Journal of Alzheimer's Disease 79 (2021) 1533–1546
DOI 10.3233/JAD-200906
IOS Press

Quantitative Longitudinal Predictions of Alzheimer's Disease by Multi-Modal Predictive Learning

Mithilesh Prakash[a,*], Mahmoud Abdelaziz[b], Linda Zhang[c], Bryan A. Strange[c,d] and Jussi Tohka[a] for the Alzheimer's Disease Neuroimaging Initiative[1]
[a]*University of Eastern Finland, A.I. Virtanen Institute for Molecular Sciences, Kuopio, Finland*
[b]*Zewail City of Science and Technology, Giza, Egypt*
[c]*Department of Neuroimaging, Alzheimer's Disease Research Centre, Reina Sofia-CIEN Foundation, Madrid, Spain*
[d]*Laboratory for Clinical Neuroscience, CTB, Universidad Politécnica de Madrid, Madrid, Spain*

Handling Associate Editor: Malek Adjouadi

Accepted 3 December 2020
Pre-press 12 January 2021

Abstract.
Background: Quantitatively predicting the progression of Alzheimer's disease (AD) in an individual on a continuous scale, such as the Alzheimer's Disease Assessment Scale-cognitive (ADAS-cog) scores, is informative for a personalized approach as opposed to qualitatively classifying the individual into a broad disease category.
Objective: To evaluate the hypothesis that the multi-modal data and predictive learning models can be employed for future predicting ADAS-cog scores.
Methods: Unimodal and multi-modal regression models were trained on baseline data comprised of demographics, neuroimaging, and cerebrospinal fluid based markers, and genetic factors to predict future ADAS-cog scores for 12, 24, and 36 months. We subjected the prediction models to repeated cross-validation and assessed the resulting mean absolute error (MAE) and cross-validated correlation (ρ) of the model.
Results: Prediction models trained on multi-modal data outperformed the models trained on single modal data in predicting future ADAS-cog scores ($MAE_{12,\ 24\ \&\ 36\ months} = 4.1$, 4.5, and 5.0, $\rho_{12,\ 24\ \&\ 36\ months} = 0.88$, 0.82, and 0.75). Including baseline ADAS-cog scores to prediction models improved predictive performance ($MAE_{12,\ 24\ \&\ 36\ months} = 3.5$, 3.7, and 4.6, $\rho_{12,\ 24\ \&\ 36\ months} = 0.89$, 0.87, and 0.80).
Conclusion: Future ADAS-cog scores were predicted which could aid clinicians in identifying those at greater risk of decline and apply interventions at an earlier disease stage and inform likely future disease progression in individuals enrolled in AD clinical trials.

Keywords: Alzheimer's disease, machine learning, magnetic resonance imaging, multi-modal imaging, neuropsychology

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

*Correspondence to: Mithilesh Prakash, PhD, A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, P.O.B. 1627, FI-70211 Kuopio, Finland. E-mail: mithilesh.prakash@uef.fi.

INTRODUCTION

Alzheimer's disease (AD) is an irreversible and multi-factorial neurodegenerative disease with a progressive decline in cognitive abilities [1]. AD affects several tens of millions of people globally and is widely studied [2]. Yet, the pathogenesis of AD remains unclear [3]. Cognitive tests, brain volumetry from magnetic resonance imaging (MRI), amyloid load, and glucose consumption levels from positron emission tomography (PET), and protein analysis of cerebrospinal fluid (CSF) provide valuable and complementary disease markers to chart the disease progression [4, 5]. Qualitative and manual analysis of these markers to diagnose patients could be potentially aided by automated algorithms [6].

The classification based on clinical diagnosis places an individual into normal, mild cognitive impairment (MCI), or AD groups [7, 8]. Memory loss (either self-reported or by an associate) is observed during the initial stages of AD [9]. Declining cognitive skills is also common and can potentially lead to dementia [10]. Hence, the disease progression must be carefully monitored at the earliest stages [11]. AD risk factors include sociodemographic factors (e.g., increasing age and fewer years of education), genetic (*APOE* alleles count), and patient medical and family history [12]. A clinical diagnosis of AD is currently a challenge due to a lack of clear diagnostic markers of AD and overlapping clinical features with other dementia types [13, 14]. At postmortem, AD is characterized by the presence of amyloid-β peptide plaques and accumulations of tau proteins in the brain histology samples [15].

The progressive nature of AD makes diagnosing an individual into any of the discrete groups a challenging proposition [16, 17]. Conventional diagnostics analyze changes in MRI, CSF, and cognitive biomarkers [18, 19], but this could be inefficient as the changes can be slow and difficult to detect [20, 21]. The change in these biomarkers is nonlinear with the clinical progression of AD, further complicating longitudinal tracking. Therefore, quantifying and tracking the condition of the patient by continuous measures such as ADAS-cog scores has been advocated [22–27]. ADAS-cog is widely used clinically (to measure language, memory, praxis, and other cognitive abilities) and provides an accurate description of the cognitive state on a continuous scale, making it an ideal choice for predicting the disease progression [28, 29]. The availability of standardized multi-modal data and corresponding longitudinal ADAS-cog scores from research organizations, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) project, has enabled the development of novel techniques for tracking AD progression by employing machine learning [30]. However, predicting ADAS-cog scores has been reported as very difficult [31]. In the recent Alzheimer's Disease Prediction of Longitudinal Evolution (TADPOLE) Challenge (https://tadpole.grand-challenge.org/), forecasts of clinical diagnosis and ventricle volume were very good, whereas, for ADAS-cog, no team participating in the challenge was able to generate forecasts that were significantly better than simple baselines.

Previous studies on predictive modeling of ADAS-Cog 13 scores include Yang et al. who developed a model that replicated the observed progression of ADAS-Cog 13 scores and used this as a more precise estimate of the pathologic stage [22]. Bhagwat et al. utilized ADAS-cog scores to classify subjects into sub-classes [26] and Moore et al. predicted future ADAS-Cog scores based on longitudinal data (at least 4 periods of data were required) [24]. Lei et al. utilized MRI data to predict ADAS-cog scores [32], while Utsumi et al. predicted ADAS-cog scores with multimodal data for up to 24 months from baseline on a limited number of subjects ($n = 100$) [25]. Instead, we will utilize multi-modal data (various imaging modalities, CSF biomarkers, cognitive tests) at the baseline to predict ADAS-Cog scores on a continuous scale for up to 36 months from baseline and study the relevance of each modality for prediction in a standardized setting.

We will test several multivariate regression techniques, such as partial least squares regression (PLSR) [33, 34], support vector regression (SVR) [35, 36], and random forest regression (RF) [24, 37], that enable modeling complex relationships between baseline multi-modal ADNI data (predictors) and future ADAS-cog 13 scores. The multivariate nature of the modeling is desirable for the ADAS-cog score trajectory analysis due to the complementary nature of the AD measures. The resulting trajectory predictions could alert clinicians to prescribe appropriately (once disease modifying interventions are available). Moreover, knowing the likely future trajectory of the disease will provide a benchmark with which to test clinical evolution in patients enrolled in clinical trials.

