



OPEN

Prediction of tau accumulation in prodromal Alzheimer's disease using an ensemble machine learning approach

Jaeho Kim^{1,2,4,5}, Yuhyun Park^{2,3}, Seongbeom Park², Hyemin Jang^{2,4,5}, Hee Jin Kim^{2,4,5}, Duk L. Na^{2,4,5,6,7}, Hyejoo Lee^{2,4,5}✉ & Sang Won Seo^{2,3,4,5,7}✉

We developed machine learning (ML) algorithms to predict abnormal tau accumulation among patients with prodromal AD. We recruited 64 patients with prodromal AD using the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. Supervised ML approaches based on the random forest (RF) and a gradient boosting machine (GBM) were used. The GBM resulted in an AUC of 0.61 (95% confidence interval [CI] 0.579–0.647) with clinical data (age, sex, years of education) and a higher AUC of 0.817 (95% CI 0.804–0.830) with clinical and neuropsychological data. The highest AUC was 0.86 (95% CI 0.839–0.885) achieved with additional information such as cortical thickness in clinical data and neuropsychological results. Through the analysis of the impact order of the variables in each ML classifier, cortical thickness of the parietal lobe and occipital lobe and neuropsychological tests of memory domain were found to be more important features for each classifier. Our ML algorithms predicting tau burden may provide important information for the recruitment of participants in potential clinical trials of tau targeting therapies.

Mild cognitive impairment (MCI) refers to a transitional state between normal ageing and Alzheimer's disease (AD)^{1,2}. The rapid development of molecular imaging methods has enabled the detection of amyloid- β (A β) using positron emission tomography (PET) in the MCI stages. Previous studies have shown that 40–60% of MCI patients are A β positive (A β +) on PET, a characteristic of prodromal AD^{3–5}. Recently, these patients with prodromal AD have also been classified into fast and slow decliners according to their downstream biomarker status⁶. Especially, considering that the presence of neurofibrillary tangles (NFTs) formed by tau is a highly predictive indicator for cognitive decline⁷, it is important to develop methods to detect tau uptake in prodromal AD in vivo. In this regard, a recent AV-1451 PET study⁸, which investigated the NFT burden in the brain, reported that patients with prodromal AD exhibit in-vivo Braak stages ranging from I/II to V/VI. However, like all PET techniques, its clinical utility in medical practice has been limited because of its cost, availability, and safety, as there are risks regarding radiation exposure⁹. Regardless, the prediction of tau uptake remains an important goal, as the expectation is that future treatment strategies may target tau protein.

The exponential growth of computing power with massive data sets has led to machine learning (ML) being an alternative analytics method for clinical decision making and for searching for new relationships between disease and symptoms. Random forest (RF)¹⁰ and gradient boosting machine (GBM)¹¹ are commonly used ML methods¹² that have been outperformed consistently in many large-scale studies¹³. In addition, unlike other ML predictions with routinely used performance measures, tree-based ML provides clinically useful information, such as the relative importance of the clinical features and whether they are related positively or negatively. However, these interpretable ML methods have not been used for classifying tau burden in previous studies.

¹Department of Neurology, Dongtan Sacred Heart Hospital, Hallym University College of Medicine, Hwaseong-si, Gyeonggi-do, Republic of Korea. ²Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Republic of Korea. ³Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University School of Medicine, Suwon, Republic of Korea. ⁴Neuroscience Center, Samsung Medical Center, Seoul, Republic of Korea. ⁵Samsung Alzheimer Research Center, Samsung Medical Center, Seoul, Republic of Korea. ⁶Stem Cell and Regenerative Medicine Institute, Samsung Medical Center, Seoul, Republic of Korea. ⁷Department of Health Sciences and Technology, SAIHST, Sungkyunkwan University, Seoul, Republic of Korea. ✉email: hyejoo271@gmail.com; sangwonseo@empas.com

	Total (N = 64)	
	A + T- (N = 30)	A + T+ (N = 34)
Demographics		
Age, years	72.1 ± 7.2	70.8 ± 7.4
Sex, % female*	8 (26.7%)	20 (58.8%)
Education, years*	16.6 ± 3.1	15.3 ± 2.1
MCI stage, % late MCI	11 (36.7%)	16 (47.1%)
MMSE	26.9 ± 3.9	24.7 ± 5.0
ADAS-cog13*	19.7 ± 8.8	28.3 ± 8.4
Biomarkers		
APOE4 carriers	19 (60.3%)	25 (73.5%)
HV mm ³ *	5776.8 ± 865.0	4854.7 ± 1077.0
AV45 PET SUVR	1.33 ± 0.16	1.44 ± 0.23
Cth_Frontal mm*	3.15 ± 0.12	3.04 ± 0.15
Cth_Temporal mm*	3.37 ± 0.17	3.21 ± 0.20
Cth_Parietal mm*	3.14 ± 0.16	2.94 ± 0.18
Cth_Cingulate mm*	3.35 ± 0.17	3.24 ± 0.23
Cth_Occipital mm*	3.14 ± 0.18	2.99 ± 0.21
Cth_Global mm*	3.19 ± 0.13	3.04 ± 0.16

Table 1. Demographics and biomarkers of A + MCI participants. *MCI* mild cognitive impairment, *MMSE* mini-mental state examination, *PET* Positron emission, *ADAS-cog* The Alzheimer's disease assessment scale-cognitive, *HV* hippocampal volume, *Cth* cortical thickness, *A* amyloid, *T* tau. * $P < 0.05$ between A + T- and A + T+.

