0.87 [0.83,0.90]
RIDGE	0.54 [0.49,0.58]	0.57 [0.53,0.61]	0.56 [0.52,0.60]	0.48 [0.41,0.55]	0.61 [0.56,0.66]

<https://doi.org/10.1371/journal.pone.0321671.t002>

freescale method was employed, where the mean was subtracted from each variable and divided by the standard deviation. Additionally, the BSWiMS procedure was run for 20 Cox bootstrap estimations using the training set, with performance evaluated on the testing set. Continuous variables were transformed into z-scores, while dichotomous variables remained unchanged, facilitating the generation of [Table 4](#). The summary method of BSWiMS analyzed the hazard ratios of each fitted model, reporting their 95% confidence intervals for the top features associated with conversion, as detailed in [Table 4](#).

Statistical analysis and software

This subsection is a summary of the statistical analysis and software employed throughout this study.

The statistical software used was R. The statistical analysis can be summarized as follows:

- Cox models building: 6 Cox models, built from 3 open-source ML packages.
 - BSWiMS (via FRESA.CAD R package). Bagging a set of Cox models.
 - CoxNet (via cv.glmnet R package). Fitting of 2 different Cox models with a 10-fold cross-validation each, with alpha value of 1 for LASSO regularization and alpha value of 0 for ridge regularization.
 - GSPDAS (via BeSS R package). Aims for the selection of best set of features. 3 Cox models built by using 3 different configurations, GS, BIC and EBIC.
- Performance metrics and validation on Cox models
 - RHOCV (via FRESA.CAD R package). This ensures a fair comparison.
 - ROC plots, AUC and CI (via pROC R package). CI of 95%.
 - DCA and KM (via survminer R package) used to indicate the prediction net benefit and predicted survival groups respectively. Statistical significance was evaluated using the log-rank test.

Results

The preceding section detailed the materials and methods used to construct and evaluate various machine learning (ML) models aimed at predicting the conversion from mild cognitive impairment (MCI) to Alzheimer's disease (AD). This section presents the findings obtained by applying these models to the TADPOLE dataset.

The objective was to identify the optimal ML model and feature selection method for accurately forecasting the progression from MCI to AD. To achieve this, we assessed the performance of six distinct ML algorithms using metrics such as the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, and specificity.

ML feature subset selection method

The performance of six machine learning (ML) models was evaluated on independent test sets. [Table 2](#) summarizes the results, including ROC AUC, the c-index, the accuracy, the sensitivity, and the specificity of all six evaluated ML methods on the testing sets. [Fig 3](#) shows the ROC curve of all the evaluated ML methods, decision curve analysis (DCA) and the Kaplan-Meier of the BSWiMS, BESS:BIC, and LASSO test results. The BSWiMS model showed a test performance with AUC 0.87, 95% CI [0.85, 0.90], SEN 0.69, 95% CI [0.63, 0.76], and SPE 0.87 at 95% CI [0.83, 0.90]. The LASSO model also showed an equivalent performance, with AUC 0.88 at 95% CI [0.85, 0.90], SEN 0.64 at 95% CI [0.57, 0.71], and SPE 0.87 at 95% CI [0.83, 0.90]. The BSWiMS and LASSO were statistically similar, with the advantage that BSWiMS models' coefficients are highly interpretable. This is due to the rigorous process of how BSWiMS is performed, in which each stage of the bagging process selects the most statistically relevant features that explains the desired outcome. The BeSS models and RIDGE regression did not produce results as good as the LASSO and BSWiMS. These models were statistically inferior to the BSWiMS methods; furthermore, the DCA analysis indicated an inferior benefit to predict subjects that will convert to AD. The worst-performing method was the RIDGE method, whose ROC AUC was 0.54 at 95% CI [0.49, 0.58]. All models were able to stratify the high-low risk groups with statistical significance.

Features source and Cox-models

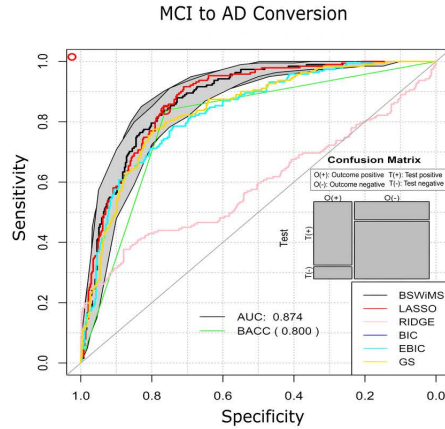
To evaluate the impact of different data sources on model performance, Cox proportional hazards models were trained and evaluated using five distinct sets of features. [Table 3](#) reports the performance metrics of these Cox models, and [Fig 4](#) shows the ROC AUC for all five sources as well as the DCA analysis and Kaplan-Meier of the cognitive and CSF features. The model developed using CSF and qMRI data showed the lowest classification performance, with an AUC of 0.78 at 95% CI [0.74, 0.82] and 0.79 at 95% CI [0.75, 0.83] respectively. The cognitive model showed the best performance of 0.84 at 95% CI [0.81, 0.87] that was statistically superior to MRI and CSF models ($p < 0.001$). The models that used multiple data sources were statistically superior to the cognitive-only model. ROC AUC improved to 0.85, 95% CI [0.82, 0.89] for the qMRI-Cognitive model and to 0.87, 95% CI [0.85, 0.90] for the model that used all the data sources. All these models that fused data sources were statistically superior ($p < 0.01$) to the models that only used a subset of features. The Kaplan-Meier analysis indicated that the model that used all the data sources ([Fig 3](#)) showed subjects classified as low risk had a nice five-year non-conversion rate (87%) when compared to the high-risk group that by year 5 more than 80% of them had converted to AD. Calibration plots for LASSO, BSWiMS, BESS:BIC models can be found at.

Biomarker evaluation

To investigate the specific biomarkers contributing to the improved risk prediction, we analyzed the coefficients of the Cox proportional hazards models. [Table 4](#) shows the beta coefficients and their corresponding hazard ratios with their corresponding 95% confidence intervals of the features that statistically improved the Cox model for the accurate high-risk, low-risk classification task. The table also shows that all data sources have features that aid the classification. The top feature was the Rey Auditory Verbal Learning Test, which indicated that higher scores indicate subjects that will remain stable with a z-standardized HR of 0.527 95% CI [0.400, 0.695]. The top CSF biomarker was the $A\beta$, HR 0.679 95% CI [0.541, 0.853]; the $APOE \epsilon 4$ showed an HR 2.063 95% CI [1.357, 3.136]. The top qMRI feature was the mean volume of left entorhinal gray matter, whose HR was 0.689 95% CI [0.557, 0.852].

The graph in [Fig 5A](#) displays a heatmap of the top features, highlighting those positively associated with conversion and those with a negative association. The graph in [Fig 5C](#) illustrates the associations of variables in predicting conversion, revealing two main clusters: the first includes $A\beta$, FAQ, and ADAS13, while the second comprises tau (p-tau), $APOE \epsilon 4$, and ADAS11. The plots on the left side report the model's performance on the testing set. The ROC curve, DCA, and Kaplan-Meier analysis demonstrate that the model, built using standardized features, exhibits performance closely aligned with predictions from the RHOCV procedure.

