Alzheimer's
Research & Therapy

RESEARCH

Open Access

Machine learning prediction of future amyloid beta positivity in amyloid-negative individuals

Elaheh Moradi[1*], Mithilesh Prakash[1], Anette Hall[2,3], Alina Solomon[2,3,4], Bryan Strange[5,6], Jussi Tohka[1] and for the Alzheimer's Disease Neuroimaging Initiative

Abstract

Background The pathophysiology of Alzheimer's disease (AD) involves β-amyloid (Aβ) accumulation. Early identification of individuals with abnormal β-amyloid levels is crucial, but Aβ quantification with positron emission tomography (PET) and cerebrospinal fluid (CSF) is invasive and expensive.

Methods We propose a machine learning framework using standard non-invasive (MRI, demographics, APOE, neuropsychology) measures to predict future Aβ-positivity in Aβ-negative individuals. We separately study Aβ-positivity defined by PET and CSF.

Results Cross-validated AUC for 4-year Aβ conversion prediction was 0.78 for the CSF-based and 0.68 for the PET-based Aβ definitions. Although not trained for the clinical status-change prediction, the CSF-based model excelled in predicting future mild cognitive impairment (MCI)/dementia conversion in cognitively normal/MCI individuals (AUCs, respectively, 0.76 and 0.89 with a separate dataset).

Conclusion Standard measures have potential in detecting future Aβ-positivity and assessing conversion risk, even in cognitively normal individuals. The CSF-based definition led to better predictions than the PET-based definition.

Keywords Machine learning, Amyloid beta, Conversion prediction, Alzheimer's disease, Mild cognitive impairment

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

*Correspondence:
Elaheh Moradi
elaheh.moradi@uef.fi
[1] A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio 70150, Finland
[2] Institute of Clinical Medicine/Neurology, University of Eastern Finland, Kuopio, Finland
[3] Division of Clinical Geriatrics, Center for Alzheimer Research, Karolinska Institute, Stockholm, Sweden
[4] Ageing Epidemiology Research Unit, School of Public Health, Imperial College London, London, UK
[5] Laboratory for Clinical Neuroscience, Center for Biomedical Technology, Universidad Politécnica de Madrid, IdISSC, Madrid, Spain
[6] Reina Sofia Centre for Alzheimer's Research, Madrid, Spain

Introduction

Alzheimer's disease (AD) is a common neurodegenerative disease with a complex and unclear pathway and a long prodromal phase. The progressive and irreversible nature of AD highlights the need for detecting early changes in the brain that occur decades before dementia. Research on amyloid, tau, and neurodegeneration (ATN) biomarkers, in accordance with the NIA-AA 2018 framework [1], has focused on tracking AD progression through beta-amyloid (Aβ) and Tau protein accumulations. These biomarkers are associated with neurodegeneration and cognitive decline [2]. Identifying amyloid burden in cognitively normal individuals holds promise for identifying those at risk of developing AD [3], and it is expected to become the standard for prescribing Aβ-targeted drugs [4]. Currently, cerebrospinal fluid (CSF) and positron emission tomography (PET) with 18F-labeled amyloid tracers are established methods for confirming the presence and measuring the extent of Aβ accumulation in the brain [5]. PET can detect metabolic and biochemical alterations in the brain deviating from normal. The clearance efficiency of Aβ protein aggregates can be detected in CSF [6]. CSF peptides ($A\beta_{1\text{-}42}$) and hyperphosphorylated tau are correlated with amyloid plaques and neuronal tangles observed in brain autopsies [7] and are linked to cognitive decline, offering valuable insights into early detection. Despite their significance, PET and CSF are not widely available. CSF can cause discomfort and are invasive, while PET involves exposure to radiation and requires specialized equipment and personnel. Additionally, PET and CSF results can disagree because they measure different aspects of amyloid pathology [8, 9]. Therefore, promoting the use of universally available data, such as demographics and cognitive scores, is important to advance AD research [10, 11].

Standardized and longitudinal datasets such as ADNI (Alzheimer's Disease Neuroimaging Initiative), provide valuable resources for developing machine learning (ML) models with different feature combinations to study PET and CSF biomarkers. These ML models can highlight the important modalities for identifying at-risk individuals transitioning from normal control (NC) to mild cognitive impairment (MCI) and also potentially to dementia by tracking changes in Aβ-positivity states. However, there are limited studies addressing the future predictability of Aβ-positivity using widely available measures [12]. Additionally, the comparative analysis of model performance based on categorizing individuals as Aβ-positive or Aβ-negative using either CSF or PET biomarkers remains understudied.

This study aims to develop an ML-based approach to predict the conversion to Aβ-positivity in individuals who are Aβ-negative using data available widely in clinical settings. We utilize demographic data (age, gender, education), APOE4 (genetic), neuropsychological scores, and MRI-derived brain volumes to predict whether an individual with Aβ-negative will convert to Aβ-positive (referred to as Aβ-Converter) or remain Aβ-negative (referred to as Aβ-Stable) over a 4-year period. Analyses are performed separately and in parallel for grouping individuals based on CSF and PET biomarkers. We analyze the role of different data types in the predictive performance of the model in each CSF and PET-based cohort. Additionally, we examine whether baseline CSF/PET measures improve the predictive performance of the model and how that affects the contribution of cognitive measures and MRI biomarkers. We also investigate the role of baseline PET measures in predicting Aβ-positivity based on CSF (Aβ42) and vice versa, the role of baseline CSF measures in predicting Aβ-positivity based on PET. Finally, we evaluate our model's predictive capability for conversion to MCI/dementia in healthy and MCI individuals. Our study extends to demonstrate that both classification and regression approaches display similar trends.

Materials and methods

ADNI data

Data used in this work were obtained from the ADNI (http://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 goal has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see http://www.adni-info.org.

This study included participants from all phases of ADNI with baseline demographics, APOE4, psychological test results, and MRI biomarkers, who also had available longitudinal CSF measures or 18 F-florbetapir (AV45) PET measures.

CSF and PET measures

CSF samples were collected and processed as previously described [13]. The concentration of Aβ42, pTau, and tTau in CSF was measured using the fully automated Elecsys immunoassays (Roche Diagnostics, Basel, Switzerland) by the ADNI biomarker core (University of Pennsylvania, Philadelphia, PA). We obtained these measures from the ADNI depository (UPENNBIOMK9_04_19_17.csv, UPENNBIOMK10_07_29_19.csv, UPENNBIOMK12_01_04_21.csv). Several cutoffs have been proposed and evaluated for determining Aβ-positivity based on CSF Aβ-42 measure, such as 980 pg/mL determined through ROC analyses with FBP PET as the

endpoint [14]. We selected a cutoff 880 pg/mL, which was derived based on the BioFINDER study, accounting for pre-analytic differences and consequently validated in ADNI [15]. We selected this cutoff for two distinct reasons: (1) among suggested cutoffs it maximizes the number of participants, which is important due to limited participant numbers, particularly in the Aβ-Converter group; (2) the cutoff has been developed independently of ADNI data but validated on ADNI, which we consider as important characteristic [15, 16].

ADNI PET acquisition and processing protocols are described in previously published methods [17, 18] (https://adni.loni.usc.edu/methods/pet-acquisition/, https://adni.loni.usc.edu/methods/pet-analysis-method/pet-analysis/). We obtained global and regional 18 F-florbetapir SUVR (standardized uptake value ratios) values from the UCBERKELEYAV45_04_26_22.csv table downloaded from the ADNI website (https://adni.loni.usc.edu/). To determine Aβ-positivity, we used a cutoff value of 1.11 to the global SUVR, i.e., summary florbetapir cortical SUV normalized by whole cerebellum SUV [19].

Demographics, APOE4 , cognitive measures, and MRI

The ADNI baseline demographics (age, gender, years of education), APOE4, neuropsychological test results, and MRI biomarkers were obtained from ADNIMERGE.csv table downloaded from the ADNI website (http://adni.loni.usc.edu/). We used RAVLT (Rey's Auditory Verbal Learning Test) Immediate, RAVLT Learning, RAVLT Forgetting, RAVLT Percent Forgetting, ADAS13 (Alzheimer Disease Assessment Scale-13 items), ADASQ4 (ADAS Delayed Word Recall), MMSE (Mini-Mental State Examination), LDELTOTAL (Logical Memory Delayed Recall Total), TRABSCOR (Trail Making Test Part B), FAQ (Functional Assessment Questionnaire), and CDRSB (Clinical Dementia Rating - Sum of Boxes) as cognitive measures. These standard measures, which are widely used in assessing the cognitive and functional performance of dementia patients, are explained in the ADNI General Procedures Manual (http://adni.loni.usc.edu/wp-content/uploads/2010/09/ADNI_GeneralProceduresManual.pdf). As MRI biomarkers, we used the volumetric measures derived from FreeSurfer [20, 21], which are listed in the ADNIMERGE table. These measures include volumes of ventricles, hippocampus, whole brain, entorhinal, fusiform, middle temporal gyrus, and intra-cranial volume (ICV). Following [22], we used ICV as a separate feature, instead of normalizing other volumetric MRI measures with it. We did not consider quality control (QC) on the FreeSurfer segmentations [21], but included all the participants and brain regions despite their FreeSurfer QC score. However, QC on original MRI data [23] was considered. We obtained FreeSurfer measures from the ADNI depository (UCSFFSX_11_02_15.csv, UCSFFSX51_ADNI1_3T_02_01_16.csv, UCSFFSX51_11_08_19.csv, UCSFFSX6_08_17_22.csv).

Omitting the FreeSurfer QC had no major effect on the predictive accuracy, but increased the number of samples. A similar conclusion was previously reached by Gómez-Sancho et al. [24] indicating that whether or not ADNI FreeSurfer segmentations were quality controlled was immaterial to MRI-based MCI-to-dementia conversion prediction.

Study cohorts

To develop our classification models for predicting progression toward Aβ-positivity, we assigned participants into two overlapping cohorts based on the availability of longitudinal CSF measures or 18 F-florbetapir (AV45) PET measures. Specifically, we compiled a CSF-based cohort and a PET-based cohort. The subject selection procedure for each cohort is visualized in Fig. 1. The labeling of participants was constrained by the availability of data and the established cut-off values.

