Received: 3 February 2021 | Revised: 15 June 2021 | Accepted: 15 June 2021

DOI: 10.1002/alz.12427

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

RESEARCH ARTICLE

Comparing data-driven and hypothesis-driven MRI-based predictors of cognitive impairment in individuals from the Atherosclerosis Risk in Communities (ARIC) study

Ramon Casanova[1] | **Fang-Chi Hsu**[1] | **Ryan T. Barnard**[1] | **Andrea M. Anderson**[1] | **Rajesh Talluri**[2] | **Christopher T. Whitlow**[3] | **Timothy M. Hughes**[4] | **Michael Griswold**[2] | **Kathleen M. Hayden**[5] | **Rebecca F. Gottesman**[6] | **Lynne E. Wagenknecht**[7] | **for the Alzheimer's Disease Neuroimaging Initiative**[†]

[1] Department of Biostatistics and Data Science, Wake Forest School of Medicine, Winston-Salem, USA

[2] University of Mississippi Medical Center, Jackson, Mississippi, USA

[3] Department of Radiology, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA

[4] Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA

[5] Department of Social Sciences and Health Policy, Wake Forest School of Medicine, Winston-Salem, USA

[6] Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[7] Divison of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA

Correspondence
Ramon Casanova, Department of Biostatistics and Data Science, Wake Forest School of Medicine, Winston-Salem, North Carolina 27157, USA.
Email: Casanova@wakehealth.edu

[†]Alzheimer's Disease Neuroimaging Initiative: 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

Abstract

Introduction: A data-driven index of dementia risk based on magnetic resonance imaging (MRI), the Alzheimer's Disease Pattern Similarity (AD-PS) score, was estimated for participants in the Atherosclerosis Risk in Communities (ARIC) study.

Methods: AD-PS scores were generated for 839 cognitively non-impaired individuals with a mean follow-up of 4.86 years. The scores and a hypothesis-driven volumetric measure based on several brain regions susceptible to AD were compared as predictors of incident cognitive impairment in different settings.

Results: Logistic regression analyses suggest the data-driven AD-PS scores to be more predictive of incident cognitive impairment than its counterpart. Both biomarkers were more predictive of incident cognitive impairment in participants who were White, female, and apolipoprotein E gene (*APOE*) ε4 carriers. Random forest analyses including predictors from different domains ranked the AD-PS scores as the most relevant MRI predictor of cognitive impairment.

Conclusions: Overall, the AD-PS scores were the stronger MRI-derived predictors of incident cognitive impairment in cognitively non-impaired individuals.

KEYWORDS
AD-PS, Alzheimer's disease, ARIC, machine learning, MRI, random forest

1 | INTRODUCTION

Despite intense research into novel biomarkers of dementia derived from blood and positron emission tomography (PET) imaging,[1,2] biomarkers from structural magnetic resonance imaging (MRI) remain an area of great interest. MRI captures cumulative damage caused by pathological processes over time[3,4] and is (1) less expensive than PET and less invasive than obtaining cerebrospinal fluid; (2) readily available in large legacy databases where other Alzheimer's disease (AD) biomarkers were not collected; (3) characterizes neurodegeneration within the Amyloid/Tau/Neurodegeneration (A/T/N) model[5]; and (4) can define severity and progression of brain disease.

The development of MRI-based biomarkers of dementia risk remains an active area of research[6–10] that continues to produce new innovations. Some MRI biomarkers are guided by expert knowledge. Racine and colleagues proposed the personalized AD cortical thickness index.[11] They used a composite measure estimated as the average cortical thickness of nine regions believed to be early targets of AD to predict progression from mild cognitive impairment (MCI) to dementia. Brickman and colleagues proposed a measure of degenerative and cerebrovascular pathology,[12] which correlated with amyloid beta (Aβ) PET imaging and cerebrospinal fluid levels of total tau, phosphorylated tau, and A$\beta_{1\text{-}42}$ and predicted incident cognitive impairment. Wu et al. investigated the value of different MRI measures as risk factors for incident MCI and AD[13] in the Atherosclerosis Risk in Communities (or ARIC) cohort, reporting that both brain tissue atrophy and vascular lesions contribute to dementia and cognitive impairment in ARIC. These approaches have in common the use of hypothesis-based composite measures that include several brain regions susceptible to AD. Some composites are volumetric and others are based on the cortical thickness average of the hypothesized brain regions.

Several groups have proposed MRI data-driven biomarkers based on machine learning methods.[6,7,9] Very few have been systematically deployed in the context of AD and related dementias. The Spatial Pattern of Abnormality for Recognition of Early Alzheimer's Disease index is a better known example of a data-driven index of AD risk, which has been applied to different problems in AD.[7,14–16] We introduced the Alzheimer's Disease Pattern Similarity (AD-PS) scores using high-dimensional machine learning methods.[10,17–19] This work was extended to the Women's Health Initiative Memory Study (WHIMS) MRI cohort,[20] where AD-PS scores were associated with incident cognitive impairment, age, and global cognitive function. Scores were consistent with the relative trajectories of global cognitive function in WHIMS women over 10 years of follow-up.[21] WHIMS AD-PS scores as a measure of neuroanatomic risk of dementia have been linked to air pollution.[22,23]

To date, there have been relatively few comparisons between data-driven and hypothesis-driven MRI indices as predictors of incident cognitive impairment, particularly in diverse cohorts. This work pursues several objectives: (1) to extend AD-PS scores to the ARIC cohort and evaluate their associations with incident cognitive impairment in a diverse cohort of cognitively nonimpaired individuals; (2) to evaluate the relative merit of AD-PS scores compared to a hypothesis-driven composite volumetric measure of several brain regions susceptible to AD available in ARIC; and (3) to perform exploratory stratified analyses across sex, race, and apolipoprotein E gene (*APOE*) ε4 carrier status to evaluate the impact of these factors on the AD-PS scores and the composite volumetric measure when predicting incident cognitive impairment.

HIGHLIGHTS

- A data-driven score was estimated via machine learning for Atherosclerosis Risk in Communities (ARIC) participants.
- The training data set was composed of Alzheimer's Disease Pattern Similarity (ADNI) magnetic resonance (MR) images.
- The score was a strong predictor of cognitive impairment in a diverse cohort.
- It outperformed an anatomically defined composite volumetric measure.
- The score outperformed other MRI measures when predicting cognitive impairment.

