Prediction of the conversion from subjective cognitive decline to mild cognitive impairment or Alzheimer's disease by integrating imputation techniques and machine learning based on multimodal clinical data

Yiping Qian[1] | Andrew Doig[2] | Sofia Michopoulou[3] | Fumie Costen[1] | for the Alzheimer's Disease Neuroimaging Initiative

[1]Department of Electrical & Electronic Engineering, University of Manchester, Manchester, UK

[2]Division of Molecular & Cellular Function, University of Manchester, Manchester, UK

[3]Department of Medical Physics, University of Southampton, Southampton, UK

Correspondence

Fumie Costen, Department of Electrical & Electronic Engineering, University of Manchester, Manchester, United Kingdom.

Email: fumie.costen@manchester.ac.uk

Abstract

Introduction

Subjective Cognitive Decline (SCD) is an early indicator of Alzheimer's disease (AD), and accurate identification of high-risk individuals is essential for timely intervention. Predicting the progression of SCD remains challenging, particularly in resource-limited clinical settings where advanced neuroimaging is not routinely available.

Methods

We developed a predictive framework to predict 10-year conversion using multimodal clinical data, including fluid biomarkers, cognitive assessments, apolipoprotein E4 (APOE4), and demographics, using 2,423 visit-level samples from 635 participants in the Alzheimer's Disease Neuroimaging Initiative (ADNI). After participant-level splitting and preprocessing for missingness and redundancy, 7 machine learning models and 2 heterogeneous ensembles were constructed and evaluated using standard performance metrics. Finally, the main model was further interpreted using calibration analysis, SHapley Additive exPlanations (SHAP)-based interpretation, and robustness analysis.

Results

Support vector machine trained on plasma-based data with MICE imputation showed the best balanced performance (balanced accuracy = 0.70, sensitivity = 0.72, specificity = 0.68, F1-score = 0.50, MCC =

0.33, AUC = 0.73). SHAP revealed executive function, episodic memory, plasma Aβ40 and p-tau, and APOE4 as key predictors.

Discussion

Our framework identifies individuals at high risk of SCD progression using cost-effective, nonimaging data. This makes it scalable to primary care settings and underserved populations.

Keywords: Alzheimer's disease; subjective cognitive decline; mild cognitive impairment; machine learning; disease progression.

1 | INTRODUCTION

Subjective Cognitive Decline (SCD) refers to a self-perceived and persistent decline in cognition, typically memory, without measurable cognitive deficits on standard neuropsychological tests [1]. It is recognized as the earliest symptomatic stage of Alzheimer's disease (AD), preceding mild cognitive impairment (MCI) and dementia [2]. Longitudinal studies have indicated that individuals with SCD have a 20–40% risk of progression to MCI or AD over a 10-year period [3, 4].

Early identification of high-risk individuals is essential for timely interventions and disease prevention. Clinically, amyloid targeting therapies have shown greater efficacy in early stages of AD. In Clarity AD, a phase 3 trial, lecanemab reduced amyloid and slowed cognitive decline over 18 months in people with early symptomatic AD [5]. In TRAILBLAZER-ALZ 2, donanemab slowed clinical progression, with stronger effects among participants with low-to-intermediate baseline tau burden than among those with high baseline tau [6].

Several nonimaging clinical measures are used to predict cognitive decline in individuals with SCD. These include cerebrospinal fluid (CSF) biomarkers such as amyloid beta 42 (Aβ42), Aβ40, tau, and phosphorylated tau (p-tau), which reflect AD pathology and support the amyloid-tau-neurodegeneration (ATN) classification framework [7–9]. Blood-based biomarkers such as plasma Aβ42 and Aβ40, p-tau, and neurofilament light chain (NfL) have emerged as promising and practical noninvasive alternatives for detecting early AD pathology [10]. Cognitive assessments, including the Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA), along with demographics (age, gender, and education level), also contribute to evaluating cognitive status and trajectories [4, 11, 12].

Combining machine learning (ML) can predict progression from SCD to MCI or AD by capturing complex nonlinear patterns in data [4, 13]. However, current approaches face limitations, including poor handling of missing data, overfitting due to redundant features, and limited interpretability. Some studies rely on costly magnetic resonance imaging (MRI) and positron emission tomography (PET) scans [14, 15], which limit their feasibility for large-scale or resource-constrained screening scenarios. While some studies have employed low-cost data such as demographics and cognitive tests, their predictive performance has often been suboptimal [13]. In light of these challenges, developing accurate, interpretable, and cost-effective methods for predicting SCD progression is necessary.

To address these gaps, this study develops an interpretable ML framework, tailored to individuals with SCD, to predict 10-year conversion to MCI or AD using multimodal clinical data. The 10-year horizon is chosen to cover the entire prodromal phase of AD, which typically includes 5–10 years from SCD to MCI [4, 13, 16], and an additional 3–5 years to AD [1]. In contrast to prior work that often relies on single-modality clinical data, the proposed framework incorporates clinically available nonimaging data, including fluid biomarkers, cognitive assessments, demographics, and apolipoprotein E4 (APOE4), while explicitly treating missingness and redundancy as central aspects of the framework design.

This study makes three contributions. First, a visit-level labeling strategy is introduced for individuals with SCD, in which each visit is prospectively assigned a 10-year conversion label according to subsequent diagnoses. Since some participants contributed more than one longitudinal visit, participant-level data splitting was used to ensure that all longitudinal visits from the same participant remained within the same split, avoiding information leakage between training and test sets. Second, a systematic preprocessing pipeline is implemented that combines correlation filtering, benchmarking of six imputation methods, and multicollinearity diagnostics. This pipeline explicitly addresses missing data and feature redundancy during model development. Third, the framework compares multiple linear, nonlinear, and ensemble classifiers across different data configurations, varying by data modality and imputation method, and uses probability

calibration, SHapley Additive exPlanations (SHAP) based interpretation, and a robustness analysis for further exploration.

By jointly considering data modalities, imputation strategies, and models, the proposed framework provides a transparent evaluation of nonimaging clinical data for long-term SCD progression prediction. It may support the development of a clinically feasible solution for the prediction of early progression in individuals with SCD.

2 | MATERIALS AND METHODS

2.1 | Data Description

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (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. The original goal of ADNI was to test whether serial magnetic MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. The current goals include validating biomarkers for clinical trials, improving the generalizability of ADNI data by increasing diversity

in the participant cohort, and providing data concerning the diagnosis and progression of Alzheimer's disease to the scientific community. For up-to-date information, see adni.loni.usc.edu.