For the CSF-based cohort, we selected participants with available baseline CSF data (N= 1415). Among these participants, we first selected the initially Aβ-negative individuals (N = 726) and then we excluded participants (1) with an initial diagnosis of dementia (N= 35), (2) without available follow-up CSF sample (N = 345), (3) with a follow-up duration of fewer than 4 years without conversion to Aβ-positivity (N = 172), and (4) with an unstable Aβ+/Aβ-status during the follow up (N= 12). Finally, among the remaining 162 individuals, we defined Aβ-negative individuals who converted positive within the available follow-up time period as Aβ-Converter (N=52) and individuals who remained Aβ-negative for more than 4 years as Aβ-Stable (N = 110). In order to balance the dataset, we selected a set of auxiliary data from the initially Aβ-positive group. This step was crucial because the number of participants in the Aβ-Converter group was considerably lower than in the Aβ-Stable group, which could negatively impact the classification performance. Specifically, we selected individuals from the initially Aβ-positive group (N= 689) whose Aβ42 values were close to the cutoff used for classifying participants into the Aβ-positive and Aβ-negative groups, i.e., individuals with Aβ42 > 680 pg/mL and Aβ42 < 880 pg/mL (N= 244). The cut-off points of 680 and 880 were chosen for auxiliary data selection in such a way that a reasonable number of samples were obtained to balance the data set. We further excluded 47 individuals with an initial dementia diagnosis, 102 individuals without a follow-up CSF sample, and 17 individuals who returned to Aβ-negative status during the available follow-up. The

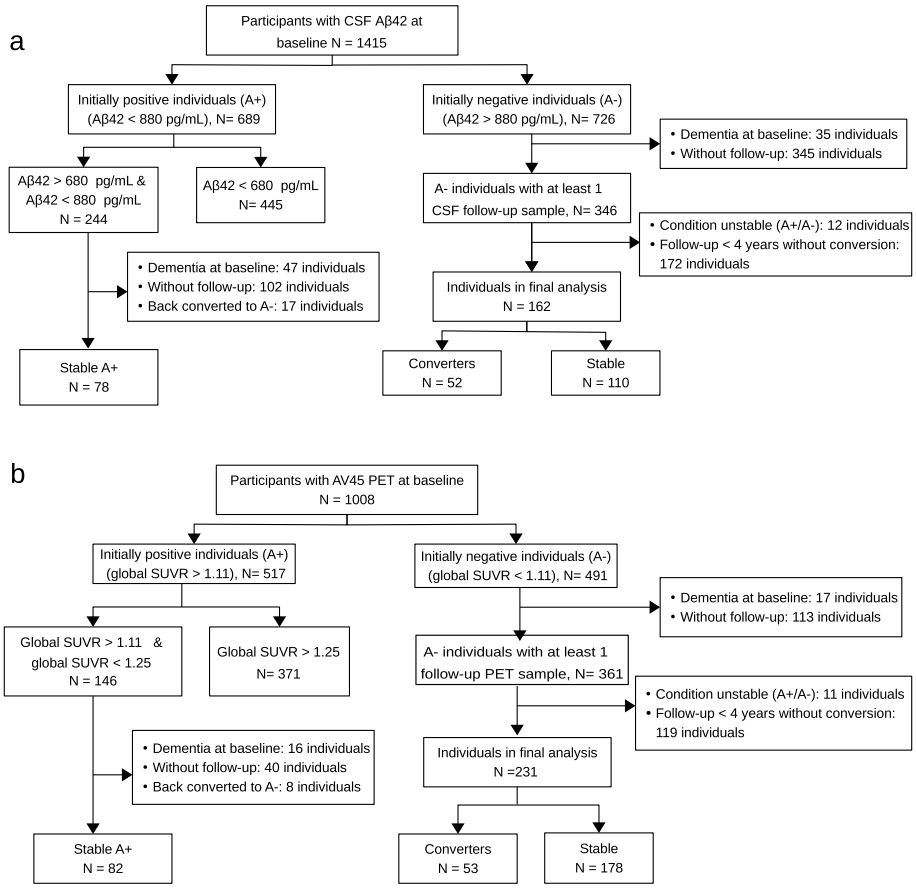

Fig. 1 Data selection procedure for predicting progression to Aβ-positivity for **a** CSF-cohort and **b** PET-cohort. The data from Stable A+ individuals, who are Aβ-positive at baseline with their Aβ values close to the cutoff point, are employed only during the training of the model and serve as Aβ -Converters for balancing the dataset during the training. They are not used in the testing of the model

remaining participants (N=78) are used as Aβ-Converters, but only during the training phase for balancing the dataset. This approach helped to create a more balanced dataset, which was necessary to develop a robust classification model.

The same procedure was applied for selecting the PET-based cohort, which resulted in a dataset of 53 individuals as Aβ-Converters and 178 individuals as Aβ -Stables. To create an auxiliary dataset from the initially Aβ-positive group, cut-off points of 1.11 and 1.25 were used. As a result, 82 individuals were included in the auxiliary dataset for use as Aβ-Converters only during the training phase. The selection of these cut-off points was done to ensure a reasonable number of samples were obtained to balance the dataset.

It is important to emphasize that participants who are Aβ-positive at baseline, with their Aβ values close to the cutoff point, i.e., 78 participants in CSF-cohort and 82 participants in PET-cohort, are employed only during the training of the model and serve as Aβ -Converters for balancing the dataset during the training. They are not used in the testing of the model.

A follow-up period of 4 years was chosen for the Aβ -Stable group in order to exclude participants with shorter follow-up duration. This decision was made to prevent potential false negatives while ensuring a reasonable number of participants within the Aβ-Stable group. Increasing the follow-up period beyond 4 years would have led to a marked reduction in the number of participants.

To develop our regression models for predicting future Aβ42, we only selected subjects with available baseline and longitudinal CSF data. Among those, we eliminated participants with dementia diagnosis at baseline and then we restricted our selection to participants with CSF follow-up samples for at least 4 years. If an individual had multiple follow-up samples after 4 years, the first sample (closest to 4 years) was used as the future sample for the analysis. Two hundred fifty-three individuals were included in our dataset as a result. Similarly, we selected

385 individuals for estimating future global SUVR based on PET data. The baseline characteristics of study cohorts are presented in Table 1 and participant's RIDs are available as Supplementary material.

For independent validation of our models, we selected a fully non-overlapping cohort from the one used for training. The independent validation was done to evaluate our CSF-based and PET-based Aβ-positivity conversion prediction models for MCI/dementia conversion prediction in healthy and MCI individuals. Since Aβ plays a significant role in the initiation of the AD process, the Aβ-positivity conversion prediction model should be able to detect AD-related cognitive decline even though it is not specifically developed to predict conversion to MCI/dementia. To this end, we selected two different datasets for assessing each CSF/PET-based model: The first dataset was for evaluating the future conversion of MCI/dementia in CN individuals and the second dataset was for the evaluation of future conversion of dementia in MCI individuals. For the evaluation of the CSF-based model, we considered all ADNI participants with a baseline diagnosis of CN or MCI who were not utilized for training the CSF-based approach and who had available baseline demographics, APOE4, neuropsychological test results, and MRI biomarkers. We determined two groups based on baseline and longitudinal diagnosis labels, regardless of CSF/PET biomarker status: (a) Stable group: CN/MCI individuals who have remained stable for 5 years or more following baseline, (b) Converter group: (1) Individuals with CN diagnosis at baseline who later develop MCI or dementia (the last two diagnoses must be MCI or dementia), and (2) MCI individuals who later develop to dementia (the last two diagnoses must be dementia). This resulted in a dataset of 67 converters and 131 stables for conversion prediction in CN individuals and a dataset of 211 converters and 126 stables in conversion prediction in MCI individuals (Table 2). The same procedure was used to select data for the PET-based model evaluation, resulting in a dataset of 74 converters and 95 stables for conversion prediction in CN individuals and a dataset of 234 converters and 90 stables in conversion prediction in MCI individuals (Table 3).

Machine learning framework

We developed a classifier based on the ridge logistic regression (RLR) approach [26] to predict the conversion of Aβ-positivity within 4 years in Aβ-negative individuals in both the CSF-based cohort and PET-based cohort. The framework of the classification procedure is shown in Fig. 2.

The number of participants in our PET-cohort and CSF-cohort differs because individuals in each cohort were selected based on the availability of CSF/PET data (Fig. 1). To ensure a fair comparison between the CSF and PET experiments, we decided to include an equal number of participants for CSF-based and PET-based

Table 1 Characteristics of the study cohorts: Age, education, Aβ42, and global SUVR measures are reported as mean(standard deviation). CN: cognitively normal, SMC: subjective memory concern (participants with self-reported significant memory concern), MCI: mild cognitive impairment, EMCI: early MCI, LMCI: late MCI. Classification of EMCI and LMCI is done by ADNI based on the WMS-R Logical Memory II Story A score. The specific cutoff scores were as follows (out of a maximum score of 25): EMCI was assigned for a score of 9–11 for 16 or more years of education, a score of 5–9 for 8–15 years of education, or a score of 3–6 for 0–7 years of education. LMIC was assigned for a score of ≤ 8 for 16 or more years of education, a score of ≤ 4 for 8–15 years of education, or a score of ≤ 2 for 0–7 years of education [25]

	CSF-cohort				PET-cohort			
Baseline characteristics	**Aβ-Converter**	**Aβ-Stable**	**Auxiliary data**	**Regression-cohort**	**Aβ-Converter**	**Aβ-Stable**	**Auxiliary data**	**Regression-cohort**
Sample size (*N*)	52 (32%)	110 (68%)	78	253	53 (23%)	178 (77%)	82	385
ADNI1/ADNIGO/ADNI2/ADNI3	15/6/30/1	25/18/67/0	27/7/41/3	81/35/137/0	0/7/44/2	0/31/140/7	0/8/66/8	0/69/308/8
Age, years	73.7 (7.9)	71.4 (7.0)	73.11 (6.7)	72.1 (6.7)	71.9 (6.9)	70.2 (6.8)	72.4 (7.1)	71.4 (6.7)
Sex, M/F	25/27	59/51	38/40	138/115	26/27	92/86	38/44	194/191
Education, years	16.9 (2.4)	16.3 (2.6)	16.1 (2.8)	16.3 (2.7)	16.6 (2.5)	16.7 (2.5)	16.2 (2.7)	16.4 (2.5)
APOE4 (0/1/2)	32/15/5	96/14/0	32/36/10	163/73/17	32/18/3	149/26/3	40/36/6	245/116/tr24
CN/SMC/EMCI/LMCI	16/5/12/19	51/12/26/21	19/6/18/35	95/18/62/78	28/7/13/5	65/24/64/25	17/20/28/17	123/57/139/66
Aβ42 (CSF-cohort)/global SUVR (PET-cohort)	1036.4 (182.2)	1791.4 (526.4)	754.4 (54.8)	1196.3 (661.5)	1.05 (0.04)	1.00 (0.04)	1.18 (0.04)	1.44 (0.20)

Table 2 Characteristics of the validation cohorts for performance validation of CSF-based model. Age and education are reported as mean(standard deviation)

	CN-cohort		MCI-cohort	
Baseline characteristics	**Converter-CN**	**Stable-CN**	**Converter-MCI**	**Stable-MCI**
Sample size (*N*)	67 (34%)	131 (66%)	211 (63%)	126 (37%)
ADNI1/ADNIGO/ ADNI2/ADNI3	32/0/31/4	57/0/73/1	145/9/48/19	41/40/45/0
Age, years	76.2 (5.0)	72.9 (5.4)	74 (6.7)	70.7 (7.1)
Sex, M/F	41/26	58/73	129/82	78/48
Education, years	16.3 (2.7)	16.4 (2.8)	15.8 (2.8)	16.0 (2.3)
APOE4 (0/1/2)	42/23/2	101/30/0	68/105/38	84/36/6
CN/SMC/EMCI/LMCI	51/16/0/0	105/26/0/0	0/0/24/187	0/0/74/52