RESEARCH IN CONTEXT

1. **Systematic review**: The authors reviewed the literature using traditional methods like PubMed. Although the applications of machine learning to the analysis of structural magnetic resonance imaging (MRI) are abundant, comparisons of data-driven and hypothesis-driven biomarkers as predictors of incident cognitive impairment in diverse cohorts are rare.
2. **Interpretation**: Our analyses suggest that our machine learning–generated data-driven score is more predictive of incident cognitive impairment than a hypothesis-driven composite volumetric measure, including the volumes of several regions susceptible to Alzheimer's disease (AD) in early stages. This work also suggests the effectiveness of generalization of neurodegeneration patterns across imaging databases produced via machine learning inference.
3. **Future directions**: Our work highlights the potential of machine learning to generate AD biomarkers derived from the data. We plan to continue this work by investigating the use of more sophisticated methods from the field of artificial intelligence, larger sample sizes of the training data sets, and other types of information.

2 | MATERIALS AND METHODS

Two data sets were utilized for this study. ARIC is the main target cohort and ADNI MRI data was used to train machine learning algorithms to generate AD-PS scores when provided with MRI data from ARIC participants.

The ADNI was launched in 2003 as a public-private partnership, led by principal investigator Michael W. Weiner, MD, USA. The primary goal of ADNI 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. The ADNI study provides a rich and well-characterized cohort of cognitively normal participants and AD patients, which we have used actively in our previous work.[24–26] The ADNI data are described in the Supplemental materials.

The ARIC study began in 1987, funded by the National Heart, Lung, and Blood Institute (NHLBI). From 1987 through 1989, a total of 15,792 mostly White and African American participants aged 45-64 years were recruited from four field centers located in Forsyth County, NC, USA; Jackson, MS, USA; Minneapolis suburbs, MN, USA; and Washington County, MD, USA. Using probability sampling, each ARIC field center recruited 4000 individuals aged 45-64 years from a defined population in their community. Only African Americans were recruited in Jackson, MS, USA; the remaining sites reflected local populations, mostly White in Minneapolis and Washington County and both races in Forsyth County. The institutional review boards from all centers approved ARIC protocols; participants provided written consent for their study participation and for use of their genetic data. To date, there have been seven examinations; relevant to this work are visit 5 (2011 to 2013) and visit 6 (2016 to 2017).

3 | ARIC DATA

3.1 | ARIC-Cognitive evaluation

The ARIC cognitive assessment used in visits 5 was described previously.[27] Briefly, ARIC obtained cognitive evaluations at visits 5 and 6, with a mean follow-up of 4.86 years. The three cognitive instruments administered beginning with ARIC visit 2 and used in visit 5 include: the Delayed Word Recall Task (DWRT), Digit Symbol Substitution (DSS) from the Wechsler Adult Intelligence Scale -Revised (WAIS-R), and a Word Fluency Test. Z scores for each test were estimated using mean and standard deviations from visit 2. Factor scores were used for global cognition, and for three previously derived cognitive domain: executive function, language, and memory.[28–30]

The cognitive status (non-impaired, MCI, or dementia) of participants who attended visits 5 and 6 was classified using a standardized algorithm based on cognitive assessment and verified by expert committee review, using information from in-person cognitive batteries, the Clinical Dementia Rating scale, and functional questionnaires completed by participants and/or informants. Because the goal of this study is to evaluate early detection of dementia risk using imaging biomarkers, only cognitively non-impaired individuals (CNI) at visit 5 were included in our analyses. Cognitively non-impaired was defined as not meeting criteria for MCI or dementia (Supplementary materials). MCI type and the etiology of dementia were not adjudicated at visit 6.

3.2 | ARIC-MRI

Structural brain images were obtained using 3-T MRI scanners (Siemens Verio [Maryland site], Siemens Skyra [North Carolina study center], Siemens Trio [Minnesota site], and Siemens Skyra [Mississippi site]) as described previously.[31] We used cortical volumes of regions of interest, estimated using the FreeSurfer system (Laboratory for Computational Neuroimaging) available in the ARIC database: frontal, temporal, occipital, parietal, deep gray matter, ventricular, total brain volume (TBV), and a composite of brain regions volumes susceptible to AD, including hippocampus, parahippocampal, entorhinal, inferior parietal lobule, precuneus, and cuneus. This last measure has been referenced in previous work as "AD-signature."[13,32] Due to growing evidence indicating these areas could be the target of other brain diseases like limbic-predominant age-related TDP-43 encephalopathy,[33] we will use the term "composite volumetric measure of regions susceptible to AD" or simply "composite volumetric measure (CVM)." We used white matter hyperintensity (WMH) volumes and volumes of several brain regions (Table S2) as predictors in some of the analyses and the intracranial volumes (ICV) to adjust for differences in brain sizes. MRI scans from 839 ARIC CNI were available.

3.3 | ARIC PET data

A subset of participants without dementia, ages 67-88 years, were imaged using 18-florbetapir PET at three sites (Maryland; North Carolina; and Mississippi) during visit 5. The details of 18-florbetapir PET image processing and co-registration with MRI, carried out at the Johns Hopkins University reading center, were described previously[34] and can be found in the Supplementary materials. A global cortical measure of florbetapir uptake was used as a weighted average (based on region of interest size) of the orbitofrontal, prefrontal, and superior frontal cortices; the lateral temporal, parietal, and occipital lobes; and the precuneus, the anterior cingulate, and the posterior cingulate. This measure is called global cortical standardized uptake value ratio (GC SUVR). We included in our analyses standardized upload value ratios (SUVRs) from individual regions linked in prior studies to AD: medial temporal, amygdala, hippocampus, anterior cingulate, posterior cingulate, caudate, putamen, and thalamus.[29] An automated region for cerebellum gray matter was used as a reference. Here we used data from 193 participants who were adjudicated as CNI on visit 5.