We hypothesized that the multivariate regression techniques are well suited for multi-factorial diseases and that the progression of AD, as indicated by ADAS-cog scores in subsequent timelines, can be accurately predicted. Furthermore, the inclusion of

baseline ADAS-cog scores could improve the predictions of the model in subsequent follow-ups.

METHODS

ADNI dataset

Data in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu/). In addition to the various summary tables directly provided by ADNI, we used summary tables prepared for the TADPOLE grand challenge based on ADNI data at https://tadpole.grand-challenge) [31, 38]. The data are from the TADPOLE tables if not otherwise stated. (Specific variable names are provided as Supplementary Table 1). The ADNI project started in 2003 as a public-private partnership, led by PI Michael W. Weiner, MD. The main objective of ADNI is to evaluate the application of serial MRI, PET, other biological markers, and clinical and neuropsychological assessment in a multi-modal approach to determine the longitudinal progression of MCI and early AD.

We utilized pre-processed ADNI data because of the standardized processing pipeline that ensured the quality of the data. This multi-modal data is readily available for other researchers enabling a direct comparison of the study results. For making the predictions, we utilized only the data from the baseline visit from the following modalities: MRI, FDG-PET, AV45-PET, neuropsychological and behavioral assessments (NePB), CSF biomarkers as well as dementia risk factors. From each modality, we selected a few variables that have been found important in the previous literature. The predictor variables and the reasons for their selection are explained in Table 1.

Subjects

The characteristics of subjects recruited in the ADNI dataset are described in detail here http://adni.loni.usc.edu/. The trends of the ADAS-cog 13 scores utilized in this study are provided in Supplementary Figure 1 and the details of subject characteristics are provided in Table 2. There are fewer subjects in follow-up visits than in the baseline visit due to subject attrition and missing data. Our purpose was to create a simple and standardized testing environment to truly assess the importance of predictor variables from different modalities. Therefore, we only retained subjects with complete data (see Table 2 for a summary of the data utilized) and no data imputation was necessary. Note that some subjects change diagnostic status over the follow-up period. The roster identification (RID) numbers of the included subjects are provided as comma-separated values in the Supplementary Material.

Multivariate regression analysis

We employed multivariate regression to predict ADAS-cog scores based on predictor variables detailed above. We considered four different prediction tasks: predicting ADAS-cog score at baseline and at 12, 24, or 36 months after the baseline. In all of these tasks, the predictor variables are from the baseline visit. The group of features (predictors) used for regression are denoted by the column vectors X_i, $(i = 0, 1, \ldots, L)$, where L is the number of features (Fig. 1). The ADAS-cog scores (dependent variable or response variable) are denoted by the column vector Y.

We employed widely used machine learning techniques including PLSR [39], SVR [40], and RF and created prediction models [41]. Additionally, a genetic algorithm (GA) was utilized to rank the variables in the order of importance in the multi-modal case [42]. A summary of these methods is provided in Table 3.

Regression modeling

The prediction of the ADAS scores (at baseline, 12 months, 24 months, and 36 months) was performed by employing PLSR, SVR, and RF. Both single modal (each modality of Table 1 alone) and multi-modal predictors (all modalities combined, Table 1) were considered. All the predictors were from the baseline visit. Under single modalities Age, years of education (Edu), number of *APOE* ε4 alleles (*APOE*) were exactly 1 variable each, CSF had 3 variables, AVF45-PET had 4 variables, NePB and FDG had 5 variables each, MRI had 9 variables and hence the multi-modal model had a total of 29 variables (Supplementary Table 1). All variables were assumed to be continuous and we standardized the variables to be zero-mean and unit standard deviation.

Performance metrics and evaluation

We evaluated the prediction models using 5-fold repeated cross-validation with 10 repeats, see Figs. 1 and 2. The models were evaluated in terms of cross-validated correlation coefficient (ρ) and the mean absolute error (MAE) between the actual ADAS-cog 13 scores and its model-predicted values. In addition to the average MAE and correlation coefficient, we

Table 1
Modalities summary

Modality	Short description of the features used in the regression models.	Feature count
MRI	We used 9 features: intracranial volume (ICV), and volumes of the hippocampus, entorhinal cortex, and lateral ventricles as well as the latter four divided by the ICV. These features were selected based on previous studies [48]. We included volumes divided by the ICV as it is unclear whether raw or ICV-corrected volumes are better predictors of dementia [48, 49]	9
AV-45 PET	We used standardized uptake values (SUVs) in four regions: frontal cortex, cingulate, lateral parietal cortex, and lateral temporal cortex. The AV-45 PET measures amyloid-beta load in the brain. We used regional SUV ratios processed according to the UC Berkeley protocol [50–52]. Each AV-45 PET scan was co-registered to the corresponding MRI and the mean AV-45 uptake within the regions of interest and reference regions was calculated. Regions of interest were composites of frontal regions, anterior/posterior cingulate regions, lateral parietal regions, and lateral temporal regions [53]	4
FDG-PET	We used average SUVs in five brain regions: bilateral angular gyri, bilateral posterior cingulate gyri, and bilateral inferior temporal gyri. The FDG PET data measures glucose consumption and is shown to be strongly related to dementia and cognitive impairment when compared to normal control subjects [52, 54, 55].	5
CSF proteins	The baseline CSF $A\beta_{42}$, t-tau, and p-tau were used as CSF features [56].	3
Neuropsychology and behavioral (NePB) assessments	We selected to include 5 NePB scores as NePB features: the summary score from Mini-Mental State Examination (MMSE) [57], three summary scores of Rey's auditory verbal learning test (RAVLT; learning, immediate, and percent forgetting) [58], and a summary score from the functional activities Questionnaire (FAQ) [59].	5
Risk factors: age, education, and *APOE*	We considered age, the number of *APOE* $\varepsilon 4$ alleles, and the years of education. With aging, normal cognitive decline is an accepted phenomenon, but lower education and lower cerebral metabolic activity could accelerate the normal decline [60]. The *APOE* $\varepsilon 4$ allele, present in approximately 10–15% of people, increases the risk for late-onset AD and lowers the age of onset. One copy of $\varepsilon 4$ ($\varepsilon 3/\varepsilon 4$) can increase risk by 2-3 times while homozygotes ($\varepsilon 4/\varepsilon 4$) can be at 12 times increased risk [61]. We coded *APOE* $\varepsilon 4$ status of absence, single copy, or homozygous as 0, 1, and 2, respectively.	3
ADAS-cog scores	The ADAS-cog 13 task scale was one such improvement on the original ADAS-cog 11, with additional memory and attention/executive function tasks [62]. The final 13 tasks test verbal memory (3 tasks), clinician-rated perception (4 tasks), and general cognition (6 tasks). It was found to perform better than the ADAS-cog 11 at discriminating between MCI and mild AD patients, as well as have better sensitivity to treatment effects in MCI [63]. As the ADAS-cog 13 fully encompasses the ADAS-cog 11 tasks, it is also backward compatible. As such, we used the ADAS-cog 13 scale for our study as a continuous quantitative measure of a subject's disease status. The scores at baseline, 12-month, 24-month, and 36-month timelines were obtained from the ADNI dataset (Table S.2). The value (0 to 85) of these scores is lowest for the normal control group and increases with disease progression and the scores are highest for AD subjects.	1

