Journal of Alzheimer's Disease xx (20xx) x–xx
DOI 10.3233/JAD-191038
IOS Press



# Predicting Amyloid-β Levels in Amnestic Mild Cognitive Impairment Using Machine Learning Techniques

Ali Ezzati[a,b,*], Danielle J. Harvey[c], Christian Habeck[d], Ashkan Golzar[e], Irfan A. Qureshi[a,f], Andrea R. Zammit[a], Jinshil Hyun[a], Monica Truelove-Hill[f], Charles B. Hall[e], Christos Davatzikos[f] and Richard B. Lipton[a,b] for the Alzheimer's Disease Neuroimaging Initiative[1]
[a]*Department of Neurology, Albert Einstein College of Medicine, Bronx, NY, USA*
[b]*Department of Neurology, Montefiore Medical Center, Bronx, NY, USA*
[c]*Department of Public Health Sciences, University of California-Davis, Davis, CA, USA*
[d]*Cognitive Neuroscience Division, Department of Neurology, Columbia University, New York, NY, USA*
[e]*Element AI, Montreal, Quebec, QC, Canada*
[f]*Biohaven Pharmaceuticals, New Haven, CT, USA*
[g]*Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA, USA*

Accepted 25 November 2019

**Abstract**.
**Background:** Amyloid-β positivity (Aβ+) based on PET imaging is part of the enrollment criteria for many of the clinical trials of Alzheimer's disease (AD), particularly in trials for amyloid-targeted therapy. Predicting Aβ positivity prior to PET imaging can decrease unnecessary patient burden and costs of running these trials.
**Objective:** The aim of this study was to evaluate the performance of a machine learning model in estimating the individual risk of Aβ+ based on gold-standard of PET imaging.
**Methods:** We used data from an amnestic mild cognitive impairment (aMCI) subset of the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort to develop and validate the models. The predictors of Aβ status included demographic and ApoE4 status in all models plus a combination of neuropsychological tests (NP), MRI volumetrics, and cerebrospinal fluid (CSF) biomarkers.
**Results:** The models that included NP and MRI measures separately showed an area under the receiver operating characteristics (AUC) of 0.74 and 0.72, respectively. However, using NP and MRI measures jointly in the model did not improve prediction. The models including CSF biomarkers significantly outperformed other models with AUCs between 0.89 to 0.92.
**Conclusions:** Predictive models can be effectively used to identify persons with aMCI likely to be amyloid positive on a subsequent PET scan.

Keywords: Alzheimer's disease, amyloid imaging, machine learning, mild cognitive impairment, predictive analytics



*Correspondence to: Ali Ezzati, MD, Albert Einstein College of Medicine, 1225 Morris Park Avenue, Bronx, NY, 10461, USA. Tel.: +1 718 430 3885; Fax: +1 718 430 3870; E-mail: aezzati@einstein.yu.edu.

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

## INTRODUCTION

Cerebral amyloid-β (Aβ) deposition is a hallmark pathologic change in Alzheimer's disease (AD) and is believed to precede dementia by many years [1]. In the last decade, many clinical trials have tried to use targeted therapies to lower brain Aβ, but all these trials have failed to achieve significant effects on clinical endpoints [2–4]. Major proposed reasons for failure include clinical heterogeneity of participants, selection of an inappropriate biological target (i.e., merely reducing amyloid production or aggregation cannot modify disease progression) [5], enrollment of individuals based on unreliable criteria, and inclusion of individuals who did not have increased cerebral Aβ and were unlikely to have had AD pathology [6].

To address some of these limitations, the new NIA-AA Research Framework has proposed to use biomarkers of Aβ deposition, pathologic tau, and neurodegeneration [AT(N)] to diagnose AD and decrease heterogeneity in research study samples. Similarly, more recent clinical trials have used biomarkers of amyloid status measured in cerebrospinal fluid (CSF) or in the brain using positron emission tomography (PET) [7]. While amyloid PET is considered non-invasive, and may be more reliable than CSF biomarkers [8], its utility in both research and clinical practice has been limited. Factors that have prevented widespread use of PET imaging in research and practice include availability, economic factors (high costs, not being covered by insurance), and patient or caregiver's concerns (safety, burden, tolerance, and radiation exposure) [9].

Recruitment of eligible amnestic mild cognitive impairment (aMCI) patients is a major bottleneck in conducting secondary prevention trials; as few as 10–20% of potential MCI patients are actually trial-eligible [10]. In addition, only 40–60% of aMCI patients are likely to be Aβ positive based on the current gold standard of amyloid PET, which further limits the number of trial-eligible individuals [11]. Without using any predictive models, to establish Aβ positivity, all enrolled participants (based on initial clinical diagnostic criteria) require amyloid PET imaging at the time of screening. Therefore, predicting Aβ positivity prior to PET imaging can decrease unnecessary patient burden and costs of running the trials.