Table 3 Characteristics of the validation cohorts for performance validation of PET-based model. Age and education are reported as mean(standard deviation)

	CN-cohort		MCI-cohort	
Baseline characteristics	**Converter-CN**	**Stable-CN**	**Converter-MCI**	**Stable-MCI**
Sample size (*N*)	74 (44%)	95 (56%)	234 (72%)	90 (28%)
ADNI1/ADNIGO/ ADNI2/ADNI3	48/0/22/4	69/0/24/2	165/8/53/8	47/25/18/0
Age, years	75.9 (4.5)	74.6 (5.6)	74.1 (6.6)	72.5 (7.3)
Sex, M/F	46/28	42/53	147/87	34/56
Education, years	16.1 (2.8)	16.5 (2.8)	15.9 (2.8)	15.7 (3.0)
APOE4 (0/1/2)	48/24/2	75/19/1	80/112/42	53/33/4
CN/SMC/EMCI/LMCI	63/11/0/0	80/15/0/0	0/0/24/187	0/0/22/212

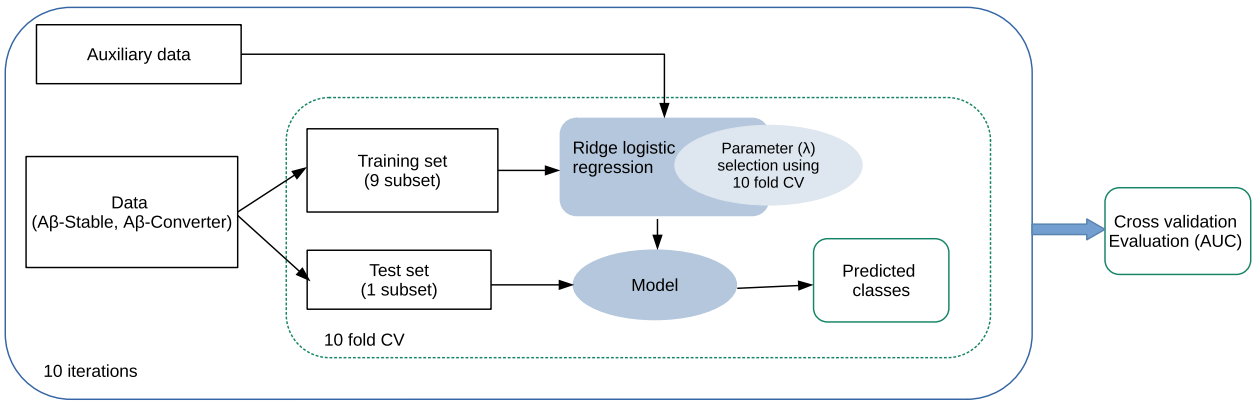

Fig. 2 Schematic representation of the classification framework

model training. Given that the PET-based dataset contains more individuals, we randomly subsampled it to achieve the same number of participants as in the CSF-based model for the training phase.

We designed three models with different feature combinations for each PET-based and CSF-based Aβ-positivity prediction task: The first model was trained using only demographic data and APOE4, the second model was trained using neuropsychological test results in addition to demographic data and APOE4, and the third model was trained using all demographic data, APOE4, neuropsychological test results, and MRI biomarkers.

In addition to the classification task, we developed a regression strategy based on ridge linear regression [27] to predict future Aβ42 (CSF-based) and global SUVR (PET-based) values from multimodal data similar to the

classification task with three models using various feature combinations. By applying a regression model, we were able to eliminate the effects of the cutoff point for classifying participants into Aβ+/Aβ- groups on the performance.

Moreover, as presented in Table 1, the participants in the CSF-cohort originated from different ADNI cohorts, including ADNI1. The participants in ADNI1 underwent 1.5-Tesla (1.5-T) T1-weighted MRI images, whereas participants in other ADNI cohorts underwent 3T MRI scans. To account for differences in MRI biomarkers caused by different field strengths, we applied ComBat [28], as a domain adaptation method to the MRI biomarkers before performing the actual machine learning procedure. ComBat employs an empirical Bayes method for batch correction in microarray expression data, and it has also been successfully utilized for domain adaptation in imaging data [29–31]. We then compared the results obtained with and without applying ComBat. However, the addition of a domain adaptation step using ComBat approach did not affect the results. Therefore, we excluded it and applied our machine learning approach to the original MRI biomarkers.

The Supplementary Table S 1 summarizes the experiments reported in different subsections of the " Results" section. The codes for classification and regression analyses are available at "https://github.com/ElahehMoradi/AB-positivity-prediction".

Implementation and performance evaluation

We used two nested cross-validation loops (10-fold for each loop) to evaluate the model's performance and estimate the model's parameter (λ) to prevent overfitting and maximize performance. First, an external 10-fold cross-validation was implemented in which samples were randomly divided into 10 subsets with the same proportion of each class label (stratified cross-validation). At each step, a single subset was left for testing and the remaining subsets were used for training. Again, the training set was divided into 10 subsets used to select the model's parameter (λ). The optimal parameters were selected in classification according to the misclassification error and in regression according to MSE(the mean square error) across the 10-fold of the inner loop. The performance of the model was then evaluated based on mainly AUC (area under the receiver operating characteristic curve) in classification analyses and correlation score in regression analyses in the test subset of the outer loop. We also provided balanced accuracy, sensitivity, and specificity for the classification analyses and mean absolute error for regression tasks in Supplementary materials. The reported results in the " Results" section are averages over 10 nested 10-fold CV runs. Repeated CV was used to reduce variability due to the partitioning of the data.

To compare the AUCs of two learning models, we used the Delong test on the results of a computation run with the median AUC. Comparison of correlation coefficient was tested using methods described by Diedenhofen and colleagues [32]. All the analyses were done using R (version 4.1.1), with the following packages: glmnet [33], caret [34], sva [35], cocor [36], pROC [37], Daim [38], ggplot2 [39], and complexheatmap [40].

Results

Predicting PET and CSF Aβ-positivity in Aβ-negative individuals from multimodal data excluding PET and CSF baseline measures

We predicted the progression to Aβ-positivity based on CSF and PET data. As explained in the "Methods" section, we designed three models with different feature combinations for each classification of PET-based and CSF-based future Aβ-positivity prediction as well as for the regression model of predicting future Aβ42 and global SUVR values. Figure 3 shows the results of all these computational analyses. These results are the average over 10 times repeated 10-fold cross-validation analyses for each method.

According to the findings (Fig. 3a, b, c, d; Fig. S1), CSF-based models outperformed PET-based models in both classification and regression tasks. The average AUC value across 10 computation runs was 0.78 with a 95% confidence interval (95% CI) of 0.70 to 0.85 with all demographic, APOE4, neuropsychological test results, and MRI biomarkers for predicting CSF-based progression to Aβ-positivity and 0.61 (95% CI of 0.53 to 0.70) for predicting PET-based progression to Aβ-positivity. The average correlation score for predicting future Aβ42 was 0.45 (95% CI of 0.37 to 0.53), and the average correlation score for predicting future global SUVR was 0.40 (95% CI of 0.31 to 0.49), with all demographic, APOE4, neuropsychological test results, and MRI biomarkers.

We explored the characteristics of individuals not correctly classified by the Aβ-positivity prediction model, particularly regarding their Aβ-values to determine if misclassified cases are closer to the positivity cutoff. For CSF-based experiments, we used the model with all measures, including demographics, APOE4, neuropsychological test results, and MRI, as it demonstrated superior performance. In PET-based experiments, the model with only demographics and APOE4 measures was chosen, exhibiting the best performance in the PET-based model. Figure S 3 presents a box plot of the Aβ-value (Aβ42 for CSF and global SUVR for PET) for the last visit, utilized for labeling in each group. However, no significant differences in Aβ-values were observed between