In a previous study, worse performance on the domain-specific neuropsychological tests was associated with a greater ¹⁸F-AV1451 uptake in key regions implicated in memory, visuospatial function, and language¹⁴. In the combined prodromal AD and AD dementia group, increased tau PET uptake and reduced cortical thickness were associated with worse performance on a variety of neuropsychological tests¹⁵. Altogether, these biomarkers seem to be the potential features of classifiers predicting tau burdens. In particular, different models with various combinations of biomarkers are needed because not all cohorts/centres have access to all biomarkers.

In the present study, we aimed to develop a model to predict tau burdens in the prodromal AD using multimodal biomarkers. We hypothesized that ML could provide an objective, unbiased estimator for classifying tau positivity as an alternative statistical method. We developed and validated several RF and GBM models with various combinations of variables in order to account for the various clinical environments. Variable importance and partial dependency plot (PDP) were also assessed to identify the most relevant features and their relationship to tau burden.

Results

Demographics and clinical characteristics of participants. The demographic information of the participants is summarised in Table 1. The A + T+ group had a higher percentage of female participants examined compared to the A + T- group (58.8% vs. 26.7%, $p = 0.009$). The A + T- group showed a higher number of years of education than the A + T+ group (16.6 ± 3.1 years vs. 15.3 ± 2.1 years, $p = 0.045$). There were no differences in age ($p = 0.463$) and frequency of APOE4 carriers ($p = 0.374$) between the A + T- and A + T+ groups.

The A + T+ group showed a lower hippocampal volume (4854.7 ± 1077.0 mm³ vs. 5776.8 ± 865.0 mm³, $p < 0.001$) and decreased cortical thickness in all lobes ($p = 3.2 \times 10^{-5}$ to 0.042) compared to the A + T- group. The A + T+ group also showed a higher score on The Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog) 13-item scale than the A + T- group (28.3 ± 8.4 vs. 19.7 ± 8.8, $p < 0.001$).

Model performance to classify tau positivity. Table 2 presents the performance metrics of GBM and RF for the six different models. The GBM resulted in an AUC of 0.61 (95% CI 0.579–0.647) with the baseline model 1 and a higher AUC of 0.817 (95% CI 0.804–0.830) with model 2 that included NP variables. The highest AUC was 0.86 (95% CI 0.839–0.885) achieved with model 6 (Fig. 1a). The RF had an AUC of 0.59 (95% CI 0.562–0.608) with model 1, 0.77 (95% CI 0.758–0.795) with model 2 and the highest AUC of 0.82 (95% CI 0.808–0.839) using model 6 (Fig. 1b).

The relative variable importance and PDP. The relative feature importance from each predictor of model 6 is shown in Fig. 2, indicating the highest contribution to the prediction of tau positivity. In the GBM model, cortical thickness of the parietal lobe was the most important feature followed by the neuropsychological test of memory domains, cortical thickness of the occipital lobe, and number cancellation test score. The important features identified by RF were similar to those identified in the GBM model, such as the cortical thickness of the parietal lobe, the neuropsychological test of memory domains, cortical thickness of the occipital lobe, and word recognition score. As expected, according to the PDP plot, cortical thickness and memory scores are

Model		GBM			RF		
		AUC	95% CI		AUC	95% CI	
Model 1	Baseline (age, sex, education) + MCI stage	0.613375	0.57972	0.64703	0.585162	0.562303	0.608021
Model 2	Baseline + MCI stage + NP test	0.817137	0.803992	0.830282	0.776385	0.758101	0.794668
Model 3	Baseline + MCI stage + NP test + APOE4	0.814002	0.795381	0.832623	0.818182	0.80336	0.833003
Model 4	Baseline + MCI stage + NP test + APOE4 + suvr_fdg	0.867294	0.849891	0.884696	0.801463	0.790315	0.812611
Model 5	Baseline + MCI stage + NP test + APOE4 + HV/ICV	0.825496	0.807022	0.843971	0.792059	0.771281	0.812836
Model 6	Baseline + MCI stage + NP test + APOE4 + cortical thickness	0.862069	0.839286	0.884852	0.823407	0.807749	0.839064

Table 2. ML models with different combinations of biomarkers using the total dataset. *AUC* Area under the curve, *MCI* mild cognitive impairment, *APOE4* Apolipoprotein E 4 genotype, *NP* Neuropsychological, *GBM* gradient boosting machine, *RF* random forest, *suvr_fdg* Standardized uptake value ratio of fluorodeoxyglucose-positron emission tomography, *HV/ICV* Hippocampal volume/Intracranial volume.