In addition, were a treatment to become available for the prevention of AD in persons with aMCI, implementation in clinical practice might be difficult. Amyloid PET would be an expensive option for identifying individuals eligible for treatment. One option might be to develop and use risk prediction models and screening algorithms similar to what has been used in cardiovascular disease [12] or various types of cancer [13, 14]. Using this approach, data gathered at lower cost (e.g., neurocognitive tests and MRIs) could be used to predict Aβ positivity. Amyloid PET would be performed in a selected subgroup of individuals predicted to have a positive amyloid scan. Machine learning (ML) techniques provide a promising method for predicting amyloid positivity. These approaches are specifically designed to predict outcomes and provide a feasible approach for exploiting and managing complex and high-dimensional data [15, 16]. Developing practical predictive models can drive a major shift in clinical care and for both primary and secondary prevention purposes [17–21].

The primary goal of this study was to compare the relative sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of different combinations of features (demographics, *APOE* ε4 status, neuropsychological tests, MRI volumetrics, and CSF biomarkers) used in a ML model to predict PET Aβ positivity. The model was developed in a subsample of the Alzheimer's Disease Neuroimaging Initiative (ADNI) aMCI population and was subsequently validated using an independent sample from the same cohort. Considering that availability and associated burden and costs of each of these measures is different (e.g., MRIs require staying still for long periods and lumbar puncture is an invasive procedure), we evaluated the predictive value of each of the multimodal features separately and jointly.

## METHODS

### *Study design and participants*

The data used for this analysis were downloaded from the ADNI database (http://www.adni.loni.usc.edu) in March 2019. The ADNI is an ongoing cohort, which was launched in 2003 as a public–private partnership. The individuals included in the current study were initially recruited as part of ADNI-GO, and ADNI-2 between 2009 and 2013. This study was approved by the Institutional Review Boards (IRB) of all participating institutions. Informed written consent was obtained from all participants at each site.

A total of 369 participants diagnosed with MCI who were enrolled in ADNI-GO, and ADNI-2 were eligible for this study. Eligible individuals completed baseline visit and had MRIs and amyloid PET imaging in the same wave of study. All ADNI participants with the diagnosis of MCI, were diagnosed as having amnestic MCI; this diagnostic classification required Mini-Mental State Examination (MMSE) scores between 24 and 30 (inclusive), a memory complaint, objective memory loss measured by education-adjusted scores on the Wechsler Memory Scale Logical Memory II, a Clinical Dementia Rating (CDR) of 0.5, absence of significant impairment in other cognitive domains, essentially preserved activities of daily living, and absence of dementia. Participants whose scans failed to meet quality control or had unsuccessful image analysis were excluded from this study.

*Study measures*

*Neuropsychological data*

Neuropsychological (NP) tests included the MMSE, the 11-item Alzheimer's Disease Assessment Scale cognitive subscale (ADAS-cog), and Logical Memory II [22–24]. These tests were available for all participants in ADNI studies from the beginning of cohort and therefore, they were not a limiting factor for inclusion of participants in this study. All NP measures were entered into models as continuous variables.

*APOE gene status*

*APOE* ε4 allele (ApoE4) frequency was available for all participants included in this study. ApoE4 status was defined as ApoE4-negative (–) if they carried no ApoE4 allele or ApoE4-positive (+) if they carried at least one ApoE4 allele.

*MRI acquisition and preprocessing*

MRIs were obtained across different sites of the ADNI study with a unified protocol (For more information, please see http://adni.loni.usc.edu/). MRI data were automatically processed using the FreeSurfer software package (available at http://surfer.nmr.mgh.harvard.edu/) by the Schuff and Tosun laboratory at the University of California-San Francisco as part of the ADNI shared data-set. FreeSurfer methods for identifying and calculation of regional brain volume are previously described in detail [25]. Volumes of 47 regions of interests (ROIs), derived from FreeSurfer software, were used as MRI indicators. For the purpose of this study, volume of all regions of interest (ROIs) were normalized for total intracranial volume (TICV) and the ratio of ROIs' volume (ROIv) to TBV [i.e., (ROIv/TICV) x mean whole population ROIv] was used in the analyses and reported throughout the manuscript unless otherwise specified.

*PET imaging acquisition and preprocessing*

Florbetapir PET images were obtained across different sites of ADNI study with a unified protocol (For more information, please see http://adni.loni.usc.edu/methods/pet-analysis-method/pet-analysis/) Data were processed at the Jagust lab at University of California, Berkeley. Details of the methods used to process PET images have been previously described [26]. In brief, a native-space MRI scan for each subject was segmented and parcellated with FreeSurfer to define cortical grey matter regions of interest (frontal, anterior/posterior cingulate, lateral parietal, lateral temporal) that make up a summary cortical ROI. In addition, five reference regions were created (cerebellar grey matter, whole cerebellum, brainstem/pons, eroded subcortical white matter, and a composite reference region). Subsequently each PET scan was coregistered to the corresponding MRI and the mean Florbetapir uptake within the cortical and reference regions were computed. A Florbetapir SUVR was calculated by averaging across the four cortical regions and dividing this summary ROI by the uptake in the whole cerebellum. To establish Amyloid positivity or negativity, a Florbetapir SUVR cutoff of 1.11 was used as recommended by previous studies [27]. For the purpose of this study, we only used the first Florbetapir PET scan obtained from each participant.