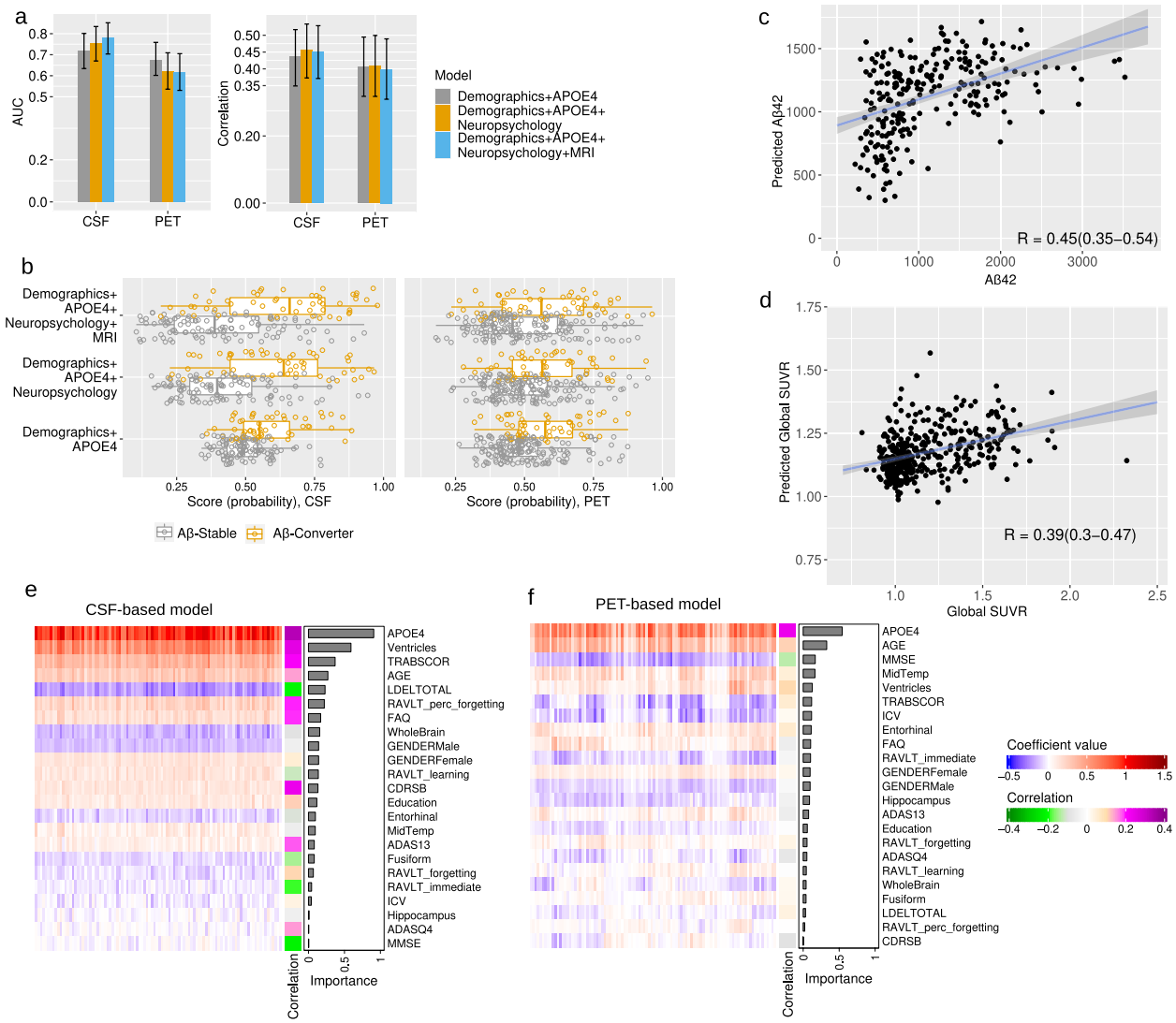

Fig. 3 Predicting future Aβ-positivity from multimodal data excluding PET and CSF baseline measures: **a** Bar plots showing the average AUC and average correlation score across 10 computation runs for CSF-based and PET-based models, with 95% confidence intervals error bars. **b** Distribution of probability score derived by RLR for PET-based and CSF-based prediction in Aβ-Stable and Aβ-Converter groups. **c**, **d** Scatter plot for estimation of Aβ42 (**c**) and global SUVR (**d**) derived by ridge linear regression (with demographics, APOE4, neuropsychology, and MRI biomarkers). The results in **b**, **c**, and d are from 1 computation run with median performance. **e**, **f** Heatmap of coefficient values across 10 runs of 10-fold CV (100 models) for CSF-based models (**e**) and pet-based models (**f**), with a single column heatmap representing the correlation score between each variable and the label (Aβ-Stable, Aβ-Converter), and a bar graph showing the importance of each predictor calculated by the mean of the absolute value of regression coefficients derived by RLR. There are 100 columns in the heatmaps, with each column representing the coefficient values for one model. The performance of predicting Aβ-positivity in Aβ-negative individuals was higher with CSF-cohort compared to PET-cohort, suggesting the higher relevance of CSF data for conversion prediction. ADAS13:Alzheimer Disease Assessment Scale, 13 items, ADASQ4: ADAS Delayed Word Recall, MMSE: Mini-Mental State Examination score, RAVLT: Rey's Auditory Verbal Learning Test, LDELTOTAL: Logical Memory Delayed Recall Total, TRABSCOR: Trail Making Test Part B, FAQ: Functional Assessment Questionnaire, CDRSB: Clinical Dementia Rating-Sum of Boxes, ICV: intracranial volume, MidTemp: middle temporal gyrus

correctly classified and misclassified individuals in both CSF-based and PET-based experiments. We also examined their diagnosis at baseline; however, there were no significant differences observed between correctly classified and misclassified individuals.

The addition of neuropsychological test results and MRI biomarkers to the demographics and APOE4 data increased the performance of the CSF-based Aβ-positivity conversion prediction model as shown in Fig. 3a. The average AUC value for predicting Aβ-positivity in

Aβ-negative individuals increased from 0.72 (95% CI of 0.63 to 0.80) to 0.78 (95% CI of 0.70 to 0.85), although the improvement was not statistically significant ($p = 0.18$) according to Delong's test. The average correlation value for predicting future aβ42 value also increased slightly, from 0.44 (95% CI of 0.35 to 0.52) to 0.45 ($p = 0.50$). However, adding neuropsychological test results and MRI biomarkers to demographic and APOE4 measures did not improve the performance of the PET-based models. Our results showed that including these variables decreased the performance of the PET-based Aβ-positivity conversion prediction model. The average AUC value decreased from 0.68 (95% CI of 0.60 to 0.76) to 0.61 (95% CI of 0.53 to 0.70). The average correlation value for predicting future global SUVR value also slightly decreased, from 0.41 (95% CI of 0.32 to 0.49) to 0.40 (95% CI of 0.31 to 0.49).

We investigated the contribution of different variables in CSF-based and PET-based models, accounting for all demographics, APOE4, neuropsychology test results, and MRI biomarkers (Fig. 3e and f, Fig. S2). Figure 3e and f show the coefficient values for 100 RLR classification models, derived from 10 runs of 10-fold CV, with a single column heatmap representing the correlation score between each variable and the label (Aβ-Stable, Aβ-Converter), and a bar graph indicating the significance of each variable, calculated by taking the mean of the absolute values of the regression coefficients.

In PET-based models, the primary contributors were APOE4 and age (as shown in Fig. 3f), whereas in CSF-based models (Fig. 3e), several variables from different data types are significantly contributing, with APOE4 being the most significant, followed by the volume of ventricles and TRABSCORE (Trail Making Test Part B). Additionally, the correlation panels in Fig. 3e and f, Tables S 2, and S 3 show a stronger correlation between various neuropsychological test results and MRI biomarkers with CSF-based labeling compared to PET-based labeling. This explains the higher contribution of different variables in CSF-based modeling, as well as the superior performance of CSF-based modeling.

The difference in the performance of the PET-based and CSF-based Aβ positivity is quite interesting. However, it is important to note that CSF and PET measures capture distinct aspects of the underlying biology, and, of course, the choice of CSF measure for defining Aβ positivity significantly influences the results. We investigated the correlation between Aβ42 and the Aβ42/Aβ40 ratio with PET global SUVR using ADNI data. The correlation between Aβ42 and PET global SUVR was −0.59, whereas the correlation between the Aβ42/Aβ40 ratio and PET global SUVR measure was −0.73 (Fig. S4). However, we defined Aβ-positivity based on Aβ42 alone, since the Aβ40 measure was only available for a small number of participants. The dataset with the Aβ42/Aβ40 ratio is rather limited, comprising 13 Aβ-Converters and 63 Aβ-Stable cases, with the converter group being particularly underrepresented. Due to the small size, we are concerned that it may not be sufficient for the development of a robust predictive model, thus putting the reliability of the results at risk.

Predicting PET and CSF Aβ-positivity in Aβ-negative individuals from multimodal data including PET and CSF baseline measures

We investigated the use of baseline CSF measures, i.e., Aβ 42, pTau, and Tau measures, to predict future CSF-based Aβ-positivity and the use of baseline PET measures, i.e., global and regional SUVR measures, to predict future PET-based Aβ-positivity. Our goal was to assess the validity of baseline CSF and PET measures for predicting future conversion. To gain insight into the contribution of other data types besides CSF baseline and PET baseline measures, we designed four models with different feature combinations. We further developed a regression-based approach to predict the changes (future - baseline) in the Aβ42 and global SUVR values. We predicted the changes in Aβ42/global SUVR values rather than their future values because of the high correlation between the baseline and future values of these measures. Figure 4 and Fig. S5 show the results of all these computational analyses.

To enhance the robustness and reliability of our results, we chose to exclude individuals with atypical changes: a significant decrease in PET-based Aβ values and a notable increase in CSF-based Aβ values with increasing age. Individuals who showed a marked reduction in PET-based Aβ values and a pronounced rise in CSF-based Aβ values over a 4-year period were considered atypical. Such cases could potentially indicate issues with data acquisition. Specifically, we removed 11 individuals from the CSF regression cohort with an increase in Aβ42 higher than 500 pg/mL, and similarly, 11 individuals were removed from the PET regression cohort with a decrease in global SUVR greater than 0.1. Although the number of participants with such changes in Aβ values was relatively small, their removal had a clear impact on the model performance of the CSF-based model but not of the PET-based model. In the CSF-based model, excluding outliers significantly improved the model's effectiveness. Specifically, in predicting changes in Aβ42 over 4 years using CSF measures, the correlation score improved from 0.21 to 0.44 upon removing outliers. Conversely, in PET-based analyses for predicting changes in global SUVR based on baseline PET measures, the correlation score slightly decreased after outlier exclusion, from 0.30 to 0.27. The outlier removal was clearly more important