We utilized longitudinal data from ADNI, focusing on individuals diagnosed with SCD. Given the self-reported nature of SCD, some individuals who consider themselves cognitively normal may in fact meet SCD criteria [17, 18]. To mitigate this, we initially included 223 baseline SCD cases and 412 cognitively normal individuals, each with a baseline and at least one follow-up visit suitable for 10-year conversion prediction. Total follow-up duration ranged from 6 to 204 months.

Each visit (baseline or follow-up) received a prospective label relative to the visit date, classified as "converted" if progression to MCI or AD occurred within the next 10 years, and "stable" otherwise. Consequently, visits from the same participant could carry different labels. The final dataset comprised 2,423 visit-level samples from 635 participants. At the visit level, 543 visits (22.41%) were labeled as "converted", and 1,880 visits (77.59%) were labeled as "stable". At the participant level, 128 participants (20.16%) converted to MCI or AD during follow-up, whereas 507 participants (79.84%) remained clinically stable.

To evaluate how modality combinations affect prediction, we constructed three multimodal datasets (A, B, and C) to capture complementary clinical information. TABLE 1 presents a detailed overview of the variables included in each dataset and their missingness statistics. The three datasets share common

demographic, cognitive, and genetic features but differ in fluid biomarkers. Dataset A includes CSF biomarkers (Aβ42, tau, and p-tau), B includes plasma biomarkers (Aβ42/40, NfL, and p-tau181), and C combines CSF and plasma biomarkers, providing the most comprehensive fluid biomarker profile among the three datasets.

2.2 | Data Preprocessing

We implemented a customized three-step preprocessing to improve data quality. This comprised (1) correlation filtering to remove redundant features, (2) evaluation and application of imputation strategies for missing data, and (3) multicollinearity checks for feature refinement.

2.2.1 | Correlation filtering

To identify redundant features, we conducted a correlation analysis by computing a hybrid correlation matrix to handle mixed-type features (numerical and categorical features), as shown in FIGURE 1(A). Specifically, we combined Spearman's rank correlation for numeric–numeric feature pairs [19], Cramér's V for categorical–categorical feature pairs [20], and asymmetric normalized mutual information (NMI) for categorical–numeric feature pairs [21].

We identified feature pairs with correlation coefficient > 0.8. For each pair, we assessed the ability of each feature to predict the binary target (stable vs. converted) using logistic regression (LR), evaluated by area under the receiver operating characteristic curve (AUC) [22] and Akaike information criterion (AIC) [23]. We then removed features with lower predictive contribution. Biomarkers with established biological relevance were not removed solely because of high pairwise association. This model-guided reduction preserved informative and nonredundant predictors.

To evaluate whether the prespecified correlation threshold of 0.8 provided a reasonable balance for redundancy screening, a threshold sensitivity analysis was conducted using thresholds of 0.7, 0.8 and 0.9. This analysis compared the number of high-correlation pairs and the number of features involved in high-correlation pairs across thresholds.

2.2.2 | Data Imputation Methods

After correlation filtering, we evaluated six imputation methods separately within each dataset (A, B, and C in TABLE 1). The workflow is illustrated in FIGURE 1(B). We first isolated complete-case subsets (rows without missing values). To reduce the influence of extreme multivariate observations during imputation method evaluation, outliers were identified and removed using Isolation Forest with a contamination rate of

5% [24]. This yielded A0 (N = 205), B0 (N = 103), and C0 (N = 78), which formed the basis for evaluating imputation methods.

To simulate a missing-at-random (MAR) scenario [25], we randomly masked 30% of the data in each subset, creating MAR A0, MAR B0, and MAR C0. We applied six imputation methods to these MAR subsets, including random forest (RF) imputation [26], multivariate imputation by chained equations (MICE) [27], MissForest [28], autoencoder-based (AE) imputation [29], generative adversarial imputation networks (GAIN) [30], and k-nearest neighbors (KNN) imputation [31]. We evaluated the methods with two complementary strategies. First, we visualized the distributional fidelity of each imputation method by the kernel density estimation (KDE) difference (Δ KDE) between imputed and original values. Second, we conducted residual analysis by comparing imputed values with the corresponding original values and reporting mean error, standard deviation (Std), and root mean squared error (RMSE) [32]. We retained the two representative imputation methods and applied them to the original datasets A, B, and C, resulting in six imputed datasets, denoted as A-Imp1, A-Imp2, B-Imp1, B-Imp2, C-Imp1, and C-Imp2. The concrete methods corresponding to Imp1 and Imp2 are reported in the Results section. This evaluation, anchored to real complete data, reduced methodological bias and information distortion.

2.2.3 | Multicollinearity Check

For each of the six imputed datasets, we conducted multicollinearity diagnostics to further refine the feature sets and reduce the risk of instability in model estimation. The process is illustrated in FIGURE 1(C). We computed variance inflation factor (VIF) for all predictors [33]. High VIF typically signals redundancy or collinearity. We excluded features with VIF > 10 using an iterative procedure [34]. At each iteration, the feature with the highest current VIF was removed, and VIF values were recalculated for the remaining features until all retained features had VIF values below the threshold. This step targeted one-to-many linear dependencies not captured by pairwise inspection. It was performed after imputation to identify and remove latent redundancy previously masked by missingness. To assess the influence of the VIF threshold, thresholds of 5, 10, and 15 were also examined as a feature-level sensitivity analysis. This step ensured the inclusion of nonredundant predictors, thereby enhancing the numerical reliability of the resulting models.

2.3 | Participant-level Data Split

To prevent information leakage, the train/test split was performed at the participant level. Once a participant was assigned to either the training set or the test set, all visits from that participant were included only in that set. The split was stratified by participant-level conversion status, defined as whether a participant had

at least one visit labeled as conversion. The same participant-level split was applied consistently to datasets A, B, and C.

The training set was used for model training and hyperparameter tuning. The test set was used only for final model evaluation.

2.4 | ML Models

We assessed the predictive capacity of multimodal clinical data for SCD-to-MCI/AD conversion by developing a structured ML framework on the six refined imputed datasets.