3.4 | Estimation of the AD-PS scores

The general approach is depicted in Figure 1. MRI scans from both ARIC and ADNI (Table S1) were aligned to a common template (derived

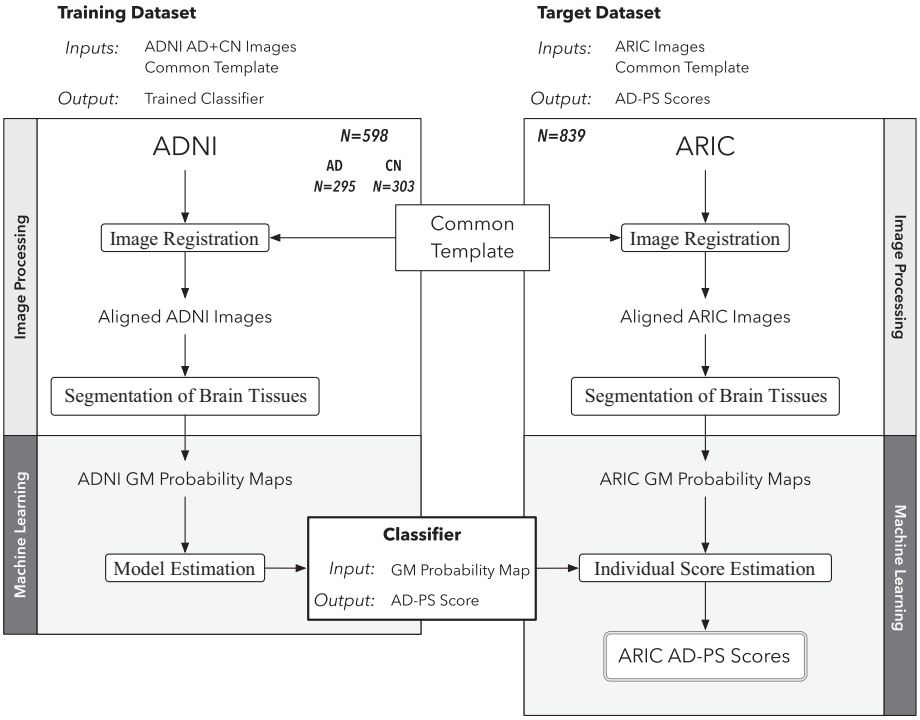

FIGURE 1 General approach to estimate Alzheimer's disease pattern similarity (AD-PS) scores for ARIC MRI images

from ADNI images) using image-processing tools available in the Advanced Normalization Tools (see image processing in Supplementary materials). Next we used high-dimensional machine learning methods to estimate the AD-PS scores. Details of the machine learning algorithms were published previously.[10,19,20,22] Briefly, a regularized logistic regression (RLR) classifier was estimated in a voxel-wise manner using the gray matter probability maps (resulting from the image processing described above) from CN and AD participants available in the training data set (ADNI in our case). The weights $\hat{\beta}$ estimated after solving the optimization problem associated with the RLR classifier are used to estimate conditional probabilities of AD according to the MRI scan. To estimate the optimal values of the regularization parameters, we combined nested 10-fold cross-validations and grid search. Probabilities we refer to as AD-PS scores were computed as the mean values of five repetitions of the computations, to account for variability due to random partitioning of cross-validation that occurred during model estimation. The scores are computed for the target data set, in this case ARIC.

3.5 | Analyses

We performed logistic regression analyses considering incident cognitive impairment (either MCI or dementia) as the outcome and AD-PS scores or the CVM described above as main independent variables, fitted in separate models. Other variables included age, race-center, sex, education, and ICV, where race-center combined information from both race and study center. We estimated area under the curve (AUC) based on 10-fold cross-validation. Similar analyses were performed stratifying by sex, race and *APOE* ε4 carrier status. The Delong method[35] was used to evaluate significance of the increase in performance resulting by adding each biomarker (AD-PS or CVM) to the basic model based only on covariates. We performed random forests (RF)[36] classification analyses (see Supplementary materials for RF details) to investigate relative importance of both MRI biomarkers when predicting incident cognitive impairment including AD-PS scores, the CVM, several MRI variables derived using FreeSurfer (described above), cognitive, clinical, demographics, and *APOE* ε4 data (Table S2). The derived MRI variables were scaled by dividing them by their corresponding ICV.

Similar logistic regression and RF analyses were performed using a subset (N = 193) of CNI with MRI and amyloid PET data available. This allowed us to compare performance of the anatomical measures derived from MRI with measures derived from 18-florbetapir PET. RF models were constructed using the same predictors as described above but adding several Aβ amyloid PET measures to the model from brain areas previously linked to AD (see the ARIC PET section above).

All RF analyses were performed using methods for imbalanced classification available in the R package randomForestSRC.[37,38] We selected ntree = 3000 and AUC as the splitting rule, both of which are recommended for imbalanced learning. Other parameters were set to the default values. The permutation index was used to determine variable importance. The pROC R package was used to estimate AUCs and confidence intervals as a measure of performance. For variable selection we used a strategy proposed by Strobl for computing an RF permutation index,[39] who suggested discarding as noise the variables with negative permutation index and the positive ones with absolute values less than the amplitude of the negative score with maximum amplitude.[40,41]

TABLE 1 Demographic characteristics, *APOE* ε4 carrier status, and MRI measures at visit 5 for CNI participants with MRI ($N = 839$) and for those with MRI and amyloid PET ($N = 193$) by cognitive status at visit 6