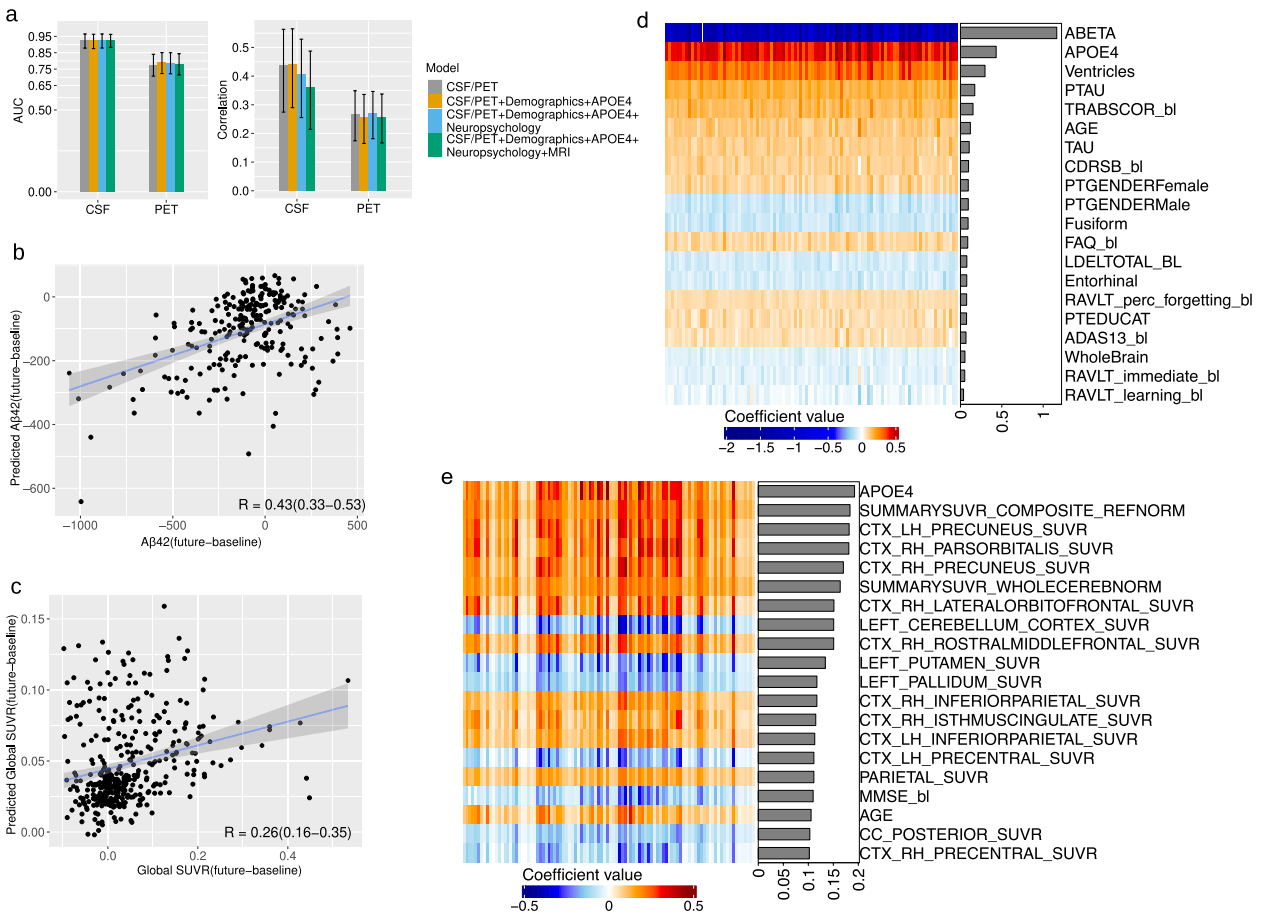

Fig. 4 Predicting future Aβ-positivity from multimodal data including PET and CSF baseline measures: **a** Bar plots showing the average AUC and average correlation score (predicting the difference between future and baseline Aβ42/global SUVR measures) across 10 computation runs for CSF-based and PET-based models, with 95% confidence intervals error bars, CSF/PET stands for CSF baseline measures (Aβ42, pTau, Tau) for predicting CSF-based Aβ-positivity and PET measures (global and regional) for predicting PET-based Aβ-positivity. **b**, **c** Scatter plot for estimation of the difference between future and baseline Aβ42 (**c**) and the difference between future and baseline global SUVR (**d**) derived by ridge linear regression (with CSF/PET measures). The results are from 1 computation run with median performance. **d**, **e** Heatmap of coefficient values across 10 runs of 10-fold CV (100 models) for CSF-based classification model (**d**) and PET-based regression model (**e**), with the bar graphs showing the importance of each predictor calculated by the mean of the absolute value of regression coefficient. The results of experiments without PET and CSF baseline measures are shown in Fig. 3

for the CSF-based model probably due to the existence of obvious outliers in CSF values but not in PET values.

As expected, using baseline CSF measures to predict future CSF-based Aβ-positivity and using baseline PET measures for predicting future PET-based Aβ-positivity resulted in relatively high performance. Specifically, for CSF-based Aβ-positivity prediction, the average AUC significantly increased to 0.93 when using only baseline CSF measures, compared to an AUC of 0.78, the highest achieved in experiments without CSF baseline data. Similarly, in the context of PET-based Aβ-positivity prediction, the average AUC increased to 0.78 with only baseline PET measures from 0.68, the best performance without baseline PET data (as detailed in the " Predicting PET and CSF Aβ-positivity in Aβ-negative individuals from multimodal data excluding PET and CSF baseline measures" section).

However, CSF-based prediction outperformed PET-based prediction in terms of performance (as shown in Fig. 4a). The average AUC value over 10 computation runs was 0.93 (95% CI of 0.88 to 0.97) for CSF-based prediction, whereas the average AUC value was 0.78 (95% CI of 0.71 to 0.84) for PET-based prediction. Moreover, the average correlation value for predicting the changes in Aβ42 was 0.44 (95% CI of 0.27 to 0.56), whereas for predicting the changes in global SUVR value, the average correlation score was 0.27 (95% CI of 0.17 to 0.35). Interestingly, adding other data types did not improve the

performance of classification and regression tasks in both CSF-based and PET-based analyses (Fig. 4a, Fig. S5).

We explored the impact of various variables, including demographics, APOE4, neuropsychological test results, and MRI biomarkers, with baseline CSF/PET measures in both PET-based and CSF-based models. Figure 4d and e visualizes the coefficient values for the top 20 predictors. Although the addition of additional data types did not improve overall model performance, the visualization of coefficient values emphasized substantial contributions from specific features, particularly APOE4, when integrated into the model. Particularly in the CSF-based model (Fig. 4d), other features from neuropsychological test results and MRI measures also exhibited significant contributions. Despite the absence of performance improvement, these findings emphasize the importance of specific features beyond baseline CSF/PET measures.

Predicting CSF-based future Aβ-positivity from CSF and PET baseline measures

We further extended our analysis by predicting CSF-based future Aβ-positivity based on baseline CSF measures and baseline PET measures. We selected only individuals with available PET baseline measures in the CSF-based cohort, resulting in a dataset of 84 Aβ-Stables, 36 Aβ-Converters, and 50 Aβ-positive as auxiliary data (for using as Aβ-Converters only in the training phase to balance dataset). We developed a classification model to predict future Aβ-positivity with three different feature combinations: the first model was trained based on CSF measures only, the second model based on PET measures only, and the third model based on both CSF and PET measures. The results are shown in Fig. 5a and Fig. S6. As anticipated, using baseline CSF measures for predicting CSF-based future Aβ-positivity resulted in improved performance compared to using baseline PET measures. Interestingly, the addition of baseline PET measures with baseline CSF measures did not improve the performance of the classification model for predicting future Aβ-positivity. The average AUC was 0.93 (95% CI of 0.87 to 0.98) using only CSF measures, 0.82 (95% CI of 0.72 to 0.90) using only PET measures, and 0.91 (95% CI of 0.85 to 0.96) using both CSF and PET measures.

To predict changes in Aβ42, first, we selected 169 individuals from the CSF regression-cohort who had available PET baseline measures. We then removed six individuals as outliers, with the changes in Aβ42 higher than 500. Again we designed the regression model for predicting the changes in Aβ42 using three different feature combinations: CSF measures only, PET measures only, and the combination of PET and CSF measures. The resulting average correlation value were 0.41 (95% CI of 0.23 to 0.55) using only CSF measures, 0.12 (95% CI of 0.004 to 0.25) using only PET measures, and 0.36 (95% CI of 0.20 to 0.50) using both CSF and PET measures. Similar to our classification results, the best predictive performance was achieved using only baseline CSF measures.

Predicting PET-based future Aβ-positivity from CSF and PET baseline measures

We continued our investigation by predicting PET-based future Aβ-positivity on baseline CSF and PET measures. Again, we limited the PET-based dataset to individuals with available CSF baseline measures, resulting in a dataset of 161 Aβ-Stables and 45 Aβ-Converters, and 72 Aβ-positive as auxiliary data (for using as Aβ-Converters only in the training phase to balance dataset). We developed a classification model to predict future PET-based Aβ-positivity with three different feature combinations: The first model was trained based on CSF measures only, the second model was based on PET measures only, and the third model was based on both CSF and PET measures. The results are shown in Fig. 5b–e and Fig. S7. Surprisingly, the use of baseline CSF measures to predict future PET-based Aβ-positivity resulted in improved performance compared with the use of baseline PET measures, the average AUC value increased from 0.79 (95% CI of 0.72 to 0.85) to 0.85 (95% CI of 0.78 to 0.91) (p-value = 0.18). In addition, a combination of CSF and PET measures provided similar performance to the use of CSF measures alone, with an average AUC value of 0.83 (95% CI of 0.76 to 0.88). The p-value was 0.049 between the model based on PET measures alone and the model with the combination of PET and CSF measures.

To predict changes in global SUVR, first, we selected 353 individuals from the PET regression-cohort who had available CSF baseline measures. We then removed 10 individuals as outliers, with the changes in global SUVR less than −0.1. Finally, the remaining 343 individuals were used for predicting the changes in global SUVR with different feature combinations. The results are shown in Fig. 5b and Fig. S7. Unlike the classification results, the average correlation value was rather near to one another in all three different feature combinations. The average correlation value was 0.30 (95% CI of 0.22 to 0.39) with CSF measures alone, 0.32 (95% CI of 0.22 to 0.42) with PET measures alone, and 0.33 (95% CI of 0.22 to 0.43) with CSF and PET measures combined.

We further investigated the contribution of different CSF and PET measures in both classification and regression models. Figure 5d and e show the coefficient values of the 20 variables with the highest importance calculated by taking the mean of the absolute values of the coefficients. Notably, Aβ42 was the most important variable in both classification and regression models (Fig. 5d and e). Furthermore, the two important CSF measures,

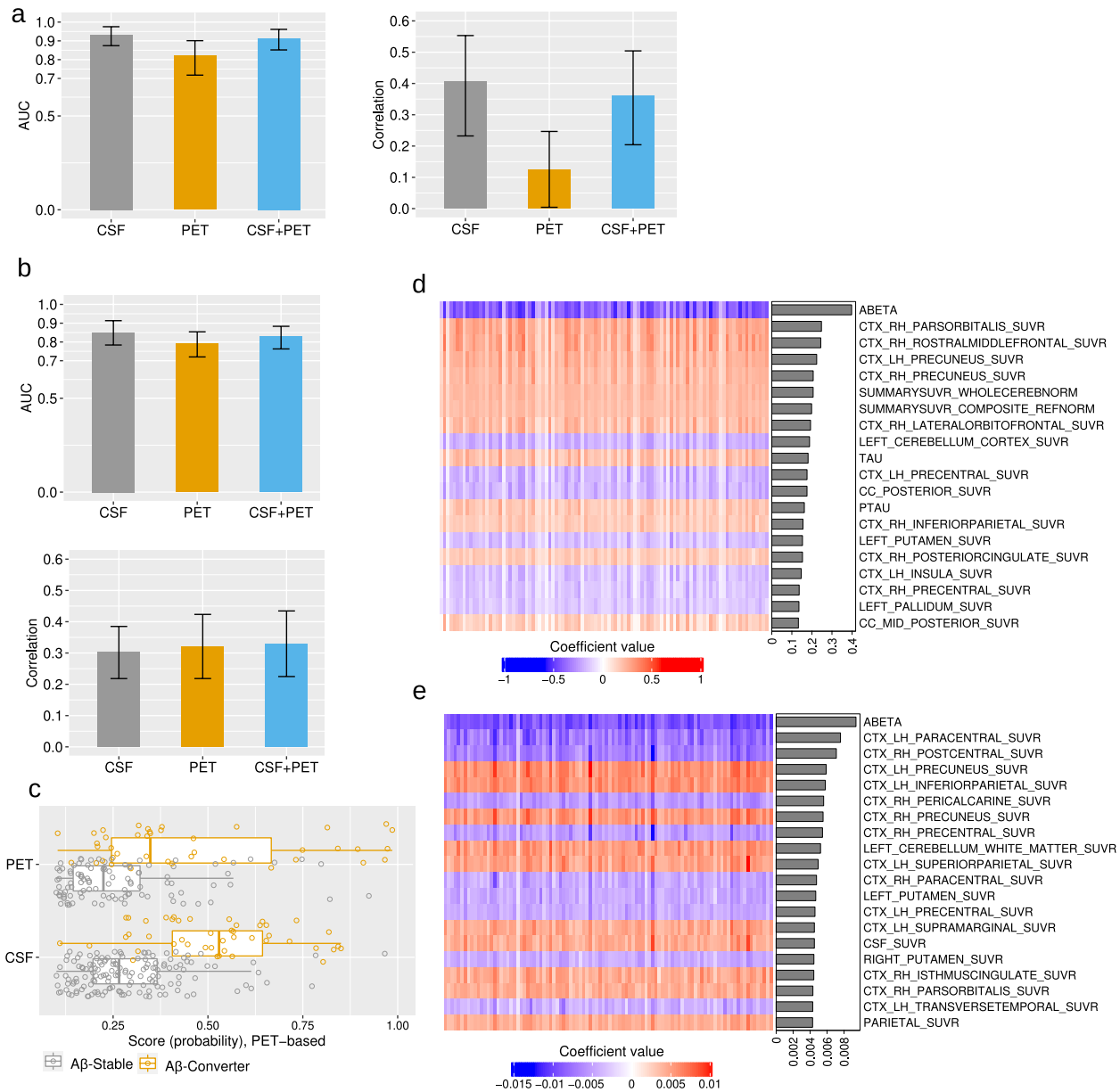

Fig. 5 Predicting future CSF-based and PET-based Aβ-positivity from baseline CSF and PET measures: **a** Bar plots showing the average AUC and average correlation score across 10 computation runs for predicting future CSF-based amyloid positivity, with 95% confidence intervals error bars. **b** Bar plots showing the average AUC and average correlation score across 10 computation runs for predicting future PET-based amyloid positivity, with 95% confidence intervals error bars. **c** Distribution of probability score derived by RLR for PET-based prediction in Aβ-Stable and Aβ-Converter groups with CSF and PET baseline measures. **d**, **e** Heatmap of coefficient values across 10 runs of 10-fold CV (100 models) for PET-based classification model (**d**) and pet-based regression model (**e**), with the bar graphs showing the importance of each predictor calculated by the mean of the absolute value of regression coefficient. The regression model is designed for predicting the difference between future and baseline global SUVR measures