The modeling workflow is illustrated in FIGURE 1(D). On each training set, we developed 7 individual classifiers, including LR, linear discriminant analysis (LDA), support vector machines (SVM) [35], decision tree (DT), KNN, RF [36], and XGBoost [37]. Hyperparameters were tuned by grid search with cross-validation. These classifiers span linear, nonlinear, and tree-based ensemble paradigms, enabling broad comparative evaluation.

To leverage the complementary strengths of the classifiers, we constructed 2 heterogeneous ensembles [38, 39]. A soft voting classifier, which averages class probabilities from base learners and aggregates predictions, and a two-layer stacking classifier with LR as a meta-learner. LR was chosen for simplicity, interpretability, and low overfitting risk in stacking architectures for structured data [40]. More complex

nonlinear meta-learners, such as gradient boosting or neural networks, were not explored because the number of converted participants was limited, and such models could increase the risk of overfitting at the meta-learning stage.

We evaluated 7 individual classifiers and 2 heterogeneous ensembles to test whether combining learners improved performance. Performance on test sets was quantified by balanced accuracy, sensitivity, specificity, F1-score, matthews correlation coefficient (MCC), and AUC [41–43]. This multi-model, multi-metric framework supports rigorous comparison across datasets, imputation choice, and algorithms to identify suitable strategies for predicting SCD progression.

After the model comparison, one model was selected as the main model for additional evaluation. The selected model was further assessed for probability calibration using a calibration curve and Brier score [44]. Model interpretation was performed using SHAP, including a summary plot and a dependence plot for a key predictor [45]. To examine whether the model was sensitive to specific predictors that may reflect spurious associations, a robustness analysis was performed by removing selected predictors and retraining the same model configuration.

3 | RESULTS

3.1 | Correlation Filtering Results

Following Section 2.2.1, a threshold sensitivity analysis was conducted using thresholds of 0.7, 0.8, and 0.9. As summarized in TABLE 2(a), a threshold of 0.7 identified a broad redundancy set. In contrast, a threshold of 0.9 was more conservative. The threshold of 0.8 provided a balanced criterion, supporting its use in the main analysis.

At the threshold of 0.8, 31 high-correlation pairs were identified in A, 15 in B, and 31 in C. These pairs mainly involved overlapping cognitive tests, derived memory measures, and everyday cognitive (Ecog) function measures reported by participants or study partners. For each pair, the retained feature was selected based on AUC/AIC comparisons and clinical considerations.

As shown in TABLE 2(b), 14 redundant features were removed from each dataset, reducing the number of predictors from 73 to 59 in A, from 72 to 58 in B, and from 78 to 64 in C.

3.2 | Evaluation of Imputation Methods

We evaluated six imputation methods and used p-tau as a benchmark in each dataset (A with CSF p-tau, B with plasma p-tau, and C with CSF p-tau). It was selected for its biological relevance to SCD progression

despite 68—79% missingness. While variables with this level of missingness are typically excluded, p-tau was retained to explore effective strategies for handling significant missingness in neurodegenerative research.

FIGURE 2 shows comparative results. FIGURE 2(A)–2(C) report Δ KDE between original and imputed p-tau in subsets A0, B0, and C0. In A0, RF, MissForest, and MICE curves are closer to zero, indicating better distributional fidelity, while other methods, especially AE and KNN, exhibit larger deviations. In C0, MissForest shows relatively small deviations in the lower-to-mid range, whereas other methods show larger peaks or troughs. In the mid-to-high range, MICE and RF show smaller deviations and remain closer to zero than other methods. For context, in ADNI, the Elecsys CSF p-tau cut-off is 24 pg/mL [46]. Around this cut-off in A0 and C0, MissForest, MICE and RF show smaller absolute Δ KDE than other methods. In B0, no method shows a consistently near-zero curve across the full range. MICE shows smaller deviations than the other methods over much of the low-to-mid range, but the curve shows no clear systematic pattern. This irregular pattern may reflect local sparsity or uneven density in the observed plasma p-tau distribution.

FIGURE 2(D) summarizes residual statistics. RF shows the lowest RMSE in A0 (2.12) and B0 (7.87), while MissForest achieves the lowest RMSE in C0 (3.40). MissForest also shows competitive residual performance in A0 (2.17) and B0 (8.59). MICE achieves relatively low RMSE in B0 (8.03) and C0 (5.42),

although its RMSE in A0 (3.25) is higher than RF and MissForest. AE, GAIN, and KNN generally show larger RMSE or larger distributional deviations in the Δ KDE comparison.

Overall, RF, MissForest, and MICE provide relatively balanced performance between distributional fidelity and residual behavior. Although RF achieves low residual errors, it is closely related to MissForest as a tree-based method, and retaining both would provide limited methodological complementarity. Therefore, MissForest was retained as the tree-based representative, while MICE was retained as a complementary chained imputation method for final imputation.

Accordingly, MissForest and MICE were applied to impute datasets A, B, and C, producing six imputed datasets (A-MICE, A-MissForest, B-MICE, B-MissForest, C-MICE, and C-MissForest).

3.3 | Multicollinearity Check Results

In TABLE 2(c), the VIF threshold sensitivity analysis showed that a stricter threshold of 5 identified 45–51 features exceeding the corresponding VIF threshold before iterative removal, whereas a more permissive threshold of 15 identified 39–48 features. The threshold of 10 identified 42–50 features, providing an intermediate feature set between the stricter threshold and the more permissive threshold. This supported the use of 10 as the primary threshold for balancing multicollinearity reduction and feature retention.

Using the primary threshold of VIF > 10, features with extreme VIFs were removed, typified by $VIF_{\text{baseline MMSE}} > 2000$, as well as features with high VIFs due to overlap in cognitive domains, such as ECog questionnaire and learning and memory assessments. As shown in TABLE 2(d), iterative VIF filtering retained 20 and 18 features in A-MICE and A-MissForest, 23 and 21 in B-MICE and B-MissForest, and 25 and 22 in C-MICE and C-MissForest, respectively. The resulting feature sets provided leaner inputs for ML models with lower variance inflation due to multicollinearity.

3.4 | Data Distribution After Participant-level Split

As shown in TABLE 3, the participant-level split produced a training set of 444 participants and a test set of 191 participants. The participant-level conversion proportions were similar between the training and the test set, with 89 converted participants (20.05%) in the training set and 39 converted participants (20.42%) in the test set. At the visit level, the training set contained 1,694 visits, including 387 conversion-labeled visits (22.85%), whereas the test set contained 729 visits, including 156 conversion-labeled visits (21.40%). No participant overlap was observed between the training and the test set.