*CSF biomarkers*

CSF samples were batch processed by the ADNI Biomarker Core at the University of Pennsylvania School of Medicine and CSF tau, p-tau$_{181p}$, and A$\beta_{1\text{-}42}$ were measured for all participants with CSF sample [28]. These data were available for 335 participants (90.5% of the whole sample) and sections of data analysis that required CSF measures were limited to these participants. CSF measures were included as continuous variables in ML models. However, for the purpose of simplicity, in Table 1 individuals were classified according to CSF concentration thresholds (tau: >93 pg/mL; p-tau$_{181p}$: >23 pg/mL; A$\beta_{1\text{-}42}$<192 pg/mL) previously estab-

Table 1
Demographics and clinical characteristics of study participants according to group

| Variables | Participants Group/Subgroup | | | | | |
|---|---|---|---|---|---|---|
| | Training set (N = 185) | | | Validation (test) set (N = 184) | | |
| | Aβ- | Aβ+ | p | Aβ- | Aβ+ | p |
| N, % | 72 (38.9) | 113 (61.1) | – | 72 (39.2) | 112 (60.8) | – |
| Age[a] | 69.5 (7.0) | 72.2 (7.2) | 0.014 | 69.4 (7.7) | 72.4 (6.5) | 0.006 |
| Men, N (%) | 39 (54.2) | 55 (48.7) | 0.469 | 39 (54.2) | 68 (60.7) | 0.385 |
| Education, y | 16.1 (2.7) | 16.3 (2.9) | 0.677 | 16.4 (2.3) | 16.2 (2.6) | 0.490 |
| ApoE4 carrier[b], N (%) | 16 (22.2) | 74 (65.5) | <0.001 | 15 (20.8) | 70 (62.5) | <0.001 |
| CDR-SB | 1.3 (0.9) | 1.5 (0.9) | 0.110 | 1.3 (0.8) | 1.7 (0.9) | 0.006 |
| MMSE | 28.4 (1.5) | 27.9 (1.8) | 0.041 | 28.5 (1.5) | 27.8 (1.6) | 0.006 |
| ADAS-cog | 7.9 (3.6) | 9.3 (4.0) | 0.019 | 7.5 (3.3) | 10.3 (4.7) | <0.001 |
| LM | 8.0 (2.2) | 6.7 (3.3) | 0.002 | 8.7 (2.0) | 6.6 (3.4) | <0.001 |
| Hippocampal volume | 3.8 (0.6) | 3.5 (0.6) | 0.003 | 3.8 (0.6) | 3.4 (0.5) | <0.001 |
| Low CSF $A\beta_{1\text{-}42}$ (%) | 7 | 84 | <0.001 | 18 | 88 | <0.001 |
| High CSF Tau (%) | 4 | 46 | <0.001 | 03 | 42 | <0.001 |
| High CSF p-$tau_{181p}$ (%) | 38 | 87 | <0.001 | 47 | 85 | <0.001 |

[a]Values are means ± SD unless otherwise stated. [b]Percentage of individuals carrying at least one E4 allele. CDR-SB, Clinical Dementia Rating scale Sum of Boxes; MMSE, Mini-Mental State Exam; ADAS, Alzheimer's Disease Assessment Scale; LM, logical memory-delayed recall. Hippocampal volume is reported in cubic centimeter.

lished to maximize sensitivity and specificity of autopsy confirmed AD [29].

*Data analysis*

*Training and validation samples*

The training and validation of the ML model was performed by using the split half method. For this purpose, participants were randomly split into two independent samples with approximately equal number of Aβ- and Aβ+ based on PET imaging. One sample was used as training data-set and the other sample was used for validation of models. This validation method enables the generalization of the trained ML model to data that have never been presented to the ML algorithms previously.

*Selection of feature-sets (indicators)*

Demographics (age, sex, and education), ApoE4 status, NP tests, all available volumetric MRI measures (FreeSurfer outputs), and all CSF biomarkers mentioned above were used as features in the predictive models. We chose 7 different feature-sets and compared the performance of ML models which used these feature-sets for classification. In addition to demographics and ApoE4, models include the following features: Model 1) NP tests; Model 2) MRI volumetrics; Model 3) CSF biomarkers; Model 4) NP tests plus MRI volumetric; Model 5) NP tests plus CSF biomarkers; Model 6) MRI volumetric plus CSF biomarkers; Model 7) NP tests, MRI volumetric plus CSF biomarkers.

*Machine learning model*

Analysis and computation of ML methods were conducted using MATLAB ©(version 2018b). We used Ensemble Linear Discriminant (ELD) ML models for the purpose of classification and pattern recognition. EDL is among the family of classification methods known as ensemble learning, in which the output of an ensemble of simple and low-accuracy classifiers trained on subsets of features are combined (e.g., by weighted average of the individual decisions), so that the resulting ensemble decision rule has a higher accuracy than that obtained by each of the individual classifiers [30, 31]. In this work, we combined linear discriminant functions (i.e., hyperplanes that dichotomize the samples based on subsets of features) in order to construct the ensemble classifier. To avoid overfitting, we trained the models for a maximum of 100 cycles. We monitored the learning curve and picked the cycle with the lowest misclassification rate for termination of the training. The parameters for the models were optimized automatically through the hyperparameter optimization process in MATLAB.