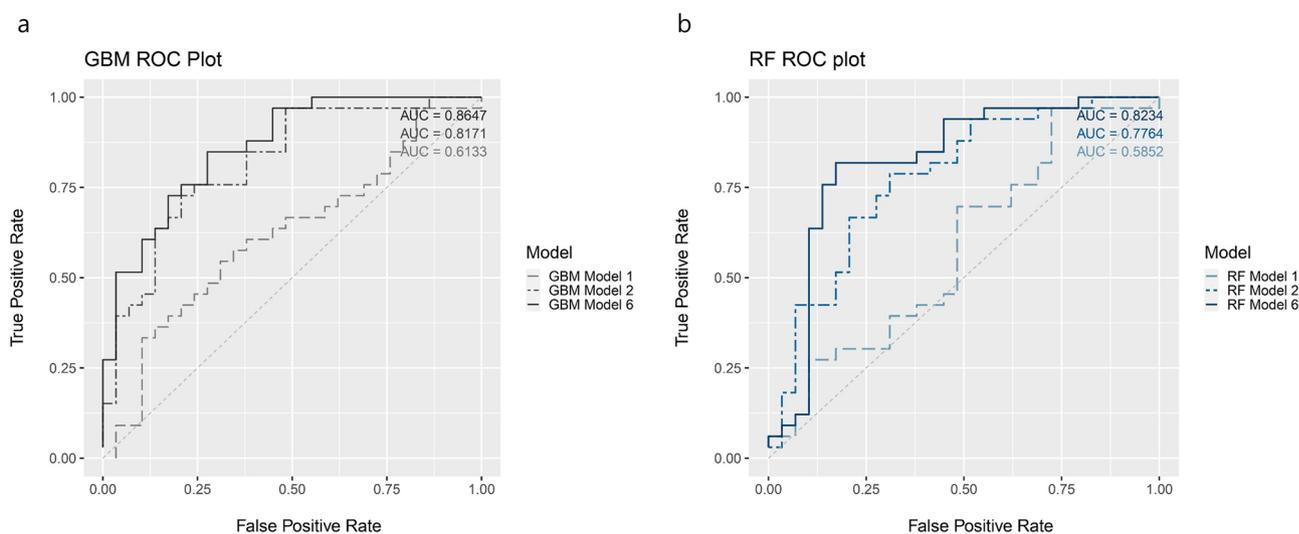


Figure 1. The ROC curve of the proposed method (a) GBM (b) RF. *ROC* receiver operating characteristic, *AUC* area under the curve, *GBM* gradient boosting machine, *RF* random forest. model 1 = Baseline (Age, Sex, Education) + MCI stage; model 2 = Baseline + MCI stage + NP test; model 6 = Baseline + MCI stage + NP test + APOE4 + cortical thickness. Each plot was made using the R software.

negatively related to the tau accumulation. Additional details regarding the influential variable ranking through model 2 to model 6 are included in Supplementary Table S1.

Discussion

In the present study, we developed and compared ML approaches for prediction of the brain tau burden in prodromal AD patients using multimodal biomarkers based on the ADNI dataset. We found that the GBM with multi-model biomarkers showed a good predictive performance. Especially, the important features in predicting the brain tau burden in prodromal AD patients involved brain structures and neuropsychological results that are responsible for memory. We also found that the GBM with baseline demographics and neuropsychological results showed a reasonable predictive performance. Furthermore, the RF had performance similar to that of the GBM. Therefore, our approaches predicting the tau burden may provide important information for the recruitment of participants in potential clinical trials of tau-targeting therapies, which is helpful to reduce failure in screening. We have developed six models for tau positivity with various combinations of input features reflecting the clinical practice. To construct a model relying on prediction performance alone may lead to underestimating the cost of acquisition and accessibility of the clinical resources. Thus, models from this study maximize the potential applicability of our models in any medical conditions and possibly provide efficient use for deploying cost-effective interventions.

We found that 53.1% of patients with prodromal AD showed a significant tau uptake, which was consistent with that seen in previous studies. Previously, Ossenkoppelle et al. set tau PET-positive (61.4%) as a Youden Index derived cut-off^{16,17} and Maass et al. set significant tau PET uptake as Braak ROI-based staging using a regression-based conditional interference tree approach⁸. Especially in the ADNI data, a significant tau uptake in prodromal AD with conditional inference method analysis was similar (57.9%) to our result.