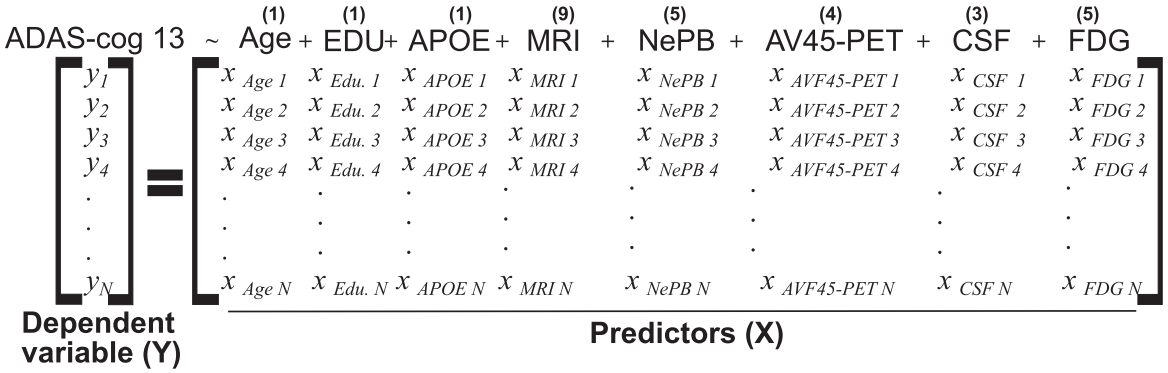

Fig. 1. Regression modeling structure. Single modality uses one predictor at a time while multi-modal uses all the predictors as indicated above. The sample size for baseline (N = 757), 12 months (N = 629), 24 months (N = 563), and 36 months (N = 314) were different due to missing values (cohort attrition). The predictors consist of age at baseline, years of formal education (Edu.), *APOE* ε4 status (absence, single copy or homozygous coded as 0, 1, and 2 respectively), MRI-derived parameters, neuropsychiatric and behavioral assessment (NePB), AV45-PET measurements, CSF biomarkers (amyloid-β, tau, p-tau) and FDG-PET measures. The number of features is indicated above each modality abbreviations. All the variables were considered as continuous and standardized to be zero-mean and unit standard deviation.

Table 2
Summary of the subject demographics and ADAS-cog 13 scores

	Baseline			12 months later			24 months later			36 months later		
	NC	MCI	AD	NC	MCI	AD	NC	MCI	AD	NC	MCI	AD
No. of Subjects	241	405	111	169	337	117	195	271	89	40	209	61
M/F	112/129	226/179	68/43	89/80	181/156	76/41	103/92	147/124	53/36	22/18	101/108	39/22
Age	56–89	55–91	55–90	55–89	56–92	55–90	56–89	55–91	55–90	55–84	55–87	55–56
ADAS-cog 13 scores	0–24	3–38	14–51	1–21	1–73	9–62	0–25	2–45	8–67	0–18	0–36	8–74

computed bootstrapped 95% confidence intervals (CIs) for them utilizing a specific procedure for bootstrapping repeated cross-validation results developed by Lewis et al. [43]. These CIs approximate the generalization performance of the prediction measured by cross-validated correlations and MAEs. Likewise, we performed hypothesis testing between the prediction models using a permutation test adapted to repeated cross-validation. This test was also developed by Lewis et al. [43]. The technical description of the hypothesis test and the code are provided at https://github.com/jussitohka/repeated_CV_permutation_test.

Implementation

The analyses were performed on MATLAB 2019b (The Mathworks Inc, Natick, MA) using native machine learning functions. The PLSR was executed with *plsregress* function and the optimal number of PLS components was manually selected based on the least root mean square error for training data [33]. SVR was executed with *fitrsvm* and RF with *fitrensemble* and in both methods the models were tuned by setting *OptimizeHyperparameters* argument as *auto* [37, 44].

A genetic algorithm for variable importance

The selection of the best subset of variables (termed variable selection or feature selection) is a challenging task in machine learning. One of the various approaches for variable selection is by GA [42], which additionally provides us a ranking of the importance of individual variables for ADAS prediction. In brief, GA aims to select a subset of variables that best predict the response variable, ADAS-cog score in our case. Due to randomness inherent to GA, a different set of variables is selected during each run of the GA. By running GA r (here $r = 100$) times, we assume that the frequency of retention of a variable among r GA runs corresponds to the importance of that variable in predicting the ADAS score. A GA is a general optimization technique inspired by natural selection. GA-based variable selection encodes each possible solution to the problem as a binary vector with the same number of components as there are variables, with the value 1 in the ith component meaning that the ith variable is in the model. GA initializes with a random population of potential solutions (called chromosomes) and then evolves this population through mutation, selection, and cross-over operators guided by the fitness

Table 3
Regression methods summary and hyperparameters information

Regression Method	Short description	Hyperparameters	Optimization employed	Matlab function
Partial least squares regression (PLSR)	The difference between PLSR and ordinary least squares (OLS) lies in the fact that PLSR is essentially a dimensionality reduction method in which the initial set of features X_i, ($i = 0, 1, \ldots$ L) are mapped to a new set of features Z_k, ($k = 0, 1, \ldots$ M), where $M < L$ and k are optimized with the training data. This property is a very useful one especially when the number of available data samples is not necessarily large enough compared to the number of features, which is the case in our problem. The new set of features are obtained in a supervised manner in which the response Y is used together with the old features X_i to identify the new features Z_k [39]	PLS components	Manual selection-based model with least RMSE	*plsregress*
Support vector regression (SVR)	Instead of minimizing the sum of squared errors between the model output and the actual response as in OLS regression, SVR aims at minimizing a different type of cost function, where only residuals larger in absolute value than some margin of tolerance contribute to the cost function [36]. Towards this end, a regularization parameter is also defined to indicate the amount of penalty applied to data samples lying beyond the margin of tolerance. This regularization parameter represents the first SVR hyperparameter. Additionally, SVR provides the option to map the input data points using linear or nonlinear kernels (or basis functions) that can provide better prediction performance, especially when there is some nonlinear dependency between the ADAS scores and the set of features.	Kernel type, regularization parameter (C), and degree	Autoconfiguration option in Matlab	*fitrsvm*
Random forest regression (RF)	Random forest methods [37], unlike PLSR or SVR, utilize a combination of results of many models/trees (ensemble of models). Trees are constructed via classification and regression tree (CART) and the data is sampled via bagging and the average of the outputs used for classification or regression modeling. A comprehensive review of RF on neuroimaging data is provided by Sarica et al. [64].	Number of trees in the model, levels of trees, number of splits in subsequent levels, number of leaves in each split.	Autoconfiguration option in Matlab	*fitrensemble*

function, which, in this case, is a cross-validated mean squared error between the true ADAS-cog score and predicted ADAS-cog score.