*Training the classification model*

Data from the training sample (N = 185) were used for training of the classifier (Fig. 1). Models were trained to recognize Aβ- versus Aβ+ individuals using all sets of the features as described above. A 10-fold cross-validation procedure was used in all models for testing validity of the models. Cross-validation is an established statistical method for

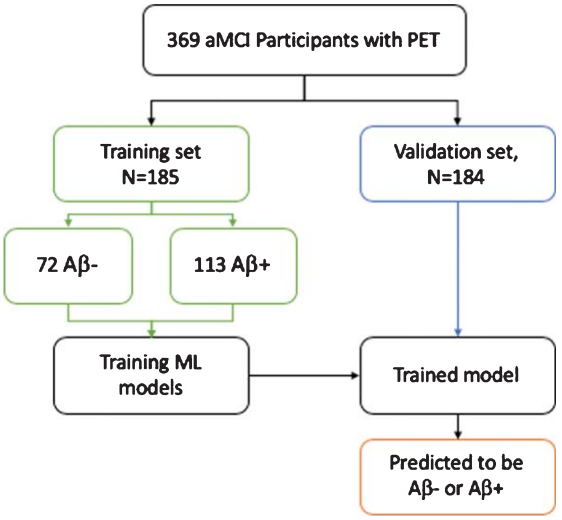


Fig. 1. Study design diagram. aMCI, amnestic mild cognitive impairment; ML, machine learning.

validating a predictive model, which involves training several parallel models, each based on a subset of the training data. Subsequently, the model performance is evaluated based on the average accuracy in predicting the labels of the omitted portion of the training data [32]. The performance of each model was calculated based on the total percentage of correct classification (accuracy), sensitivity, specificity, PPV, NPV, and area under the curve (AUC).

*Prediction of amyloid status in the validation sample*

Following training of the models, they were applied to the validation sample to predict amyloid positivity of each person (Fig. 1). Using the same feature-sets used for training of models, each individual was assigned to "predicted Aβ-" or "predicted Aβ+" groups. The performance of the predicted outcome was evaluated using the results obtained from PET imaging. Accuracy, sensitivity, specificity, PPV, and NPV for each model were estimated.

*Inverse cross-validation*

To further validate the models, we performed an inverse cross-validation by training the ML model using the half-sample that was used for prediction previously and using the half-sample that was used for training as the prediction subset. Considering that results for this analysis was very similar to the initial model (see Supplementary Tables 1 and 2) and to avoid confusion, we primarily focus on the results of the first model for the rest of this article.

*Data availability*

Data from ADNI cohort is publicly available. Programming codes used for this paper are available upon request.

## RESULTS

*Sample characteristics*

Participants with aMCI had an average age of 71.2 years (SD = 7.2) and 54.5% were men. In both subsamples (training and validation), in comparison with Aβ- subgroup, the Aβ+ subgroup was older and had less favorable performance on NP tests, had smaller hippocampal volumes and had a CSF profile that was more similar to AD. Table 1 summarizes participants' demographics and clinical characteristics.

Developing the amyloid prediction models in the training subsample

Performance of ELD models using 7 different feature-sets for classification of training sample to Aβ- or Aβ+ on PET is summarized in Table 2. In the training set, the area under the curve (AUC) of models including demographics, ApoE4, and NP tests or MRI volumetrics (models 1 and 2) were 0.74 and 0.72, respectively. The combination of NP with MRI (model 4, AUC = 0.70) did not improve the prediction. AUC of the models including demographics, ApoE4, and CSF markers alone was substantially higher (model 3, AUC = 0.86), however neither addition of NP (model 5, AUC = 0.89) or MRI (model 6, AUC = 0.90) improve the models. The combination of all measures yielded an AUC of 0.90 (model 7).

Performance of the amyloid prediction models in the validation subsample

After development of ELD models, they were applied to the data from validation sample to assign participants to Aβ- or Aβ+ (Table 3). The AUC of models including demographics, ApoE4, and NP tests or MRI volumetrics (models 1 and 2) were 0.72 and 0.71, respectively. AUC of the model including demographics, ApoE4, and CSF markers as features was higher (model 3, AUC = 0.86). Inclusion of both MRI volumetric and NP tests as features in the same model did not make a substantial change in the performance of model in comparison with models including them separately (model 4, AUC = 0.73). Models that included CSF measures (models 3, 5, 6, 7) had substantially better performance in comparison with