3.5 | Model Performance Across Datasets

We evaluated 7 individual ML models and 2 heterogeneous ensembles on the six imputed datasets to study effects of data modality, imputation strategy, and model type on predicting conversion from SCD to MCI or AD.

TABLE 4 summarizes performance across all settings. Considering the class imbalance, model selection prioritized performance on the minority class, with balanced accuracy, sensitivity, F1-score, MCC, and AUC used as the main criteria. These metrics were more informative than accuracy alone because the converted group was the minority class.

The best-performing model in Dataset A was SVM trained on A-MICE (balanced accuracy = 0.69, sensitivity = 0.72, specificity = 0.66, F1-score = 0.48, MCC = 0.31, AUC = 0.72). In Dataset B, SVM trained on B-MICE achieved the best overall performance (balanced accuracy = 0.70, sensitivity = 0.72, specificity = 0.68, F1-score = 0.50, MCC = 0.33, AUC = 0.73). This model achieved the highest F1-score among all settings and shared the highest balanced accuracy and MCC with LR on B-MICE. In Dataset C, LR trained on C-MICE showed the best trade-off between sensitivity and specificity (balanced accuracy = 0.68, sensitivity = 0.69, specificity = 0.67, F1-score = 0.48, MCC = 0.30, AUC = 0.74). Although RF trained on C-MissForest achieved a slightly higher MCC (0.31), its lower sensitivity (0.52) made it less suitable as the main model for identifying converters.

These results further indicate that data modality affected model performance. Across data modality, the plasma-based dataset B achieved the strongest overall model in the evaluation, while the combined dataset C achieved the highest AUC in several models but did not consistently improve balanced classification performance. This suggests that adding more modalities did not necessarily improve prediction.

Regarding imputation methods, MICE-imputed variants produced the best models in Datasets A, B, and C. MissForest-imputed variants showed competitive AUC, specificity, or MCC in several settings, but they often had lower sensitivity, indicating a less balanced classification of converters and stable cases.

Across model types, SVM and LR showed the most consistent performance under the criteria. LR achieved competitive performance in several settings and was close to SVM on B-MICE, whereas LDA often showed high specificity but low sensitivity, suggesting a tendency to favor the majority class under class imbalance. Tree-based models, including XGBoost and RF, also tended to show higher specificity than sensitivity, indicating that they classified stable cases more reliably but were less consistent in identifying converters. KNN demonstrated limited ability to capture complex patterns, as reflected in lower sensitivity and MCC, consistent with a prior study reporting that distance-based models often underperform when handling high-dimensional clinical data [47]. Although prior work reported gains from heterogeneous ensembles [39], we did not observe this effect, suggesting limited benefit when a strong base learner and class imbalance were present.

On the basis of the comparison, SVM trained on B-MICE was selected as the main model for subsequent calibration, SHAP-based interpretation, and robustness analyses. This model was preferred because it provided the strongest overall balance between identifying converters and avoiding excessive false positives. This balance is important in SCD progression prediction because missing future converters may have greater clinical consequences than only misclassifying stable cases.

3.6 | Selected Model Analysis

The selected B-MICE SVM model was further evaluated on the test set using participant-level bootstrap 95% confidence interval (CI). As shown in TABLE 5, the model showed meaningful discriminative ability and maintained a balanced sensitivity-specificity profile on the test set, indicating that the model did not rely exclusively on the majority class.

3.6.1 | Calibration Analysis

The calibration curve of the selected B-MICE SVM model is shown in FIGURE 3. The predicted probabilities were mainly distributed below 0.55, indicating that the model produced relatively conservative probability estimates rather than extreme high-risk predictions. The curve broadly followed the diagonal

trend, but local deviations were observed. In the lower predicted probability range, approximately 0.08–0.22, most bins lay below the ideal line, indicating mild overestimation, although this pattern was not uniform across all bins. In the intermediate range, around 0.27–0.32, the observed conversion rate exceeded the predicted probability, suggesting local underestimation. At higher observed predicted probabilities, around 0.38–0.52, the curve again fell slightly below the ideal line, indicating mild overestimation.

The Brier score of the model was 0.15 with a 95% CI of 0.12–0.18, indicating moderate probability-level prediction error. Overall, the calibration analysis suggests that the B-MICE SVM model provided usable but imperfect probability estimates, complementing its discrimination-based performance metrics.

3.6.2 | SHAP-based Interpretation

SHAP analysis was used to interpret the selected B-MICE SVM model. As shown in FIGURE 4(A), baseline ADAS-cog 13 score (ADAS13_bl), mPACC Trails B score (mPACCtrailsB), APOE4, RAVLT percentage forgetting (RAVLT_perc_forgetting), baseline mPACC Trails B score (mPACCtrailsB_bl), plasma Aβ40, and p-tau were among the main contributors to the model output. Higher ADAS13_bl, APOE4 positivity, and greater RAVLT percentage forgetting generally produced positive SHAP values, indicating a stronger contribution toward conversion class. In contrast, higher mPACCtrailsB and mPACCtrailsB_bl

values mainly produced negative SHAP values, suggesting that better cognitive performance contributed to stable class.

Taking plasma p-tau as an example, the SHAP dependence plot is shown in FIGURE 4(B). As plasma p-tau increased, its SHAP value shifted from negative to positive, with values below approximately 18–20 mainly contributing negatively and higher values contributing positively to the model output. This indicates that plasma p-tau values beyond this approximate range shifted the model output toward the conversion class. However, this threshold should be interpreted only as a model explanation, not as a clinically actionable cutoff, and clinical validation is necessary.

3.6.3 | Robustness Analysis

To examine whether the B-MICE SVM model was overly dependent on potentially confounded demographics, an additional robustness analysis was performed by removing marital status, race, and ethnicity from the input feature set. The same SVM modeling procedure was then repeated on the reduced feature set, and its performance was compared with that of the original B-MICE SVM.

As shown in TABLE 6, excluding marital status, race, and ethnicity did not materially reduce the overall performance of the B-MICE SVM. Balanced accuracy remained unchanged at 0.70, while AUC slightly increased from 0.73 to 0.75. MCC and Brier score were also unchanged at 0.33 and 0.15, respectively,

and F1 showed only a small decrease from 0.50 to 0.49. Sensitivity increased from 0.72 to 0.77, whereas specificity decreased from 0.68 to 0.64. These results indicate acceptable robustness, suggesting that the model was not primarily driven by these demographics. However, the decrease in specificity indicates that these features may still contribute to the balance between sensitivity and specificity.

