Received: 17 December 2020 | Revised: 11 May 2021 | Accepted: 30 July 2021 | Published online: 11 October 2021

DOI: 10.1002/trc2.12212

Alzheimer's & Dementia
Translational Research & Clinical Interventions

RESEARCH ARTICLE

Predicting future amyloid biomarkers in dementia patients with machine learning to improve clinical trial patient selection

Fabian H. Reith[1] | **Elizabeth C. Mormino**[2] | **Greg Zaharchuk**[1]

[1] Department of Radiology, Stanford University, Palo Alto, California, USA

[2] Department of Neurology and Neurological Sciences, Stanford University, Palo Alto, California, USA

Correspondence
Greg Zaharchuk, Stanford University, 300 Serra Mall, Stanford, CA 94305 USA.
E-mail: gregz@stanford.edu

Funding information
No external support.

Funding information
the Alzheimer's Disease Neuroimaging Initiative; National Institutes of Health, Grant/Award Number: U01 AG024904; Department of Defense, Grant/Award Number: W81XWH-12-2-0012; the National Institute on Aging; the National Institute of Biomedical Imaging and Bioengineering

Abstract

Introduction: In Alzheimer's disease, asymptomatic patients may have amyloid deposition, but predicting their progression rate remains a substantial challenge with implications for clinical trial enrollment. Here, we demonstrate an artificial intelligence approach to use baseline clinical information and images to predict changes in quantitative biomarkers of brain pathology on future images.

Methods: Patients from the Alzheimer's Disease Neuroimaging Initiative (ADNI) who underwent positron emission tomography (PET) with the amyloid radiotracer 18F-AV45 (florbetapir) were included. We identified important baseline PET image features using a deep convolutional neural network based on ResNet. These were combined with eight clinical, demographic, and genetic markers using a gradient-boosted decision tree (GBDT) algorithm to predict future quantitative standardized uptake value ratio (SUVR), an established biomarker of brain amyloid deposition. We used this model to better identify individuals with the highest positive change in amyloid deposition on future images and compared this to typical inclusion criteria for clinical trials. We also compared the model's performance to other methods such as multivariate linear regression and GBDT without imaging features.

Findings: Using 2577 PET scans from 1224 unique individuals, we showed that the GBDT with deep image features was significantly more accurate than the other approaches, reaching a root mean squared error of 0.0339 ± 0.0027 for future SUVR prediction. Using this approach, we could identify individuals with the highest 10% SUVR accumulation at rates 2- to 4-fold higher than by random pick or existing inclusion criteria.

Discussion: Predicting quantitative biomarkers on future images using machine learning methods consisting of deep image features combined with clinical data may allow better targeting of treatments or enrollment in clinical trials.

KEYWORDS
Alzheimer's disease, artificial intelligence, biomarkers, deep learning, medical imaging, personalized medicine, prognostics, radiology

1 | INTRODUCTION

While correct diagnosis of a disease entity is required to select appropriate therapies, knowledge about the disease evolution and prognosis is also important. For example, some diseases might be self-limited and require minimal or no interventions. Other diseases might have faster and more virulent time courses, pushing providers to consider higher risk or more invasive therapies. Furthermore, prognosis information plays a critical role in patient discussions and resource planning. While radiology has so far focused largely on diagnosis, an opportunity exists to use radiological information to risk-stratify patients regarding disease prognosis.

Dementia is one condition for which prognosis is especially important. Understanding the likelihood and rate of progression of this disease would be extremely helpful, not only for individual patients and families, but also to plan clinical trials. Alzheimer's disease (AD) trials face significant challenges with enrollment.[1,2] Being able to selectively recruit patients likely to progress quickly, based in part on brain imaging biomarkers such as amyloid and tau deposition, could significantly impact the design, duration, and cost of clinical trials.

Deep learning has shown much promise in classifying patients and predicting their future disease trajectories.[3–6] It has also been used at the image level to transform images, either for better image reconstruction or the synthesis of desired contrasts (i.e., predicting computed tomography [CT] from magnetic resonance imaging [MRI] to enable MR-based positron emission tomography [PET] attenuation correction[7–9]). Combining these two approaches would allow prediction of future image-based biomarkers, such as regional radiotracer uptake. In this study, we combined clinical, genetic, and imaging features at baseline and then used these to identify individuals at highest risk of rapid biomarker progression, in this case the quantitative change in amyloid beta protein deposition on future imaging studies. Using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), we show that a model combining deep learning–derived image features and gradient-boosted random forest regression outperforms existing methods, that the baseline imaging features play a large role in detecting fast progressors, and that this can be used to enrich cohorts with fast progressors.