Table 2
Performance of Ensemble Linear Discriminant (ELD) classifiers in differentiating Aβ- versus Aβ+ in training set (subsample 1)

| Feature-set | Accuracy, % | Sensitivity, % | Specificity, % | PPV, % | NPV, % | AUC |
|---|---|---|---|---|---|---|
| 1) Dem/ApoE4 + NP | 67.50 | 78.6 | 53.7 | 68.1 | 66.6 | 0.74 |
| | (60.3–74.2) | (69.5–86.1) | (42.3–64.7) | (62.3–73.3) | (56.7–75.3) | |
| 2) Dem/ApoE4 + MRI-v | 68.1 | 73.8 | 61.0 | 70.4 | 64.9 | 0.72 |
| | (60.9–74.7) | (64.2–82.0) | (49.6–71.6) | (63.9–76.1) | (56.1–72.7) | |
| 3) Dem/ApoE4 + CSF-b | 86.0 | 87.4 | 84.1 | 87.4 | 84.1 | 0.92 |
| | (80.1–90.6) | (79.4–93.1) | (74.4–91.2) | (50.7–92.0) | (76.0–89.9) | |
| 4) Dem/ApoE4 + NP + MRI-v | 67.0 | 72.8 | 59.7 | 69.4 | 63.6 | 0.70 |
| | (60.0–73.7) | (63.2–81.1) | (48.3–70.4) | (63.0–75.2) | (54.9–71.5) | |
| 5) Dem/ApoE4 + NP + CSF-b | 83.8 | 83.5 | 84.1 | 86.9 | 80.2 | 0.90 |
| | (77.6–88.8) | (74.9–90.1) | (74.4–91.3) | (80.0–91.7) | (72.2–86.4) | |
| 6) Dem/ApoE4 + MRI-v + CSF-b | 85.4 | 87.4 | 82.9 | 86.5 | 84.0 | 0.89 |
| | (79.5–90.2) | (79.4–93.1) | (73.0–90.3) | (79.9–91.2) | (75.7–89.7) | |
| 7) Dem/ApoE4 + NP + MRI-v + CSF-b | 85.4 | 88.0 | 82.3 | 85.4 | 85.4 | 0.90 |
| | (79.5–90.2) | (80.0–93) | (72.6–89.7) | (78.6–90.9) | (79.5–90.2) | |

Dem, demographics; NP, neuropsychological tests; MRI-v, all MRI volumetrics; CSF-b, all CSF biomarkers.

Table 3
Performance of Ensemble Linear Discriminant (ELD) classifiers in predicting Aβ- versus Aβ+ in validation (test) set (subsample 2)

| Indicators | Accuracy (%) | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | AUC |
|---|---|---|---|---|---|---|
| 1) Dem/ApoE4 + NP | 72.8 | 75.0 | 69.4 | 79.2 | 64.1 | 0.72 |
| | (65.8–79.1) | (65.9–82.7) | (57.5–79.7) | (72.6–84.6) | (55.6–71.8) | |
| 2) Dem/ApoE4 + MRI-v | 71.2 | 70.5 | 72.2 | 79.8 | 61.2 | 0.71 |
| | (64.1–77.6) | (61.1–78.7) | (60.4–82.1) | (72.7–85.4) | (53.3–68.4) | |
| 3) Dem/ApoE4 + CSF-b | 86.4 | 86.1 | 86.4 | 90.7 | 80.5 | 0.86 |
| | (80.6–91.0) | (84.4–94.5) | (75.9–93.1) | (84.4–94.5) | (71.8–86.9) | |
| 4) Dem/ApoE4 + NP + MRI-v | 71.7 | 68.7 | 76.4 | 81.9 | 61.1 | 0.73 |
| | (64.6–78.1) | (59.3–77.1) | (64.9–85.6) | (74.6–87.5) | (53.7–68.0) | |
| 5) Dem/ApoE4 + NP + CSF-b | 85.7 | 85.7 | 86.1 | 90.5 | 79.5 | 0.86 |
| | (80.0–90.5) | (77.8–91.6) | (75.9–93.1) | (84.3–94.5) | (70.9–86.0) | |
| 6) Dem/ApoE4 + MRI-v + CSF-b | 84.8 | 83.9 | 86.1 | 90.3 | 77.5 | 0.85 |
| | (78.7–89.6) | (75.8–90.1) | (75.9–93.1) | (84.0–94.3) | (69.0–84.1) | |
| 7) Dem/ApoE4 + NP + MRI-v + CSF-b | 85.9 | 83.0 | 90.3 | 93.0 | 77.4 | 0.87 |
| | (80.0–90.6) | (74.8–89.5) | (81.0–96.0) | (86.7–96.4) | (69.3–83.8) | |

Dem, demographics; NP, neuropsychological tests; MRI-v, all MRI volumetrics; CSF-b, all CSF biomarkers.

models that did not include them (see Table 3 for details).

## DISCUSSION

In this study, we evaluated the value of machine learning models in predicting amyloid positivity based on florbetapir PET scans. We showed that the positive predictive values of models, which used NP tests, MRI volumetrics, or CSF biomarkers were 0.72, 0.71, and 0.86, respectively. Addition of MRI measures to NP tests in the models did not lead to improvement in the prediction performance. As expected, addition of CSF measures noticeably improved performance of models.