4 | DISCUSSION

This study developed and evaluated a machine learning framework for predicting 10-year conversion from SCD to MCI or AD using multimodal nonimaging clinical data. The analysis combined visit-level prospective labeling, participant-level train-test split, systematic preprocessing for missingness and redundancy, comparison of multiple models across different data configurations, and additional evaluation of the best model through calibration, SHAP-based interpretation, and robustness analysis. Rather than identifying a single universally superior algorithm, the findings show that prediction performance depends jointly on the data modality, imputation method, feature filtering strategy, and model type.

As an initial step in the analysis, systematic preprocessing was an important component affecting the reliability and interpretability of model development. For correlation filtering, a hybrid strategy was employed to capture correlations across mixed feature types, particularly between unordered categorical features and continuous measures, without the need for preliminary imputation of missing values. This

extends beyond traditional pairwise correlation matrices, which are typically based on Pearson's correlation coefficient, limited to assessing linear relationships between pairs of continuous numerical features, and do not account for associations involving categorical features. For missing values, rather than discarding incomplete records, six imputation methods were compared based on distributional fidelity and residual error, ultimately selecting MissForest and MICE. Although Liu et al. [14] addressed missing data in SCD progression using deep learning-based imputation for imaging features, our study is, to our knowledge, among the first to systematically evaluate and compare multiple imputation methods, including traditional and ML-based approaches, on a multimodal clinical dataset in this field. For multicollinearity check, this step played a critical role in preparing the data for interpretable and stable modeling. While earlier correlation filtering targeted pairwise redundancy, multicollinearity analysis captures one-to-many linear dependencies, where a single feature can be explained by a combination of multiple other features, relationships that may not be evident through pairwise inspection alone.

The model comparison suggests that balanced classification is more informative than overall discrimination alone in this task. Because converters were the minority class, models with high specificity but low sensitivity were less suitable for identifying individuals at risk of progression. The selected B-MICE SVM model provided the most balanced profile across the main evaluation metrics, whereas several tree-based and ensemble models tended to favor stable cases. This finding indicates that more complex models do not

necessarily improve converter detection. In this setting, model balance, sensitivity to the minority class, and robustness are more relevant than maximizing a single metric.

The comparison across data modalities also indicates the potential value of plasma-based prediction. The combined dataset C contained the most complete fluid biomarker profile and achieved the highest AUC in several models, but it did not consistently improve balanced classification performance. This suggests that adding more modalities does not necessarily improve prediction when additional variables introduce missingness, redundancy, or noise. In contrast, the plasma-based B-MICE provided a suitable model. Since plasma biomarkers are less invasive than CSF biomarkers and more practical than MRI or PET in many primary screening contexts, and a previous study have shown that plasma biomarkers can contribute to the prediction of cognitive decline in cognitively unimpaired elderly populations [48], this result supports the potential value of plasma-based nonimaging models. Nevertheless, the current performance should be suitable for risk stratification or pre-screening research rather than clinical diagnosis.

Our work extends the recent study by Dolcet et al. [49], who developed ML models to predict SCD-to-MCI conversion using demographics, neuropsychological, and structural MRI-derived features in a small cohort (99 participants, including 32 converters). While their best-performing model, XGBoost with oversampling, achieved impressive results in a single holdout validation (AUC = 0.96, sensitivity = 1.00, specificity = 0.92), these metrics were unstable under repeated evaluation. Across 50 random train/test splits,

their model's median AUC dropped to 0.75 (IQR = 0.11), sensitivity to 0.83 (IQR = 0.17), and specificity to 0.77 (IQR = 0.23), revealing considerable variability across runs and a decline in generalizability. This instability was likely due to their repeated splitting approach, which relied on random partitioning, making performance evaluation sensitive to specific data splits, particularly in small cohorts. In contrast, we used larger datasets (635 participants) with a participant-level train-test split to reduce leakage across repeated visits. Although the B-MICE SVM model did not achieve higher discrimination, it provided a balanced sensitivity-specificity profile using nonimaging variables.

There is a fundamental difference in the input features. The study by Dolcet et al. [49] relied on MRI-derived gray matter volumes, with over 100 structural brain features, requiring specialized imaging infrastructure and higher acquisition costs. In contrast, B-MICE SVM used more accessible and lower-cost plasma data. This makes our approach particularly suited for implementation in resource-limited or primary care settings. Therefore, its main value lies in feasibility and cost-effectiveness rather than in outperforming imaging-based models.

Compared with imaging-heavy approaches, the present framework has a different intended role. Some previous studies used MRI or PET to predict progression from SCD to MCI or AD, which may provide strong biological information but requires specialized imaging infrastructure. The present study focused on cost-effective nonimaging features, particularly plasma biomarkers, cognitive assessments, APOE4

status, and demographics. Therefore, the current framework should be viewed as a practical nonimaging alternative for early screening, especially in situations where resources are limited and the application of structural neuroimaging is restricted.

The clinical implications of these findings are notable. Our study suggests that plasma-based non-imaging clinical data can support early prediction of SCD progress without relying on advanced imaging. This positions our best model as a viable tool for early screening and patient stratification in settings where resource constraints limit the use of structural neuroimaging.

Despite the strengths of our approach, several limitations should be considered. First, our data were derived entirely from the ADNI cohort, which may not fully represent more diverse or community-based populations. This could affect how well the model generalizes to underrepresented groups. Therefore, external validation in independent and more diverse cohorts remains necessary before clinical deployment. Second, although we focused on predicting whether individuals would convert within 10 years, our method does not account for the exact timing of conversion. Third, more complex deep generative imputation models and deep multimodal fusion models, such as transformer based or diffusion based approaches, were not included in this study. Given the structured tabular nature of the data, the limited participant number, and the absence of an external validation cohort, this study prioritized interpretable and reproducible models with lower overfitting risk.