RESEARCH IN CONTEXT

1. **Systematic Review**: We reviewed the literature of biomarker prediction with machine learning models during the development of this work (August 2020) and references to relevant work are included. Works combining deep learning–derived imaging features combined with clinical and genetic data to predict a future imaging biomarker were found to be lacking. Specifically, no studies capable of predicting future amyloid deposition were identified.
2. **Interpretation**: This study showed that combining deep learning–derived imaging features with clinical, demographic, and genetic information could be used to accurately predict the change in amyloid deposition. Using this method, it was possible to identify the fastest amyloid accumulators with a much higher frequency than by using existing techniques.
3. **Future Directions**: Identifying patients with rapid amyloid accumulation will enable better selection for clinical trials and personalized targeting of therapies, once available. The method is general, with the framework likely applicable to other future biomarker prediction tasks.

2 | METHODS

2.1 | Imaging information

We obtained all available 18F-AV45 (florbetapir, Avid Lilly) PET studies from ADNI as of August 2019. All scans were downloaded in Neuroimaging Informatics Technology Initiative (NIFTI) file format along with the University of California (UC) Berkeley AV45 analysis to obtain a standardized uptake value ratio (SUVR) values based on a reference region consisting of cerebellum, brainstem/pons, and eroded white matter (SUMMARYSUVR_COMPOSITE_REFNORM).[10–12] Higher SUVR reflects more amyloid deposition in supratentorial cortical regions. For amyloid positivity, we used an SUVR threshold of 0.79 as suggested by the ADNI researchers. We selected all patients with multiple scans and calculated the interval SUVR change from baseline (ΔSUVR).

2.2 | Clinical and genetic information

For model development, we also included several clinical and genetic features, including patient age, sex, weight, baseline cognitive testing scores, and apolipoprotein E (*APOE*) gene status. Two cognitive tests were included: the Mini-Mental State Examination (MMSE)[13] and the Functional Activities Questionnaire (FAQ) total score.[14] We also included the polymorphic expression of the *APOE* gene,[15] as the *APOE* genotype is known to strongly affect amyloid deposition.[16] To assess performance of the model in different clinical cohorts, we examined clinical status using the Clinical Dementia Rating (CDR) score if it was made ±50 days of the baseline PET scan.

2.3 | Prediction models

2.3.1 | Linear regression

We performed multivariate regression using the StatsModels library [17] in Python, which fits the following equation:

$$y_i = \beta_0\ constant + \beta_1 x_{i1} + \beta_1 x_{i2} + \cdots + \beta_n x_{in}$$

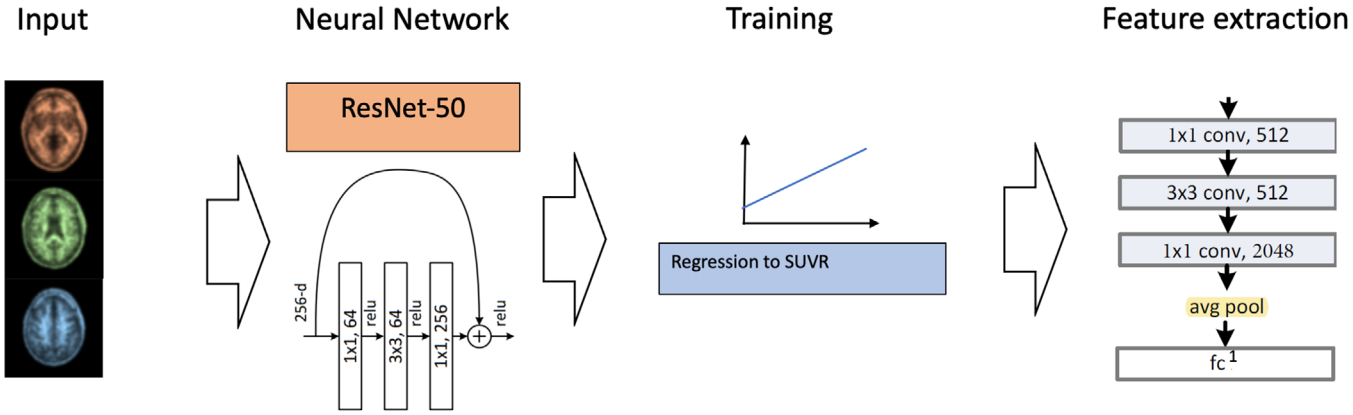

FIGURE 1 Overview of ResNet-50 training procedure. Three central slices are fed into the input color channels. The ResNet algorithm is modified to perform regression on standardized uptake value ratio (SUVR) rather than classification. Finally, for the testing of ΔSUVR prediction, we extract 2048 deep features, the results of the average pool operation (scalar values)