i.e., Aβ42 and TAU, rank among the top ten most important variables. These findings demonstrate the importance of CSF measures in predicting PET-based future Aβ-positivity and also provide an explanation of why the classification model based on baseline CSF measures outperformed the model based on baseline PET measures.

It is essential to emphasize that, in utilizing baseline PET measures, we incorporated all available regional SUVR values, even though some may not provide significant information. In contrast, CSF consists of only three main features. This distinction may influence the significance of the contribution of PET global SUVR to the

modeling process compared to the contribution of CSF Aβ42 measure. To assess the importance of the baseline global SUVR measure in predicting PET-based Aβ-positivity, we devised two distinct classification models. The first exclusively featured PET global SUVR and CSF Aβ42 as predictors, while the second included all CSF measures (Aβ42, PTAU, and TAU) alongside five PET baseline SUVR measures-specifically, global SUVR (SUMMARYSUVR_WHOLECEREBNORM) and four additional SUVR measures used in creating the global SUVR (FRONTAL_SUVR, CINGULATE_SUVR, PARIETAL_SUVR, TEMPORAL_SUVR). The results are shown in Fig. S8. The AUC of the first model with CSF Aβ42 and PET global SUVR as predictors was 0.81, and the AUC for all three CSF measures and five PET SUVR measures was 0.82. This indicates a slight decrease compared to using all regional SUVR and CSF measures (AUC: 0.83). Although AUC performance was similar across all three models, an examination of sensitivity revealed that the model with only two predictors exhibited relatively low sensitivity and balanced accuracy (Fig. S8c). However, the model with five SUVRs and CSF measures demonstrated performance comparable to the model incorporating CSF measures with all SUVR measures. Investigating the contribution of CSF Aβ42 and PET global SUVR measures, when only five SUVRs are included, demonstrates that the importance of Aβ42 and global SUVR is nearly the same (Fig. S8a).

Additionally, we compared the performance of a model with only five SUVRs against a model with all regional SUVR measures. The AUC was 0.79 when considering all baseline regional SUVRs, and 0.73 when using only five SUVRs. These findings suggest that if CSF baseline measures are not available during the learning process, the inclusion of all SUVR measures improves the performance. Given the application of regularized logistic regression, we anticipate the model to adeptly select the most significant information from all available predictors, even when all regional SUVRs are employed (Fig. S8c).

MCI/dementia conversion prediction

We validated the relevance of our Aβ-positivity prediction models, both CSF-based and PET-based models, by applying them to an independent dataset (Tables 2 and 3) to predict changes in clinical status from cognitively normal to MCI/dementia or from MCI to dementia. The predictors were demographics, APOE4, neuropsychological test results and MRI biomarkers (the model trained in the " Predicting PET and CSF Aβ-positivity in Aβ-negative individuals from multimodal data excluding PET and CSF baseline measures" section).

We trained a model for predicting CSF-based future Aβ-positivity using all available participants (Fig. 1a, Table 1) and used it to classify CN-Stable vs. CN-Converter groups, as well as MCI-Stable vs. MCI-Converter groups. We next repeated the process for the PET-based model. The results are shown in Fig. 6. The CSF-based model performed quite well, achieving an AUC of 0.76 (95% CI of 0.68 to 0.83) for MCI/dementia conversion prediction in the CN group and an AUC of 0.89 (95% CI of 0.86 to 0.93) for dementia conversion prediction in the MCI group. However, the PET-based model had lower performance, with an AUC of 0.60 (95% CI of 0.51 to 0.69) for classifying CN-Stables vs. CN-Converters and an AUC of 0.72 (95% CI of 0.67 to 0.78) for classifying MCI-Stables vs. MCI-Converters.

Given that neither the CSF-based nor the PET-based model was specifically designed for classifying CN-Stables vs. CN-Converters/MCI-stables vs. MCI-converters, and that both models used basic and noninvasive data (demographics, APOE4, cognitive, and MRI), the CSF-based model performed exceptionally well for the MCI/dementia conversion prediction in CN and MCI groups, while the PET-based model performed well only in the MCI group. However, in the CN group, the PET-based model failed to predict conversion to MCI/dementia.

Discussion

Determining Aβ-status is crucial for the prescription of amyloid-targeted treatments in the future. This study aimed to predict conversion to Aβ-positivity in Aβ -negative individuals using data that is widely available in clinical settings. To achieve this goal, we developed a classification model based on an RLR approach incorporating demographics, APOE4, neuropsychological tests, and MRI biomarkers. We categorized participants into Aβ-positive and Aβ-negative groups, based on CSF Aβ42 and PET global SUVR. Due to categorization inconsistencies [41], we conducted separate analyses with participants classified as Aβ+/Aβ- based on CSF and PET biomarkers.

The selection of a suitable cutoff point for categorizing subjects into Aβ-positive and Aβ-negative groups is important in the prediction of Aβ-positivity in Aβ-negative individuals as it can impact the results of our analysis. Particularly, in CSF-based grouping, the selection of the cutoff point is challenging due to the presence of various cutoff values in the existing literature [15, 16]. We decided to use the cutoff 880 pg/mL, a value established from predictions made in the BioFINDER study, independently from the ADNI dataset. The selection of a cutoff point determined in independent data is statistically appropriate. Additionally, our choice of the 880 pg/mL cutoff, as opposed to larger cutoff values, allowed us

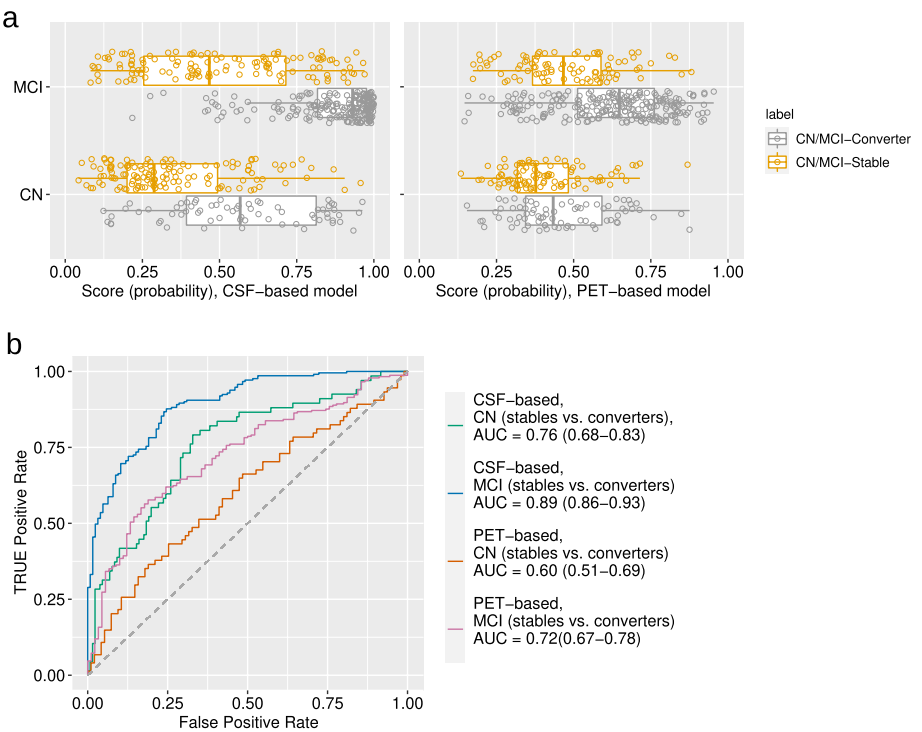

Fig. 6 MCI/dementia conversion prediction in CN and MCI groups: **a** Boxplot of probability score derived by RLR for CSF-based and PET-based model for classification of stables vs converters in CN and MCI groups. **b** ROC curves of CN and MCI subjects classification to stable and converter groups using CSF-based and PET-based models, with AUC values and 95% confidence intervals

to include a larger number of participants in the Aβ-Converter group. However, to confirm the reliability of our classification results and to eliminate any potential bias arising from the cutoff point for categorizing participants into Aβ+/Aβ− groups, we conducted regression analyses that were aligned with our classification experiments for predicting future CSF Aβ (Aβ42) or PET Aβ (global SUVR) values via a ridge linear regression. These steps collectively contribute to the robustness and credibility of our study's findings.

In our analyses, we demonstrated that features from multiple modalities, including demographics, neurophysiological scores, APOE4, and MRI biomarkers, can model the progression to Aβ-positivity detected by either CSF or PET biomarkers. Interestingly, our findings indicate that utilizing CSF Aβ42 for participant categorization resulted in more accurate predictions of future Aβ -positivity compared to PET global SUVR. Also, the CSF-based model benefited from additional features, whereas the accuracy of the PET-based model decreased with more features. In more detail, the AUC of CSF-based model increased from 0.72 with APOE and demographics as the features to 0.78 with all the features. However, the AUC of the PET-based decreased from 0.68 with APOE4 and demographics as the features to 0.61 to all the features. To comprehend the performance disparity between the PET-based and CSF-based models, we investigated the contributions of different data types in each model. We observed that cognitive scores and MRI biomarkers, along with APOE4 as the most important variable, strongly contribute to the CSF-based model. In contrast, in the PET-based model, the contribution of cognitive and MRI biomarkers was smaller, and the primary predictors were APOE4 and age.