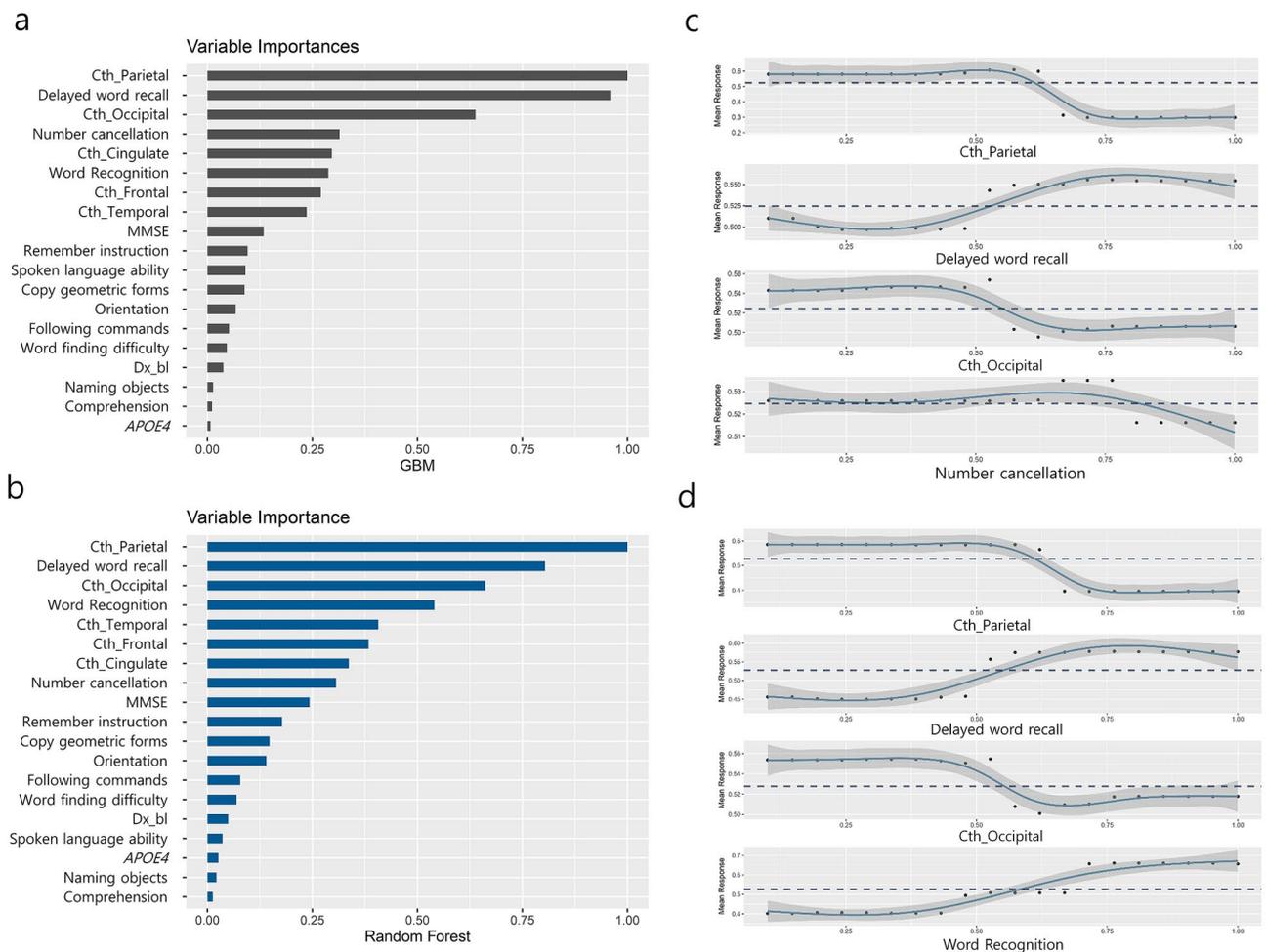


Figure 2. Variable importance plot for model 6 (a) GBM and (b) RF; and partial dependence plot for model 6 (c) GBM and (d) RF. *MMSE* mini-mental state examination, *APOE4* Apolipoprotein E 4 genotype, *Cth* cortical thickness, *GBM* gradient boosting machine, *RF* random forest. Each plot was made using the R software.

In the present study, our algorithm using demographics, neuropsychological results, *APOE4* genotype, *SUVR* of *FDG PET* and cortical thickness showed a good predictive performance for predicting tau burden in $A\beta + MCI$ populations. Previously, a study predicting *A/T/N* stages for a spectrum of individuals ranging from healthy controls to those with *MCI* and *AD* was published¹⁸. This study used structural-MRI alone and showed that the model predicted tau at 89% across the clinical diagnostic group. However, those prediction values were analysed in a healthy control to *MCI* and *AD* whereas we did it in a group of homogeneous patients, which could affect predictive performance. Furthermore, in the present study, the *GBM* with only baseline demographics and neuropsychological results showed a reasonable predictive performance.

The *GBM* and *RF* had an adequate performance predicting tau positivity. In the *GBM* method, a number of weak learners were combined to decrease bias. They generally showed a better performance with low variance data. Since our data set consists of only prodromal *AD* patients indicating a low variance, the performance of the *GBM* might be better than *RF*. In addition, interpretability, which is one of the main challenges of *ML*, was enabled by providing additional information on the model via variable importance and *PDP*. The results of this study provide evidence to consider *ML* to be a more accessible prediction tool for clinical use.

Through analysis of the impact order of the variables in each machine learning classifier, abnormalities in the cortical thickness and neuropsychological tests related to memory function were selected as important features. Our findings were consistent with a recent study showing strong relationships between increased tau pathology and reduced cortical thickness with worse performance on neuropsychological test pronounced in bilateral temporoparietal regions in prodromal *AD* and *AD* dementia¹⁵. Considering that our participants consisted of those with prodromal *AD*, our findings might be explained by the fact that memory is affected early during the course of *AD*¹⁹. Interestingly, we found that the cortical thickness in the occipital region has a strong predictive value for disease severity in prodromal *AD*. Our findings might be supported by a previous study showing that cognitive function in prodromal/early stage of *AD* is related to occipital connectivity²⁰.

We were able to conduct this study because of the availability of various clinical data through the *ADNI* because the *ADNI* is a large cohort of well-characterised subjects, and the clinical and imaging data were based on standardised protocols and analyses. However, there are a few limitations to this study. First, we set binary

limits to tau burden as only tau positivity, defined as positive when the in-vivo Braak stage was \geq III/IV, which is of particular interest since it might be considered as the transitional stage towards AD⁸. There is, however, no consensus yet on how to label tau PET scans as normal or abnormal⁸. However, the frequency of tau (+) in prodromal AD patients^{8,17} seemed to be similar to that observed in previous studies. Second, some etiologically important variables or risk factors that have previously been established in AD research were not examined. Future research should certainly take into account other variables found to be of etiological significance.