A



B

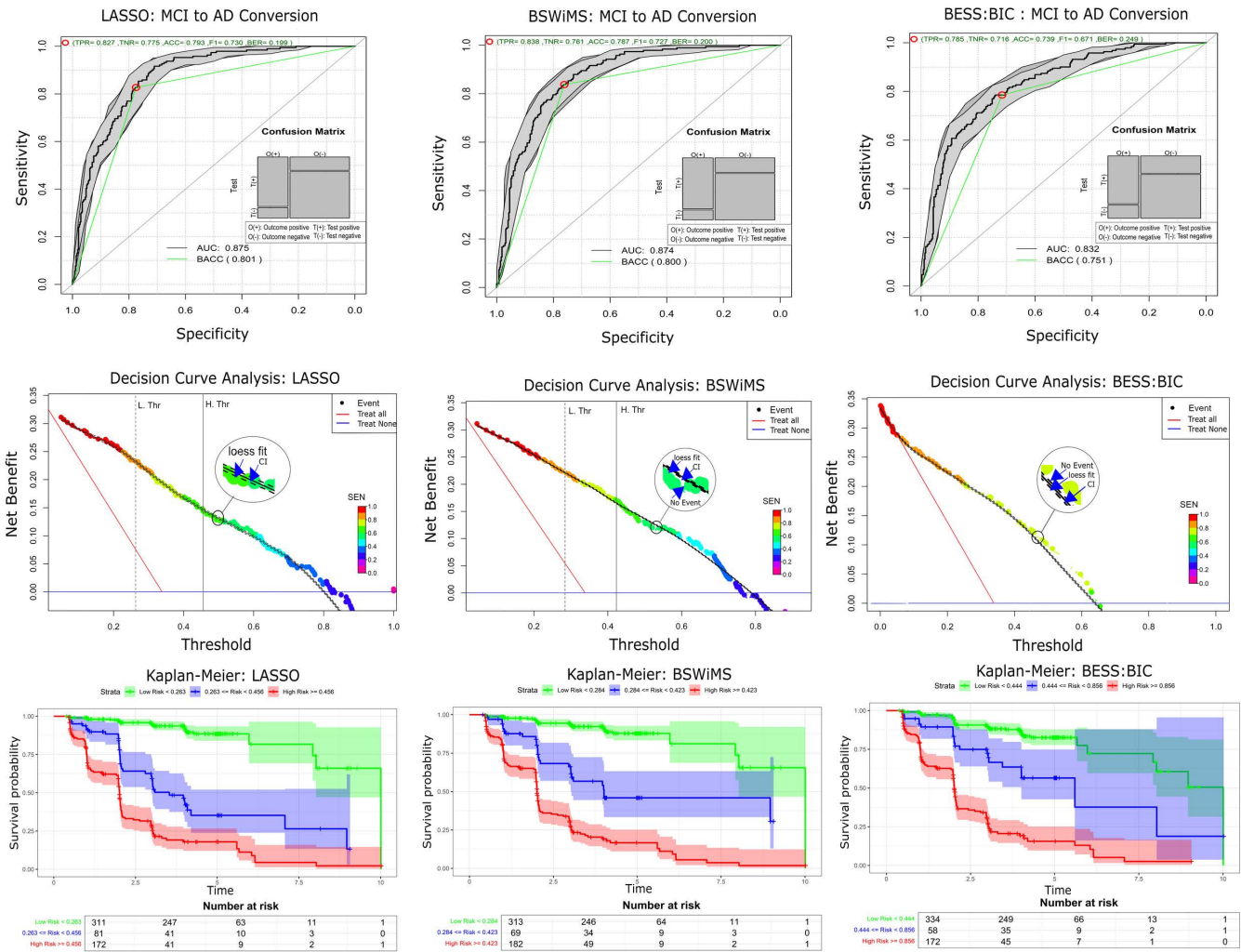


Fig 3. Evaluation and survival model analysis comparison. (A) Comparison of all ROC curves of all source-experiments. **(B)** ROC curves, Decision Curve, and Kaplan-Meier of LASSO, BSWiMS, and BESS BIC, respectively.

<https://doi.org/10.1371/journal.pone.0321671.g003>

Table 3. Source comparison on BSWiMS models.

Model	ROC AUC	c-Index	Accuracy	Sensitivity	Specificity
All	0.87 [0.85,0.90]	0.85 [0.82,0.87]	0.81 [0.78,0.84]	0.69 [0.63,0.76]	0.87 [0.83,0.90]
Cog + MRI	0.85 [0.82,0.89]	0.83 [0.81,0.86]	0.79 [0.76,0.83]	0.65 [0.58,0.72]	0.87 [0.83,0.90]
Cog	0.84 [0.81,0.87]	0.83 [0.80,0.85]	0.79 [0.75,0.82]	0.62 [0.55,0.69]	0.87 [0.83,0.90]
MRI+ APOE ε4	0.79 [0.75,0.83]	0.76 [0.73,0.80]	0.74 [0.70,0.77]	0.47 [0.40,0.54]	0.87 [0.84,0.91]
CSF+ APOE ε4	0.78 [0.74,0.82]	0.73 [0.69,0.76]	0.72 [0.68,0.76]	0.40 [0.33,0.47]	0.88 [0.85,0.92]

