Interpretable Multivariate Survival Models: Improving Predictions for Conversion from Mild Cognitive Impairment to Alzheimer’s Disease via Data Fusion and Machine Learning

Jorge Orozco-Sanchez1, Gerardo Fumagal González1\*, Diana Rosales-Gurmendi1, Joshua Farber2, Victor Treviño1, Emmanuel Martinez-Ledesma1, Antonio Martinez-Torteya3, Jose Tamez-Peña1,2

1Tecnologico de Monterrey, Monterrey, Mexico

2Qmetrics Technologies LLC, Rochester, NY, USA

3Departamento de Ingeniería, Universidad de Monterrey, Monterrey, Mexico;

**\* Correspondence:**Gerardo Alejandro Fumagal González  
gerardofumagalgzz@gmail.com

Keywords: qMRI1, Alzheimer’s disease2, Survival3, Cox4, Machine Learning5, Radiomics6

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 assessments, Apolipoprotein E4 (APOEε4), and cerebrospinal fluid (CSF) biomarkers in survival models for predicting the progression from MCI to AD. Data from 564 participants in the ADNI study, who transitioned from MCI to AD, were analyzed. The dataset included 330 features encompassing qMRI, cognitive assessments, CSF biomarkers, and APOEε4 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 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). The results demonstrated that combining qMRI features with cognitive assessments, CSF biomarkers, and APOEε4 status significantly improves the ability to predict MCI-to-AD progression. These findings highlight the value of integrating multiple data sources in Cox survival models for the early identification of individuals at risk for Alzheimer’s disease.

# Introduction

Alzheimer's disease 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. The most accurate AD diagnostic test requires a histopathological evaluation of brain tissue by autopsy or biopsy (3). Without a biopsy, the diagnosis of a normal patient is defined as possible or probable AD according to patient reports, cognitive observation, and symptomatology (4). Hence, understanding the process and each of the stages of AD is essential for developing effective treatments. It is noteworthy that mild cognitive impairment (MCI) is regarded as a transitional stage between normal aging and AD (5), therefore in the context of conversion of mild cognitive impairment (MCI) to AD, some biomarkers have been identified in the AD literature. The beta-amyloid and the tau proteins from the cerebrospinal fluid (CSF) (6,7), the scores in cognitive assessments (8), and the polymorphism of the fourth allele of the apolipoprotein E (APOE-4) (9) have been recognized as validated risk factors for the conversion. On the other hand, some studies have presented that clinical information and imaging biomarkers can be used to predict patients who will undergo conversion from those who will not (10). Imaging biomarkers related to AD have been found in PET and MRI and have a clear association with the evolution and presence of AD (11) and with the conversion of MCI to AD (12). 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 qMRI features in survival models, have not been fully studied.

CSF biomarkers are increasingly being used to support a diagnosis of Alzheimer’s disease, especially to determine the difference between AD from non-Alzheimer's disease dementia (13–15). Different biomarkers can be extracted from CSF. The most established in the usual clinical practice include beta-amyloid 1-42 (), total tau (t-tau), and phosphorylated tau at threonine 181 (p-tau) (13). All of them have been previously described in AD pathogenesis. Despite their role in AD pathogenesis, the direct measurement of these CSF biomarkers lacks diagnostic accuracy (15,16).

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 (17,18). 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 (19) and others proposed the combination of patterns of brain atrophy and cognitive assessment scores, the latter offering the highest predictive power for the conversion (20).

Other tests available for the understanding of the AD process are magnetic resonance imaging (MRI) and positron-emitting tomography (PET). These imaging modalities can visualize direct changes in brain structure, function, sugar metabolism, and AD build-up (21,22). Therefore, AD-related imaging biomarkers with a clear association with the early development and presence of AD have been discovered in MRI and PET (11). Furthermore, imaging biomarkers have been associated with the conversion from MCI to AD (23–27). Consequently, MRI and PET are commonly used to monitor the progression of the disease and to detect the current stage of neuronal degeneration (28).

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4, the polymorphic allele of the apolipoprotein E, 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 diagnosing accuracy of AD, and to improve the predicting power of MCI to AD conversion (6).

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 their 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 (37). L1 regularization via the LASSO also allows subset selection (38). 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 (38,39). Finally, researchers usually rely on traditional statistical methods controlled for false discovery rate (FDR) for feature selection (FS) (40,41).

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, and SPDAS, (42) and FDR-controlled univariate filtering for building suitable survival models. 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. In addition, an important aim 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.

# Materials and Methods

This study is based on the TADPOLE challenge standard data sets (https://tadpole.grand-challenge.org). The TADPOLE sets were derived from the ADNI study (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary objective of ADNI has been to test whether MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). For up-to-date information, see [www.adni-info.org](http://www.adni-info.org)

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, qMRI and PET data, and APOE4 status. Detailed information regarding the rationale and the features contained in the TADPOLE challenge can be found Marinescu et.al (41). In this study, we screened for features that were common among the TADPOLE individuals (>35%). This criterion removed all the PET measurements.