Another limitation of our study is relatively small number of samples. There has been no consensus on the measure to estimate the effective sample size for machine learning models. Additionally, acquisition for disease-specific data is still limited and relatively small in clinical practice. Therefore, our result needs to be addressed and clarified using a larger sample size in future studies. Despite these limitations, machine learning with the rigorously well-defined framework proposed here may be useful to explore the nature of heterogeneous tau pathology in the prodromal stage of AD and to examine the relationship between clinical information, neuropsychological profiles, and brain imaging. Developing a better understanding of the algorithms and integration of machine learning into clinical practice is therefore a critical step to support the development of general population prediction models in the prodromal stage of AD.

In conclusion, our ML algorithms for predicting the brain tau burden in prodromal AD showed good accuracy, it can be a useful tool to screen study populations for targeted tau therapies and predict disease severity and prognosis. Future studies are warranted to evaluate tau burden in the transitional stage and account for other significant etiological variables.

Methods

Participants. Our study population primarily consisted of subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI)-3. A full list of inclusion/exclusion criteria is described in detail at <http://adni.loni.usc.edu/methods/documents/>. All participants provided written informed consent, and all protocols were approved by each participating site's institutional review board. The authors obtained approval from the ADNI Data Sharing and Publications Committee for data use and publication. In addition, all methods were implemented in accordance with the approved guidelines. Briefly, MCI participants had a subjective memory complaint with a Clinical Dementia Rating (CDR) score of 0.5 (Petersen et al., 2010). The stage of MCI (early and late) patients was determined using the Wechsler Memory Scale (WMS) Logical Memory II; Early MCI (EMCI) subjects must have education-adjusted scores between approximately 0.5 and 1.5 SD below the mean of cognitively normal adults (on delayed recall of one paragraph from WMS Logical Memory II). All subjects gave written informed consent prior to participation⁶.

In this study, we included MCI patients who underwent 3.0 T MRI scanning, ¹⁸F-AV45 (florbetapir) PET, and AV1451 (flortaucipir) PET at baseline. As of March 2019, a total of 133 patients met this qualification, and their baseline diagnoses were EMCI (n = 76) and late MCI (LMCI, n = 57). Among these, we included in the present study patients with A β positivity on AV45 PET, which was defined as standardised uptake value ratios (SUVR) above a cut-off value of 1.11 (Landau et al., 2013; Landau et al., 2012) (37 with EMCI and 27 with LMCI) (Fig. 3).

Clinical data collection: feature space. Basic demographics and clinical data were extracted from the ADNIMERGE dataset from the ADNI database (<http://adni.loni.usc.edu/>) in March 2019. Extracted clinical data included the presence of the APOE4 genotype, hippocampal volume (HV), total intracranial volume (ICV), ¹⁸F-fluorodeoxyglucose (FDG) standardised uptake value ratio (SUVR) (Average FDG SUVR of bilateral angular, inferior temporal, and posterior cingulate regions; AD signature regions relative to the pons/vermis reference region) (Landau et al., 2011a; Landau et al., 2010), AV45 SUVR (Average AV45 SUVR of the frontal, anterior cingulate, precuneus, and parietal cortex relative to the cerebellum) and AV1451 SUVR (Average AV1451 Braak I/II, Braak III/IV and Braak V/VI). The detailed protocols for image processing have been described in previous studies (Bittner et al., 2016; Hsu et al., 2002; Landau et al., 2011b) and in the ADNI methods section at <http://adni.loni.usc.edu/>.

Definition of Tau abnormality: outcome. We defined participants as having an abnormal "T" (T+) if their in-vivo Braak stage was III/IV or greater by a conditional inference tree approach. This approach embeds decision tree-structured regression models to determine in-vivo Braak staging based on AV1451 uptake, as suggested by a previous study²¹. The regression model assigned participants with a mean Braak V/VI ROI AV1451 SUVR > 1.267 to in-vivo Braak stage V/VI. The remaining participants underwent the same procedure, using first Braak III/IV (> 1.207) and then Braak I/II (> 1.142) ROIs, leaving the remaining participants in in-vivo Braak stage 0. This conditional inference tree approach thus classified all participants into either Braak V/VI, Braak III/IV, Braak I/II or Braak stage 0 groups.

Cortical thickness measurement. In order to obtain local cortical thickness measurements for each subject, all T1 volume scans were processed by the CIVET pipeline (version 2.1.0) developed at the Montreal Neurological Institute for fully automated structural image analysis. In brief, using a linear transformation, native MRI images were registered to the MNI-152 template²². The N3 algorithm was used for correction of intensity non-uniformity caused by the inhomogeneities in the magnetic field. The next step is to perform the tissue classification into white matter (WM), grey matter (GM), cerebrospinal fluid (CSF), and background (BG) based on the T1-weighted image. The brain is split into the left and right hemispheres for the purpose of surface extraction. The surfaces of the inner and outer cortices were automatically extracted using the Constrained Laplacian-based Automated Segmentation with Proximities (CLASP) algorithm²³. The inner and outer surfaces had the same number of vertices, and there was a close correspondence between the counterpart vertices of the inner and