		Cognitive Status at Visit 6		
	Total	**Normal**	**MCI**	**Dementia**
N	839	691	122	26
Age mean (SD)	75.3 (5.0)	74.9 (5.0)	76.5 (4.7)	79.4 (4.0)
Gender				
Female	555 (66.2%)	435 (63.1%)	78 (63.9%)	21 (80.8%)
Male	316 (34.8%)	256 (36.9%)	44 (35.1%)	5 (19.2%)
Race				
Black	272 (32.4%)	227 (32.9%)	31 (25.4%)	14 (53.9%)
White	567(67.6%)	464 (67.1%)	91 (74.6%)	12 (46.1%)
Education				
Basic	98 (11.7%)	72 (10.4%)	20 (16.4%)	6 (23.1%)
Intermediate	306 (36.5%)	264 (38.2%)	34 (27.9%)	8 (30.8%)
Advanced	435 (51.8%)	355 (51.4%)	68 (55.7%)	12 (46.1%)
APOE ε4 carrier status				
ε4 carrier (%)	224 (27.7%)	171 (24.8%)	45 (36.9%)	8 (30.8%)
AD-PS scores, mean (SD)	0.20 (20.3)	0.18 (0.18)	0.30 (0.25)	0.47 (0.33)
Composite volume, mean (SD)	59893.1 (6674.0)	60270.5 (6654.3)	58943.2 (6349.4)	54321 (5935.1)
MRI and amyloid PET Subset				
		Cognitive Status at Visit 6		
	Total	**Normal**	**MCI**	**Dementia**
N	193	174	16	3
Age mean (SD)	75.0 (5.2)	74.8 (5.3)	76.6 (4.8)	79.3 (4.5)
Gender				
Female	121 (62.7%)	109 (62.6%)	10 (62.5%)	2 (66.7%)
Male	72 (37.3%)	65 (37.4%)	6 (37.5%)	1 (33.3%)
Race				
Black	78 (40.4%)	69 (39.7%)	6 (37.5%)	3 (100%)
White	115 (59.6%)	105 (63.3%)	10 (62.5%)	0 (0%)
Education*				
Basic	23 (11.9%)	18 (10.3%)	4 (25%)	1 (33.4%)
Intermediate	83 (43.0%)	75 (43.1%)	8 (50%)	0 (0.0%)
Advanced	87 (45.1%)	81 (46.6%)	4 (25%)	2 (66.6%)
APOE ε4 carrier status				
Yes (%)*	50 (25.9%)	41 (23.6%)	8 (50%)	1 (33.3%)
AD-PS scores, mean (SD)	0.19 (0.19)	0.18 (0.18)	0.37 (0.20)	0.39 (0.42)
AD-PS scores, mean (SD)	0.19 (0.19)	0.18 (0.18)	0.37 (0.20)	0.39 (0.42)
Composite volume, mean (SD)	59588 (6448)	59923 (6414)	57737 (5842)	50040 (1546)
Global cortical SUVR, mean (SD)	1.23 (0.20)	1.21 (0.17)	1.44 (0.31)	1.44 (0.35)

Abbreviations: GED, General Educational Development test.
Education levels – Basic (<12 years), Intermediate (completed high school or GED), Advanced (some college).

4 | RESULTS

Age, race, education, *APOE* ε4 carrier status, CVM, and AD-PS scores at visit 5 for 839 CNI with MRI are presented in Table 1, distributed by cognitive status as adjudicated at visit 6. The average AD-PS scores of CNI estimated at visit 5, increases with the severity of the future cognitive impairment classification as adjudicated in visit 6.

In Table 2, results for the incident impairment analyses based on logistic regression using the full sample ($N = 839$) and stratifications by race, sex, and *APOE* ε4 carrier status are presented. Overall, AUC

TABLE 2 The predictive value of the two MRI measures in CNI participants at visit 5 with MRI and for those with MRI and amyloid PET

Group	Effect	*N*	Cases (*Dem.)	AUC	95% CI	**P
Full Sample	AD-PS	839	148 (26)	0.692	[0.687-0.697]	.0004
	CVM			0.654	[0.650-0.658]	.036
White	AD-PS	567	103 (12)	0.688	[0.681-0.695]	.006
	CVM			0.672	[0.664-0.680]	.10
Black	AD-PS	272	45 (14)	0.645	[0.625-0.665]	.10
	CVM			0.596	[0.576-0.616]	.08
Male	AD-PS	305	49 (5)	0.661	[0.643-0.679]	.08
	CVM			0.615	[0.597-0.633]	.17
Female	AD-PS	*534*	99 (25)	0.685	[0.677-0.693]	.003
	CVM			0.645	[0.637-0.653]	.21
APOE ε4 carriers	AD-PS	224	53 (8)	0.744	[0.728-0.760]	.06
	CVM			0.692	[0.674-0.710]	.37
APOE ε4 Non-carriers	AD-PS	615	95 (18)	0.668	[0.659-0.677]	.002
	CVM			0.617	[0.609-0.625]	.07
MRI and amyloid PET Subset						
Group	**Effect**	***N***	**Cases (Dem)**	**AUC**	**95% CI**	****P**
Full Sample	AD-PS	193	19 (3)	0.737	[0.714-0.760]	.024
	CVM			0.633	[0.605-0.661]	.19
	GC SUVR			0.672	[0.633-0.711]	.18

*Dem – Number of dementia cases.
**P-value - Delong test comparing AUCs produced by AD-PS or CVM plus covariates with respect to the basic model containing only covariates.

for incident cognitive impairment was 0.692 for AD-PS and 0.654 for the CVM. Consistently, AD-PS scores were significantly more predictive of incident cognitive impairment than the CVM (see also Table S3). Both MRI-based biomarkers were more predictive of incident cognitive impairment in participants who were White, female, and *APOE ε4* carriers. Spearman correlation between both measures was rho = −0.13 ($P < .001$).

For participants with both MRI and PET ($N = 193$), we present the demographic information at visit 5 (Table 1) and logistic regression results for the full sample (Table 2). The model including AD-PS scores, generated the larger AUC when predicting incident cognitive impairment. For completeness, the results across different stratifications of the data set are presented in the Supplementary materials in Table S4, although the sample size in several cases was small.

Figure 2 shows the rank of predictors based on the RF permutation index resulting from the incident cognitive impairment analysis with 839 CNI at visit 5. The horizontal red line defines the threshold for variable selection according to the Strobl criterion.[39] AD-PS scores were the most relevant predictor of incident cognitive impairment followed by global cognition and memory domain scores, age, CVM, WFT z-score, and ventricular volume. Performance of the classifier was AUC = 0.735 (95% confidence interval [CI] 0.692-0.780. The results of the incident impairment analysis in the PET sample ($N = 193$) are shown in Figure 3. Results were driven mostly by five predictors: three from florbetapir PET(anterior cingulate, posterior cingulate, and global cortical SUVRs) followed by two derived from MRI (AD-PS scores and TBV). The classifier performance was AUC = 0.825 (95% CI 0.726-0.934). Several complementary analyses were performed. Linear regression models were fitted to evaluate associations of both MRI biomarkers with memory and executive function scores. The AD-PS scores were in all cases significantly associated with these cognitive measures cross-sectionally and longitudinally, whereas the CVM was not associated with memory function cross-sectionally and to executive function longitudinally (see Table S5). Logistic regression analyses (amyloid+ vs amyloid−) adjusted for age, sex, education, race, and ICV were fitted to investigate associations of both MRI measures with amyloid PET in CNI, but results were not significant in either case (not presented).