To assess the difference between groups in the study and to evaluate the importance of adding features to the model, we divided the features into three major groups. The features detailing levels of three different proteins, , t-tau, and p-tau were included and studied in a group labeled as CSF features. These measures were obtained during baseline evaluation at the University of Pennsylvania Medical Center. Next, 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. Seven assessments were used: Clinical Dementia Rating Sum of Boxes (CDRSB) (44), Alzheimer’s Disease Assessment Scale (ADAS) (45), Mini-Mental State Examination (MMSE) (46), Rey's Auditory Verbal Learning Test (RAVLT) (47), and Functional Assessment Questionnaire (FAQ) (48). The 346 qMRI measurements provided by the University of California San Francisco (UCSF) were included and labeled as Radiomic features. UCSF used FreeSurfer Version 4.3, for the analysis of the MRI data sets (49). Finally, sex and APOE4 status was also included in this group.

Figure 1 shows the subject selection process, and the key features considered in this paper. The figure highlights that the original ADNI dataset included 1737 subjects with a total of 1907 features. However, just 330 features from 1115 subjects had complete observations. In this study we explored only subjects that with an MCI diagnosis at baseline; henceforth, we only analyzed 564 MCI subjects with 330 features in this paper. Among the MCI subjects, 191 converted from MCI to AD, and 373 remained stable during the observation period. Table I shows the demographics of the three groups. The age and gender of the three groups were statistically similar.

## 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 left and right sides of the brain were described by the mean and absolute differences, and relative differences.

To explore the set 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, , of the subject given the observed feature vector , and the unknown baseline hazard , i.e.:

(1)

where is the vector of coefficients. The fitting is commonly performed by a maximum likelihood estimation method providing the 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 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 aimed at building Cox models (50–54). In this paper, we evaluated three open-source ML packages. These packages provided us with nine different strategies for building Cox models. We complemented the results by comparing them to a Cox-based univariate filter. The first ML method was BSWiMS, part of the FRESA.CAD R package (50), 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 (55). The second evaluated package was the gmlnet 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 (37). Changing the elastic net penalty leads to Cox models with different coefficients and prediction behaviors. 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 three alpha values: alpha=1 (Feature selection with LASSO regularization), alpha=0 (Ridge or L2 regularization), and alpha =0.95 (Elastic net solution). The selected features for Ridge and Elastic Net were based on selecting features on which beta coefficients were not close to zero. The third evaluated package was the BeSS R package. This package implements the Primal-Dual Active set algorithm aiming for the selection of the best set of features of the Cox model. Like CoxNet, BeSS can use different strategies for set selection. The default configuration uses the GPDAS algorithm (38). The second BeSS option is SPDAS based on the Akaike information criterion (56). The third option was to run SPDAS algorithm using the Bayesian information criterion (BIC) as a target metric. Finally, we added three more Cox models based on filter strategies: BSWiMS, LASSO, and univariate COX. We used BSWiMS and LASSO selected features to fit the Cox model using the CoxPH method of the Survival R package. The univariate Cox model was based on models containing all the features that had a significant univariate association with the survival outcome. The univariate association was found by first fitting univariate Cox models to the survival outcome and then selecting only the features that had statistically significant association after correcting for false discovery rate (FDR) (57). In summary, we evaluated six different Cox models: BSWiMS, LASSO, RIDGE, BESS:BIC, BESS:EBIC, and BESS:GS.

## ML Validation and Evaluation

The main aim of this paper is the comprehensive evaluation of the Cox models for the prediction of the MCI to AD conversion. To achieve this goal, we constructed a fair playground where the same repeated holdout cross-validation (RHOCV) approach was used for the evaluation of the different ML strategies on the different sets of features provided by the TADPOLE challenge. 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 (55). 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. Predictions are calculated considering the aforementioned, Cox Model in Equation 1. First, we will use a linear predictor which is just the addition of all the explanatory variables. Then, risk predictions consider the second term of the risk equation by applying the exponential function to the linear predictions (.

The estimated coefficients on each training set were used to get the subject i linear predictions of the holdout set at each repetition:

(2)

Once the linear predictions were obtained for each repetition, the test results were ensembled by computing the median prediction for each subject as . The ensemble prediction was used to divide the subjects into high-risk (HR:) and low-risk (LR:) groups. 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 (58). 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).

(3)

(4)

(5)

(6)

(7)

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

(8)

where the indicator function if , and 0 otherwise. is the number of ordered pairs. is the median of the predicted survival time, and is the actual observed time of the uncensored subject . The values of the c-index range from 0 to 1, where 1 implies a perfect concordance between observed and predicted times.

The visualization of the predicted survival groups, high-risk vs. mid-risk, and low-risk, was done using Kaplan-Meier (KM) plots of survminer R package (59). The statistical significance of the difference between the survival groups was evaluated by the Logrank test (60):

(9)

(10)

where is the actual number of events, is the number of subjects below rank, and is the number of subjects at high risk.

## Experiments for the Evaluation of ML Feature Selection Method

The first set of experiments consisted in evaluating six different ML methodologies to construct Cox models. BSWiMS, BESS with BIC, EBIC and GS, and Logistic Regression with L1 penalization (LASSO), and L2 penalization (RIDGE). All six methods explored the construction of the optimal Cox model using all the available features.