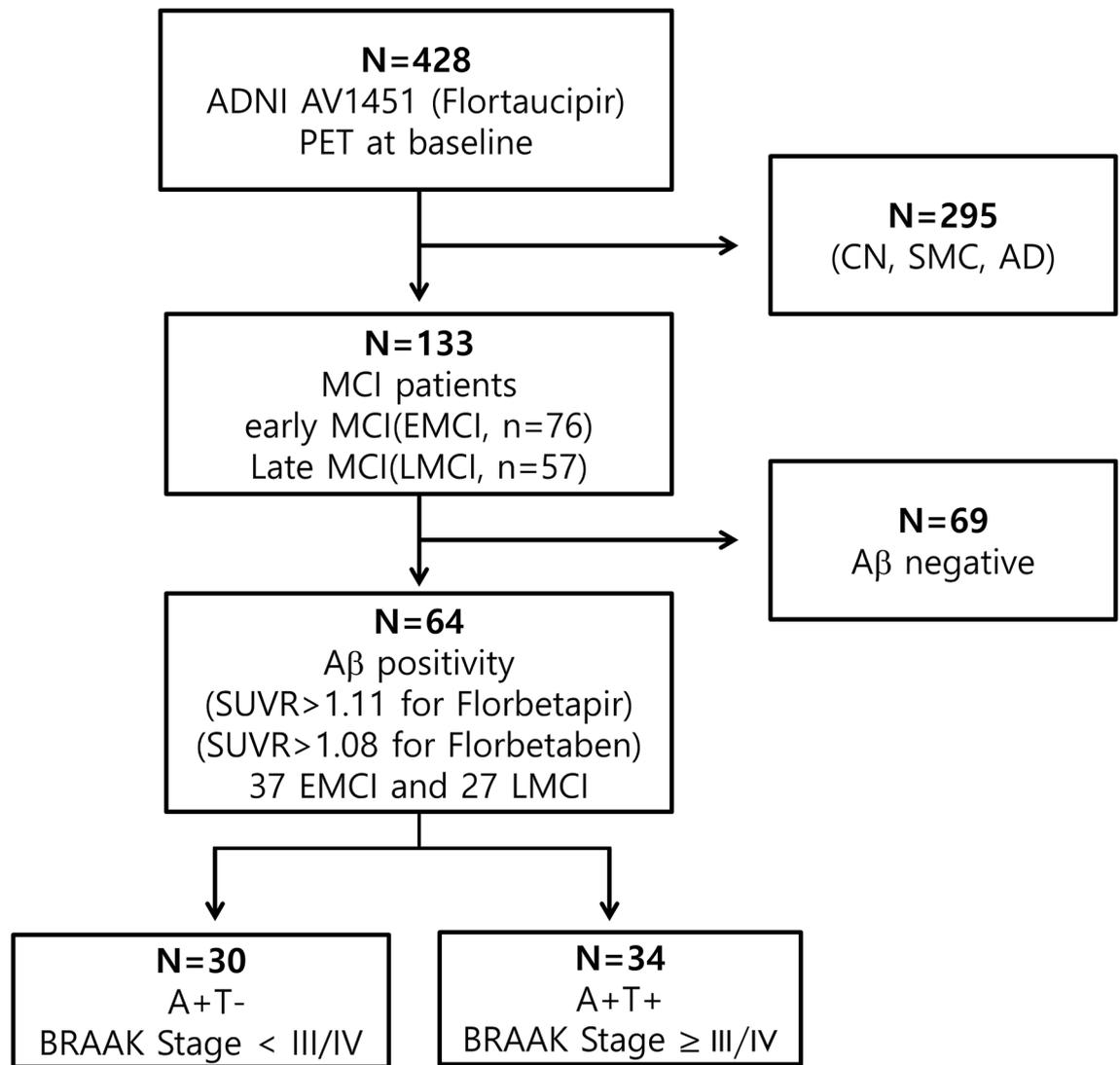


Figure 3. Flowchart showing selection of participants included in the study. *N* number, *CN* cognitively normal, *SMC* subjective memory complaints, *MCI* mild cognitive impairment, *PET* Positron emission tomography, *SUVR* Standardised uptake value ratio, *AD* Alzheimer’s dementia, *A* amyloid, *T* tau.

outer cortical surfaces²⁴. The cortical thickness was defined as the Euclidean distance between the linked vertices of the inner and outer surfaces; there were 40,962 vertices in each hemisphere in the native space^{23,25}.

Cortical thickness values were calculated in native brain spaces rather than in Talairach spaces because of the limitations of linear stereotaxic normalisation²⁶. Intracranial volume (ICV) is defined as the total volume of grey matter, white matter, and cerebrospinal fluid. We calculated ICV by measuring the total volume of the voxels within the brain mask²⁷. Brain masks were generated using the FMRIB (Functional Magnetic Resonance Imaging of the Brain) Software Library (FSL) bet algorithm. Since cortical surface models were extracted from MRI volumes transformed into stereotaxic space, cortical thickness was measured in the native space by applying an inverse transformation matrix to the cortical surfaces and reconstructing them in native space²⁵.

To measure hippocampal volume (HV), we used an automated hippocampus segmentation method using a graph cut algorithm combined with an atlas-based segmentation and morphological opening as described in an earlier study²⁸.

Machine learning algorithms. To examine changes in prediction accuracy according to the different combinations of predictors, we developed six models. We derived two tree-based ML algorithms: GBM and RF. GBM²⁹ is a tree ensemble model that generates a strong prediction model from weak learners, typically decision trees. The RF was proposed by Breiman³⁰ builds a tree ensemble predictor with multiple decision trees, in which the predictions of multiple trees are aggregated by averaging or majority voting³¹.