5 | DISCUSSION

The AD-PS score, a data-driven MRI biomarker of dementia risk, was estimated for ARIC participants. ARIC is a diverse cohort containing one of the largest MRI databases ever collected among African Americans. Using data from CNI, AD-PS scores were strong predictors of incident cognitive impairment over 4.86 years of follow-up.

An important goal was to understand the relative merit of our data-driven score with respect to a hypothesis-driven CVM based on regions susceptible to AD. Areas contributing to the AD-PS scores

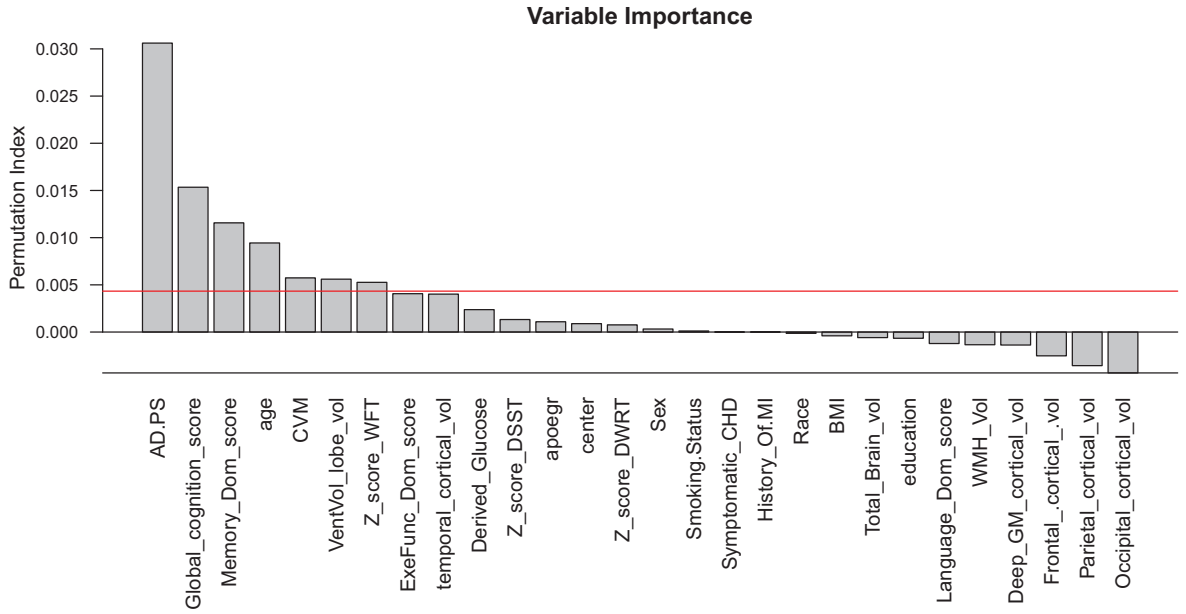

FIGURE 2 Rank of predictors based on the RF permutation index resulting from the incident cognitive impairment analysis with 839 cognitive non-impaired participants at visit 5. The horizontal red line defines the threshold for variable selection. The AD-PS score was the more relevant predictor. Performance of the classifier was AUC = 0.735 [0.692-0.780] CI (95%)

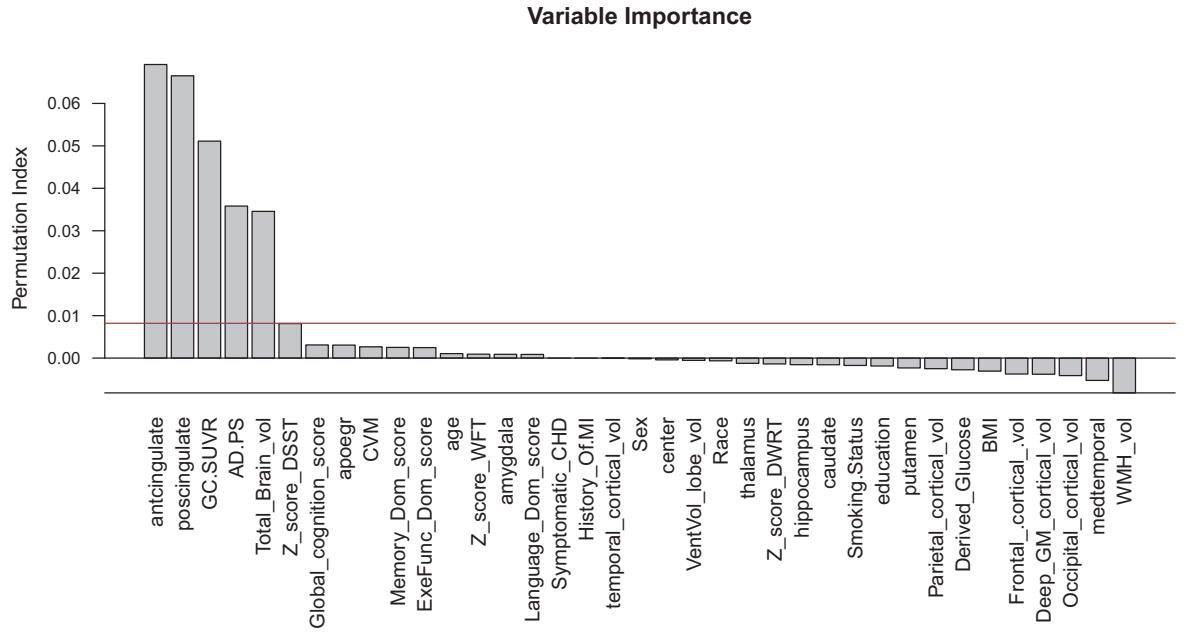

FIGURE 3 Rank of predictors based on the RF permutation index resulting from the incident cognitive impairment analysis of 193 cognitively non-impaired ARIC PET study participants at visit 5 adding to MRI, cognitive, clinical, and demographic data information derived from Florbetapir-18 PET. The horizontal red line defines the threshold for variable selection. The AD-PS scores and total brain volume ranked as the two most relevant MRI measures after anterior cingulate, posterior cingulate and global cortical SURVs, which were the three most relevant predictors in this analysis. Performance of the classifier was AUC = 0.834 [0.735-0.932] CI (95%)

estimation are selected by the algorithm from the gray matter tissue in a voxel-wise manner (see Figure S1 in Supplementary materials). We investigated the relative value of these two MRI-based biomarkers for prediction of incident cognitive impairment using (1) parsimonious logistic regression models and (2) high-dimensional RF models that included other MRI, cognitive, demographic, genetic, and clinical measures. Logistic regression analyses based on individuals with MRI (N = 839) showed that models including the AD-PS scores were very often significantly more predictive of incident cognitive impairment than the CVM. In analyses stratified by race, sex, and *APOE* ε4,

we observed that both MRI metrics were simultaneously more predictive of incident cognitive impairment in White, female, and *APOE* ε4 carriers. Complementary analyses were performed to further investigate relationships between AD-PS and CVM (see Table S3). When both biomarkers were included in the same model, the CVM associations with incident cognitive impairment became nonsignificant in all cases.