## Experiments for the Evaluation of Features Source and Cox-Models

We executed five new independent RHOCV experiments to explore the impact of data sources. Experiment I used CSF and APOE4 features. Experiment II studied six Cognitive Assessment features. Experiment III constructed models from the MRI and APOE4. Experiment IV used Cognitive Assessment and MRI. Experiment V used all the features from CSF, Cognitive Assessment, MRI, and APOE4 as part of the feature set. Each experiment consisted of the independent execution of the RHOCV using the BSWiMS fiprocedure 20 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.

## Results

Six models were evaluated for MCI to AD conversion. The BSWiMS model showed excellent performance with AUC 0.87, SEN 0.69, SPE 0.87, and a positive net benefit within a threshold range of 0.2–0.7, peaking at 0.4. The LASSO model also showed strong performance, with AUC 0.88, SEN 0.64, and SPE 0.87. Figure 4 shows the decision curves for both models. Clinically, the BSWiMS and LASSO models maximized clinical benefit by identifying high-risk subjects while minimizing false positives. In contrast, the BESS model showed weaker results, with negative net benefit and high variability across thresholds, indicating instability in clinical application. Similarly, the Kaplan-Meier analysis divided patients into three risk groups according to the BSWiMS model (Figure 4). Survival probability decreased over time for all groups, but the decline was most pronounced in the high-risk group, indicating a higher probability of MCI-to-AD conversion compared to other groups.

The data source was the top contributor to Cox modeling performance. The model developed using CSF data in Experiment I showed the weakest high-risk/low-risk classification performance, with an AUC of 0.78 (95% CI 0.74-0.82) and exhibited the highest specificity at 0.88 (95% CI 0.85-0.92), although had the lowest sensitivity at 0.40 (95% CI 0.33-0.47). In contrast, models using all features in Experiment V had the best classification performance, with an AUC of 0.87 (95% CI 0.85-0.90).

A comprehensive feature analysis across all data sources identified eight common features among the top-performing models. These features, detailed in Figure 3 and Table 2, included several cognitive measures (RAVLT immediate, RAVLT perc\_forgetting, RAVLT learning, ADAS13, ADAS11, and FAQ) and neuroimaging markers (CV –volume Cortical Parcellation). Notably, APOEe4 was not among the top features. Among the MRI-derived features, three (CV –CV-volume cortical Parcellation) exhibited a value below 1, suggesting a protective effect against MCI-to-AD conversion. Conversely, two features with limited prior research, M\_ST31CV (Volume cortical Parcellation of Left Inferior Parietal (LIP)) and RD\_ST31TA (Relative Diff – -cortical thickness of LIP), demonstrated values above 1, indicating a higher risk of conversion. Further analysis of MRI features in Experiment III revealed 22 replicated MRI features from different ML approaches, with 7 relating to structural sizes. The most significant volumetric feature was the overall volume of the cortical Parcellation cortex. Thirteen features displayed large left-right asymmetry, with the greatest difference in the Thickness of the Left Inferior Parietal region.

## Discussion

These findings show which sources of biomarker data are most useful for predictive modeling and offer significant insights into the development from MCI to AD. Recent research has reaffirmed the importance of CSF biomarkers, such as phosphorylated tau (p-tau) and amyloid-β (Aβ), in identifying early pathological alterations linked to AD. CSF tau and amyloid levels, along with APOE4 status, offer strong predictive indicators for the transition from MCI to AD, as shown by Palmqvist, S., et al. (58). This is in line with the lowest classification performance shown in our investigation in Experiment I (58), which employed CSF alone. Volumetric analysis of MRIs, especially of the hippocampus, is still a reliable indicator of the course of AD. One of the most accurate measures of cognitive impairment in MCI patients is structural MRI biomarkers, particularly hippocampus atrophy, according to Mito, R., et al. (59), this validates our study's results, which showed that MRI features were quite predictive, particularly when paired with cognitive tests.

The advantage of combining multiple biomarker modalities has been increasingly recognized. Leuzy, A., et al. (60), found that integrating CSF, MRI, and cognitive data significantly improves the prediction of AD progression, aligning with the superior performance of our Experiment V, where we combined all biomarker modalities. Their study highlights the importance of multi-modal approaches for capturing the full spectrum of AD pathology. Machine learning algorithms have proven to be highly effective in integrating diverse biomarkers for AD prediction. Liu, M., et al. (61) showed that models like LASSO and other machine learning techniques provide superior predictive power when combining MRI and cognitive data, echoing our findings of strong c-index values in Experiment IV.

The integration of multiple biomarker modalities—CSF, APOE4, MRI, and cognitive assessments—has proven to enhance the accuracy of predicting MCI-to-AD conversion. Machine learning models such as LASSO and BSWiMS further demonstrate their capacity to capture complex biomarker interactions, improving diagnostic precision and enabling more personalized treatment strategies. These findings underscore the critical importance of multi-modal biomarker integration in advancing early AD diagnosis and patient care.