<https://doi.org/10.1371/journal.pone.0321671.t003>

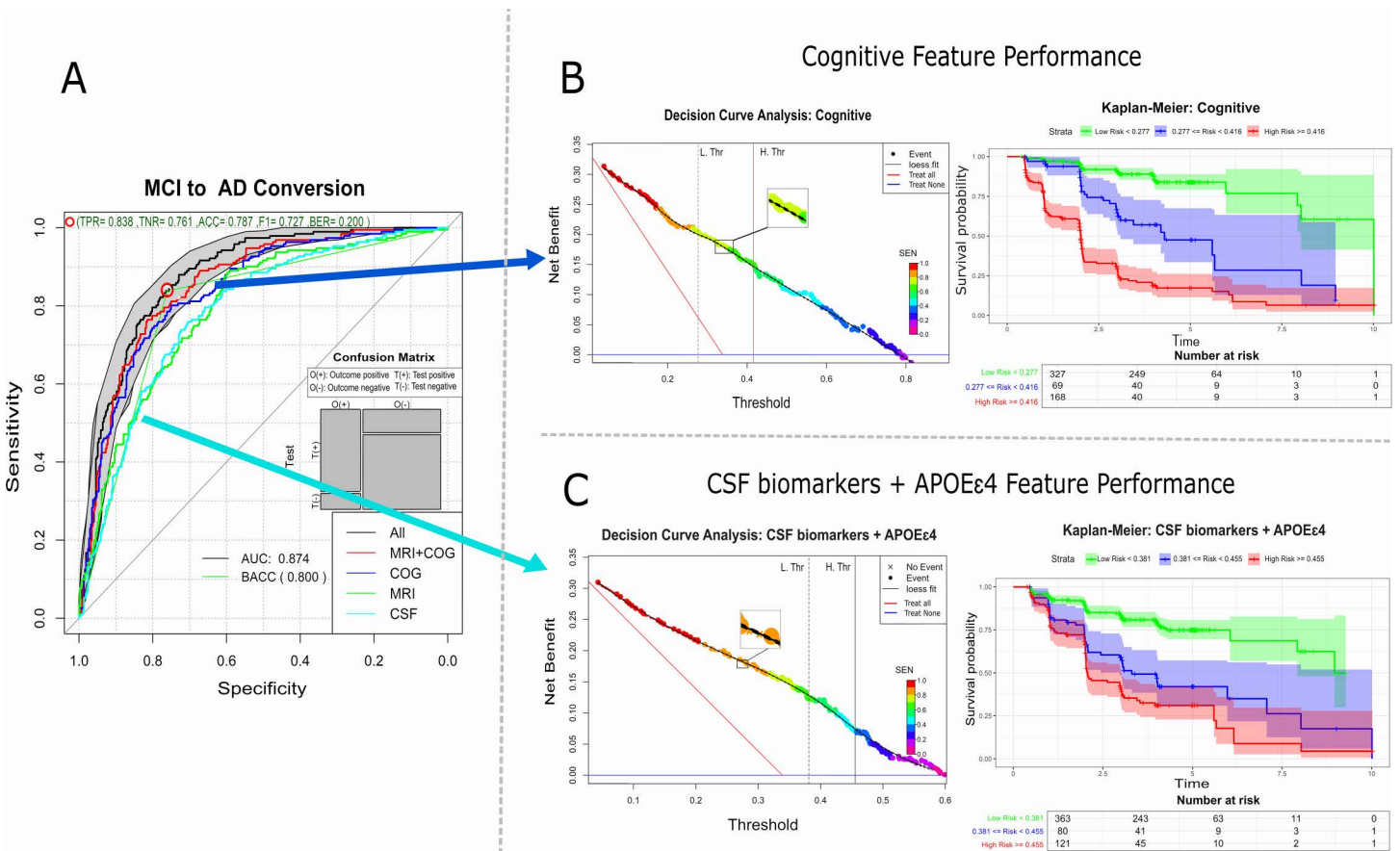


Fig 4. ROC curves, DCA and Kaplan-Meier. (A) ROC curves of method comparison experiment. **(B, C)** Plots shows the DCA and Kaplan-Meier performance on cognitive and CSF feature performance.

<https://doi.org/10.1371/journal.pone.0321671.g004>

Building on this performance evaluation, the calibration analysis of the Cox models revealed that the BSWiMS model exhibited excellent alignment, suggesting superior clinical reliability, whereas the LASSO model displayed poor calibration, with significant deviations between predicted and observed risks (particularly in high-risk ranges). This finding implies that the Bayesian framework of BSWiMS may more effectively handle uncertainty and sparse predictors in neurodegeneration modeling. Additional details are provided in the supplementary material.

Table 4. Top Features characteristics that intervene in the conversion of MCI to Alzheimer like relevant biomarkers.

Feature	Estimate	HR [95%CI]	Source	Description
RAVLT_immediate	-0.344	0.527 [0.400, 0.695]	Cognitive Function	Rey Auditory Verbal Learning Test: Immediate
FAQ	0.246	1.581 [1.273, 1.963]	Cognitive Function	Functional Activities Questionnaire
ABETA	-0.208	0.679 [0.541, 0.853]	CSF	Amyloid Beta
PTAU	0.014	1.399 [1.165, 1.680]	CSF	p-Tau
ADAS13	0.124	1.501 [1.130, 1.993]	Cognitive Function	Alzheimer's Disease Assessment Scale-Cog13
TAU	0.314	1.465 [1.224, 1.753]	CSF	Tau
ADAS11	0.191	1.789 [1.463, 2.189]	Cognitive Function	Alzheimer's Disease Assessment Scale-Cog11
M_ST24CV	-0.102	0.689 [0.557, 0.852]	qMRI	Mean: Volume (Cortical Parcellation) of Left Entorhinal
M_ST29SV	-0.117	0.643 [0.500, 0.826]	qMRI	Mean: Volume (WM Parcellation) of Left Hippocampus
APOE ϵ 4	0.315	2.063 [1.357, 3.136]	Genomic	APOE ϵ 4
RAVLT_perc_forgetting	0.071	1.684 [1.224, 2.319]	Cognitive Function	Rey Auditory Verbal Learning Test: forgetting
RD_ST34CV	-0.089	0.779 [0.683, 0.888]	qMRI	Relative Dif: Volume (Cortical Parcellation) of Left Isthmus Cingulate
M_ST13TA	-0.204	0.659 [0.572, 0.760]	qMRI	Mean: Cortical Thickness Average of Left Bankssts
M_ST26CV	0.000	0.749 [0.620, 0.905]	qMRI	Mean: Volume (Cortical Parcellation) of Left Fusiform
M_ST11SV	-0.156	0.736 [0.625, 0.868]	qMRI	Mean: Volume (WM Parcellation) of Left Accumbens Area
M_ST12SV	-0.037	0.711 [0.581, 0.869]	qMRI	Mean: Volume (WM Parcellation) of Left Amygdala
M_ST56CV	-0.023	0.782 [0.694, 0.881]	qMRI	Mean: Volume (Cortical Parcellation) of Left Superior Frontal
M_ST56SA	-0.216	0.693 [0.586, 0.819]	qMRI	Mean: Surface Area of Left Superior Frontal
M_ST39CV	0.000	0.873 [0.814, 0.936]	qMRI	Mean: Volume (Cortical Parcellation) of Left Medial Orbitofrontal
MMSE	-0.040	0.745 [0.610, 0.911]	Cognitive Function	Mini-Mental State Examination
M_ST59CV	0.000	0.780 [0.693, 0.876]	qMRI	Mean: Volume (Cortical Parcellation) of Left Supramarginal
RD_ST15TS	-0.060	0.773 [0.661, 0.903]	qMRI	Relative Dif: Cortical Thickness Standard Deviation of Left Caudal Middle Frontal
M_ST13CV	0.000	0.786 [0.682, 0.907]	qMRI	Mean: Volume (Cortical Parcellation) of Left Bankssts
M_ST14TS	0.059	1.252 [1.094, 1.433]	qMRI	Mean: Cortical Thickness Standard Deviation of Left Caudal Anterior Cingulate
M_ST59SA	-0.085	0.754 [0.645, 0.883]	qMRI	Mean: Surface Area of Left Supramarginal
RD_ST49SA	-0.087	0.804 [0.699, 0.925]	qMRI	Relative Dif: Surface Area of Left Postcentral
RD_ST34SA	-0.102	0.750 [0.619, 0.911]	qMRI	Relative Dif: Surface Area of Left Isthmus Cingulate
M_ST39SA	0.000	0.940 [0.904, 0.978]	qMRI	Mean: Surface Area of Left Medial Orbitofrontal