Future research will extend this study in several directions. We will focus on external validation and more clinically aligned modeling. Independent cohorts should be used to test whether the selected plasma-based model maintains performance across different populations. Larger independent cohorts should also be used to compare the present framework with more advanced deep generative imputation and multimodal fusion approaches. We will explore time-to-event modeling, such as Cox proportional hazards models or flexible survival learners, to estimate risk over time rather than only whether conversion occurs. We will develop a clinician-in-the-loop decision support tool that surfaces case-level explanations, enables interactive adjustment of thresholds and feature inclusion, and recomputes risk in real time.

In conclusion, this study presents an interpretable and cost-efficient framework for predicting 10-year conversion from SCD to MCI or AD. By integrating multimodal clinical data, applying rigorous imputation and dimensionality control, and employing transparent ML strategies, the B-MICE SVM model was identified as the most balanced model. The model showed balanced predictive performance and was further assessed using calibration analysis, SHAP-based interpretation, and robustness analysis. All findings suggest that nonimaging clinical and plasma biomarkers support early prediction for SCD progression and contribute to the future development of practical tools for individualized dementia prevention research.

CREDIT AUTHOR CONTRIBUTION STATEMENT

Yiping Qian: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Andrew Doig:** Conceptualization, Validation, Writing - Review & Editing, Supervision. **Sofia Michopoulou:** Conceptualization, Validation, Writing - Review & Editing, Supervision. **Fumie Costen:** Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

ETHICS AND CONSENT STATEMENT

This study is a secondary analysis of de-identified data obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The ADNI study protocol was approved by the institutional review board/ethics committee at each participating site, and written informed consent was obtained from all participants (or their legally authorized representatives) at each site. No new human participant recruitment, intervention, or data collection was performed for the present work. Access to ADNI data was granted through an approved ADNI online data access application and acceptance of the ADNI Data Use Agreement, and all analyses were conducted in accordance with ADNI data use policies.

DATA AND CODE AVAILABILITY STATEMENT

The data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Due to privacy or ethical restrictions, ADNI data are not publicly downloadable without registration. Qualified investigators may apply for access through the ADNI website (adni.loni.usc.edu) by completing the required online data access application and agreeing to the ADNI Data Use Agreement. The raw ADNI data cannot be redistributed by the authors. The analysis code and derived processed outputs supporting the findings of this study are not deposited in a public repository, but are available from the corresponding author upon reasonable request, subject to ADNI data use policies and applicable privacy restrictions.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGMENTS

Data collection and sharing for the Alzheimer's Disease Neuroimaging Initiative (ADNI) is funded by the National Institute on Aging (National Institutes of Health Grant U19AG024904). The grantee organization is the Northern California Institute for Research and Education. In the past, ADNI has also received funding from the National Institute of Biomedical Imaging and Bioengineering, the Canadian Institutes of Health Research, and private sector contributions through the Foundation for the National Institutes of Health (FNIH) including generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; BristolMyers 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.

REFERENCES

[1] Y.-W. Cheng, T.-F. Chen, M.-J. Chiu, From mild cognitive impairment to subjective cognitive decline: conceptual and methodological evolution, Neuropsychiatric Disease and Treatment 13 (2017) 491–498, https://doi.org/10.2147/NDT.S123428.

[2] A. Rostamzadeh, L. Bohr, M. Wagner, C. Baethge, F. Jessen, Progression of subjective cognitive decline to mci or dementia in relation to biomarkers for alzheimer disease: a meta-analysis, Neurology 99 (17) (2022) e1866–e1874, https://doi.org/10.1212/WNL.0000000000201072.

[3] M. A. Fernández-Blázquez, M. Ávila-Villanueva, F. Maestú, M. Medina, Specific features of subjective cognitive decline predict faster conversion to mild cognitive impairment, Journal of Alzheimer's Disease 52 (1) (2016) 271–281, https://doi.org/10.3233/JAD-150956.

[4] L. Yue, D. Hu, H. Zhang, J. Wen, Y. Wu, W. Li, L. Sun, X. Li, J. Wang, G. Li, et al., Prediction of 7-year's conversion from subjective cognitive decline to mild cognitive impairment, Human Brain Mapping 42 (1) (2021) 192–203, https://doi.org/10.1002/hbm.25216.

[5] C. H. Van Dyck, C. J. Swanson, P. Aisen, R. J. Bateman, C. Chen, M. Gee, M. Kanekiyo, D. Li, L. Reyderman, S. Cohen, et al., Lecanemab in early alzheimer's disease, New England Journal of Medicine 388 (1) (2023) 9–21, https://doi.org/10.1056/NEJMoa2212948.

[6] J. R. Sims, J. A. Zimmer, C. D. Evans, M. Lu, P. Ardayfio, J. D. Sparks, A. M. Wessels, S. Shcherbinin, H. Wang, E. S. Monkul Nery, E. C. Collins, P. Solomon, S. Salloway, L. G. Apostolova, O. Hansson, C. Ritchie, D. A. Brooks, M. Mintun, D. M. Skovronsky, Donanemab in early symptomatic Alzheimer disease: The TRAILBLAZER-ALZ 2 randomized clinical trial, JAMA 330 (6) (2023) 512–527, https://doi.org/10.1001/jama.2023.13239.

[7] M. Scarth, I. Rissanen, R. J. Scholten, M. I. Geerlings, Biomarkers of alzheimer's disease and cerebrovascular lesions and clinical progression in patients with subjective cognitive decline: A systematic review, Journal of Alzheimer's Disease 83 (3) (2021) 1089–1111, https://doi.org/10.3233/JAD-210218.

[8] S. Wolfsgruber, A. Polcher, A. Koppara, L. Kleineidam, L. Frölich, O. Peters, M. Hüll, E. Rüther, J. Wiltfang, W. Maier, et al., Cerebrospinal fluid biomarkers and clinical progression in patients with subjective cognitive decline and mild cognitive impairment, Journal of Alzheimer's Disease 58 (3) (2017) 939–950, https://doi.org/10.3233/JAD-161252.

[9] C. R. Jack Jr, D. A. Bennett, K. Blennow, M. C. Carrillo, B. Dunn, S. B. Haeberlein, D. M. Holtzman, W. Jagust, F. Jessen, J. Karlawish, et al., Nia-aa research framework: toward a biological definition of alzheimer's disease, Alzheimer's & dementia 14 (4) (2018) 535–562, https:

//doi.org/10.1016/j.jalz.2018.02.018.