The study demonstrated notable differences in AUC values across the five experimental conditions, highlighting the varying predictive power of different biomarker combinations. Experiment V, which integrated all biomarker modalities (CSF, APOE4, 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 outcome is consistent with recent studies such as Leuzy et al. (60), which showed that combining multiple biomarker sources offers a more comprehensive understanding of disease pathology and leads to improved predictive performance. Conversely, models based on single modalities, such as Experiment I (CSF and APOE4, AUC = 0.78) and Experiment III (MRI and APOE4, AUC = 0.79), displayed weaker predictive power, suggesting that isolated biomarkers do not capture the full complexity of MCI-to-AD progression.

The LASSO model exhibited excellent performance in our study, particularly in Experiment IV (Cognitive and MRI), where it achieved a specificity of 0.85 (95% CI 0.82-0.89). LASSO's ability to handle high-dimensional data and select the most relevant features likely contributed to its robust performance. Recent research by Liu et al. (61) also 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 for capturing subtle patterns indicative of early AD progression. The balance between sensitivity and specificity achieved by LASSO in this study further underscores its clinical utility, as it provides reliable predictions while minimizing false positives.

Similarly, the BSWiMS model also performed remarkably well, especially in its ability to balance sensitivity and specificity, yielding an AUC of 0.87, similar to LASSO. Its strong net benefit in a threshold range of 0.2-0.7 further reinforces its reliability in clinical applications. As noted by Mito et al. (59), models that integrate structural MRI with other biomarkers, like BSWiMS, tend to provide higher predictive accuracy, particularly for identifying patients at high risk of AD conversion. The BSWiMS model's ability to stratify patients into distinct risk groups, as seen in the Kaplan-Meier analyses, 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.

Comparing our methodology with a deep learning survival model, Mirabnahrazam et al. (64), 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 models such as LASSO and BSWiMS achieved similar results, 0.79 and 0.81 respectively.

**Final model feature analysis**

Table 2 highlights the top features selected for our final model, many of which are well-established predictors of MCI-to-AD conversion, consistent with findings in recent studies and reviews. Notable examples include p-Tau, Tau, APOE4, Amyloid Beta, the Functional Activities Questionnaire, the Rey Auditory Verbal Learning Test, and specific brain regions such as the entorhinal and hippocampal cortices, as identified by Delli et.al (65), Varatharajah and Chen (62,66)

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. For example, Varatharajah (62), 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 on the volume from cortical parcellation on left isthmus cingulate as well as 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 by Muhammed et al.(63) identified that the majority of 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, one study by Lei et al (67), used hippocampal volume as an MRI feature alongside A-beta as a CSF feature, while another study by Frolich et al (68), 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, and genomic data, aligning with the multimodal approaches observed in related research.

## 4.3 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 it is the probable diagnosis of AD with probable errors in 10% to 15% of cases (68). 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 free-surfer 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 (68). These key limitations indicate that the presented findings must be confirmed on cohorts from different countries and ethnicities.

## Conclusion

This study shows that multi-modal biomarker integration and machine learning models can accurately predict the progression of moderate cognitive impairment (MCI) to Alzheimer's disease. We improved diagnostic accuracy significantly by combining CSF fluid, APOE4 genotype, MRI, and cognitive tests over single-modality models. In particular, the LASSO and BSWiMS models performed admirably in identifying complicated biomarker relationships and stratifying patients into separate risk categories. Notably, the discovery of unique MRI features, such as cortical thickness standard deviation, emphasizes the necessity of investigating alternative MRI metrics other than usual averages. 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 Alzheimer's disease diagnosis and risk management. Furthermore, the discovery of novel 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 Alzheimer's disease.

**Conflict of Interest**

*The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest*.

**Author Contributions**

**Funding**

**Acknowledgments**

This work was partially supported by Secretaría de Educación Superior, Ciencia, Tecnología e Innovación part of Gobierno de la República del Ecuador and by Strategic Research Group of Bioinformatics for Clinical Diagnosis from Tecnólogico de Monterrey. 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.