We employed the GA-PLS toolbox by Leardi et al. to create PLS models [45]. The *gapls* function from the toolbox was used and the frequency of retention of variables in the models was recorded with default settings [45]. The GA internally employs cross-validation and root mean square error (RMSECV) on PLS models to detect the change in this error and decide to either retain, delete, and/or carry over a subset of variables to the next iteration.

RESULTS

As depicted in Fig. 2, we created single modal and multi-modal regression models and compared their performance. The comparison (Fig. 3) shows that multi-modal based prediction models outperform single modality-based prediction models consistently in

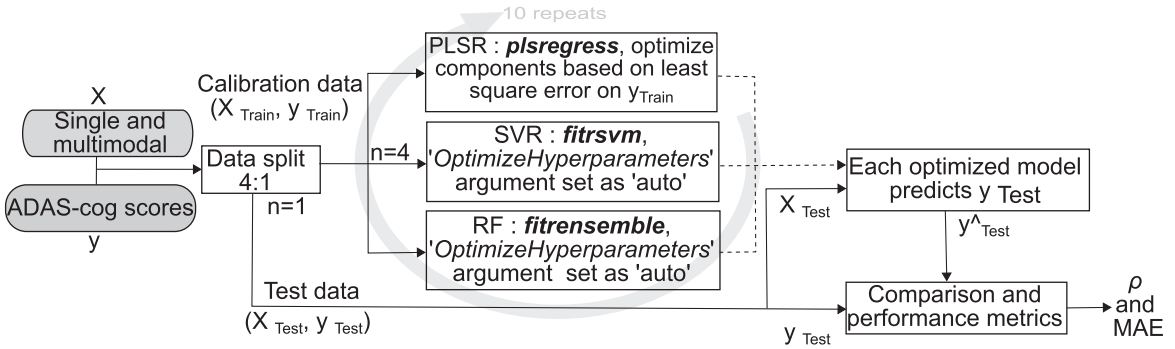

Fig. 2. Schematic of regression modeling. X is single or multi-modal predictors and Y is the target value to be predicted. We utilized 5-fold cross-validation repeated 10 times to account for the random assignment of subjects to different folds. Partial least squares regression (PLSR), support-vector regression (SVR), and random forest regression (RF) models were trained and tuned based on training folds and evaluated on test folds. The utilized Matlab function and hyperparameter tuning are shown in italics. Cross-validated correlation (ρ) and mean absolute error (MAE) metrics were employed and average performance for 10 runs computed.

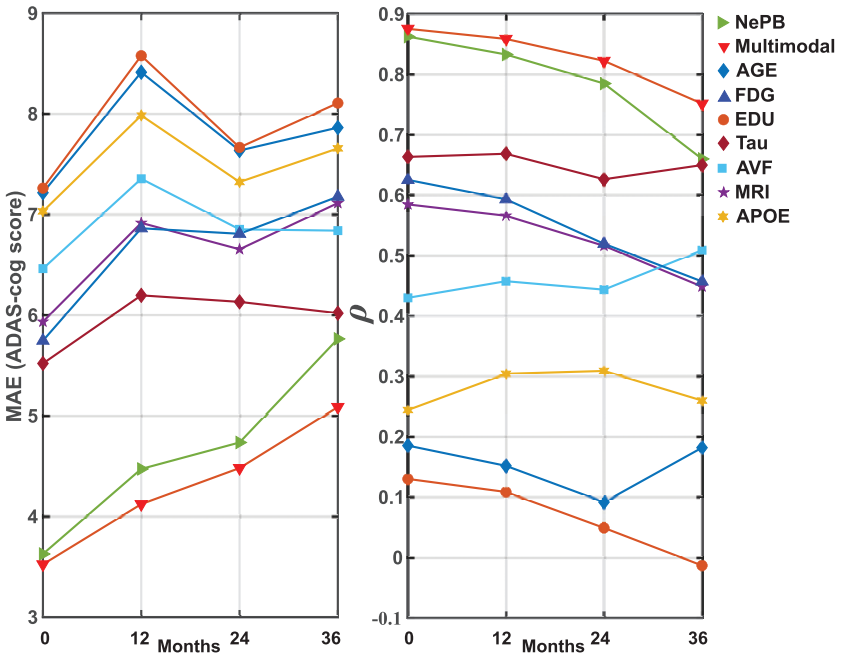

Fig. 3. Comparison of single and multi-modal dataset performance—collapsing across diagnostic status—with partial least squares regression. The performance measures cross-validation correlation (ρ) and mean absolute error (MAE) for ADAS-cog scores are plotted for predictions at 0, 12, 24, and 36 months in advance.