It is essential to highlight that CSF Aβ42 and PET global SUVR measures, utilized in defining Aβ-positivity, employ different mechanisms for detecting Aβ protein. PET imaging reveals the presence of amyloid plaques in the brain, whereas CSF analysis is associated with the clearance of amyloid from the brain. Low CSF amyloid levels may indicate inefficient clearance, leading to brain amyloid accumulation [42, 43]. Although we used the same ML framework for predicting future Aβ-positivity, direct comparison of the results is challenging. However, our findings suggest that predicting future CSF-based Aβ -positivity is relatively easier compared to PET-based Aβ -positivity. In a previous study, Jagust and Landau [44] explored the factors influencing the transition from Aβ -negative to Aβ-positive, as determined by PET measures. Their findings indicated that age, baseline PET,

and being a female APOE4 carrier were associated with an increased risk of conversion. Interestingly, in another study conducted by Elman et al. [45], cognitive tests were found to predict conversion to Aβ-positivity, which was defined based on CSF and/or PET. The discrepancy between these findings can be understood based on the results of the current study, where prediction performance improved by incorporating cognitive measures for CSF-based cohort but not in the PET-based cohort.

Determining Aβ-positivity using CSF biomarker can be done either based on Aβ42 levels and the Aβ42/Aβ40 ratio. In this study, we defined Aβ-positivity solely based on Aβ42, as the Aβ40 measure was only available for a small number of participants. However, recent studies have emphasized the relevance of the Aβ42/Aβ40 ratio in defining Aβ-positivity using CSF biomarker [46], as it exhibits a stronger correlation with PET-based Aβ measure compared to Aβ42 alone. Moreover, the correlation between the ratio of Aβ42/Aβ40 and global SUVR is higher than the correlation between Aβ42 and the global SUVR measure in ADNI data. We speculate that utilizing the Aβ42/Aβ40 ratio for predicting CSF-based Aβ -positivity may yield results that align more closely with PET-based predictions. Nevertheless, due to the limited number of available samples, we cannot conduct experiments to confirm this hypothesis.

Previous studies have primarily focused on detecting Aβ-positivity at the time of the study [47–51] (baseline prediction). These studies have employed either machine learning algorithms or statistical analyses to identify Aβ-positivity at baseline and explore the relationships between cognitive measures, various biomarkers, and Aβ -positivity. For example, Palmqvist et al. [47] developed a model utilizing demographics, cognitive tests, white matter lesions, APOE, and plasma biomarkers (Aβ42/Aβ40, tau, and neurofilament light chain) to detect Aβ-positivity at baseline. Their model achieved an AUC of 0.80-0.82 when trained on BIOFINDER dataset [48, 52] and validated on ADNI data. Similarly, another study by the same group [48], detected baseline Aβ-positivity using demographic, APOE, and cognitive information and achieved an AUC of 0.65 in cognitively healthy individuals. However, we are aware of only two studies that have focused on predicting future Aβ-positivity [12, 45]. The first study, by Elman et al. [45], examined the association of baseline cognitive measures with progression to Aβ-positivity in Aβ-negative individuals. The second study, by Park et al. [12], employed machine learning algorithms to predict future Aβ-positivity in Aβ-negative individuals. In Park et al. [12] study, they used a PET biomarker for subject classification in Aβ+/Aβ- groups and developed a classifier with baseline age, gender, APOE4 genotype, and PET SUVR measures in ADNI data. They achieved a cross-validated AUC of 0.67 using basic demographic and genetic factors, which improved to 0.84 when including PET SUVR measures. Our work has a similar objective to the study by Park et al. [12], but we focused on widely available non-invasive measures to predict future Aβ conversion. Moreover, a key characteristic of our study was to assess the future predictability of Aβ-positivity determined based on either CSF and PET biomarkers whereas Park et al. [12] only considered a PET-based definition.

We validated the relevance of our prediction models for future Aβ-positivity in an independent dataset composed of separate ADNI participants, predicting clinical status changes from CN to MCI/dementia or from MCI to dementia. The CSF-based model performed well, achieving an AUC of 0.76 for MCI/dementia conversion prediction in CN individuals and an AUC of 0.89 for the dementia conversion prediction in MCI individuals. However, the PET-based model performed worse in predicting conversion, reaching an AUC of 0.60 for CN to MCI/dementia conversion and an AUC of 0.72 for MCI to dementia conversion. The CSF-based model's excellent performance in predicting MCI/dementia conversion in a completely independent dataset is intriguing, indicating its potential for detecting AD/dementia-related changes at an early stage. Although the model was not specifically designed to classify CN/MCI-Converter versus CN/MCI-Stable groups, the CSF-based model demonstrated strong performance comparable to existing studies designed for such classification tasks [53–55].

We analyzed the use of baseline CSF and PET measures for predicting future Aβ-positivity. Including baseline scores significantly improved the performance of the prediction models. Notably, when baseline CSF/PET measures were available, the addition of other data types such as cognitive scores and MRI biomarkers contributed very little value. Moreover, utilizing baseline CSF/PET measures enabled the prediction of the change/rate in Aβ42 and global SUVR values which are rather hard to model [56]. However, the invasive nature, high cost, and limited availability of CSF/PET measures restrict their usage. An alternative approach could be to predict future Aβ-positivity using predicted CSF and PET measures derived from other variables. This strategy could offer more robust insights, particularly when substituting less accessible measures with more readily available or cost-effective options. A promising area for future research lies in exploring the prediction of PET and CSF values from blood measures, which are generally more accessible.

Moreover, we investigated the role of baseline CSF and PET measures in both predicting CSF-based future Aβ -positivity as well as in predicting PET-based future Aβ -positivity. Interestingly, baseline CSF measures showed

superior predictive ability for future PET-based Aβ-positivity compared to baseline PET measures. A closer look at the coefficient values in the PET-based model revealed that Aβ42 made the greatest contribution to the model. Recent studies suggest that AD-related alterations are detectable earlier in CSF than in PET [57, 58], which may explain the superior performance of baseline CSF measures in predicting PET-based Aβ-positivity.

Our study has several limitations. First, the sample sizes were small due to our specific inclusion criteria, although the data came from one of the largest longitudinal dementia-prediction cohorts (ADNI). To compensate for the small sample size, we evaluated our CSF-based/PET-based models in an independent dataset for MCI/dementia conversion prediction. Second, the sizes of Aβ-Stable and Aβ-Converter groups were different, which we partially addressed by introducing auxiliary data for label balancing during model training. Third, the CSF-based and PET-based analyses used different datasets, challenging direct performance comparison. To address this, we included an equal number of participants for CSF-based and PET-based model training to facilitate a fairer comparison. Fourth, we utilized a single ML approach to combine various data types and develop our models. While this approach worked well for CSF-based analyses, a more advanced approach may be needed for PET-based future Aβ-positivity prediction. Last, the models to predict future Aβ-positivity in Aβ-negative individuals were not evaluated on an independent database. While the validation of the prediction models in an independent dataset composed of separate ADNI participants, predicting clinical status changes from CN to MCI/dementia or from MCI to dementia, provides a partial remedy, it does not completely overcome the limitation.

Conclusion

We developed ML-based models to predict future Aβ-positivity in Aβ-negative individuals, determined based on either CSF or PET biomarkers. The CSF-derived dichotomization achieved better predictive performance (AUC = 0.78) compared to PET dichotomization (AUC = 0.68). The discrepancy in performance may be attributed to the different mechanisms of Aβ detection and the discordance between the two biomarkers. Further research is required to understand this discrepancy. However, by using non-invasive measures including demographics, APOE4, neuropsychological scores, and MRI biomarkers, our CSF-based Aβ-positivity conversion prediction model performed well in identifying Aβ-Stables vs Aβ-Converters with an AUC 0f 0.78, as well as in identifying CN-Converters vs. CN-Stables (AUC = 0.76) and MCI-Stables vs. MCI-Converters (AUC = 0.89). These findings demonstrate the significance of neuropsychological and MRI biomarkers to detect the risk of conversion to AD, even in cognitively normal individuals. The detection may occur before current thresholds for Aβ-positivity are reached, providing an opportunity for early intervention.

Abbreviations

Aβ	Beta-amyloid
AD	Alzheimer's disease
ADAS13	Alzheimers Disease Assessment Scale-13 items
ADASQ4	ADAS Delayed Word Recall
ADNI	Alzheimer's Disease Neuroimaging Initiative
APOE4	Apolipoprotein E ε4
ATN	Amyloid, tau, and neurodegeneration
AUC	Area under the curve
CDRSB	Clinical Dementia Rating Sum of Boxes
CI	Confidence interval
CSF	Cerebrospinal fluid
EMCI	Early mild cognitive impairment
ICV	Intra-cranial volume
LDELTOTAL	Logical Memory Delayed Recall Total
LMCI	Late mild cognitive impairment
FAQ	Functional Assessment Questionnaire
MCI	Mild cognitive impairment
ML	Machine learning
MMSE	Mini-Mental State Examination
MRI	Magnetic resonance imaging
NC	Normal control
PET	Positron emission tomography
RAVLT	Reys Auditory Verbal Learning Test
RLR	Ridge logistic regression
SUVR	Standardized uptake value ratios
TRABSCOR	Trail Making Test Part B

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13195-024-01415-w.

Supplementary material 1.

Supplementary material 2.

Acknowledgements

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 (http://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.

Authors' contributions
EM and JT conceptualized and designed the study. EM performed the analysis and wrote the manuscript. JT and MP contributed to the writing. AH, BS, and AS contributed to the study design and revised the paper. The authors read and approved the final manuscript.

Funding
This research has been supported by European Research Council (grant 804371), The Academy of Finland (grants 346934 and 349184), and a grant 351849 from the Academy of Finland under the frame of ERA PerMed ("Pattern-Cog"), a grant AC21-2/00023 from Spain's Instituto de Salud Carlos III with NextGenEU European funding from the "Plan de Recuperación, Transformación y Resiliencia (PRTR)".

Availability of data and materials
Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (https://adni.loni.usc.edu/). Details about data access are detailed there. The participant's RIDs are provided in Supplementary materials.