**References**

1. Alzheimer’s Association. 2018 Alzheimer’s disease facts and figures. Alzheimer’s & Dementia. 2018 Mar 1;14(3):367–429.
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 & Dementia. 2011 May 1;7(3):280–92.
3. Joachim CL, Morris JH, Selkoe DJ. Clinically diagnosed Alzheimer’s disease: Autopsy results in 150 cases. Ann Neurol. 1988 Jul 1;24(1):50–6.
4. 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.
5. Lin SY, Lin PC, Lin YC, Lee YJ, Wang CY, Peng SW, et al. The Clinical Course of Early and Late Mild Cognitive Impairment. Front Neurol [Internet]. 2022 May 16 [cited 2024 Jun 17];13:685636. Available from: www.frontiersin.org
6. Orozco-Sanchez J, Tamez-Peña JG. Prediction of MCI to AD Risk of Conversion Survival Models: qMRI vs CSF Measures and Cognitive Assessments. In: Medical Imaging. 2020.
7. Jung NY, Kim ES, Kim H-S, et al. Comparison of Diagnostic Performances Between Cerebrospinal Fluid Biomarkers and Amyloid PET in a Clinical Setting. Journal of Alzheimer's Disease: JAD. IOS Press; 2020.
8. 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. Alzheimer’s and Dementia. 2013 Feb;9(1 SUPPL.).
9. 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 (1979). 1993;
10. Martínez-Torteya A, Treviño V, Tamez-Peña JG. Improved Diagnostic Multimodal Biomarkers for Alzheimer’s Disease and Mild Cognitive Impairment. Biomed Res Int. 2015 Jan 1;2015:1–11.
11. 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 Feb 1;266(2):583–91.
12. 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 May 24;10(MAY):341799.
13. Paterson RW, Toombs J, Slattery CF, Nicholas JM, Andreasson U, Magdalinou NK, et al. Dissecting IWG-2 typical and atypical Alzheimer’s disease: insights from cerebrospinal fluid analysis. J Neurol. 2015 Dec 1;262(12):2722–30.
14. Doherty CM, Forbes RB. Diagnostic lumbar puncture. Ulster Medical Journal. 2014;83(2):93–102.
15. Paterson RW, Slattery CF, Poole T, Nicholas JM, Magdalinou 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 Mar 20;10(1).
16. Maddalena A, Papassotiropoulos A, Müller-Tillmanns B, Jung HH, Hegi T, Nitsch RM, et al. Biochemical diagnosis of Alzheimer disease by measuring the cerebrospinal fluid ratio of phosphorylated tau protein to beta-amyloid peptide42. Arch Neurol. 2003 Sep;60(9):1202–6.
17. Cullen B, O’Neill B, Evans JJ, Coen RF, Lawlor BA. A review of screening tests for cognitive impairment. Vol. 78, Journal of Neurology, Neurosurgery and Psychiatry. BMJ Publishing Group; 2007. p. 790–9.
18. Kirsebom BE, 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.
19. 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. 25(1):73–84.
20. Da X, Toledo JB, Zee J, Wolk DA, Xie SX, Ou Y, et al. Integration and relative value of biomarkers for prediction of MCI to AD progression: spatial patterns of brain atrophy, cognitive scores, APOE genotype and CSF biomarkers. Neuroimage Clin. 2014;4:164–73.
21. Smith JJ, Sorensen AG, Thrall JH. Biomarkers in Imaging: Realizing Radiology’s Future. Radiology. 2003 Jun 1;227(3):633–8.
22. McEvoy LK, Brewer JB. Quantitative structural MRI for early detection of Alzheimer’s disease. Expert Rev Neurother. 2010 Nov;10(11):1675–88.
23. Fox NC, Schott JM. Imaging cerebral atrophy: normal ageing to Alzheimer’s disease. The Lancet. 2004 Jan 31;363(9406):392–4.
24. Jack CR, Shiung MM, Weigand SD, O’Brien PC, Gunter JL, Boeve BF, et al. Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnestic MCI. Neurology. 2005 Oct 25;65(8):1227–31.
25. Henneman WJP, Sluimer JD, Barnes J, van der Flier WM, Sluimer IC, Fox NC, et al. Hippocampal atrophy rates in Alzheimer disease: added value over whole brain volume measures. Neurology. 2009 Mar 17;72(11):999–1007.
26. Suk HI, Shen D. Deep learning-based feature representation for AD/MCI classification. Med Image Comput Comput Assist Interv. 2013;16(Pt 2):583–90.
27. 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.
28. Khan A, Corbett A, Ballard C. Emerging treatments for Alzheimer’s disease for non-amyloid and non-tau targets. Expert Rev Neurother. 2017 Jul 3;17(7):683–95.
29. Mashal Y, Abdelhady HG, Iyer AK. Comparison of Tau and Amyloid-β Targeted Immunotherapy Nanoparticles for Alzheimer’s Disease. Vol. 12, Biomolecules. MDPI; 2022.
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 (Switzerland). 2018 Feb 3;18(2).
31. 33. Sarica A, Aracri F, Bianco MG, Arcuri F, Quattrone A, Quattrone A. Explainability of random survival forests in predicting conversion risk from mild cognitive impairment to Alzheimer’s disease. Brain Inform [Internet]. 2023 Dec 1 [cited 2024 Jun 17];10(1):1–17. Available from: https://link.springer.com/articles/10.1186/s40708-023-00211-w
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.
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. Journal of Alzheimer’s Disease. 2015 May 7;46(1):167–78.
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.
35. Guo B, Yi N. A Scalable and Flexible Cox Proportional Hazards Model for High-Dimensional Survival Prediction and Functional Selection. In: Proceedings of the 2022 Annual Meeting. 2022.