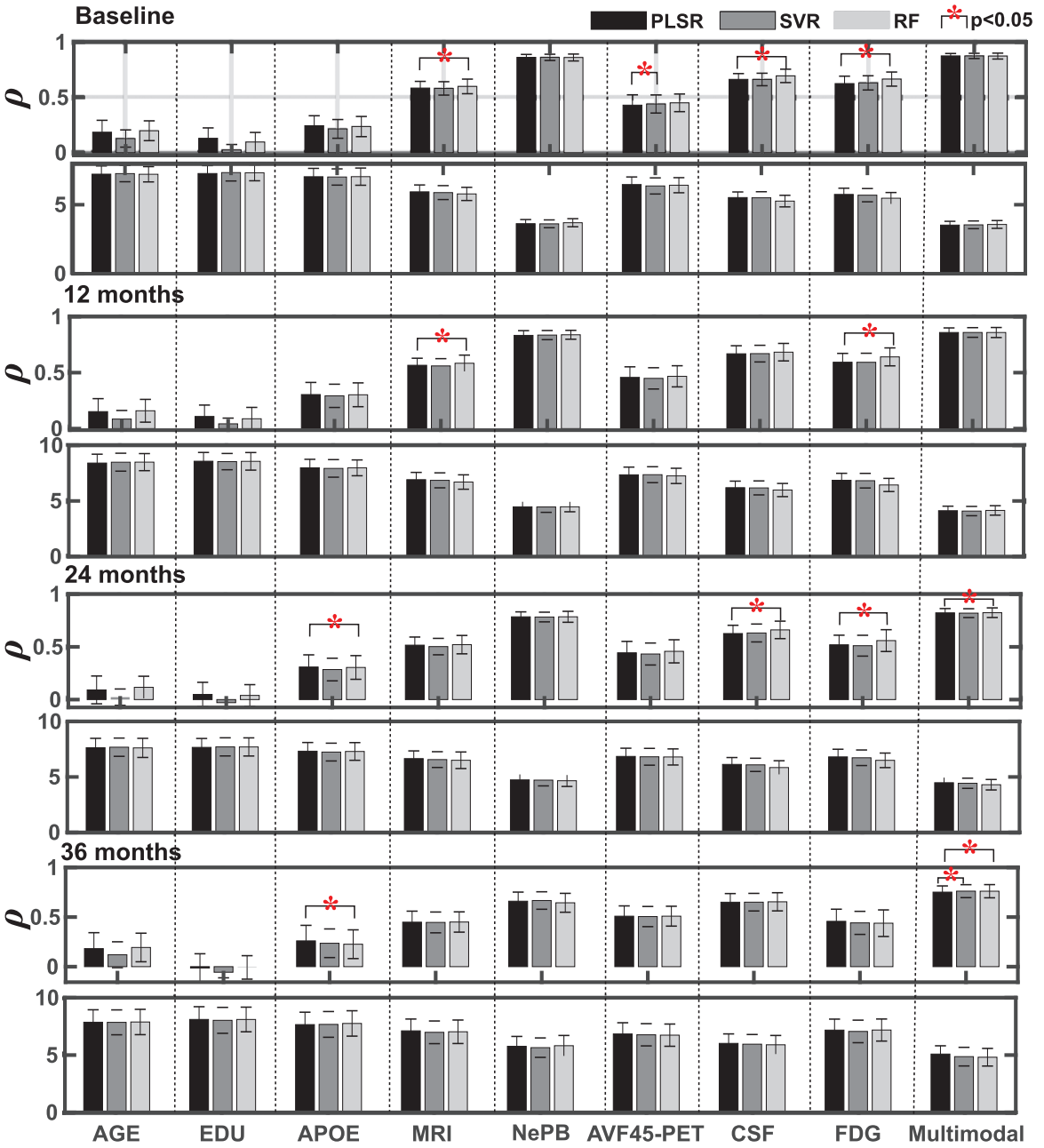

Fig. 4. Comparison of single and multi-modal dataset performance—collapsing across diagnostic status—with partial least squares regression (PLSR), support-vector regression (SVR), and random forest regression (RF). The performances are shown for cross-validation correlation (ρ) and mean absolute error (MAE).

all periods (baseline and subsequent 12, 24, and 36-month follow-up) in all studied subject-groups (i.e., collapsing over diagnostic categories). The correlations between the predicted ADAS-Cog 13 scores based on multi-modal data and that of the observed scores at 12, 24, and 36 months were 0.88, 0.82, and 0.75, respectively. The performance comparison (Fig. 4) shows that the differences among PLSR, SVR, and RF were not significant (i.e., $p > 0.05$), except for some instances where PLSR underperformed compared to RF (baseline and 12 months: MRI, CSF, and FDG; 24 and 36 months: *APOE* and multi-modal). However, PLSR models were computationally faster and performed consistently.

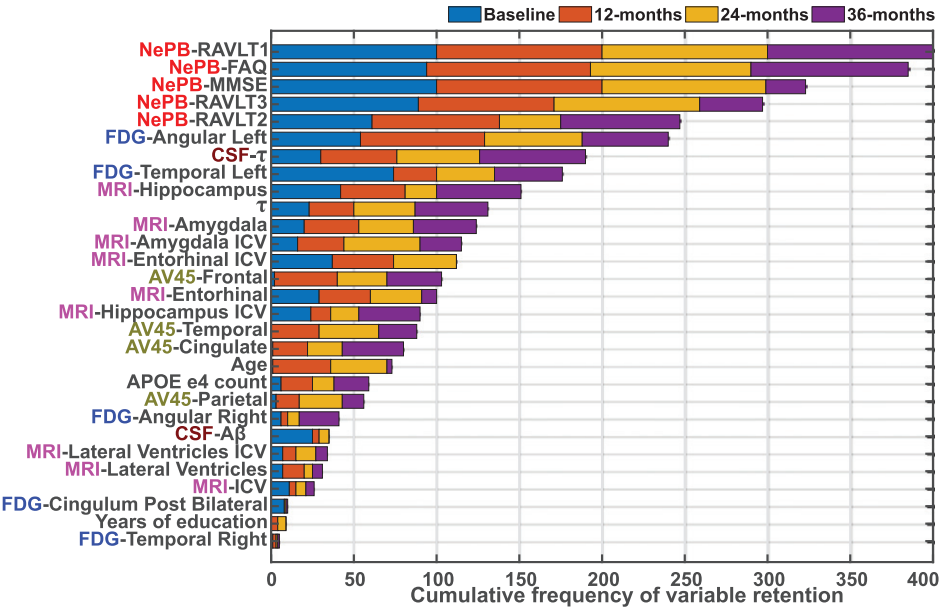

Fig. 5. Genetic algorithm-based importance of parameters in correlations as observed for 100 runs for every period. The frequency indicates the proportional contribution in ADAS-cog 13 score prediction. The modality group is prefixed to the variable names.

By analyzing the importance of measures (Fig. 5) contributing to PLSR's correlation we observe that the NePB were most important and consistent across periods for predicting ADAS score, followed by CSF and MRI biomarkers. Despite the association of age at baseline, years of education (Edu.), and *APOE* ε4 status with AD risk, these parameters were found to be least important, perhaps because these factors are somehow reflected in other parameters. By contrast, the importance of amyloid and tau increased when predictions were made 36 months in advance (Supplementary Figure 2). Additionally, metabolic activity in the temporal right and left sides were on the opposite ends of the importance in the ADAS-cog score predictions.

Grouping data based on diagnosis at baseline (Fig. 6) and analyzing the prediction performances, we generally observed low correlations when a single modal data was employed to predict ADAS-score. However, we observed that NePB, as a single modal predictor, showed the best predictive performance, in line with the fact that the to-be-predicted variable (ADAS-Cog 13) also contains NePB outcomes. However, the multi-modal approach performs better than the NePB modality for MCI and AD groups especially during 24- and 36-month periods. Due to the high variation in ADAS scores in AD groups, the correlation (ρ) and MAE were not inversely proportional to each other.

Our multi-modal based prediction models with the inclusion of baseline ADAS-cog scores were better ($\rho = 0.80$ to 0.90, Fig. 7 and Supplementary Figure 3) than the prediction models based only on baseline ADAS-cog scores ($\rho = 0.75$ to 0.87). The inclusion of the ADAS-cog score with other baseline multi-modal predictors was observed with improvements ($p = 0.002$) in the correlations. Overall, the models predicted well for all periods and this can be observed when we compare the mean predicted values versus the actual mean values (Fig. 8).

DISCUSSION

We presented a multi-modal regression approach to quantitatively track the progression of AD and showed that it outperforms the conventional single modal approach. Quantification of AD aids clinicians in decisions with treatment and a multi-modal approach ensures that the prediction models consider all biomarkers contributing to the disease condition. Furthermore, conventional classification of patients into normal, MCI, or AD could be avoided as a clear distinction among the group is a challenging task [46].