To investigate the relative performance of both MRI biomarkers in the presence of multiple variables from different domains we used RF, a state-of-the-art machine learning method well known to the predictive modeling community. The permutation index is perhaps the most popular variable importance measure it provides. It evaluates decreases in model performance when a given variable is randomly permuted. If the variable is important, the model performance will decrease and vice versa. The RF analyses performed using MRI, cognitive, clinical, demographic, and *APOE* ε4 carrier status predictors showed the AD-PS score to be a more relevant predictor of incident cognitive impairment than the CVM. Overall AD-PS scores were the most relevant predictor in the model followed by global cognition and memory domain cognitive scores, age, and ventricular volume. Complementary RF analyses dropping one of the two measures (AD-PS or CVM) at a time from the full model showed that only the removal of AD-PS scores led to significant differences in model performance AUC = 0.717 (95% CI 0.672-0.763) ($P < .05$).[35]

To investigate the performance of the AD-PS scores and the CVM with respect to amyloid PET SUVRs, we took advantage of data collected in the ARIC PET study.[34] Similar analyses with CNI ($N = 193$) were repeated. The results showed the anterior and posterior cingulate and global cortical SUVRs to be more relevant predictors of incident cognitive impairment followed by AD-PS scores. Intriguingly, the other MRI predictor that survived the threshold was TBV, a traditional biomarker of brain tissue atrophy. This is possible due to differences between the data sets (eg, sample sizes, % of cases, demographic differences, and so on) and requires further confirmation.

The anterior and posterior cingulate cortex are part of the default mode network,[42] which has been linked to early changes related to AD.[43] Development of Aβ plaques in the anterior and posterior cingulate in cognitively normal individuals has been reported previously.[44,45] Nadkarni and colleagues used the Amyloid/Tau/Neurodegeneration/Vascular framework[5] to evaluate associations of Aβ amyloid burden, white matter hyperintensities, and fluorodeoxyglucose-PET with incident MCI in cognitively normal individuals.[46] They found that baseline Aβ positivity alone or combined hypo-metabolism positivity was associated with incident amnestic MCI. Similarly, Burnham and colleagues, using data from 573 participants in the Biomarker and Lifestyle (AIBL) study (mean age = 73 years), evaluated their clinical progression over 6 years of follow-up based on Aβ PET and hippocampal volume.[47] They reported that Aβ burden was a risk factor for cognitive decline and progression from preclinical to symptomatic stages of the disease, with neurodegeneration acting as a compounding factor. Similar to other groups, we have used SUVRs of several brain ROIs as predictors.[48,49] However, a more common practice is to use only a composite of different ROIs.

The increasing role of the less expensive plasma biomarkers[2] raises questions about the role of MRI biomarkers in the future. According to the ADRD research framework,[5] Aβ, tau, and MRI biomarkers reside in different domains (A, T, and N, respectively). MRI provides valuable information related to neurodegeneration and cerebrovascular disease. In clinical practice, MRI biomarkers, like the AD-PS scores, can provide complementary information about abnormalities in the brain. In addition, prediction of future events is difficult in any field, and prediction of future cognitive impairment is not an exception due to the complexity of the processes that lead to it. Accurate prediction of incident cognitive impairment will require complex models including predictors from different sources. In the foreseeable future, MRI will be an important component of the diagnostic process[16] and MRI-derived biomarkers will contribute valuable information for mathematical models designed to predict incident cognitive impairment.

Our study is not without limitations. The ADNI cohort is highly selected and not representative of the general population. The majority of participants used in the training data set are White (>90%), and this must be considered when interpreting results across race. Ideally, the training data set would be more representative of the ARIC cohort. However, this is the second database to which AD-PS scores have been extended via machine learning inference. In both cases with very different populations, the scores are predictive of incident cognitive impairment. We believe that the AD-PS scores will benefit from a much larger sample size of the training data set. Our comparison of the AD-PS scores with respect to the composite measure was restricted to the volumetric version available in ARIC. However, other hypothesis-driven volumetric composites or composites based on the average of cortical thickness of several regions instead of volumes[12] could be designed, which could be more sensitive measures to future cognitive decline. Our data set did not contain AD blood-based biomarkers such as p-tau 181,[2] so we were unable to evaluate their relative performance with respect to AD-PS scores. The sample sizes of participants with florbetapir PET who developed cognitive impairment were relatively small, which could impact our analyses. Most African American participants were from one site. Simulation studies have reported permutation index biases in some situations.[50] However, we do not expect this to affect the comparison between AD-PS scores and the CVM or the model's performance.

Finally, despite being based on a training data set (ADNI) with an excellent clinical characterization of AD, AD-PS scores are most likely capturing mixed pathology because AD often coexists or overlaps with other brain diseases or due to classification not verified by biomarkers or confirmed at autopsy. Some researchers have suggested that mixed pathology biomarkers have the potential to be better predictor of future clinical outcomes relative to a biomarker of a specific pathology.[12,51]

6 | CONCLUSIONS

We estimated a data-driven MRI index of dementia risk we call AD-PS scores in the ARIC study. Scores were predictive of incident cognitive

impairment adjudicated ≈4.86 years after the MRI data were collected. Overall, the data-driven AD-PS score outperformed the CVM. RF analyses using variables derived from MRI, amyloid PET, cognitive testing, genetic, and demographic data showed the AD-PS scores to be the most relevant predictor of incident cognitive impairment among the MRI variables but following several amyloid PET measures. Our work supports the potential of data-driven biomarkers of dementia. Future work will investigate early signs of dementia risk according to AD-PS scores and define cut-off values for clinical diagnosis. More important, we will refine our machine learning methodology by investigating the use of very large training data sets, how to include imaging information from large amounts of MCI individuals in model inference, investigate other outcomes to train the machine learning algorithms, the use of Deep Learning and Manifold Learning methods, its applications to other available imaging databases, and extensions or combinations with other imaging modalities.