36. Wen C, Zhang A, Quan S, Wang X. BeSS: An R Package for Best Subset Selection in Linear, Logistic and CoxPH Models. 2017 Sep 19;
37. Tamez-Pena J, Martinez-torteya A, Alanis I. Package ‘FRESA.CAD’ Feature Selection Algorithms for Computer Aided Diagnosis. 2016.
38. Bichindaritz I, Englebert C, Regua A, Kotula L. Feature Selection and Case-Based Reasoning for Survival Analysis in Bioinformatics. The Thirty-First International Flairs Conference. 2018 May 10;
39. Aguirre-Gamboa R, Martinez-Ledesma E, Gomez-Rueda H, Palacios R, Fuentes-Hernandez I, Sánchez-Canales E, et al. Efficient Gene Selection for Cancer Prognostic Biomarkers Using Swarm Optimization and Survival Analysis. Curr Bioinform. 2016 Jun 15;11(3):310–23.
40. Schwarz G. Estimating the Dimension of a Model. The Annals of Statistics. 1978 Mar;6(2):461–4.
41. Marinescu R V., Oxtoby NP, Young AL, Bron EE, Toga AW, Weiner MW, et al. TADPOLE Challenge: Prediction of Longitudinal Evolution in Alzheimer’s Disease. 2018 May 10;
42. Morris JC. The clinical dementia rating (cdr): Current version and scoring rules. Neurology. 1993;43(11):2412–4.
43. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer’s disease. American Journal of Psychiatry. 1984;141(11 (1356-1364),).
44. Wiggins M, Price C. Mini-Mental State Examination (MMSE). In: A Compendium of Tests, Scales and Questionnaires. 2020.
45. Bean J. Rey Auditory Verbal Learning Test, Rey AVLT. In: Encyclopedia of Clinical Neuropsychology. New York, NY: Springer New York; 2011. p. 2174–5.
46. 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. 2017;13:415–27.
47. Reuter M, Schmansky NJ, Rosas HD, Fischl B. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage. 2012 Jul 16;61(4):1402–18.
48. Tamez-pena JG, Martinez-torteya A, Alanis I. Package ‘FRESA.CAD’ Feature Selection Algorithms for Computer Aided Diagnosis. 2016.
49. Collett D. Modelling survival data in medical research. Chapman & Hall/CRC; 2003. 391 p.
50. Tibshirani R. Regression Shrinkage and Selection via the Lasso Robert Tibshirani. Journal of the Royal Statistical Society Series B (Methodological). 1996;
51. Wen C, Zhang A, Quan S, Wang X. BeSS: An R Package for Best Subset Selection in Linear, Logistic and CoxPH Models. 2017 Sep 19;
52. Tamez-Pena J. Feature Selection and the BSWiMS Method. 2018.
53. Nishii R. Asymptotic Properties of Criteria for Selection of Variables in Multiple Regression. The Annals of Statistics. 1984;
54. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Vol. 57, Journal of the Royal Statistical Society. Series B (Methodological). WileyRoyal Statistical Society; 1995. p. 289–300.
55. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011 Mar 17;12:77.
56. Kassambara A, Kosinski M, Biecek P. R. 2017. Package ’ survminer ’.
57. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep. 1966 Mar;50(3):163–70.
58. Palmqvist S, Hansson O, Zetterberg H, et al. Discriminative Accuracy of Plasma Phospho-Tau217 for Alzheimer’s Disease vs Other Neurodegenerative Disorders. Vol. 324, JAMA. American Medical Association; 2021. p. 772–781.
59. Mito R, Li QX, Doecke J, et al. Combining MRI and PET Biomarkers Improves Prediction of Progression from Mild Cognitive Impairment to Alzheimer’s Disease. Vol. 13, Alzheimer’s Research & Therapy. BioMed Central; 2021. p. 99.
60. Leuzy A, Mattsson-Carlgren N, Palmqvist S, et al. Biomarker-Based Prediction of Progression in Preclinical Alzheimer’s Disease. Vol. 18, Alzheimer’s & Dementia. Wiley; 2022. p. 573–583
61. Liu M, Cheng D, Wang X, et al. Multimodal Neuroimaging Feature Selection with Consistent Metric Constraint for Diagnosis of Alzheimer’s Disease. Vol. 60, Medical Image Analysis. Elsevier; 2020. p. 101625.
62. 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. Vol. 9, Scientific Reports. Nature Publishing Group; 2019. p. 2235.
63. Muhammed Niyas KP, Thiyagarajan P. A Systematic Review on Early Prediction of Mild Cognitive Impairment to Alzheimer’s Using Machine Learning Algorithms. Vol. 4, International Journal of Intelligent Networks. Elsevier; 2023. p. 74–88.
64. Mirabnahrazam G, Ma D, Beaulac C, et al. Predicting Time-to-Conversion for Dementia of Alzheimer's Type Using Multi-Modal Deep Survival Analysis. Vol. 121, Neurobiology of Aging. Elsevier; 2023. p. 139–156.
65. Delli Pizzi S, Punzi M, Sensi SL. Functional Signature of Conversion of Patients with Mild Cognitive Impairment. Vol. 74, Neurobiology of Aging. Elsevier; 2019. p. 21–37.
66. Chen Y, Qian X, Zhang Y, et al. Prediction Models for Conversion from Mild Cognitive Impairment to Alzheimer’s Disease: A Systematic Review and Meta-Analysis. Vol. 14, Frontiers in Aging Neuroscience. Frontiers Media SA; 2022.
67. Lei, B., Yang, P., Wang, T., Chen, S., & Ni, D. (2017). Relational-regularized discriminative sparse learning for Alzheimer’s disease diagnosis. *IEEE transactions on cybernetics*, *47*(4), 1102-1113.
68. Frölich, L., Peters, O., Lewczuk, P., Gruber, O., Teipel, S. J., Gertz, H. J., ... & Kornhuber, J. (2017). Incremental value of biomarker combinations to predict progression of mild cognitive impairment to Alzheimer’s dementia. *Alzheimer's research & therapy*, *9*, 1-15.