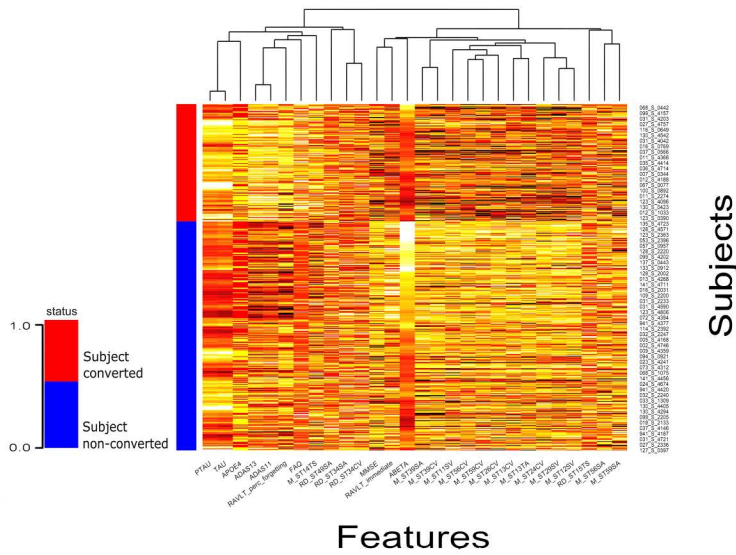
<https://doi.org/10.1371/journal.pone.0321671.t004>

Discussion

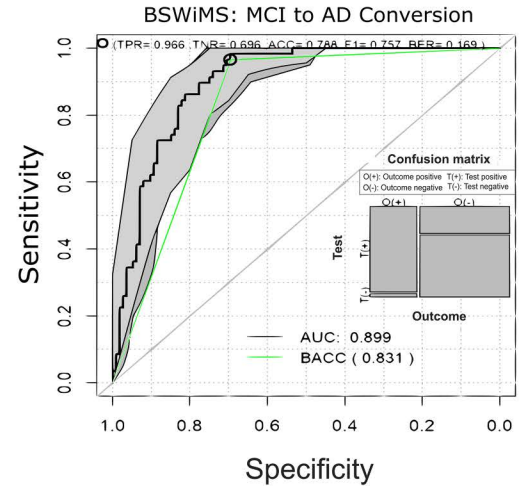
In this section, we explore the potential of machine learning techniques combined with Cox models, with the aim of predicting the conversion from MCI to AD using multimodal data from subjects with MCI. As mentioned in the introduction, combining biomarkers from CSF, cognitive assessments, MRI features, and *APOE* ϵ 4 status allows for a more comprehensive understanding of the dynamics associated with disease progression.

We validated that the integration of multiple biomarker modalities—CSF, *APOE* ϵ 4, MRI, and cognitive assessments in Cox models for predicting MCI-to-AD conversion is effective. Our validation was based on a stringent hold-out cross-validation scheme that randomly repeated the model training 50 times and analyzed the test results by advanced performance analysis methods like decision curve analysis and Kaplan-Meier plots. This analysis indicated that machine learning models such as LASSO and BSWiMS demonstrated their capacity to capture complex biomarker interactions,

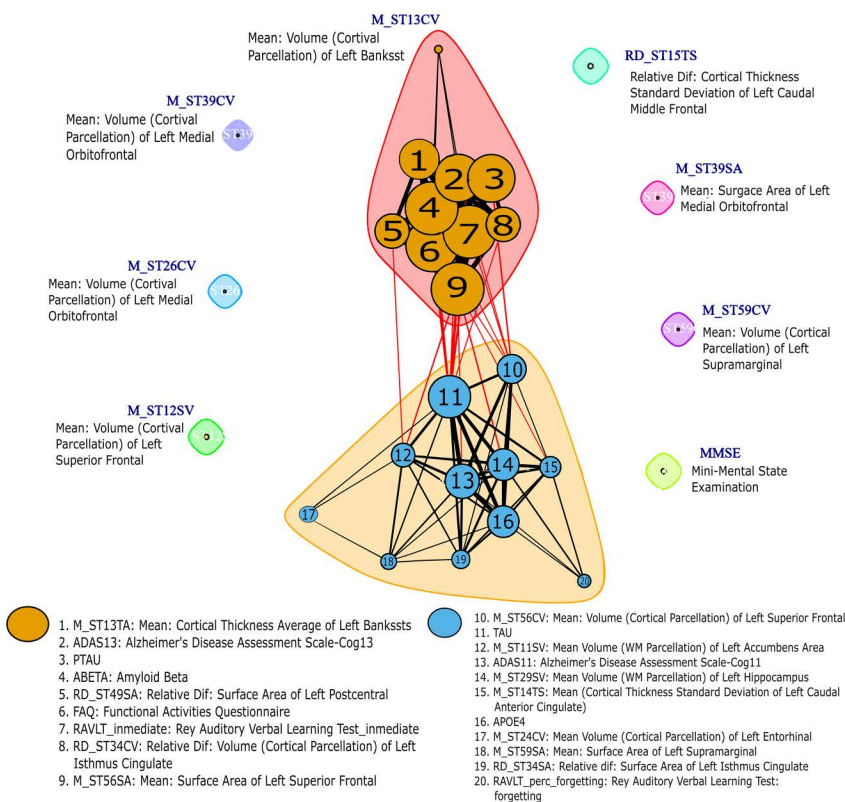
A Train: Selected Features



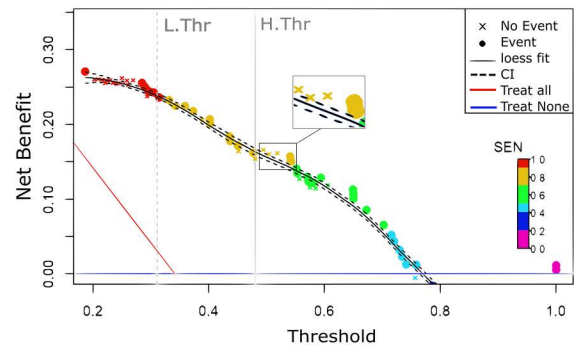
B Test: Model Evaluation



C MCI to AD Conversion



Decision Curve Analysis: Test



Kaplan-Meier: Test

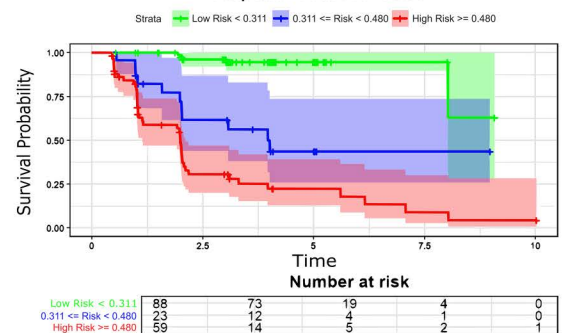


Fig 5. Feature selection and model evaluation for MCI to dementia conversion prediction. (A) Heat map of selected features at the training phase; features are classified from subject conversion or non-conversion. **(B)** Evaluation of the predictions on the testing set. ROC, DCA and Kaplan-Meier plots. **(C)** Graph clusters from the different features used in training.

<https://doi.org/10.1371/journal.pone.0321671.g005>

improving diagnostic precision and enabling more personalized treatment strategies. Recent research [55], highlighted the utility of LASSO in AD prediction, particularly when used with multimodal data, supporting our findings. The model's capability to integrate cognitive and imaging biomarkers makes it well-suited to capture subtle patterns indicative of early AD progression.