Declarations

Ethics approval and consent to participate
The ADNI study was approved by the Institutional Review Board (IRB) of each participating site and was conducted in accordance with Federal Regulations. For participants in ADNI protocols, written informed consent was obtained and the study was conducted in accordance with the Declaration of Helsinki.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 25 August 2023 Accepted: 11 February 2024
Published online: 27 February 2024

References

1. Jack CR Jr, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimers Dement. 2018;14(4):535–62.
2. Scheltens P, De Strooper B, Kivipelto M, Holstege H, Chételat G, Teunissen CE, et al. Alzheimer's disease. Lancet. 2021;397(10284):1577–90.
3. Villemagne VL, Doré V, Burnham SC, Masters CL, Rowe CC. Imaging tau and amyloid-β proteinopathies in Alzheimer disease and other conditions. Nat Rev Neurol. 2018;14(4):225–36.
4. Fagan AM. What does it mean to be 'amyloid-positive'? Brain. 2015;138(3):514–6.
5. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7(3):263–9.
6. Tarasoff-Conway JM, Carare RO, Osorio RS, Glodzik L, Butler T, Fieremans E, et al. Clearance systems in the brain–implications for Alzheimer disease. Nat Rev Neurol. 2015;11(8):457–70.
7. Nelson PT, Jicha GA, Schmitt FA, Liu H, Davis DG, Mendiondo MS, et al. Clinicopathologic correlations in a large Alzheimer disease center autopsy cohort: neuritic plaques and neurofibrillary tangles" do count" when staging disease severity. J Neuropathol Exp Neurol. 2007;66(12):1136–46.
8. Young CB, Mormino EC. Prevalence rates of amyloid positivity–updates and relevance. JAMA Neurol. 2022;79(3):225–7.
9. Zwan M, van Harten A, Ossenkoppele R, Bouwman F, Teunissen C, Adriaanse S, et al. Concordance between cerebrospinal fluid biomarkers and [11C] PIB PET in a memory clinic cohort. J Alzheimers Dis. 2014;41(3):801–7.
10. Contador J, Vargas-Martínez AM, Sanchez-Valle R, Trapero-Bertran M, Lladó A. Cost-effectiveness of Alzheimer's disease CSF biomarkers and amyloid-PET in early-onset cognitive impairment diagnosis. Eur Arch Psychiatry Clin Neurosci. 2023;273(1):243–52.
11. Wittenberg R, Knapp M, Karagiannidou M, Dickson J, Schott JM. Economic impacts of introducing diagnostics for mild cognitive impairment Alzheimer's disease patients. Alzheimers Dement Transl Res Clin Interv. 2019;5:382–7.
12. Park CJ, Seo Y, Choe YS, Jang H, Lee H, Kim JP, et al. Predicting conversion of brain β-amyloid positivity in amyloid-negative individuals. Alzheimers Res Ther. 2022;14(1):129.
13. Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann Neurol. 2009;65(4):403–13.
14. Yosypyshyn D, Kučikienė D, Ramakers I, Schulz JB, Reetz K, Costa AS. Clinical characteristics of patients with suspected Alzheimer's disease within a CSF Aß-ratio grey zone. Neurol Res Pract. 2023;5(1):1–11.
15. Hansson O, Seibyl J, Stomrud E, Zetterberg H, Trojanowski JQ, Bittner T, et al. CSF biomarkers of Alzheimer's disease concord with amyloid-β PET and predict clinical progression: a study of fully automated immunoassays in BioFINDER and ADNI cohorts. Alzheimers Dement. 2018;14(11):1470–81.
16. Bucci M, Chiotis K, Nordberg A, Initiative ADN. Alzheimer's disease profiled by fluid and imaging markers: tau PET best predicts cognitive decline. Mol Psychiatry. 2021;26(10):5888–98.
17. Jagust WJ, Landau S, Shaw L, Trojanowski J, Koeppe R, Reiman E, et al. Relationships between biomarkers in aging and dementia. Neurology. 2009;73(15):1193–9.
18. Joshi AD, Pontecorvo MJ, Clark CM, Carpenter AP, Jennings DL, Sadowsky CH, et al. Performance characteristics of amyloid PET with florbetapir F 18 in patients with Alzheimer's disease and cognitively normal subjects. J Nucl Med. 2012;53(3):378–84.
19. Landau S, Thomas B, Thurfjell L, Schmidt M, Margolin R, Mintun M, et al. Amyloid PET imaging in Alzheimer's disease: a comparison of three radiotracers. Eur J Nucl Med Mol Imaging. 2014;41:1398–407.
20. Fischl B. FreeSurfer. Neuroimage. 2012;62(2):774–81.
21. Hartig M, Truran-Sacrey D, Raptentsetsang S, Simonson A, Mezher A, Schuff N, et al. UCSF freesurfer methods. San Francisco: ADNI Alzheimers Disease Neuroimaging Initiative; 2014.
22. Karaman BK, Mormino EC, Sabuncu MR, Initiative ADN. Machine learning based multi-modal prediction of future decline toward Alzheimer's disease: an empirical study. PLoS ONE. 2022;17(11):0277322.
23. Jack CR Jr, Barnes J, Bernstein MA, Borowski BJ, Brewer J, Clegg S, et al. Magnetic resonance imaging in Alzheimer's disease neuroimaging initiative 2. Alzheimers Dement. 2015;11(7):740–56.
24. Gomez-Sancho M, Tohka J, Gomez-Verdejo V, Initiative ADN, et al. Comparison of feature representations in MRI-based MCI-to-AD conversion prediction. Magn Reson Imaging. 2018;50:84–95.
25. Edmonds EC, McDonald CR, Marshall A, Thomas KR, Eppig J, Weigand AJ, et al. Early versus late MCI: improved MCI staging using a neuropsychological approach. Alzheimers Dement. 2019;15(5):699–708.
26. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1.
27. McDonald GC. Ridge regression. Wiley Interdiscip Rev Comput Stat. 2009;1(1):93–100.
28. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118–27.
29. Radua J, Vieta E, Shinohara R, Kochunov P, Quidé Y, Green MJ, et al. Increased power by harmonizing structural MRI site differences with the ComBat batch adjustment method in ENIGMA. NeuroImage. 2020;218:116956.
30. Orlhac F, Eertink JJ, Cottereau AS, Zijlstra JM, Thieblemont C, Meignan M, et al. A guide to ComBat harmonization of imaging biomarkers in multicenter studies. J Nucl Med. 2022;63(2):172–9.
31. Fortin JP, Cullen N, Sheline YI, Taylor WD, Aselcioglu I, Cook PA, et al. Harmonization of cortical thickness measurements across scanners and sites. Neuroimage. 2018;167:104–20.

32. Diedenhofen B, Musch J. cocor: a comprehensive solution for the statistical comparison of correlations. PLoS ONE. 2015;10(4):0121945.
33. Hastie T, Qian J, Tay K. An Introduction to glmnet. CRAN R Repositary; 2021.
34. Kuhn M. Building predictive models in R using the caret package. J Stat Softw. 2008;28:1–26.
35. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28(6):882–3.
36. Diedenhofen B, Diedenhofen MB. Package 'cocor'. Comprehensive R Archive Network; 2016. https://cran.r-project.org/web/packages. Accessed 24 Mar 2022.
37. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. Package "pROC'. Technical Report. 2017. https://cran.r-project.org/web.
38. Potapov S, Adler W, Hofner B, Lausen B. Package 'daim'. Comput Stat Data Anal. 2009;53(3):718–29.
39. Wickham H. ggplot2. Wiley Interdiscip Rev Comput Stat. 2011;3(2):180–5.
40. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847–9.
41. Blennow K, Mattsson N, Schöll M, Hansson O, Zetterberg H. Amyloid biomarkers in Alzheimer's disease. Trends Pharmacol Sci. 2015;36(5):297–309.
42. Li Y, Rusinek H, Butler T, Glodzik L, Pirraglia E, Babich J, et al. Decreased CSF clearance and increased brain amyloid in Alzheimer's disease. Fluids Barriers CNS. 2022;19(1):1–9.
43. Sala A, Nordberg A, Rodriguez-Vieitez E, Initiative ADN. Longitudinal pathways of cerebrospinal fluid and positron emission tomography biomarkers of amyloid-β positivity. Mol Psychiatry. 2021;26(10):5864–74.
44. Jagust WJ, Landau SM, Initiative ADN, et al. Temporal dynamics of β -amyloid accumulation in aging and Alzheimer disease. Neurology. 2021;96(9):e1347–57.
45. Elman JA, Panizzon MS, Gustavson DE, Franz CE, Sanderson-Cimino ME, Lyons MJ, et al. Amyloid-β positivity predicts cognitive decline but cognition predicts progression to amyloid-β positivity. Biol Psychiatry. 2020;87(9):819–28.
46. Zetterberg H. Biofluid-based biomarkers for Alzheimer's disease-related pathologies: an update and synthesis of the literature. Alzheimers Dement. 2022;18(9):1687–93.
47. Palmqvist S, Insel PS, Zetterberg H, Blennow K, Brix B, Stomrud E, et al. Accurate risk estimation of β-amyloid positivity to identify prodromal Alzheimer's disease: cross-validation study of practical algorithms. Alzheimers Dement. 2019;15(2):194–204.
48. Insel PS, Palmqvist S, Mackin RS, Nosheny RL, Hansson O, Weiner MW, et al. Assessing risk for preclinical β-amyloid pathology with APOE, cognitive, and demographic information. Alzheimers Dement Diagn Assess Dis Monit. 2016;4:76–84.
49. Ezzati A, Harvey D, Golzar A, Habeck C, Qureshi I, Zammit A, et al. Risk estimation of Amyloid Beta positivity on PET with Cognitive, CSF Biomarkers, volumetric MRI, APOE, and Demographic information, vol 515. AAN Enterprises; 2020.
50. Tosun D, Veitch D, Aisen P, Jack Jr CR, Jagust WJ, Petersen RC, et al. Detection of β-amyloid positivity in Alzheimer's Disease Neuroimaging Initiative participants with demographics, cognition, MRI and plasma biomarkers. Brain Commun. 2021;3(2):fcab008.
51. Li C, Liu M, Xia J, Mei L, Yang Q, Shi F, et al. Predicting Brain Amyloid-β PET Grades with Graph Convolutional Networks Based on Functional MRI and Multi-Level Functional Connectivity. J Alzheimers Dis. 2022;(Preprint):1–15.
52. Palmqvist S, Schöll M, Strandberg O, Mattsson N, Stomrud E, Zetterberg H, et al. Earliest accumulation of β-amyloid occurs within the default-mode network and concurrently affects brain connectivity. Nat Commun. 2017;8(1):1214.
53. Strange BA, Zhang L, Sierra-Marcos A, Alfayate E, Tohka J, Medina M. Predicting the future development of mild cognitive impairment in the cognitively healthy elderly. bioRxiv. 2020:2020–07. https://doi.org/10.1101/2020.07.30.227496.
54. Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J, Initiative ADN, et al. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. Neuroimage. 2015;104:398–412.
55. Palmqvist S, Tideman P, Cullen N, Zetterberg H, Blennow K, Initiative ADN, et al. Prediction of future Alzheimer's disease dementia using plasma phospho-tau combined with other accessible measures. Nat Med. 2021;27(6):1034–42.
56. Jack CR Jr, Wiste HJ, Weigand SD, Therneau TM, Lowe VJ, Knopman DS, et al. Predicting future rates of tau accumulation on PET. Brain. 2020;143(10):3136–50.
57. Palmqvist S, Zetterberg H, Mattsson N, Johansson P, Minthon L, Blennow K, et al. Detailed comparison of amyloid PET and CSF biomarkers for identifying early Alzheimer disease. Neurology. 2015;85(14):1240–9.
58. Palmqvist S, Mattsson N, Hansson O, Initiative ADN. Cerebrospinal fluid analysis detects cerebral amyloid-β accumulation earlier than positron emission tomography. Brain. 2016;139(4):1226–36.

Publisher's Note