Table : List of various abbreviations and acronyms used throughout the paper.

|  |  |
| --- | --- |
| **Abbreviation** | **Definitoin** |
| Ab | amyloid beta 1-42 |
| ACC | accuracy |
| AD | Alzheimer’s disease |
| ADAS | Alzheimer’s Disease Assessment Scale |
| ADNI | Alzheimer's Disease Neuroimaging Initiative |
| APOE 4 | apolipoprotein E4 |
| AUC | area under the curve |
| BESS | Best subset selection |
| BIC | Bayesian information criterion |
| BSWiMS | Bootstrap Stepwise Model Selection |
| CDRSB | Clinical Dementia Rating Sum of Boxes |
| CI | confidence intervals |
| CSF | cerebrospinal fluid |
| FAQ | Functional Assessment Questionnaire |
| FDR | false discovery rate |
| FS | future selection |
| GPDAS | golden section primal-dual active set |
| HR | hazard ratios |
| ICV | Intracranial volume |
| KM | Kaplan-Meier |
| LASSO | Least Absolute Shrinkage and Selection Operator |
| LIP | Left Inferior Parietal |
| MCI | mild cognitive impairment |
| ML | machine learning |
| MMSE | Mini-Mental State Examination |
| MRI | magneto resonance |
| PET | Positron-emitting tomography |
| p-tau | phosphorylated tau |
| RAVLT | Rey's Auditory Verbal Learning Test |
| RHOCV | repeated holdout cross-validation |
| ROC | receiver operating characteristic |
| SEN | sensitivity |
| SL | Statistical learning |
| SPDAS | sequential primal-dual active set |
| SPE | specificity |
| TADPOLE | The Alzheimer’s Disease Prediction Of Longitudinal Evolution |
| TN | true negative |
| TP | true positive |
| UCSF | University of California San Francisco |

A flowchart of a flowchart

Description automatically generated

a)

A flowchart of a computer program

Description automatically generated

b)

**Figure 1:** Data inclusion and exclusion diagram: a) Feature selection, 1907 features were identified from ADNI-TADPOLE, 1575 PET, and longitudinal analysis features with Free Surfer were excluded, then features with lost data were excluded to include a total of 330 features at the end; b) Subject selection, a baseline of 1737 subjects from the same database was obtained, 622 subjects with missing data or few observations were excluded.

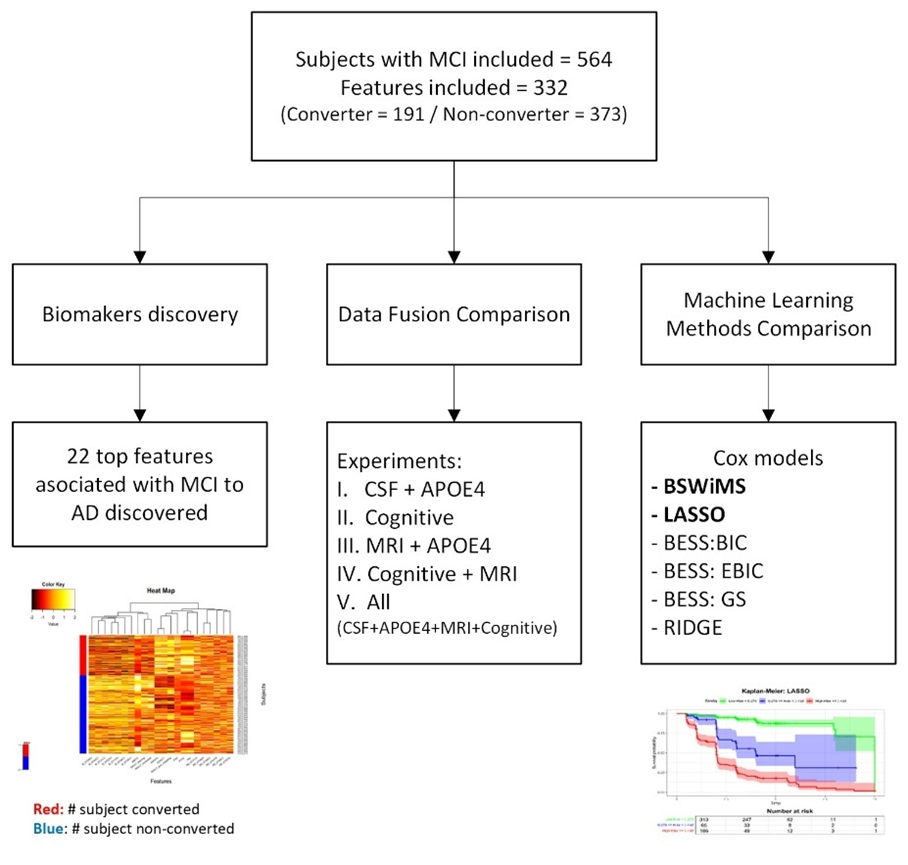


Figure : Diagram, the process of analysis of discovered biomarkers, data fusion (CSF: cerebrospinal fluid, MRI: magnetic resonance imaging, Cog: Cognitive, APOE4: apolipoprotein E4) by cross-validation and Machine Learning methods comparison evaluating the performance of 6 Cox models