Similarly, the BSWiMS model also showed top performance, especially in its ability to be the best model to predict conversion at 90% sensitivity, yielding an AUC of 0.87 with a sensitivity of 0.69 95% CI [0.63, 0.76]. Furthermore, BSWiMS showed a strong net benefit in a threshold range of 0.2–0.7, further reinforcing its reliability in clinical applications. As noted in a study [56], models that integrate structural MRI with other biomarkers, such as BSWiMS, tend to provide higher predictive accuracy, particularly for identifying patients at high risk of AD conversion. The ability of the BSWiMS model to stratify patients into distinct risk groups, as seen in Kaplan-Meier analysis, makes it a valuable tool for early diagnosis and risk management in clinical practice. Overall, the performance of both LASSO and BSWiMS models validates the importance of multimodal approaches for accurate and clinically relevant AD prediction.

Moreover, the study demonstrated differences in AUC values across the five experimental data-source conditions, highlighting the varying predictive power of different combinations of biomarkers. The integrated model with all biomarker modalities (CSF, *APOE* ϵ 4, MRI, and cognitive data) achieved the highest AUC of 0.87, 95% CI [0.85, 0.90], indicating that multimodal integration significantly improves classification accuracy. This result is consistent with recent studies [55,57], showing that the combination of multiple sources of biomarkers offers a more comprehensive understanding of disease pathology and leads to improved predictive performance. Conversely, models based on single modalities, such as CSF + *APOE* ϵ 4, and MRI + *APOE* ϵ 4, displayed a lower predictive power, suggesting that single-source biomarkers do not capture the full complexity of MCI-to-AD progression.

Comparing our methodology, with a deep learning survival model, such as the study of [58], worked with a deep learning-based survival model (DeepSurv) which included demographics, cognitive tests, genetic data, CSF biomarkers, and MRI measures as their features showed an accuracy of 0.83, while our best model BSWiMS, achieved similar results: 0.81, 95% CI [0.78, 0.84], with the added advantage that the BSWiMS Cox model is highly interpretable.

Likewise, this study validated many of the well-established predictors of MCI-to-AD conversion. Notable examples [59–61], include p-Tau, Tau, *APOE* ϵ 4, $A\beta$, the Functional Activities Questionnaire (FAQ), the Rey Auditory Verbal Learning Test (RAVLT), and specific brain regions such as the entorhinal and hippocampal cortices.

While our model aligns with recent studies in identifying several MRI features associated with MCI-to-AD conversion, it also highlights additional MRI features that emerged as significant, despite not being traditionally recognized as relevant indicators for this progression. A research study [60], identified the average cortical thickness in the superior frontal, medial orbitofrontal, caudal anterior cingulate, and isthmus cingulate regions as significant predictors. In contrast, our study highlights the significance of the mean volume in the cortical parcellation of the left superior frontal region, the mean surface area of the left medial orbitofrontal region, the mean cortical thickness standard deviation of the left caudal anterior cingulate, and relative difference in the volume from cortical parcellation on the left isthmus cingulate as well as the mean surface area of the same region. These differences highlight the importance of looking beyond average cortical thickness, considering instead localized asymmetries and alternative measurements, such as surface area and mean volume within cortical parcellations.

In a broader context of feature relevance, a systematic review [62], identified that most studies prioritize MRI features, particularly entorhinal and hippocampal volumes, as key indicators for MCI-to-AD conversion. These features are also highlighted as critical components in our final model. Furthermore, the review indicates that many researchers combine structured data from multiple modalities. For instance, a study [63], used hippocampal volume as an MRI feature alongside $A\beta$ as a CSF feature, while another study [64], included hippocampal volume, tau, and $A\beta$ as features for predicting

conversion. These studies align closely with our findings, as we incorporate both hippocampal and entorhinal volumes as features. Additionally, our model leverages multimodal data by integrating MRI, CSF, cognitive assessment, and genomic data, aligning with the multimodal approaches observed in related research.

Limitations

The results presented in this work are limited to four key aspects. First, patient misdiagnosis is present, hence affecting feature selection and model building. The AD diagnosis used in this study is not definitive, but according to the NINCDS-ADRDA [65], it is the probable diagnosis of AD with probable errors in 10% to 15% of cases. Second, the presented findings were based on the ADNI cohort and measurements; therefore, it is biased toward the environmental factors present in the US and the Caucasian race. Third, qMRI results were based on FreeSurfer analysis; hence, changes in analytical tools may produce different results. Fourth, limited research has been done regarding the comparison of analytical techniques for automatic prediction of time to event within AD [66]. Additionally, the biomarker analysis is constrained by the specific version of the ADNI dataset used, which includes individual measurements of $A\beta$ but lacks the CSF $A\beta$ 42/40 ratio. This choice aligns with the biomarker profiling outlined by [67], where $A\beta$ was identified as a critical marker for AD diagnosis, showing decreased CSF levels in affected individuals compared to controls. While this biomarker has proven effective in our study, the absence of the $A\beta$ 42/40 ratio, a metric increasingly recognized for its enhanced accuracy in reflecting amyloid pathology and predicting MCI-to-AD progression, represents a limitation. Incorporating this ratio could improve predictive precision, as suggested by recent research. Future efforts leveraging updated ADNI datasets, which may include the $A\beta$ 42/40 ratio, could enable complementary analyses to better capture underlying amyloid pathology and refine diagnostic capabilities, building on the foundation established by related research [67,68].

To address these limitations comprehensively, future research should validate the findings across diverse cohorts from different countries and ethnicities, cross-validate FreeSurfer analyses with alternative pipelines, incorporate more patients with biomarker-confirmed AD diagnoses, and establish benchmarks for analytical techniques. These steps would enhance the robustness and applicability of the proposed study's outcomes.

Conclusion

This study showed that multimodal biomarker integration and machine learning methods for the construction of interpretable Cox survival models are effective strategies in predicting the progression of mild cognitive impairment to AD. We improved diagnostic accuracy by combining CSF fluid, *APOE* ϵ 4 genotype, MRI, and cognitive tests over single-modality models. In particular, the LASSO and BSWiMS models performed well in identifying complicated biomarker relationships and stratifying patients into separate risk categories. Notably, the discovery of unique MRI features, such as the standard deviation of cortical thickness, emphasizes the necessity of investigating alternative MRI metrics other than usual averages. In addition, the importance of showing metrics such as sensitivity and specificity can further aid the clinical context in determining which features are taken into account for the best prediction of conversion from MCI to AD, by valuing this metrics trade-offs, true positive values and true negative values, a more personalized approach in AD diagnosis could be proposed. Our findings are consistent with previous research that has highlighted the usefulness of multimodal approaches and specific biomarkers such as p-Tau and hippocampal atrophy. Overall, this study provides strong evidence that our proposed methodology is clinically effective in early AD diagnosis and risk management. Furthermore, the discovery of the aforementioned MRI characteristics indicates the possibility of improving diagnostic criteria and establishing more targeted therapy strategies. Future studies should validate these findings in larger, more diverse populations and investigate the possibility of adding additional biomarkers, such as genetic and epigenetic variables. Finally, the findings from this study may help to design more accurate and tailored diagnostic methods for AD.