2.4 | Deep learning

We trained a convolutional neural network (CNN) to predict amyloid PET SUVR,[18] using methods described in Reith et al.[19] Of note, the CNN is not trained to predict future SUVR change, but instead learned image features associated with baseline SUVR. In brief, the ResNet-50 architecture[20] was used. Network input was three centrally located slices. Standard ResNet ends with a layer for distinguishing 1000 differing classes, but we modified it for SUVR prediction (a regression task). The final layer was changed to a single output without an activation function. The cost function was the mean squared error between predicted and true SUVR using the ADAM optimizer.[21] We applied the best-performing hyperparameters for training on current SUVR[19] and settled on an initial learning rate of 0.0001, 30 epochs, with 10x decrease of learning rate every 10 epochs. Training time was 22 minutes. The model was pretrained using the ImageNet dataset of natural images.[22] After training, we used PyTorch to extract the last layer's activations. This resulted in 2048 numbers (features) for each individual PET scan (Figure 1).

Becauase the goal was to predict SUVR change based on baseline patient information and the ResNet-50-derived features, we trained this network on baseline images only. The training set consisted of 1441 amyloid PET scans (610 baseline scans of subjects used subsequently for testing, 489 baseline scans of subjects with only a single PET study, and 342 scans from 125 subjects with multiple scans). We used a cross-validation testing design (described later) such that none of the follow-up scans used for training were from patients that were evaluated for ΔSUVR in the test set. A smaller ResNet training dataset consisting of 831 scans was also tested to demonstrate the effect of larger training sets and the details and results are found in the supporting information. The test set consisted of follow-up amyloid PET scans not used in training (n = 1136 scans from 610 subjects; details in Figure S2 in supporting information).

2.5 | Gradient boosting decision tree

To combine clinical/genetic features and deep imaging features, we used a gradient boosting decision tree (GBDT) algorithm, specifically the LightGBM implementation, to predict SUVR change.[23,24] Details of the training procedure can be found in the supporting information.

We tested GBDT models with and without deep learning–based PET features to assess the importance of the images. The GBDT model with activations is based on the ResNet training we performed as described above (8 clinical + 2048 deep features). To determine the importance of deep activations, we tested two additional models. One type used the same GBDT in which the true deep activations were replaced with random numbers, reflecting no impact of the imaging features (8 clinical + 2048 random activations). These random numbers were drawn from a normal distribution consistent with the mean and variance of the ResNet features themselves. The second method was to train a GBDT using only the eight available clinical features. A schematic of the different machine learning (ML) methods can be found in Figure S1 in supporting information. More details of the GBDT models are also in the supporting information.

2.6 | Data analysis

We analyzed GBDT performance via root mean squared error (RMSE) for ΔSUVR in all follow-up scans (1136 scans in 610 individuals). We performed 5-fold cross-validation and present the average RMSE. For cross-validation purposes, we divide our dataset into five distinct parts. Each part has its unique subjects, meaning no subject can be found in more than one of the five distinct parts, guaranteeing that there is no subject shared by both the training and test sets. We analyzed the statistical significance of model design choices with linear mixed effects models and Wilcoxon rank sum tests, as appropriate. For these measures, we compared the squared error of ML system predictions.

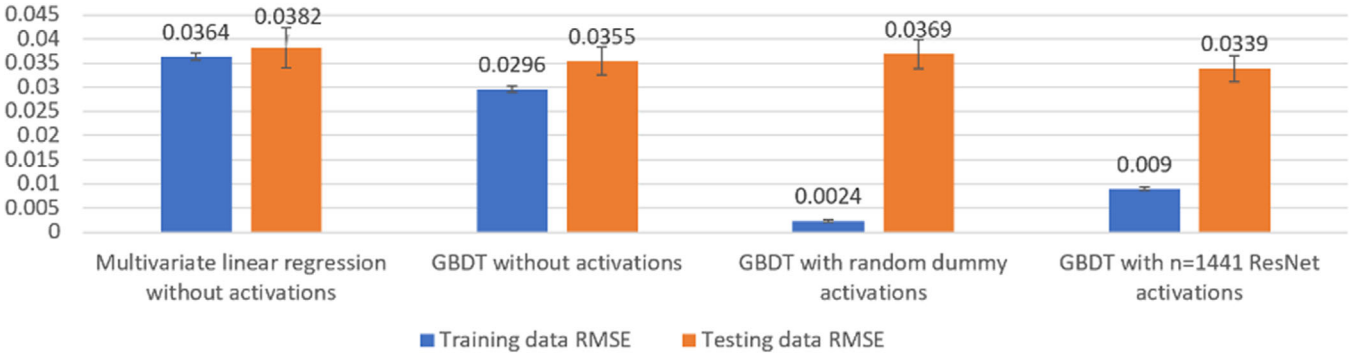

FIGURE 2 Training and test set performance. Root mean-squared error (RMSE) between prediction and the true ΔSUVR (standardized uptake value ratio) is shown, so lower values represent better model performance. The gradient-boosted decision tree (GBDT) using the image-based activations had the best performance.