The classification of subjects based on a few modalities has been the focus of most recent studies. Although high classification accuracy (>80%)

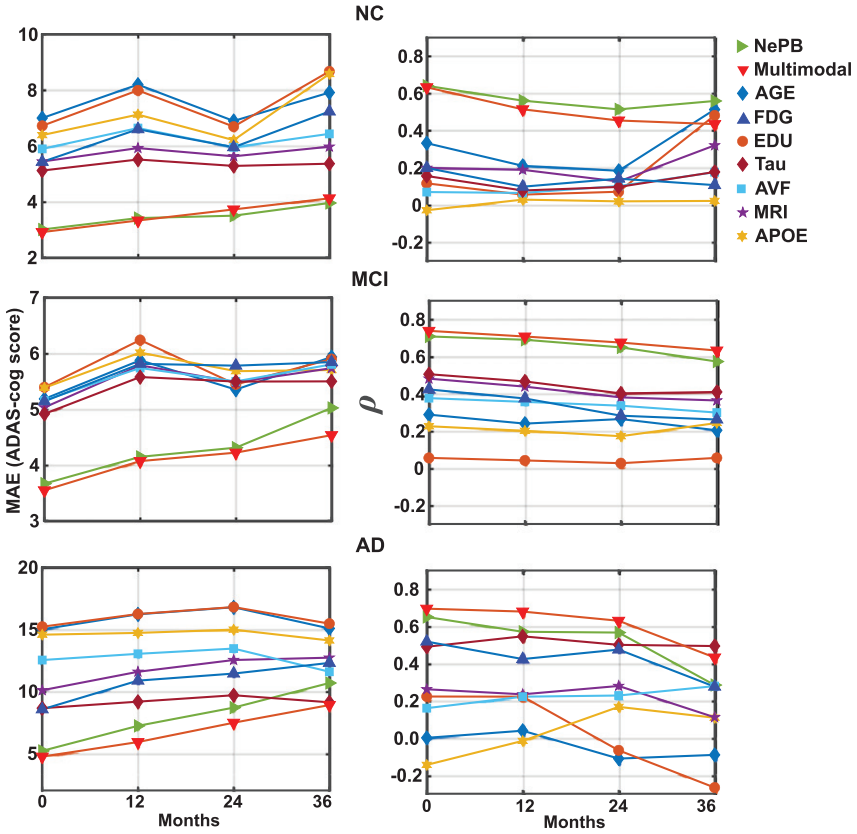

Fig. 6. Performance of PLSR on single- and multi-modal data stratified according to baseline clinical diagnosis [normal cognition (NC), mild cognitive impairment (MCI), and Alzheimer's disease (AD)]. The performance measures cross-validation correlation (ρ) and mean absolute error (MAE) for ADAS-cog scores are plotted for predictions at 0, 12, 24, and 36 months in advance.

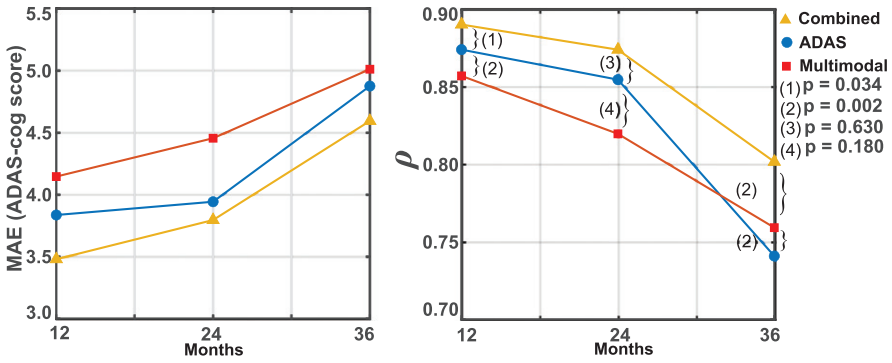

Fig. 7. Performance comparison of prediction models utilizing only ADAS scores versus multi-modal data with and without the combination with baseline ADAS scores. The *p*-values correspond to pair-wise differences between the three prediction models at different periods. The comparison of the models was based on a permutation test and is detailed in the Supplementary Material [43].

has been reported [17], we speculate that the impact of mislabeling a subject in the wrong category (and hence, wrong therapy prescribed) is higher than the error in predicting ADAS-cog scores (<5 units). Additionally, ADAS-cog scores are easy to interpret and follow the longitudinal tracking of AD progression. In agreement with the classification-based studies [7], the multi-modal approach outperforms the single modal approach. Furthermore, our multi-modal approach shows that ADAS-cog scores are

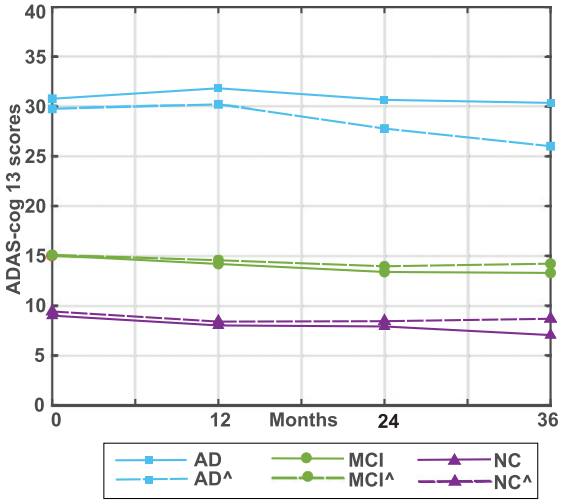

Fig. 8. The mean of actual ADAS-cog 13 scores of subjects over 36 months is plotted (solid line) for the 3 diagnostic groups (normal cognition (NC), mild cognitive impairment (MCI), and Alzheimer's disease (AD)). The learning model predicted scores for these periods are also plotted (dashed lines) for each diagnostic group (model predictions indicated with a caret (∧) symbol).

conducive to longitudinal predictions contrary to Marinescu et al. [31], where ADAS-cog scores were concluded not to be predictable. We, however, acknowledge that studies were not set up equally as we used cross-sectional predictors instead of longitudinal predictors and our setup aims for standardized comparison of predictors from several modalities while Marinescu et al. aimed to model the patient stratification in clinical trials.

Clinically, NePB tests and ADAS-cog scores measure the subject's cognitive abilities and this similarity was showcased with the observance of higher correlations (Fig. 3 and Supplementary Figure 3). As the precise pathophysiology and relative contribution of different pathogenic factors to AD at different phases of disease progression are currently still under investigation, the results advocate that instead of manually estimating the best markers, a multi-modal approach is beneficial. However, we acknowledge that the variable selection methods can be utilized to select the best AD measures (or create sparse models) utilized in multi-modal modeling further improving the robustness of the prediction model.

The multivariate techniques (i.e., PLSR, SVR, and RF) were observed to perform very similarly in their predictions but the computation times were different, and this prompted us to favor PLSR. Other nonlinear regression techniques could improve the current results [47]. The subject attrition during follow-ups may have diminished the predictive performance of the model.

CONCLUSION

ADAS-cog 13 scores reflect the current cognitive state of individuals, and through multivariate regression and a multi-modal dataset, our results show that quantitative longitudinal prediction of AD progression is possible. Thus, the automated multi-modal approach could help clinicians make timely decisions for interventions at all stages of AD and inform likely disease progression at the start of clinical trials.

ACKNOWLEDGMENTS

This study was funded by the Research Committee of the Kuopio University Hospital Catchment Area for the State Research Funding (5041778) and The Finnish Foundation for Technology Promotion (8193, 6227) as well as The Academy of Finland (grant 316258 to JT).