Supporting information

S1 File. Supplementary material. Additional information. Experimental data, definitions of abbreviations, and calibration plots are provided in the Supplementary Material.

(PDF)

S2 Checklist. Inclusivity in global research. Completed checklist of ethical, cultural, and scientific considerations for research conducted outside the authors' home country.

(PDF)

Acknowledgments

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: <https://adni.loni.usc.edu/data-samples/adni-data/#AccessData> (List of investigators is located in the pdf named "Group Acknowledgements List").

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Author contributions

Conceptualization: Jose Tamez-Peña.

Data curation: Jorge Orozco.

Formal analysis: Jorge Orozco, Jose Tamez-Peña.

Investigation: Diana Sofia Rosales-Gurmendi, Gerardo Alejandro Fumagal-González, Jorge Orozco.

Methodology: Diana Sofia Rosales-Gurmendi.

Software: Jorge Orozco, Jose Tamez-Peña.

Supervision: Joshua Farber, Victor Treviño, Antonio Martinez-Torteya, Jose Tamez-Peña.

Validation: Victor Treviño, Emmanuel Martinez-Ledesma, Fabiola Rosales-Gurmendi.

Visualization: Emmanuel Martinez-Ledesma.

Writing – original draft: Diana Sofia Rosales-Gurmendi, Gerardo Alejandro Fumagal-González, Jorge Orozco.

Writing – review & editing: Diana Sofia Rosales-Gurmendi, Gerardo Alejandro Fumagal-González, Antonio Martinez-Torteya, Jose Tamez-Peña.

References

1. Association A. 2018 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*. 2018;14(3):367–429. <https://doi.org/10.1016/j.jalz.2014.02.001>
2. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement*. 2011;7(3):280–92. <https://doi.org/10.1016/j.jalz.2011.03.003> PMID: [21514248](https://pubmed.ncbi.nlm.nih.gov/21514248/)
3. DeTure MA, Dickson DW. The neuropathological diagnosis of Alzheimer's disease. *Mol Neurodegener*. 2019;14(1):32. <https://doi.org/10.1186/s13024-019-0333-5> PMID: [31375134](https://pubmed.ncbi.nlm.nih.gov/31375134/)
4. Porsteinsson AP, Isaacson RS, Knox S, Sabbagh MN, Rubino I. Diagnosis of Early Alzheimer's Disease: Clinical Practice in 2021. *J Prev Alzheimers Dis*. 2021;8(3):371–86. <https://doi.org/10.14283/jpad.2021.23> PMID: [34101796](https://pubmed.ncbi.nlm.nih.gov/34101796/)
5. Taha HB. Alzheimer's disease and related dementias diagnosis: a biomarkers meta-analysis of general and CNS extracellular vesicles. *npj Dement*. 2025;1(1). <https://doi.org/10.1038/s44400-024-00002-y>
6. Trejo-Lopez JA, Yachnis AT, Prokop S. Neuropathology of Alzheimer's Disease. *Neurotherapeutics*. 2022;19(1):173–85. <https://doi.org/10.1007/s13311-021-01146-y> PMID: [34729690](https://pubmed.ncbi.nlm.nih.gov/34729690/)
7. Lin S-Y, Lin P-C, Lin Y-C, Lee Y-J, Wang C-Y, Peng S-W, et al. The Clinical Course of Early and Late Mild Cognitive Impairment. *Front Neurol*. 2022;13:685636. <https://doi.org/10.3389/fneur.2022.685636> PMID: [35651352](https://pubmed.ncbi.nlm.nih.gov/35651352/)
8. Orozco-Sanchez J, Tamez-Peña J. Prediction of MCI to AD risk of conversion survival models: qMRI vs CSF measures and cognitive assessments. In: *Medical Imaging 2020: Computer-Aided Diagnosis*, 2020. 96. <https://doi.org/10.1117/12.2549301>
9. Jung N-Y, Kim ES, Kim H-S, Jeon S, Lee MJ, Pak K, et al. Comparison of Diagnostic Performances Between Cerebrospinal Fluid Biomarkers and Amyloid PET in a Clinical Setting. *J Alzheimers Dis*. 2020;74(2):473–90. <https://doi.org/10.3233/JAD-191109> PMID: [32039853](https://pubmed.ncbi.nlm.nih.gov/32039853/)
10. Raghavan N, Samtani MN, Farnum M, Yang E, Novak G, Grundman M, et al. The ADAS-Cog revisited: novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. *Alzheimers Dement*. 2013;9(1 Suppl):S21–31. <https://doi.org/10.1016/j.jalz.2012.05.2187> PMID: [23127469](https://pubmed.ncbi.nlm.nih.gov/23127469/)
11. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*. 1993;261(5123):921–3. <https://doi.org/10.1126/science.8346443> PMID: [8346443](https://pubmed.ncbi.nlm.nih.gov/8346443/)
12. Shaffer JL, Petrella JR, Sheldon FC, Choudhury KR, Calhoun VD, Coleman RE, et al. Predicting cognitive decline in subjects at risk for Alzheimer disease by using combined cerebrospinal fluid, MR imaging, and PET biomarkers. *Radiology*. 2013;266(2):583–91. <https://doi.org/10.1148/radiol.12120010> PMID: [23232293](https://pubmed.ncbi.nlm.nih.gov/23232293/)
13. Salvatore C, Cerasa A, Castiglioni I. MRI Characterizes the Progressive Course of AD and Predicts Conversion to Alzheimer's Dementia 24 Months Before Probable Diagnosis. *Front Aging Neurosci*. 2018;10:135. <https://doi.org/10.3389/fnagi.2018.00135> PMID: [29881340](https://pubmed.ncbi.nlm.nih.gov/29881340/)
14. Paterson RW, Toombs J, Slattery CF, Nicholas JM, Andreasson U, Magdalina NK, et al. Dissecting IWG-2 typical and atypical Alzheimer's disease: insights from cerebrospinal fluid analysis. *J Neurol*. 2015;262(12):2722–30. <https://doi.org/10.1007/s00415-015-7904-3> PMID: [26410752](https://pubmed.ncbi.nlm.nih.gov/26410752/)
15. Doherty CM, Forbes RB. Diagnostic Lumbar Puncture. *Ulster Med J*. 2014;83(2):93–102. PMID: [25075138](https://pubmed.ncbi.nlm.nih.gov/25075138/)
16. Paterson RW, Slattery CF, Poole T, Nicholas JM, Magdalina NK, Toombs J, et al. Cerebrospinal fluid in the differential diagnosis of Alzheimer's disease: clinical utility of an extended panel of biomarkers in a specialist cognitive clinic. *Alzheimers Res Ther*. 2018;10(1):32. <https://doi.org/10.1186/s13195-018-0361-3> PMID: [29558979](https://pubmed.ncbi.nlm.nih.gov/29558979/)
17. Janelidze S, Stomrud E, Smith R, Palmqvist S, Mattsson N, Airey DC, et al. Cerebrospinal fluid p-tau217 performs better than p-tau181 as a biomarker of Alzheimer's disease. *Nat Commun*. 2020;11(1):1683. <https://doi.org/10.1038/s41467-020-15436-0> PMID: [32246036](https://pubmed.ncbi.nlm.nih.gov/32246036/)
18. Janelidze S, Bali D, Ashton NJ, Barthélemy NR, Vanbrabant J, Stoops E, et al. Head-to-head comparison of 10 plasma phospho-tau assays in prodromal Alzheimer's disease. *Brain*. 2023;146(4):1592–601. <https://doi.org/10.1093/brain/awac333> PMID: [36087307](https://pubmed.ncbi.nlm.nih.gov/36087307/)
19. Bouwman FH, Frisoni GB, Johnson SC, Chen X, Engelborghs S, Ikeuchi T, et al. Clinical application of CSF biomarkers for Alzheimer's disease: From rationale to ratios. *Alzheimers Dement (Amst)*. 2022;14(1):e12314. <https://doi.org/10.1002/dad2.12314> PMID: [35496374](https://pubmed.ncbi.nlm.nih.gov/35496374/)
20. Zaretsky DV, Zaretskaia MV, Molkov YI, for the Alzheimer's Disease Neuroimaging Initiative. Patients with Alzheimer's disease have an increased removal rate of soluble beta-amyloid-42. *PLOS ONE*. 2022;17(10):e0276933. <https://doi.org/10.1371/journal.pone.0276933>
21. Kirsebom B-E, Espenes R, Waterloo K, Hessen E, Johnsen SH, Bråthen G, et al. Screening for Alzheimer's Disease: Cognitive Impairment in Self-Referred and Memory Clinic-Referred Patients. *J Alzheimers Dis*. 2017;60(4):1621–31. <https://doi.org/10.3233/JAD-170385> PMID: [28984581](https://pubmed.ncbi.nlm.nih.gov/28984581/)
22. O'Connor A, Weston PSJ, Pavisic IM, Ryan NS, Collins JD, Lu K, et al. Quantitative detection and staging of presymptomatic cognitive decline in familial Alzheimer's disease: a retrospective cohort analysis. *Alzheimers Res Ther*. 2020;12(1):126. <https://doi.org/10.1186/s13195-020-00695-2> PMID: [33023653](https://pubmed.ncbi.nlm.nih.gov/33023653/)