[10] D. Mengel, E. Soter, J. M. Ott, M. Wacker, A. Leyva, O. Peters, J. Hellmann-Regen, L.-S. Schneider, X. Wang, J. Priller, et al., Blood biomarkers confirm subjective cognitive decline (scd) as a distinct molecular and clinical stage within the nia-aa framework of alzheimer´ s disease, Molecular Psychiatry 30 (7) (2025) 3150–3159, https://doi.org/10.1038/s41380-025-03021-0.

[11] V. Bessi, S. Mazzeo, S. Padiglioni, C. Piccini, B. Nacmias, S. Sorbi, L. Bracco, From subjective cognitive decline to alzheimer's disease: The predictive role of neuropsychological assessment, personality traits, and cognitive reserve. a 7-year follow-up study, Journal of Alzheimer's Disease 63 (4) (2018) 1523–1535, https://doi.org/10.3233/JAD-171180.

[12] F.-F. Pan, L. Huang, K.-l. Chen, Q.-h. Zhao, Q.-h. Guo, A comparative study on the validations of three cognitive screening tests in identifying subtle cognitive decline, BMC neurology 20 (2020) 78, https://doi.org/10.1186/s12883-020-01657-9.

[13] S. Mazzeo, M. Lassi, S. Padiglioni, V. Moschini, A. A. Vergani, M. Scarpino, G. Giacomucci, R. Buralli, C. Morinelli, C. Fabbiani, et al., A machine learning model to predict progression from subjective cognitive decline to mild cognitive impairment and dementia: a 12-year follow-up study, Alzheimer's & Dementia 19 (S24) (2023) e082276, https://doi.org/10.1002/alz.082276.

[14] Y. Liu, L. Yue, S. Xiao, W. Yang, D. Shen, M. Liu, Assessing clinical progression from subjective cognitive decline to mild cognitive impairment with incomplete multi-modal neuroimages, Medical image analysis 75 (2022) 102266, https://doi.org/10.1016/j.media.2021.102266.

[15] T. Guo, S. M. Landau, W. J. Jagust, A. D. N. Initiative, Detecting earlier stages of amyloid deposition using pet in cognitively normal elderly adults, Neurology 94 (14) (2020) e1512–e1524, https://doi.org/10.1212/WNL.0000000000009216.

[16] C. R. Jack, D. S. Knopman, W. J. Jagust, R. C. Petersen, M. W. Weiner, P. S. Aisen, L. M. Shaw, P. Vemuri, H. J. Wiste, S. D. Weigand, et al., Tracking pathophysiological processes in alzheimer's disease: an updated hypothetical model of dynamic biomarkers, The lancet neurology 12 (2) (2013) 207–216, https://doi.org/10.1016/S1474-4422(12)70291-0.

[17] F. Jessen, R. E. Amariglio, M. Van Boxtel, M. Breteler, M. Ceccaldi, G. Chételat, B. Dubois, C. Dufouil, K. A. Ellis, W. M. Van Der Flier, et al., A conceptual framework for research on subjective

cognitive decline in preclinical alzheimer's disease, Alzheimer's & dementia 10 (6) (2014) 844–852, https://doi.org/10.1016/j.jalz.2014.01.001.

[18] A. Studart, R. Nitrini, Subjective cognitive decline: The first clinical manifestation of alzheimer's disease?, Dementia & neuropsychologia 10 (03) (2016) 170–177, https://doi.org/10.1590/S1980-5764-2016DN1003002.

[19] J. H. Zar, Spearman rank correlation, Encyclopedia of biostatistics 7 (2005), https://doi.org/10.1002/0470011815.b2a15150.

[20] M. S. Ben-Shachar, I. Patil, R. Thériault, B. M. Wiernik, D. Lüdecke, Phi, fei, fo, fum: effect sizes for categorical data that use the chi-squared statistic, Mathematics 11 (9) (2023) 1982, https://doi.org/10.3390/math11091982.

[21] B. C. Ross, Mutual information between discrete and continuous data sets, PloS one 9 (2) (2014) e87357, https://doi.org/10.1371/journal.pone.0087357.

[22] J. A. Hanley, B. J. McNeil, The meaning and use of the area under a receiver operating characteristic (roc) curve., Radiology 143 (1) (1982) 29–36, https://doi.org/10.1148/radiology.143.1.7063747.

[23] H. Akaike, A new look at the statistical model identification, IEEE transactions on automatic control 19 (6) (1974) 716–723, https://doi.org/10.1109/TAC.1974.1100705.

[24] F. T. Liu, K. M. Ting, Z.-H. Zhou, Isolation forest, in: 2008 eighth ieee international conference on data mining, IEEE, 2008, pp. 413–422, https://doi.org/10.1109/ICDM.2008.17.

[25] D. B. Rubin, Inference and missing data, Biometrika 63 (3) (1976) 581–592, https://doi.org/10.1093/biomet/63.3.581.

[26] F. Tang, H. Ishwaran, Random forest missing data algorithms, Statistical Analysis and Data Mining: The ASA Data Science Journal 10 (6) (2017) 363–377, https://doi.org/10.1002/sam.11348.

[27] S. Van Buuren, K. Groothuis-Oudshoorn, mice: Multivariate imputation by chained equations in r, Journal of statistical software 45 (2011) 1–67, https://doi.org/10.18637/jss.v045.i03.

[28] D. J. Stekhoven, P. Bühlmann, Missforest—non-parametric missing value imputation for mixed-type data, Bioinformatics 28 (1) (2012) 112–118, https:

//doi.org/10.1093/bioinformatics/btr597.

[29] L. Gondara, K. Wang, Mida: Multiple imputation using denoising autoencoders, in: Pacific-Asia conference on knowledge discovery and data mining, Springer, 2018, pp. 260–272, https://doi.org/10.1007/978-3-319-93040-4_21.

[30] J. Yoon, J. Jordon, M. van der Schaar, Gain: Missing data imputation using generative adversarial nets, in: International conference on machine learning, PMLR, 2018, pp. 5689–5698.
URL https://proceedings.mlr.press/v80/yoon18a.html

[31] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, R. B. Altman, Missing value estimation methods for dna microarrays, Bioinformatics 17 (6) (2001) 520–525, https://doi.org/10.1093/bioinformatics/17.6.520.

[32] D. Yang, R. Khera, E. Chang, M. Joel, G. Janda, H. Park, S. Aneja, Comparison of machine learning approaches for missing data imputation among non-small cell lung cancer patients, International Journal of Radiation Oncology, Biology, Physics 111 (3) (2021) S115, https://doi.org/10.1016/j.ijrobp.2021.07.264.