A few studies have previously proposed different types predictive models for detecting cerebral amyloid positivity based on demographics, NP tests, MRI measures, and blood or CSF-based biomarkers [33–38]. For example, Kander et al. [34] reported AUCs of 0.59–0.67 for individual NP tests, AUC of 0.64 using all NP tests, and AUC of 0.64 for hippocampal volume. Similar to our findings, they showed that adding imaging biomarkers to NP tests in the multivariate analysis does not improve the AUC. Palmqvist et al. [36] applied a forward selection logistic regression model to demographics, ApoE4, NP tests, and white matter lesions for prediction of amyloid positivity and achieved AUCs of 0.80–0.82 in ADNI. Kim et al. [35] used similar variables and using logistic regressions, developed a nomogram that achieved predictive AUCs of 0.74–0.77.

A common limitation in the previous studies is that in many cases they have used scores of individual tests or they have relied on data from one or two modalities, which limited investigating the incremental value of combining various modalities. Understanding the joint and separate value of different feature

sets are of interest to new clinical trials as it could affect recruitment strategies due to associated cost and burden of each modality. Obtaining demographic info, NP tests and ApoE4 status is relatively easy and inexpensive; however, obtaining and processing MRIs are more burdensome (to both the patient and researcher/clinician) and obtaining CSF biomarkers is difficult considering the invasive nature of lumbar punctures. On the other hand, MRI is routinely obtained both in trials and in practice to identify or exclude structural factors that could contribute to MCI, such as mass lesions or vascular disease. Given that the MRI is part of the evaluation, the incremental cost usually arises from image processing and not image acquisition.

It is important to note that interpretation of the performance of the prediction models (and therefore their effectiveness) should be evaluated based on the clinical or research question and the clinical setting. One setting in which such models could be of use is in a primary care setting for screening, especially when an effective treatment for Aβ+ patients becomes available. In such settings, using models with the *highest sensitivity* are more suitable. Another setting that these models could be used is for enrichment of AD clinical trials in which Aβ positivity on PET scan is an enrollment criterion. In such cases, amyloid risk models with *high PPV* are the most desirable models for reducing the number of unnecessary PET scans and decreasing costs and burden of trial. For example, let's assume a trial design that requires 1000 Aβ+ aMCI participants to be enrolled and Aβ status verified using amyloid PET. Assuming that the aMCI population that participants are selected from are similar to the ADNI cohort, prevalence of Aβ+ individuals with aMCI would be 61.0%. Therefore, without use of any predictive models, 1639 individuals who have passed the initial clinical prescreening should undergo amyloid PET screening to identify 1000 Aβ+ individuals. Using a predictive model incorporating demographics, ApoE4 status, and NP (model 1 in Table 3), can decrease the number of participants to undergo PET scan to 1263 individuals (approximately 23% decrease in number of screening PET scans), and reduce the costs by >2.5 million USD (with an approximate cost of 5000 USD for acquisition and analysis of each PET scan), while concurrently decreasing the number of people undergoing this invasive and time-consuming procedure. This cost-saving calculation is in line with reports of previous studies that have suggested using predictive models to enrich clinical trials [36, 38]. It should be noted that in these studies and in our example above, the costs associated with clinical prescreening and NP testing is either ignored or it is assumed that they are obtained through an online platform at no cost. However, in practice, most clinical trials still require a clinic visit for clinical prescreening and NP testing, which costs approximately $1000 per person in USA (considerably less in Europe [39]). The number needed to screen in a design using amyloid PET predictive models is substantially higher: in the example above, clinical data and NP tests should be obtained from a total of 2193 participants to identify 1263 individuals who are predicted to be amyloid positive based on Model 1. Therefore, the costs of in-person clinical visit can potentially offset the costs of obtaining fewer PET scans. Considering that AD therapy is moving toward using drugs targeting tau or combination therapies (e.g., tau and amyloid), in the long run, such predictive models along with online prescreening tools can substantially decrease the costs of trial while decreasing the number of people undergoing invasive and time-consuming procedures. Additionally, considering the high PPV of models that include CSF biomarkers (>90%), and lower costs of obtaining and analyzing CSF (approximately $1000 in 2019), it might be a reasonable choice to replace amyloid PET data with CSF data when obtaining PET scans is not an option.

A few limitations for this study should be mentioned. First, ADNI is not a population-based study and there are strict inclusion and exclusion criteria for selection of participants, which can affect generalizability of our findings. Therefore, validating these models in other population-based studies and clinical trials' data is an essential next step. Moreover, the inclusion criteria in ADNI study may further limit the applicability of the findings presented here to a broader range of patients. This study focused on aMCI subjects and it is possible that in a broader selection of MCI population or in individuals with subjective cognitive complaints who do not meet MCI criteria, the models might show different capabilities in prediction of amyloid status. Although we showed that using our models can reduce costs of conducting research trials or clinical practice, it is difficult to estimate the imposed burden of obtaining additional tests (e.g., MRIs, lumbar punctures, etc.) on patients, caregivers, or researchers and clinicians. Ultimately, efficiency of clinical trials depends not just on reducing the cost of amyloid PET scanning but on the identification of persons who will progress in the absence of treatment and who are more

likely to respond to treatment. Similar approaches have been used extensively for conducting research in other neurodegenerative disease such as Parkinson's disease and have shown substantial potential for use. In a subsequent study, we plan to investigate the rate of progression in various groups as identified by predictive models.