K-fold CV is used to divide the data set into non-overlapping *K* partitions. *K*-1 data partitions are used as a training set where a classifier is trained, and its generalization performance is tested on the one left-out validation set. This process is repeated *K* times. We selected *K* = 10 as an empirically ideal situation since accuracy is saturated when *K* = 10. Under the CV procedure, the generalization of the predictive power and validation error

was computed. The predictive performance was estimated using the area under the receiver operating characteristic (ROC) curve (AUC) and their 95% confidence interval.

Interpretable ML: variable importance and partial dependence plot (PDP). For each optimized model examined the variable importance criterion, which measures the relative prediction power (prediction strength) by using mean decreased accuracy (MDA) or Gini index¹⁰. For each analysis, variable importance was estimated to find which independent variables were influential features for an accurate classification³². Influential variables were ranked by calculating relative importance values. In the tree-based model such as GBM and RF, when the variables split the tree, the relative importance value of that variable was estimated by the discrepancy of the squared error loss over all trees. A higher relative importance value indicated a greater influence of the variables for classifying tau positivity.

We conducted a PDP proposed by J.H. Friedman, which can provide information on whether the feature is positively or negatively correlated to the final prediction. In order to avoid over-weighted or underweighted results, a Min–Max normalisation³³ was conducted. PDP is a graphical representation tool, which can provide information on whether the feature is positively or negatively related to the final prediction, it is shown as follows.

Let \mathbf{x}_s be the space of input variables consisting of a chosen subset space and \mathbf{x}_c be the complementary space, $\mathbf{x}_s \cup \mathbf{x}_c = \mathbf{x}$. Then the functional form of approximation $\hat{f}(\mathbf{x})$ depends on both subset space

$$\hat{f}(\mathbf{x}) = \hat{f}(\mathbf{x}_s, \mathbf{x}_c), \hat{f}_c(\mathbf{x}_s) = \hat{f}(\mathbf{x}_s | \mathbf{x}_c)$$

If the dependency of the complementary space is not too strong, the average function

$$\bar{f}_s(\mathbf{x}_s) = E_{\mathbf{x}_c}[\hat{f}(\mathbf{x})] = \int \hat{f}(\mathbf{x}_s, \mathbf{x}_c) p_c(\mathbf{x}_c) d\mathbf{x}_c$$

where $p_c(\mathbf{x}_c)$ is a marginal probability density function of \mathbf{x}_c .

An alternative functional form of approximation $\hat{f}(\mathbf{x})$ becomes

$$\tilde{f}_s(\mathbf{x}_s) = E_{\mathbf{x}}[\hat{f}(\mathbf{x}) | \mathbf{x}_s] = \int \hat{f}(\mathbf{x}) p_z(\mathbf{x}_c | \mathbf{x}_s) d\mathbf{x}_c$$

Statistical analysis. For the comparison of demographic and clinical data, a two-sample t-test was used for continuous variables, and a chi-square test was used for categorical variables. All analyses were performed with R package³⁴, version 3.6.1 (R Project for Statistical Computing).

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 20 November 2020; Accepted: 17 February 2021

Published online: 11 March 2021

References

- Petersen, R. C. *et al.* Current concepts in mild cognitive impairment. *Arch. Neurol.* **58**, 1985–1992. <https://doi.org/10.1001/archneur.58.12.1985> (2001).
- Morris, J. C. *et al.* Mild cognitive impairment represents early-stage Alzheimer disease. *Arch. Neurol.* **58**, 397–405. <https://doi.org/10.1001/archneur.58.3.397> (2001).
- Okello, A. *et al.* Conversion of amyloid positive and negative MCI to AD over 3 years: An 11C-PIB PET study. *Neurology* **73**, 754–760. <https://doi.org/10.1212/WNL.0b013e3181b23564> (2009).
- Wolk, D. A. *et al.* Amyloid imaging in mild cognitive impairment subtypes. *Ann. Neurol.* **65**, 557–568. <https://doi.org/10.1002/ana.21598> (2009).
- Doraiswamy, P. M. *et al.* Flortetapir F 18 amyloid PET and 36-month cognitive decline: A prospective multicenter study. *Mol. Psychiatry* **19**, 1044–1051. <https://doi.org/10.1038/mp.2014.9> (2014).
- Jang, H. *et al.* Prediction of fast decline in amyloid positive mild cognitive impairment patients using multimodal biomarkers. *NeuroImage. Clin.* **24**, 101941. <https://doi.org/10.1016/j.nicl.2019.101941> (2019).
- Sebastian-Serrano, A., de Diego-Garcia, L. & Diaz-Hernandez, M. The neurotoxic role of extracellular tau protein. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms19040998> (2018).
- Maass, A. *et al.* Comparison of multiple tau-PET measures as biomarkers in aging and Alzheimer's disease. *Neuroimage* **157**, 448–463. <https://doi.org/10.1016/j.neuroimage.2017.05.058> (2017).
- Teipel, S. *et al.* Multimodal imaging in Alzheimer's disease: Validity and usefulness for early detection. *Lancet. Neurol.* **14**, 1037–1053. [https://doi.org/10.1016/s1474-4422\(15\)00093-9](https://doi.org/10.1016/s1474-4422(15)00093-9) (2015).
- Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
- Friedman, J. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* <https://doi.org/10.1214/aos/1013203451> (2000).
- Fernández-Delgado, M., Cernadas, E., Barro, S. & Amorim, D. Do we need hundreds of classifiers to solve real world classification problems?. *J. Mach. Learn. Res.* **15**, 3133–3181 (2014).
- Caruana, R. & Niculescu-Mizil, A. *Proceedings of the 23rd International Conference on Machine Learning* 161–168 (Association for Computing Machinery, 2006).
- Ossenkoppele, R. *et al.* Tau PET patterns mirror clinical and neuroanatomical variability in Alzheimer's disease. *Brain* **139**, 1551–1567. <https://doi.org/10.1093/brain/aww027> (2016).
- Ossenkoppele, R. *et al.* Associations between tau, Abeta, and cortical thickness with cognition in Alzheimer disease. *Neurology* **92**, e601–e612. <https://doi.org/10.1212/wnl.0000000000006875> (2019).

16. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3**, 32–35. [https://doi.org/10.1002/1097-0142\(1950\)3:1%3c32::aid-cnrcr2820030106%3e3.0.co;2-3](https://doi.org/10.1002/1097-0142(1950)3:1%3c32::aid-cnrcr2820030106%3e3.0.co;2-3) (1950).
17. Ossenkoppele, R. *et al.* Discriminative accuracy of [18F]flortaucipir positron emission tomography for Alzheimer disease vs other neurodegenerative disorders. *JAMA* **320**, 1151–1162. <https://doi.org/10.1001/jama.2018.12917%JJAMA> (2018).
18. Lang, A., Weiner, M. W. & Tosun, D. What can structural MRI tell about A/T/N staging?. *Alzheimer Dement.* **15**, P1237–P1238. <https://doi.org/10.1016/j.jalz.2019.06.4758> (2019).
19. Jahn, H. Memory loss in Alzheimer's disease. *Dialog. Clin. Neurosci.* **15**, 445–454 (2013).
20. De Marco, M., Duzzi, D., Meneghello, F. & Venneri, A. Cognitive efficiency in Alzheimer's disease is associated with increased occipital connectivity. *J. Alzheimer's Dis. JAD* **57**, 541–556. <https://doi.org/10.3233/jad-161164> (2017).
21. Scholl, M. *et al.* PET imaging of tau deposition in the aging human brain. *Neuron* **89**, 971–982. <https://doi.org/10.1016/j.neuron.2016.01.028> (2016).
22. Collins, D. L., Neelin, P., Peters, T. M. & Evans, A. C. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr.* **18**, 192–205 (1994).
23. Kim, J. S. *et al.* Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. *Neuroimage* **27**, 210–221. <https://doi.org/10.1016/j.neuroimage.2005.03.036> (2005).
24. Im, K. *et al.* Brain size and cortical structure in the adult human brain. *Cereb. Cortex* **18**, 2181–2191. <https://doi.org/10.1093/cercor/bhm244> (2008).
25. Im, K. *et al.* Gender difference analysis of cortical thickness in healthy young adults with surface-based methods. *Neuroimage* **31**, 31–38. <https://doi.org/10.1016/j.neuroimage.2005.11.042> (2006).
26. Sung, H.K. *et al.* The cortical neuroanatomy related to specific neuropsychological deficits in alzheimer's continuum. *Dement Neurocogn Disord.* **18**(3), 77–95 <https://doi.org/10.12779/dnd.2019.18.3.77> (2019).
27. Smith, S. M. Fast robust automated brain extraction. *Hum. Brain Mapp.* **17**, 143–155. <https://doi.org/10.1002/hbm.10062> (2002).
28. Kwak, K. *et al.* Fully-automated approach to hippocampus segmentation using a graph-cuts algorithm combined with atlas-based segmentation and morphological opening. *Magn. Reson. Imaging* **31**, 1190–1196. <https://doi.org/10.1016/j.mri.2013.04.008> (2013).
29. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* **2**, 1189–1232 (2001).
30. Breiman, L. *J. Random For.* **45**, 5–32 (2001).
31. Breiman, L. Bagging predictors. *Mach. Learn.* **24**, 123–140 (1996).
32. Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van Der Linde, A. Bayesian measures of model complexity and fit. *J. R. Stat.* **64**, 583–639. <https://doi.org/10.1111/1467-9868.00353> (2002).
33. Han, J., Pei, J. & Kamber, M. *Data Mining: Concepts and Techniques* (Elsevier, 2011).
34. Team, R. R: A language and environment for statistical computing. (2019).

Acknowledgements

This research was supported by a Grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare and Ministry of science and ICT, Republic of Korea (grant number: HU20C0111), funded by the Ministry of Health & Welfare, Republic of Korea (Grant number: HI19C1082); by a Grant from the Korean Health Technology R&D Project, Ministry of Health & Welfare, Republic of Korea (HI19C1132); by a fund (2018-ER6203-02) by the Research of Korea Centers for Disease Control and Prevention; by the Brain Research Program of the National Research Foundation (NRF) funded by the Ministry of Science & ICT (NRF-2018M3C7A1056512).

Author contributions

J.K.: study concept and design, acquisition, analysis and interpretation of data. Y.P.: analysis of data. S.P.: analysis of data. H.J.: acquisition of data, critical revision of manuscript. H.J.K.: critical revision of manuscript. D.L.N.: critical revision of manuscript. H.L.: study concept and design, acquisition, analysis and interpretation of data, critical revision of manuscript for intellectual content. S.W.S.: study concept and design, acquisition, analysis and interpretation of data, critical revision of manuscript for intellectual content.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-85165-x>.

Correspondence and requests for materials should be addressed to H.L. or S.W.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021