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 affliated 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 made use of the TADPOLE data sets https://tadpole.grand-challenge.org constructed by the EuroPOND consortium http://europond.eu funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 666992. The computational analysis was run on the servers provided by Bioinformatics Center, University of Eastern Finland, Finland.

Authors' disclosures available online (https://www.j-alz.com/manuscript-disclosures/20-0906r1)

AVAILABILITY OF DATA AND MATERIALS

Readers are directed to http://www.adni-info.org for detailed information on the ADNI project and the TADPOLE challenge (https://tadpole.grand-challenge.org) constructed by the EuroPOND consortium (http://europond.eu). The main codes and resulting.mat file are available on GitHub: https://github.com/mithp/ADAS_multimodal.git

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/JAD-200906

REFERENCES

[1] Gaudreault R, Mousseau N (2019) Mitigating Alzheimer's disease with natural polyphenols: A review. *Curr Alzheimer Res* **16**, 529-543.

[2] Ceyzériat K, Zilli T, Millet P, Frisoni GB, Garibotto V, Tournier BB (2020) Learning from the past: A review of clinical trials targeting amyloid, tau and neuroinflammation in Alzheimer's disease. *Curr Alzheimer Res* **17**, 112-125.

[3] Alzheimer's Association (2019) 2019 Alzheimer's disease facts and figures. *Alzheimers Dement* **15**, 321-387.

[4] Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, Trojanowski JQ, Toga AW, Beckett L (2005) Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimers Dement* **1**, 55-66.

[5] Nie L, Zhang L, Yang Y, Wang M, Hong R, Chua T-S (2015) Beyond doctors. In *Proceedings of the 23rd ACM international conference on Multimedia - MM '15* ACM Press, New York, New York, USA, pp. 591-600.

[6] Partovi S, Yuh R, Pirozzi S, Lu Z, Couturier S, Grosse U, Schluchter MD, Nelson A, Jones R, O'Donnell JK, Faulhaber P (2017) Diagnostic performance of an automated analysis software for the diagnosis of Alzheimer's dementia with 18F FDG PET. *Am J Nucl Med Mol Imaging* **7**, 12-23.

[7] Zhang D, Wang Y, Zhou L, Yuan H, Shen D (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* **55**, 856-867.

[8] Kärkkäinen M, Prakash M, Zare M, Tohka J, for the Alzheimer's Disease Neuroimaging Initiative (2020) Structural brain imaging phenotypes of mild cognitive impairment (MCI) and Alzheimer's disease (AD) found by hierarchical clustering. *Int J Alzheimers Dis* **2020**, 2142854.

[9] Perry RJ, Watson P, Hodges JR (2000) The nature and staging of attention dysfunction in early (minimal and mild) Alzheimer's disease: Relationship to episodic and semantic memory impairment. *Neuropsychologia* **38**, 252-271.

[10] Morris JC, Storandt M, Miller JP, McKeel DW, Price JL, Rubin EH, Berg L (2001) Mild cognitive impairment represents early-stage Alzheimer disease. *Arch Neurol* **58**, 397-405.

[11] Nestor PJ, Scheltens P, Hodges JR (2004) Advances in the early detection of Alzheimer's disease. *Nat Rev Neurosci* **10**, S34.

[12] Duara R, Barker WW, Lopez-Alberola R, Loewenstein DA, Grau LB, Gilchrist D, Sevush S, St. George-Hyslop PH (1996) Alzheimer's disease: Interaction of apolipoprotein E genotype, family history of dementia, gender, education, ethnicity, and age of onset. *Neurology* **46**, 1575-1579.

[13] Ryan J, Fransquet P, Wrigglesworth J, Lacaze P (2018) Phenotypic heterogeneity in dementia: A challenge for epidemiology and biomarker studies. *Front Public Health* **6**, 181.

[14] Karantzoulis S, Galvin JE (2011) Distinguishing Alzheimer's disease from other major forms of dementia. *Expert Rev Neurother* **11**, 1579-1591.

[15] Selkoe DJ (1991) The molecular pathology of Alzheimer's disease. *Neuron* **6**, 487-498.

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

[17] Rathore S, Habes M, Iftikhar MA, Shacklett A, Davatzikos C (2017) A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *Neuroimage* **155**, 530-548.

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

[19] Keihaninejad S, Zhang H, Ryan NS, Malone IB, Modat M, Cardoso MJ, Cash DM, Fox NC, Ourselin S (2013) An unbiased longitudinal analysis framework for tracking white matter changes using diffusion tensor imaging with application to Alzheimer's disease. *Neuroimage* **72**, 153-163.

[20] Jack CR, Holtzman DM (2013) Biomarker modeling of Alzheimer's disease. *Neuron* **80**, 1347-1358.

[21] Growdon JH (2001) Incorporating biomarkers into clinical drug trials in Alzheimer's disease. *J Alzheimers Dis* **3**, 287-292.

[22] Yang E, Farnum M, Lobanov V, Schultz T, Raghavan N, Samtani MN, Novak G, Narayan V, Dibernardo A (2011) Quantifying the pathophysiological timeline of Alzheimer's disease. *J Alzheimers Dis* **26**, 745-753.

[23] William-Faltaos D, Chen Y, Wang Y, Gobburu J, Zhu H (2013) Quantification of disease progression and dropout for Alzheimer's disease. *Int J Clin Pharmacol Ther* **51**, 120-131.

[24] Moore PJ, Lyons TJ, Gallacher J (2019) Random forest prediction of Alzheimer's disease using pairwise selection from time series data. *PLoS One* **14**, e0211558.

[25] Utsumil Y, Rudovicl OO, Petersonl K, Guerrero R, Picardl RW (2018) Personalized Gaussian processes for forecasting of Alzheimer's Disease Assessment Scale-Cognition Sub-Scale (ADAS-Cog13). In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* IEEE, pp. 4007-4011.

[26] Bhagwat N, Viviano JD, Voineskos AN, Chakravarty MM (2018) Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data. *PLoS Comput Biol* **14**, e1006376.

[27] Liem F, Dadi K, Engemann DA, Gramfort A, Bellec P, Cameron R, 5 C, Damoiseaux JS, Steele CJ, 5 Y, Margulies DS, Varoquaux G (2020) Predicting future cognitive decline from non-brain and multimodal brain imaging data in healthy and pathological aging. *bioRxiv* 2020.06.10.142174.

[28] Skinner J, Carvalho JO, Potter GG, Thames A, Zelinski E, Crane PK, Gibbons LE (2012) The Alzheimer's Disease Assessment Scale-Cognitive-Plus (ADAS-Cog-Plus): An expansion of the ADAS-Cog to improve responsiveness in MCI. *Brain Imaging Behav* **6**, 489-501.

[29] Kueper JK, Speechley M, Montero-Odasso M (2018) The Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog): Modifications and responsiveness in predementia populations. A narrative review. *J Alzheimers Dis* **63**, 423-444.