23. Llano DA, Laforet G, Devanarayan V, Alzheimer's Disease Neuroimaging Initiative. Derivation of a new ADAS-cog composite using tree-based multivariate analysis: prediction of conversion from mild cognitive impairment to Alzheimer disease. *Alzheimer Dis Assoc Disord*. 2011;25(1):73–84. <https://doi.org/10.1097/WAD.0b013e3181f5b8d8> PMID: 20847637
24. Atri A, Dickerson BC, Clevenger C, Karlawish J, Knopman D, Lin P-J, et al. The Alzheimer's Association clinical practice guideline for the diagnostic evaluation, testing, counseling, and disclosure of suspected Alzheimer's disease and related disorders (DETeCD-ADRD): Validated clinical assessment instruments. *Alzheimers Dement*. 2025;21(1):e14335. <https://doi.org/10.1002/alz.14335> PMID: 39713939
25. Yue L, Hu D, Zhang H, Wen J, Wu Y, Li W, et al. Prediction of 7-year's conversion from subjective cognitive decline to mild cognitive impairment. *Hum Brain Mapp*. 2021;42(1):192–203. <https://doi.org/10.1002/hbm.25216> PMID: 33030795
26. Young PNE, Estarellas M, Coomans E, Srikrishna M, Beaumont H, Maass A, et al. Imaging biomarkers in neurodegeneration: current and future practices. *Alzheimers Res Ther*. 2020;12(1):49. <https://doi.org/10.1186/s13195-020-00612-7> PMID: 32340618
27. McEvoy LK, Brewer JB. Quantitative structural MRI for early detection of Alzheimer's disease. *Expert Rev Neurother*. 2010;10(11):1675–88. <https://doi.org/10.1586/ern.10.162> PMID: 20977326
28. Gupta Y, Lee KH, Choi KY, Lee JJ, Kim BC, Kwon GR, et al. Early diagnosis of Alzheimer's disease using combined features from voxel-based morphometry and cortical, subcortical, and hippocampus regions of MRI T1 brain images. *PLoS One*. 2019;14(10):e0222446. <https://doi.org/10.1371/journal.pone.0222446> PMID: 31584953
29. Mashal Y, Abdelhady H, Iyer AK. Comparison of Tau and Amyloid- β Targeted Immunotherapy Nanoparticles for Alzheimer's Disease. *Biomolecules*. 2022;12(7):1001. <https://doi.org/10.3390/biom12071001> PMID: 35883556
30. Celaya-Padilla JM, Galván-Tejada CE, López-Monteagudo FE, Alonso-González O, Moreno-Báez A, Martínez-Torteya A, et al. Speed Bump Detection Using Accelerometric Features: A Genetic Algorithm Approach. *Sensors (Basel)*. 2018;18(2):443. <https://doi.org/10.3390/s18020443> PMID: 29401637
31. Sarica A, Aracri F, Bianco MG, Arcuri F, Quattrone A, Quattrone A, et al. Explainability of random survival forests in predicting conversion risk from mild cognitive impairment to Alzheimer's disease. *Brain Inform*. 2023;10(1):31. <https://doi.org/10.1186/s40708-023-00211-w> PMID: 37979033
32. Liu K, Chen K, Yao L, Guo X. Prediction of Mild Cognitive Impairment Conversion Using a Combination of Independent Component Analysis and the Cox Model. *Front Hum Neurosci*. 2017;11:33. <https://doi.org/10.3389/fnhum.2017.00033> PMID: 28220065
33. Zeifman LE, Eddy WF, Lopez OL, Kuller LH, Raji C, Thompson PM, et al. Voxel Level Survival Analysis of Grey Matter Volume and Incident Mild Cognitive Impairment or Alzheimer's Disease. *J Alzheimers Dis*. 2015;46(1):167–78. <https://doi.org/10.3233/JAD-150047> PMID: 25720412
34. Michaud TL, Su D, Siahpush M, Murman DL. The Risk of Incident Mild Cognitive Impairment and Progression to Dementia Considering Mild Cognitive Impairment Subtypes. *Dement Geriatr Cogn Dis Extra*. 2017;7(1):15–29. <https://doi.org/10.1159/000452486> PMID: 28413413
35. Tamez-Pena J, Martínez-Torteya A, Alanis I. Package 'FRESA.CAD' Feature Selection Algorithms for Computer Aided Diagnosis. CRAN Repository. 2016. <https://cran.r-project.org/web/packages/FRESA.CAD/index.html>
36. Bichindaritz I, Englebort C, Regua A, Kotula L. Feature selection and case-based reasoning for survival analysis in bioinformatics. In: 2018.
37. Aguirre-Gamboa R, Martínez-Ledesma E, Gómez-Rueda H, Palacios R, Fuentes-Hernández I, Sánchez-Canales E, et al. Efficient Gene Selection for Cancer Prognostic Biomarkers Using Swarm Optimization and Survival Analysis. *CBIO*. 2016;11(3):310–23. <https://doi.org/10.2174/1574893611999160610125628>
38. Schwarz G. Estimating the Dimension of a Model. *Ann Statist*. 1978;6(2). <https://doi.org/10.1214/aos/1176344136>
39. Marinescu RV, Oxtoby NP, Young AL, Bron EE, Toga AW, Weiner MW. TADPOLE Challenge: Prediction of Longitudinal Evolution in Alzheimer's Disease. The TADPOLE Challenge. 2018.
40. Morris JC. The clinical dementia rating (CDR): current version and scoring rules. *Neurology*. 1993;43(11):2412–4. <https://doi.org/10.1212/WNL.43.11.2412>
41. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry*. 1984;141(11):1356–64. <https://doi.org/10.1176/ajp.141.11.1356> PMID: 6496779
42. Folstein M, Folstein S, Folstein J. The Mini-Mental State Examination: A Brief Cognitive Assessment. In: Principles and Practice of Geriatric Psychiatry: Third Edition. John Wiley and Sons; 2010. 145–6.
43. Bean J. New York, NY: Springer New York. 2011. 2174–5.
44. Moradi E, Hallikainen I, Hänninen T, Tohka J, Alzheimer's Disease Neuroimaging Initiative. Rey's Auditory Verbal Learning Test scores can be predicted from whole brain MRI in Alzheimer's disease. *Neuroimage Clin*. 2016;13:415–27. <https://doi.org/10.1016/j.nicl.2016.12.011> PMID: 28116234
45. Collett D. Modelling survival data in medical research. Chapman & Hall/CRC. 2003.
46. Tamez-Pena J. Feature Selection and the BSWiMS Method. 2018.
47. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B (Methodological)*. 1996.
48. Wen C, Zhang A, Quan S, Wang X. BeSS: An R package for best subset selection in linear, logistic and CoxPH models. 2017.
49. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J Stat Softw*. 2011;39(5):1–13. <https://doi.org/10.18637/jss.v039.i05> PMID: 27065756