[33] C. G. Thompson, R. S. Kim, A. M. Aloe, B. J. Becker, Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results, Basic and applied social psychology 39 (2) (2017) 81–90, https://doi.org/10.1080/01973533.2016.1277529.

[34] J. H. Kim, Multicollinearity and misleading statistical results, Korean journal of anesthesiology 72 (6) (2019) 558–569, https://doi.org/10.4097/kja.19087.

[35] C. Cortes, V. Vapnik, Support-vector networks, Machine learning 20 (3) (1995) 273–297, https://doi.org/10.1007/BF00994018.

[36] L. Breiman, Random forests, Machine learning 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[37] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.

[38] J. Kunnen, M. Duchateau, Z. Van Veldhoven, J. Vanthienen, Benchmarking stacking against other heterogeneous ensembles in telecom churn prediction, in: 2020 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2020, pp. 1234–1240, https://doi.org/10.1109/SSCI47803.2020.9308188.

[39] A. Ashfaq, A. Imran, I. Ullah, A. Alzahrani, K. M. A. Alheeti, A. Yasin, Multi-model ensemble based approach for heart disease diagnosis, in: 2022 International Conference on Recent Advances in Electrical Engineering & Computer Sciences (RAEE & CS), IEEE, 2022, pp. 1–8, https://doi.org/10.1109/RAEECS56511.2022.9954490.

[40] J. Tang, J. Liang, C. Han, Z. Li, H. Huang, Crash injury severity analysis using a two-layer stacking framework, Accident Analysis & Prevention 122 (2019) 226–238, https://doi.org/10.1016/j.aap.2018.10.016.

[41] K. H. Brodersen, C. S. Ong, K. E. Stephan, J. M. Buhmann, The balanced accuracy and its posterior distribution, in: 2010 20th international conference on pattern recognition, IEEE, 2010, pp. 3121–3124, https://doi.org/10.1109/ICPR.2010.764.

[42] D. Chicco, G. Jurman, The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation, BMC genomics 21 (1) (2020) 6, https://doi.org/10.1186/s12864-019-6413-7.

[43] T. Fawcett, An introduction to roc analysis, Pattern recognition letters 27 (8) (2006) 861–874, https://doi.org/10.1016/j.patrec.2005.10.010.

[44] E. W. Steyerberg, A. J. Vickers, N. R. Cook, T. Gerds, M. Gonen, N. Obuchowski, M. J. Pencina, M. W. Kattan, Assessing the performance of prediction models: a framework for traditional and novel measures, Epidemiology 21 (1) (2010) 128–138, https://doi.org/10.1097/EDE.0b013e3181c30fb2.

[45] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Advances in neural information processing systems 30 (2017).
URL https://proceedings.neurips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions

[46] K. Blennow, L. M. Shaw, E. Stomrud, N. Mattsson, J. B. Toledo, K. Buck, S. Wahl, U. Eichenlaub, V. Lifke, M. Simon, et al., Predicting clinical decline and conversion to alzheimer's disease or dementia

using novel elecsys aβ (1–42), ptau and ttau csf immunoassays, Scientific reports 9 (1) (2019) 19024, https://doi.org/10.1038/s41598-019-54204-z.

[47] S.-I. Chiu, L.-Y. Fan, C.-H. Lin, T.-F. Chen, W. S. Lim, J.-S. R. Jang, M.-J. Chiu, Machine learning-based classification of subjective cognitive decline, mild cognitive impairment, and alzheimer's dementia using neuroimage and plasma biomarkers, ACS Chemical Neuroscience 13 (23) (2022) 3263–3270, https://doi.org/10.1021/acschemneuro.2c00255.

[48] N. C. Cullen, A. Leuzy, S. Janelidze, S. Palmqvist, A. L. Svenningsson, E. Stomrud, J. L. Dage, N. Mattsson-Carlgren, O. Hansson, Plasma biomarkers of alzheimer's disease improve prediction of cognitive decline in cognitively unimpaired elderly populations, Nature communications 12 (1) (2021) 3555, https://doi.org/10.1038/s41467-021-23746-0.

[49] M. M. Dolcet-Negre, L. Imaz Aguayo, R. García-de Eulate, G. Martí-Andrés, M. Fernández-Matarrubia, P. Domínguez, M. A. Fernández-Seara, M. Riverol, Predicting conversion from subjective cognitive decline to mild cognitive impairment and alzheimer's disease dementia using ensemble machine learning, Journal of Alzheimer's Disease 93 (1) (2023) 125–140, https://doi.org/10.3233/JAD-221002.

FIGURE AND TABLE LEGENDS

FIGURE 1 Preprocessing and model-development workflow after participant-level data split. (A) is a correlation filtering process. (B) is a benchmarking and selection process for imputation methods. (C) is a multicollinearity check process. (D) is a modeling process.

FIGURE 2 Comparison of 6 imputation methods using p-tau as an example. (A)–(C) show the Δ KDE between original and imputed p-tau values for Datasets A0, B0, and C0, respectively. (D) presents residual statistics for each imputation method across A0, B0, and C0.

FIGURE 3 Calibration curve of B-MICE SVM. The dashed diagonal line represents ideal calibration, and the plotted curve shows the agreement between the mean predicted probability and the observed conversion rate across probability bins.

FIGURE 4 SHAP-based interpretation of B-MICE SVM. (A) SHAP summary plot showing the ranked contribution of predictors to the model output. The x-axis represents SHAP values, indicating the direction (positive or negative) and magnitude of effect. The y-axis ranks features by importance. Each dot represents a sample, and the color reflects the feature value (red indicates high values, blue indicates low values). (B) SHAP dependence plot for plasma p-tau, showing the relationship between plasma p-tau 181 values and their SHAP contribution to predicted conversion. Positive SHAP values indicate contributions toward predicted conversion.

TABLE 1 Overview of variables included in each dataset (A, B and C).

TABLE 2 Feature reduction across preprocessing.

TABLE 3 Data distribution after participant-level split.

TABLE 4 Summarizes all performance metrics across models and datasets.

TABLE 5 Performance of the model (B-MICE SVM) with 95% confidence intervals (CI).

TABLE 6 Robustness analysis of the B-MICE SVM after excluding confounded demographics.