To conclude, our results indicate that predictive models can be effectively used to decrease the number of participants who need to undergo amyloid PET scans. This approach can potentially decrease the costs of the trial and also decrease the burden on patients and caregivers who are participating in the trial. By implementing a step-by-step screening (adaptable design) procedure to enroll participants in trials and using validated predictive models, we can reduce the number of screen failures due to biomarker inclusion criteria and associated costs. A similar approach can be used to improve clinical decision-making with the least associated cost and burden for treatment of patients in AD continuum when effective treatments targeted at AD pathology becomes available.

## ACKNOWLEDGMENTS

Data collection and sharing for ADNI 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.

This work was supported by a grant from the Alzheimer's Association (Ezzati, 2019-AACSF-641329), and the Leonard and Sylvia Marx Foundation. Authors of this study were supported in part by grants from National Institutes of Health NIA 2 P01 AG03949 (Lipton), NIA 1R01AG039409-01 (Lipton), NIH K01AG054700 (Zammit), and the Czap Foundation (Lipton).

Authors' disclosures available online (https://www.j-alz.com/manuscript-disclosures/19-1038r1).

## SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/JAD-191038.

## REFERENCES

[1] Jack CR Jr, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, Shaw LM, Vemuri P, Wiste HJ, Weigand SD (2013) Tracking pathophysiological processes in Alzheimer's disease: An updated hypothetical model of dynamic biomarkers. *Lancet Neurol* **12**, 207-216.

[2] Salloway S, Sperling R, Fox NC, Blennow K, Klunk W, Raskind M, Sabbagh M, Honig LS, Porsteinsson AP, Ferris S (2014) Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. *N Engl J Med* **370**, 322-333.

[3] Doody RS, Thomas RG, Farlow M, Iwatsubo T, Vellas B, Joffe S, Kieburtz K, Raman R, Sun X, Aisen PS (2014) Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease. *N Engl J Med* **370**, 311-321.

[4] Coric V, Salloway S, van Dyck CH, Dubois B, Andreasen N, Brody M, Curtis C, Soininen H, Thein S, Shiovitz T (2015) Targeting prodromal Alzheimer disease with avagacestat: A randomized clinical trial. *JAMA Neurol* **72**, 1324-1333.

[5] Karran E, Mercken M, De Strooper B (2011) The amyloid cascade hypothesis for Alzheimer's disease: An appraisal for the development of therapeutics. *Nat Rev Drug Discov* **10**, 698-712.

[6] Brashear HR (2015) Comment: Age effects on clinical trial results in Alzheimer dementia. *Neurology* **84**, 1126.

[7] Jack CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, Holtzman DM, Jagust W, Jessen F, Karlawish J (2018) NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement* **14**, 535-562.

[8] Palmqvist S, Zetterberg H, Mattsson N, Johansson P, Minthon L, Blennow K, Olsson M, Hansson O, Swedish BioFINDER Study Group (2015) Detailed comparison of amyloid PET and CSF biomarkers for identifying early Alzheimer disease. *Neurology* **85**, 1240-1249.

[9] Teipel S, Drzezga A, Grothe MJ, Barthel H, Chételat G, Schuff N, Skudlarski P, Cavedo E, Frisoni GB, Hoffmann W (2015) Multimodal imaging in Alzheimer's disease: Validity and usefulness for early detection. *Lancet Neurol* **14**, 1037-1053.

[10] Grill JD, Karlawish J (2010) Addressing the challenges to successful recruitment and retention in Alzheimer's disease clinical trials. *Alzheimers Res Ther* **2**, 34.

[11] Laforce R, Rabinovici GD (2011) Amyloid imaging in the differential diagnosis of dementia: Review and potential clinical applications. *Alzheimers Res Ther* **3**, 31.

[12] Siontis GC, Tzoulaki I, Siontis KC, Ioannidis JP (2012) Comparisons of established risk prediction models for cardiovascular disease: Systematic review. *BMJ* **344**, e3318.

[13] Skates SJ, Xu FJ, Yu YH, Sjövall K, Einhorn N, Chang Y, Bast RC, Knapp RC (1995) Toward an optimal algorithm for ovarian cancer screening with longitudinal tumor markers. *Cancer* **76**, 2004-2010.

[14] Wood DE (2015) National Comprehensive Cancer Network (NCCN) clinical practice guidelines for lung cancer screening. *Thorac Surg Clin* **25**, 185-197.

[15] Witten IH, Frank E, Hall MA, Pal CJ (2016) *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann.

[16] Obermeyer Z, Emanuel EJ (2016) Predicting the future—big data, machine learning, and clinical medicine. *N Engl J Med* **375**, 1216-1219.

[17] Shah ND, Steyerberg EW, Kent DM (2018) Big data and predictive analytics: Recalibrating expectations. *JAMA* **320**, 27-28.

[18] Ezzati A, Zammit AR, Harvey DJ, Habeck C, Hall CB, Lipton RB, Alzheimer's disease neuroimaging initiative (2019) Optimizing machine learning methods to improve predictive models of Alzheimer's disease. *J Alzheimers Dis* **71**, 1027-1036.