ACKNOWLEDGEMENTS

The Atherosclerosis Risk in Communities (ARIC) study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700005I, HHSN268201700004I). Neurocognitive data is collected by U01 2U01HL096812, 2U01HL096814, 2U01HL096899, 2U01HL096902, 2U01HL096917 from the National Institutes of Health (NIH) (NHLBI, National Institute of Neurological Disorders and Stroke (NINDS), National Institute on Aging (NIA), and National Institute on Deafness and Other Communication Disorders (NIDCD)), and with previous brain MRI examinations funded by R01-HL70825 from the NHLBI. The authors thank the staff and participants of the ARIC study for their important contributions. RC, KH, and TH receive funding from the Wake Forest Alzheimer's Disease Core Center (P30AG049638-01A1).

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

CONFLICT OF INTEREST

RC received a contract from Atrium Health paid to the Institution to do machine learning using Emergency Room data. FC was consultant for the San Diego State University and was member of a data and safety monitoring board (DSMB) at Rutgers University. KH was consultant with Fred Hutchinson Cancer Research Center and the UNC at Chapel Hill. She received support to attend meetings by the Alzheimer's Association and the NIH. KH also Participated in a DSMB at Wake Forest School of Medicine with no compensation and is editor of several journals of the Alzheimer's Association with no compensation. MG received honoraria for lecture at Johns Hopkins Biostatistics. RG received Honoraria for lectures: University of Alabama at Birmingham; University of Michigan; American College of Cardiology. She is secretary of the American Neurological Association with no compensation. CW received payment as Forensic panel consultant. He received payments from the Biogen MRI Protocol Advisory Board and Biogen US Evolving the Care Team: Treat & Monitor Advisory Board. RB, AA, RT, and TH have nothing to declare.

REFERENCES

1. La Joie R, Visani AV, Baker SL, et al. Prospective longitudinal atrophy in Alzheimer's disease correlates with the intensity and topography of baseline tau-PET. *Sci Transl Med*. 2020;12(524):eaau5732.
2. Karikari TK, Pascoal TA, Ashton NJ, et al. Blood phosphorylated tau 181 as a biomarker for Alzheimer's disease: a diagnostic performance and prediction modelling study using data from four prospective cohorts. *Lancet Neurol*. 2020;19(5):422-433.
3. Jack CR. Brain atrophy on magnetic resonance imaging as a biomarker of neurodegeneration. *JAMA Neurol*. 2016;73(10):1179-1182.
4. Vemuri P, Jack CR. Role of structural MRI in Alzheimer's disease. *Alzheimers Res Ther*. 2010;2(4):23.
5. Jack CR, Bennett DA, Blennow K, et al. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement*. 2018;14(4):535-562.
6. Vemuri P, Whitwell JL, Kantarci K, et al. Antemortem MRI based STructural Abnormality iNDex (STAND)-scores correlate with postmortem Braak neurofibrillary tangle stage. *Neuroimage*. 2008;42(2):559-567.
7. Davatzikos C, Xu F, An Y, Fan Y, Resnick SM. Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index. *Brain*. 2009;132(Pt 8):2026-2035.
8. Kloppel S, Stonnington CM, Barnes J, et al. Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. *Brain*. 2008;131(Pt 11):2969-2974.
9. Kloppel S, Stonnington CM, Chu C, et al. Automatic classification of MR scans in Alzheimer's disease. *Brain*. 2008;131(Pt 3):681-689.
10. Casanova R, Hsu F-C, Sink KM, et al. Alzheimer's disease risk assessment using large-scale machine learning methods. *PLoS One*. 2013;8(11):e77949.
11. Racine AM, Brickhouse M, Wolk DA, Dickerson BC. The personalized Alzheimer's disease cortical thickness index predicts likely pathology and clinical progression in mild cognitive impairment. *Alzheimers Dement (Amst)*. 2018;10:301-310.