[30] Zhang D, Shen D (2012) Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. *PLoS One* **7**, e33182.

[31] Marinescu RV, Oxtoby NP, Young AL, Bron EE, Toga AW, Weiner MW, Barkhof F, Fox NC, Golland P, Klein S, Alexander DC (2019) TADPOLE Challenge: Accurate Alzheimer's disease prediction through crowdsourced forecasting of future data. In *Predictive Intelligence in Medicine*, Rekik I, Adeli E, Park S, eds. PRIME 2019. Lecture Notes in Computer Science, vol 11843. Springer, Cham.

[32] Lei B, Hou W, Zou W, Li X, Zhang C, Wang T (2019) Longitudinal score prediction for Alzheimer's disease based on ensemble correntropy and spatial-temporal constraint. *Brain Imaging Behav* **13**, 126-137.

[33] de Jong S (1993) SIMPLS: An alternative approach to partial least squares regression. *Chemom Intell Lab Syst* **18**, 251-263.

[34] Chen C, Cao X, Tian L (2019) Partial least squares regression performs well in MRI-based individualized estimations. *Front Neurosci* **13**, 1282.

[35] Lahmiri S, Shmuel A (2019) Performance of machine learning methods applied to structural MRI and ADAS cognitive scores in diagnosing Alzheimer's disease. *Biomed Signal Process Control* **52**, 414-419.

[36] Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. *Stat Comput* **14**, 199-222.

[37] Breiman L (2001) Random forests. *Mach Learn* **45**, 5-32.

[38] Marinescu RV, Oxtoby NP, Young AL, Bron EE, Toga AW, Weiner MW, Barkhof F, Fox NC, Klein S, Alexander DC (2018) TADPOLE Challenge: Prediction of longitudinal evolution in Alzheimer's disease. *Predict Intell Med* **11843**, 1-10.

[39] Krishnan A, Williams LJ, McIntosh AR, Abdi H (2011) Partial least squares (PLS) methods for neuroimaging: A tutorial and review. *Neuroimage* **56**, 455-475.

[40] Zhang D, Shen D (2012) Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *Neuroimage* **59**, 895-907.

[41] Gray KR, Aljabar P, Heckemann RA, Hammers A, Rueckert D (2013) Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *Neuroimage* **65**, 167-175.

[42] Leardi R, Boggia R, Terrile M (1992) Genetic algorithms as a strategy for feature selection. *J Chemom* **6**, 267-281.

[43] Lewis JD, Evans AC, Tohka J (2018) T1 white/gray contrast as a predictor of chronological age, and an index of cognitive performance. *Neuroimage* **173**, 341-350.

[44] Breiman L (1996) Bagging predictors. *Mach Learn* **24**, 123-140.

[45] R. Leardi and A. Lupiáñez (1998) Genetic algorithms applied to feature selection in PLS regression: How and when to use them. *Chemom Intell Lab Syst* **41**, 95-207.

[46] Leyhe T, Reynolds CF, Melcher T, Linnemann C, Klöppel S, Blennow K, Zetterberg H, Dubois B, Lista S, Hampel H (2017) A common challenge in older adults: Classification, overlap, and therapy of depression and dementia. *Alzheimers Dement* **13**, 59-71.

[47] Chouldechova A, Hastie T (2015) Generalized additive model selection. *arXiv Prepr arXiv150603850*.

[48] Gómez-Sancho M, Tohka J, Gómez-Verdejo V (2018) Comparison of feature representations in MRI-based MCI-to-AD conversion prediction. *Magn Reson Imaging* **50**, 84-95.

[49] Voevodskaya O (2014) The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease. *Front Aging Neurosci* **6**, 264.

[50] Johnson KA, Sperling RA, Gidicsin CM, Carmasin JS, Maye JE, Coleman RE, Reiman EM, Sabbagh MN, Sadowsky CH, Fleisher AS, Murali Doraiswamy P, Carpenter AP, Clark CM, Joshi AD, Lu M, Grundman M, Mintun MA, Pontecorvo MJ, Skovronsky DM (2013) Florbetapir (F18-AV-45) PET to assess amyloid burden in Alzheimer's disease dementia, mild cognitive impairment, and normal aging. *Alzheimers Dement* **9**(5 Suppl), S72-83.

[51] Landau SM, Mintun MA, Joshi AD, Koeppe RA, Petersen RC, Aisen PS, Weiner MW, Jagust WJ (2012) Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Ann Neurol* **72**, 578-586.

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

[53] Mormino EC, Kluth JT, Madison CM, Rabinovici GD, Baker SL, Miller BL, Koeppe RA, Mathis CA, Weiner MW, Jagust WJ (2009) Episodic memory loss is related to hippocampal-mediated β-amyloid deposition in elderly subjects. *Brain* **132**, 1310-1323.

[54] Landau SM, Harvey D, Madison CM, Koeppe RA, Reiman EM, Foster NL, Weiner MW, Jagust WJ (2011) Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. *Neurobiol Aging* **32**, 1207-1218.

[55] Jagust WJ, Bandy D, Chen K, Foster NL, Landau SM, Mathis CA, Price JC, Reiman EM, Skovronsky D, Koeppe RA (2010) The Alzheimer's Disease Neuroimaging Initiative positron emission tomography core. *Alzheimers Dement* **6**, 221-229.

[56] Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, Blennow K, Soares H, Simon A, Lewczuk P, Dean R, Siemers E, Potter W, Lee VMY, Trojanowski JQ (2009) Cerebrospinal fluid biomarker signature in alzheimer's disease neuroimaging initiative subjects. *Ann Neurol* **65**, 403-413.

[57] 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.

[58] Presses Universitaires de France, L'examin clinique en psychologie.

[59] Pfeffer RI, Kurosaki TT, Harrah CH, Chance JM, Filos S (1982) Measurement of functional activities in older adults in the community. *J Gerontol* **37**, 323-329.

[60] Prencipe M, Casini AR, Ferretti C, Lattanzio MT, Fiorelli M, Culasso F (1996) Prevalence of dementia in an elderly rural population: Effects of age, sex, and education. *J Neurol Neurosurg Psychiatry* **60**, 628-633.

[61] Michaelson DM (2014) APOE ε4: The most prevalent yet understudied risk factor for Alzheimer's disease. *Alzheimers Dement* **10**, 861-868.

[62] Mohs RC, Knopman D, Petersen RC, Ferris SH, Ernesto C, Grundman M, Sano M, Bieliauskas L, Geldmacher D, Clark C, Thal LJ (1997) Development of cognitive instruments for use in clinical trials of antidementia drugs: Additions to the Alzheimer's disease assessment scale that broaden its scope. *Alzheimer Dis Assoc Disord* **11**(Suppl 2), S13-21.

[63] Raghavan N, Samtani MN, Farnum M, Yang E, Novak G, Grundman M, Narayan V, Dibernardo A (2013) The ADAS-Cog revisited: Novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. *Alzheimers Dement* **9** (1 Suppl), S21-31.

[64] Sarica A, Cerasa A, Quattrone A (2017) Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review. *Front Aging Neurosci* **9**, 329.