[19] Ezzati A, Zammit AR, Habeck C, Hall CB, Lipton RB, Alzheimer's disease neuroimaging initiative (2019) Detecting biological heterogeneity patterns in ADNI amnestic mild cognitive impairment based on volumetric MRI. *Brain Imaging Behav*, doi: 10.1007/s11682-019-00115-6

[20] Zammit AR, Hall CB, Bennett DA, Ezzati A, Katz MJ, Muniz-Terrera G, Lipton RB (2019) Neuropsychological latent classes at enrollment and postmortem neuropathology. *Alzheimers Dement* **15**, 1195-1207.

[21] Zammit AR, Hall CB, Katz MJ, Muniz-Terrera G, Ezzati A, Bennett DA, Lipton RB (2018) Class-specific incidence of all-cause dementia and Alzheimer's disease: A latent class approach. *J Alzheimers Dis* **66**, 347-357.

[22] Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189-198.

[23] Mirra SS, Heyman A, McKeel D, Sumi S, Crain BJ, Brownlee L, Vogel F, Hughes J, Van Belle G, Berg L (1991) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* **41**, 479-486.

[24] Mohs R (1994) *Administration and scoring manual for the Alzheimer's Disease Assessment Scale, 1994 revised edition*. The Mount Sinai School of Medicine, New York.

[25] Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, Van Der Kouwe A, Killiany R, Kennedy D, Klaveness S (2002) Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341-355.

[26] Landau SM, Lu M, Joshi AD, Pontecorvo M, Mintun MA, Trojanowski JQ, Shaw LM, Jagust WJ, Alzheimer's disease neuroimaging initiative (2013) Comparing positron emission tomography imaging and cerebrospinal fluid measurements of β-amyloid. *Ann Neurol* **74**, 826-836.

[27] Landau SM, Breault C, Joshi AD, Pontecorvo M, Mathis CA, Jagust WJ, Mintun MA (2013) Amyloid-β imaging with Pittsburgh compound B and florbetapir: Comparing radiotracers and quantification methods. *J Nucl Med* **54**, 70-77.

[28] Shaw LM, Vanderstichele H, Knapik-Czajka M, Figurski M, Coart E, Blennow K, Soares H, Simon AJ, Lewczuk P, Dean RA (2011) Qualification of the analytical and clinical performance of CSF biomarker analyses in ADNI. *Acta Neuropathol* **121**, 597-609.

[29] Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, Blennow K, Soares H, Simon A, Lewczuk P (2009) Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol* **65**, 403-413.

[30] Zhang C, Ma Y (2012) *Ensemble machine learning: Methods and applications*, Springer.

[31] Kuncheva LI (2004) *Combining pattern classifiers: Methods and algorithms*, John Wiley & Sons.

[32] Hastie T, Friedman J, Tibshirani R (2001) Model assessment and selection. In *The elements of statistical learning*. Springer, pp. 193-224.

[33] Haghighi M, Smith A, Morgan D, Small B, Huang S, Alzheimer's Disease Neuroimaging Initiative (2015) Identifying cost-effective predictive rules of amyloid-β level by integrating neuropsychological tests and plasma-based markers. *J Alzheimers Dis* **43**, 1261-1270.

[34] Kandel BM, Avants BB, Gee JC, Arnold SE, Wolk DA (2015) Neuropsychological testing predicts cerebrospinal fluid amyloid-β in mild cognitive impairment. *J Alzheimers Dis* **46**, 901-912.

[35] Kim SE, Woo S, Kim SW, Chin J, Kim HJ, Lee BI, Park J, Park KW, Kang D-Y, Noh Y (2018) A nomogram for predicting amyloid PET positivity in amnestic mild cognitive impairment. *J Alzheimers Dis* **66**, 681-691.

[36] Palmqvist S, Insel PS, Zetterberg H, Blennow K, Brix B, Stomrud E, Mattsson N, Hansson O, Alzheimer's Disease Neuroimaging Initiative (2019) Accurate risk estimation of β-amyloid positivity to identify prodromal Alzheimer's disease: Cross-validation study of practical algorithms. *Alzheimers Dement* **15**, 194-204.

[37] Lee JH, Byun MS, Yi D, Sohn BK, Jeon SY, Lee Y, Lee J-Y, Kim YK, Lee Y-S, Lee DY (2018) Prediction of cerebral amyloid with common information obtained from memory clinic practice. *Front Aging Neurosci* **10**, 309.

[38] Insel PS, Palmqvist S, Mackin RS, Nosheny RL, Hansson O, Weiner MW, Mattsson N, Alzheimer's Disease Neuroimaging Initiative (2016) Assessing risk for preclinical β-amyloid pathology with APOE, cognitive, and demographic information. *Alzheimers Dement (Amst)* **4**, 76-84.

[39] Valcárcel-Nazco C, Perestelo-Pérez L, Molinuevo JL, Mar J, Castilla I, Serrano-Aguilar P (2014) Cost-effectiveness of the use of biomarkers in cerebrospinal fluid for Alzheimer's disease. *J Alzheimers Dis* **42**, 777-788.