12. Brickman AM, Tosto G, Gutierrez J, et al. An MRI measure of degenerative and cerebrovascular pathology in Alzheimer disease. *Neurology*. 2018;91(15):e1402-e1412.
13. Wu A, Sharrett AR, Gottesman RF, et al. Association of brain magnetic resonance imaging signs with cognitive outcomes in persons with non-impaired cognition and mild cognitive impairment. *JAMA Netw Open*. 2019;2(5):e193359.
14. Habes M, Erus G, Toledo JB, et al. White matter hyperintensities and imaging patterns of brain ageing in the general population. *Brain*. 2016;139(Pt 4):1164-1179.
15. Janowitz D, Habes M, Toledo JB, et al. Inflammatory markers and imaging patterns of advanced brain aging in the general population. *Brain Imaging Behav*. 2020;14(4):1108-1117.
16. Habes M, Pomponio R, Shou H, et al. The Brain Chart of Aging: machine-learning analytics reveals links between brain aging, white matter disease, amyloid burden, and cognition in the iSTAGING consortium of 10,216 harmonized MR scans. *Alzheimers Dement*. 2021;17(1):89-102.
17. Casanova R, Whitlow CT, Wagner B, et al. High dimensional classification of structural MRI Alzheimer's disease data based on large scale regularization. *Frontiers of Neuroscience in Neuroinformatics*. 2011;5:22. Epub 2011 Oct 14.
18. Casanova R, Maldjian JA, Espeland MA. Evaluating the impact of different factors on voxel-wise classification methods of ADNI structural MRI brain images. *International Journal of Biomedical Datamining*. 2011;1:11.
19. Casanova R, Hsu F-C. Classification of structural MRI images in Alzheimer's disease from the perspective of ill-posed problems. *PLoS One*. 2012;7(10):e44877.
20. Casanova R, Barnard RT, Gaussoin SA, et al. Using high-dimensional machine learning methods to estimate an anatomical risk factor for Alzheimer's disease across imaging databases. *Neuroimage*. 2018;183:401-411.
21. Espeland MA, Carmichael O, Yasar S, et al. Sex-related differences in the prevalence of cognitive impairment among overweight and obese adults with type 2 diabetes. *Alzheimers Dement*. 2018;14(9):1184-1192.
22. Younan D, Petkus AJ, Widaman KF, et al. Particulate matter and episodic memory decline mediated by early neuroanatomic biomarkers of Alzheimer's disease. *Brain*. 2020;143(1):289-302.
23. Younan D, Wang X, Casanova R, et al. PM2.5 associated with gray matter atrophy reflecting increased Alzheimers risk in older women. *Neurology*. 2020;96(8):e1190-201.
24. Casanova R, Whitlow CT, Wagner B, et al. High dimensional classification of structural MRI Alzheimer's disease data based on large scale regularization. *Frontiers of Neuroscience in Neuroinformatics*. 2011;5:22. Epub 2011 Oct 14 p. 5-22, PMCID: PMC3193072.
25. Casanova R, Hsu F-C. Classification of structural MRI images in Alzheimer's disease from the perspective of ill-posed problems. *PLoS One*. 2012;7(10):e44877. PMCID: PMC3468621.
26. Casanova R, Hsu F-C, Sink KM, et al. Alzheimer's disease risk assessment using large-scale machine learning methods. *PLoS One*. 2013;8(11):e77949. PMCID: PMC3826736.
27. Knopman DS, Gottesman RF, Sharrett AR, et al. Mild cognitive impairment and dementia prevalence: the Atherosclerosis Risk in Communities Neurocognitive Study (ARIC-NCS). *Alzheimers Dement (Amst)*. 2016;2:1-11.
28. Gross AL, Power MC, Albert MS, et al. Application of latent variable methods to the study of cognitive decline when tests change over time. *Epidemiology*. 2015;26(6):878-887.
29. Park LQ, Gross AL, Mclaren DG, et al. Confirmatory factor analysis of the ADNI Neuropsychological Battery. *Brain Imaging Behav*. 2012;6(4):528-539.
30. Hayden KM, Jones RN, Zimmer C, et al. Factor structure of the National Alzheimer's Coordinating Centers uniform dataset neuropsychological battery: an evaluation of invariance between and within groups over time. *Alzheimer Dis Assoc Disord*. 2011;25(2):128-137.
31. Schneider ALC, Selvin E, Sharrett AR, et al. Diabetes, prediabetes, and brain volumes and subclinical cerebrovascular disease on MRI: the Atherosclerosis Risk in Communities Neurocognitive Study (ARIC-NCS). *Diabetes Care*. 2017;40(11):1514-1521.
32. Power MC, Lamichhane AP, Liao D, et al. The association of long-term exposure to particulate matter air pollution with brain MRI findings: the ARIC Study. *Environ Health Perspect*. 2018;126(2):027009.
33. Nelson PT, Dickson DW, Trojanowski JQ, et al. Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. *Brain*. 2019;142(6):1503-1527.
34. Gottesman RF, Schneider ALC, Zhou Y, et al. The ARIC-PET amyloid imaging study: brain amyloid differences by age, race, sex, and APOE. *Neurology*. 2016;87(5):473-480.
35. Delong ER, Delong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837-845.
36. Breiman L. Random forests. *Machine Learning*. 2001;45:5-32.
37. Ishwaran H, Kogalur U, Random forests for survival, regression and clas- sification (RF-SRC) R package version 1.6 http://CRANR-project.org/package=randomForestSRC . 2014.
38. O'Brien R, I H. A random forests quantile classifier for class imbalanced data. *Pattern Recognition*. 2017(90):232-249.
39. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods*. 2009;14(4):323-348.
40. Casanova R, Saldana S, Lutz MW, Plassman BL, Kuchibhatla M, Hayden KM. Investigating predictors of cognitive decline using machine learning. *J Gerontol B Psychol Sci Soc Sci*. 2020;75(4):733-742.
41. Kaup AR, Nettiksimmons J, Harris TB, et al. Cognitive resilience to apolipoprotein E epsilon4: contributing factors in black and white older adults. *JAMA Neurol*. 2015;72(3):340-348.
42. Raichle ME, Macleod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. A default mode of brain function. *Proc Natl Acad Sci U S A*. 2001;98(2):676-682.
43. Greicius MD, Srivastava G, Reiss AL, Menon V. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *Proc Natl Acad Sci U S A*. 2004;101(13):4637-4642.
44. Chételat G, La Joie R, Villain N, et al. Amyloid imaging in cognitively normal individuals, at-risk populations and preclinical Alzheimer's disease. *Neuroimage Clin*. 2013;2:356-365.
45. Aizenstein HJ, Nebes RD, Saxton JA, et al. Frequent amyloid deposition without significant cognitive impairment among the elderly. *Arch Neurol*. 2008;65(11):1509-1517.
46. Nadkarni NK, Tudorascu D, Campbell E, et al. Association between amyloid-beta, small-vessel disease, and neurodegeneration biomarker positivity, and progression to mild cognitive impairment in cognitively normal individuals. *J Gerontol A Biol Sci Med Sci*. 2019;74(11):1753-1760.
47. Burnham SC, Bourgeat P, Doré V, et al. Clinical and cognitive trajectories in cognitively healthy elderly individuals with suspected non-Alzheimer's disease pathophysiology (SNAP) or Alzheimer's disease pathology: a longitudinal study. *Lancet Neurol*. 2016;15(10):1044-1053.
48. Guo T, Shaw LM, Trojanowski JQ, Jagust WJ, Landau SM. Association of CSF Abeta, amyloid PET and cognition in cognitively unimpaired elderly adults. *Neurology*. 2020:e2075-e2085.
49. Lockhart SN, Schöll M, Baker SL, et al. Amyloid and tau PET demonstrate region-specific associations in normal older people. *Neuroimage*. 2017;150:191-199.

50. Strobl C, Boulesteix A-L, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics*. 2007;8:25.
51. Bennett DA. An MRI biomarker of mixed pathology. *Neurology*. 2018;91(15):682-683.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Casanova R, Hsu F-C, Barnard RT, et al., for the Alzheimer's Disease Neuroimaging Initiative. Comparing data-driven and hypothesis-driven MRI-based predictors of cognitive impairment in individuals from the Atherosclerosis Risk in Communities (ARIC) study. *Alzheimer's Dement*. 2021;1-11. https://doi.org/10.1002/alz.12427