Table : Demographic data with features sectioned by non-converter and converter group.

|  |  |  |  |
| --- | --- | --- | --- |
| 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 |
| APOE4 | 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 | 4.52 (2.43) | 5.08 (2.27)\*\*\* |
| 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)\*\*\* |
|  | mm | 115.15 (3.95) | 115.65 (4.38) |
|  | mm | 884.40 (21.25) | 868.01 (20.44)\*\*\* |
|  | mm | 280.03 (49.43) | 296.33 (46.40)\*\*\* |

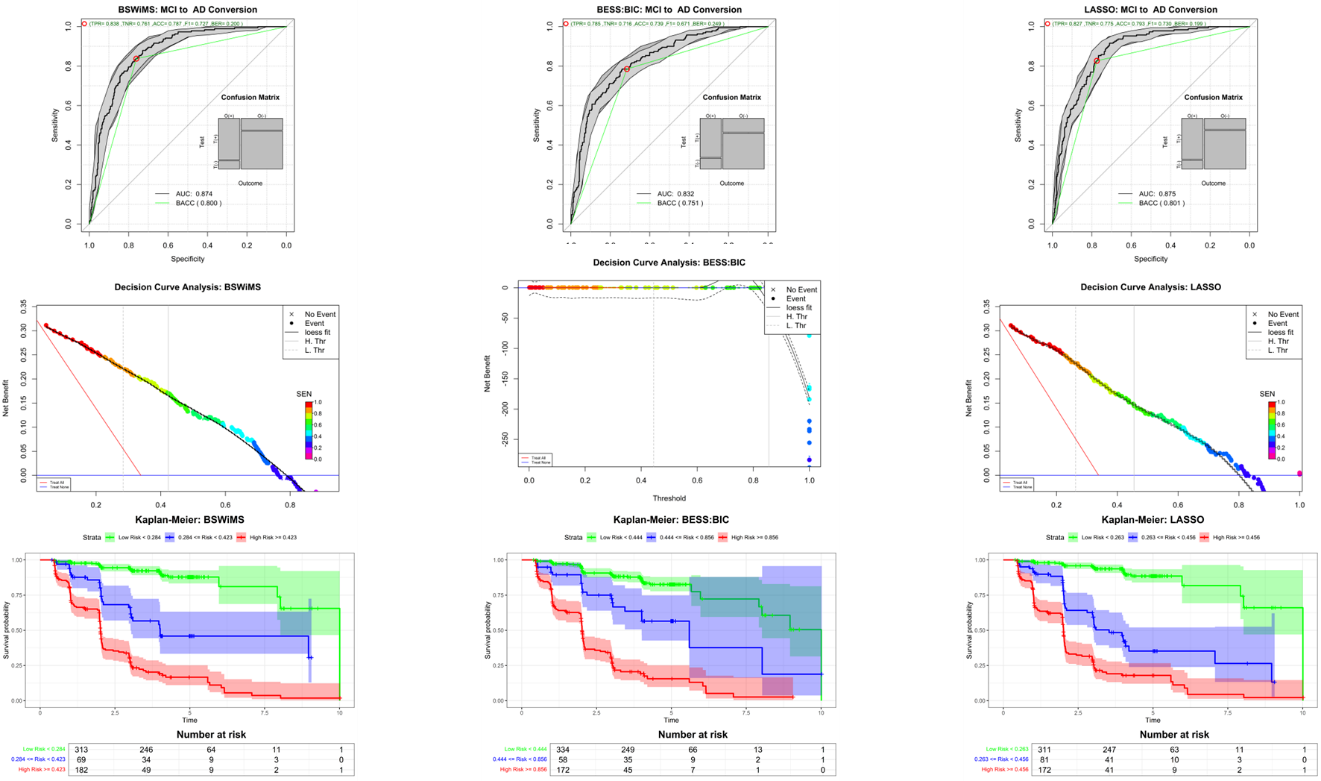


Figure : Results graphs (ROC curves, Decision Curve, and Kaplan-Meier) for the models with better performance (a) Method: BSWiMS, (b) Method: LASSO, compared with the model with worth performance (c) Method: BESS BIC

A heat map with different colors

Description automatically generated

**Red:** # subject converted

**Blue:** # subject non-converted

Figure 4: The 22 features selected by all the 3 methods in at least in the half of the iterations (vertical axis) and subjects on the horizontal axis. Red group: converter subjects, Blue group: non-converter subjects.

Table : ML 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) |
| 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) |
| 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) |
| 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) |

Table : Source Comparison

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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+APOE4 | 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+APOE4 | 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) |

Table : 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 LeftEntorhinal |
| **M\_ST29SV** | -0.117 | 0.643 (0.500-0.826) | qMRI | Mean: Volume (WM Parcellation) of Left Hippocampus |
| **APOE4** | 0.315 | 2.063 (1.357-3.136) | Genomic | APOE4 |
| **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 |