50. Kim J-H. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Computational Statistics & Data Analysis*. 2009;53(11):3735–45. <https://doi.org/10.1016/j.csda.2009.04.009>
51. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77. <https://doi.org/10.1186/1471-2105-12-77> PMID: 21414208
52. Kassambara A, Kosinski M, Biecek P. Package 'survminer': Drawing Survival Curves using 'ggplot2'. 2017. <https://CRAN.R-project.org/package=survminer>
53. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep*. 1966;50(3):163–70. PMID: 5910392
54. Tamez-Pena J, Orozco J, Sosa P, Valdes A, Nezhadmoghadam F. Ensemble of SVM, Random-Forest and the BSWiMS Method to Predict and Describe Structural Associations with Fluid Intelligence Scores from T1-Weighted MRI. *Lecture Notes in Computer Science*. Springer International Publishing. 2019. 47–56. https://doi.org/10.1007/978-3-030-31901-4_6
55. Hao X, Bao Y, Guo Y, Yu M, Zhang D, Risacher SL, et al. Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease. *Med Image Anal*. 2020;60:101625. <https://doi.org/10.1016/j.media.2019.101625> PMID: 31841947
56. Mito R, Li Q, Doecke J. Combining MRI and PET biomarkers improves prediction of progression from mild cognitive impairment to Alzheimer's disease. *Alzheimer's Research & Therapy*. 2021;13:99. <https://doi.org/10.1186/s13195-021-00804-7>
57. Leuz A, Mattsson-Carligen N, Palmqvist S. Biomarker-based prediction of progression in preclinical Alzheimer's disease. *Alzheimer's & Dementia*. 2022;18:573–83. <https://doi.org/10.1002/alz.12694>
58. Mirabnahrzazam G, Ma D, Beaulac C, Lee S, Popuri K, Lee H, et al. Predicting time-to-conversion for dementia of Alzheimer's type using multi-modal deep survival analysis. *Neurobiol Aging*. 2023;121:139–56. <https://doi.org/10.1016/j.neurobiolaging.2022.10.005> PMID: 36442416
59. Delli Pizzi S, Punzi M, Sensi SL, Alzheimer's Disease Neuroimaging Initiative. Functional signature of conversion of patients with mild cognitive impairment. *Neurobiol Aging*. 2019;74:21–37. <https://doi.org/10.1016/j.neurobiolaging.2018.10.004> PMID: 30408719
60. Varatharajah Y, Ramanan VK, Iyer R, Vemuri P, Alzheimer's Disease Neuroimaging Initiative. Predicting Short-term MCI-to-AD Progression Using Imaging, CSF, Genetic Factors, Cognitive Resilience, and Demographics. *Sci Rep*. 2019;9(1):2235. <https://doi.org/10.1038/s41598-019-38793-3> PMID: 30783207
61. Chen Y, Qian X, Zhang Y, Su W, Huang Y, Wang X, et al. Prediction Models for Conversion From Mild Cognitive Impairment to Alzheimer's Disease: A Systematic Review and Meta-Analysis. *Front Aging Neurosci*. 2022;14:840386. <https://doi.org/10.3389/fnagi.2022.840386> PMID: 35493941
62. Muhammed Niyas KP, Thiyagarajan P. A systematic review on early prediction of Mild cognitive impairment to alzheimers using machine learning algorithms. *International Journal of Intelligent Networks*. 2023;4:74–88. <https://doi.org/10.1016/j.ijin.2023.03.004>
63. Lei B, Yang P, Wang T, Chen S, Ni D. Relational-Regularized Discriminative Sparse Learning for Alzheimer's Disease Diagnosis. *IEEE Trans Cybern*. 2017;47(4):1102–13. <https://doi.org/10.1109/TCYB.2016.2644718> PMID: 28092591
64. Frölich L, Peters O, Lewczuk P, Gruber O, Teipel SJ, Gertz HJ, et al. Incremental value of biomarker combinations to predict progression of mild cognitive impairment to Alzheimer's dementia. *Alzheimers Res Ther*. 2017;9(1):84. <https://doi.org/10.1186/s13195-017-0301-7> PMID: 29017593
65. Thal LJ, Kantarci K, Reiman EM, Klunk WE, Weiner MW, Zetterberg H, et al. The role of biomarkers in clinical trials for Alzheimer disease. *Alzheimer Dis Assoc Disord*. 2006;20(1):6–15. <https://doi.org/10.1097/01.wad.0000191420.61260.a8> PMID: 16493230
66. Billichová M, Coan LJ, Czanner S, Kováčová M, Sharifian F, Czanner G. Comparing the performance of statistical, machine learning, and deep learning algorithms to predict time-to-event: A simulation study for conversion to mild cognitive impairment. *PLoS One*. 2024;19(1):e0297190. <https://doi.org/10.1371/journal.pone.0297190> PMID: 38252622
67. Shaw LM. PENN biomarker core of the Alzheimer's disease Neuroimaging Initiative. *Neurosignals*. 2008;16(1):19–23. <https://doi.org/10.1159/000109755> PMID: 18097156
68. Vuoksima E, McEvoy LK, Holland D, Franz CE, Kremen WS, Alzheimer's Disease Neuroimaging Initiative. Modifying the minimum criteria for diagnosing amnesic MCI to improve prediction of brain atrophy and progression to Alzheimer's disease. *Brain Imaging Behav*. 2020;14(3):787–96. <https://doi.org/10.1007/s11682-018-0019-6> PMID: 30511118