To assess the practical value of these SUVR predictions, we also used them to select subjects with the highest SUVR changes. The rationale is that these patients might be desirable candidates for clinical trials assessing the impact of an amyloid-lowering agent. We identified the top 10% of cases (61 individuals) with the highest ΔSUVR. In subjects with multiple follow-up scans, we selected the scan with the maximum ΔSUVR. We assessed the performance of multivariate linear regression, GBDT without imaging features, and GBDT with imaging features by calculating the % of these top progressors also predicted by the model. For example, a random selection would lead to a 10% "hit rate," while the models should be able to improve upon this if they are making more accurate predictions.

We also compared model performance with other methods of selection in two ways. The first is to randomly select patients that meet a specific criterion. We included the following groups for these tests: amyloid positive at baseline (n = 313), presence of at least one *APOE* ε4 allele (n = 237), mildly positive amyloid patients (defined as baseline SUVR between 0.79 and 0.95; n = 156), amyloid positive with at least one *APOE* ε4 allele (n = 178), and mildly amyloid positive subjects with at least one *APOE* ε4 allele (n = 70). We also examined the various models' performance in pre-selected groups often targeted in clinical trials, specifically: mildly positive amyloid patients (as defined above) and subjects with mild dementia (baseline CDR 0.5; n = 229). Because these latter datasets start from a smaller denominator, the task was to identify the top 20% fastest true ΔSUVR progressors.

3 | RESULTS

3.1 | Patient cohort

The baseline demographics and clinical features of the 610 unique subjects with 1136 follow-up scans are summarized in Table 1. The time horizon of the follow-up predictions was as follows: 1 to 3 years (n = 553), 3 to 5 years (n = 354), and 5+ years (n = 227).

TABLE 1 Baseline demographics and SUVR of the 610 patients

Clinical feature	Value, mean±SD, (IQR)
Age (yrs)	73.1±7.4 (67.9, 78.1)
Sex	46.4% female, 53.6% male
Weight (kg)	78.3±15.8 (68.0, 87.0)
APOE	ε2/ε2 0.2%, ε2/ε3 9.7%, ε3/ε3 49.3%, ε2/ε4 1.9%, ε3/ε4 32.1%, ε4/ε4 6.7%.
FAQtotal	2.4±4.7 (0, 2.0)
MMSE	27.4±3.4 (26.0, 30.0)
Baseline SUVR	0.84±0.14 (0.74, 0.96)
CDR	0: 39.9%, 0.5: 55.1%, 1: 4.1%, 2: 0.7%, 3: 0.3%
Delta time (yrs)	3.5±1.6 (2.0, 4.3)
ΔSUVR	0.016±0.038 (-0.0085, 0.037)

Abbreviations: *APOE*, apolipoprotein E; CDR, Clinical Dementia Rating; FAQ, Functional Activities Questionnaire; GBDT, gradient-boosted decision tree; IQR, interquartile range; MMSE, Mini-Mental State Examination; SD, standard deviation; SUVR, standardized uptake value ratio.
Notes Delta time and ΔSUVR are based on 1136 follow-up data points used for training and testing the GBDTs. Please note that the CDR values were not used for model training or testing but used to compare performance in clinically relevant subgroups

3.2 | ΔSUVR prediction

Visual presentation of the performance of each model is shown in Figure 2. For multivariate linear regression, we found an average RMSE of 0.0364 ± 0.0007 and 0.0382 ± 0.004 on the training and test sets, respectively. The weights and significance of each feature are shown in Table S1 in supporting information. In contrast, using the GBDT with the same inputs (8 clinical features) without deep activations, we found slightly better performance (RMSE 0.0296 ± 0.0007 train, 0.0355 ± 0.0003 test). Using random activations in place of the deep activations, the model heavily overfits the training set (0.0024 ± 0.0003) but had worse performance on the test set (0.0369 ± 0.0003). For the GBDT model incorporating the deep imaging activations, we saw the

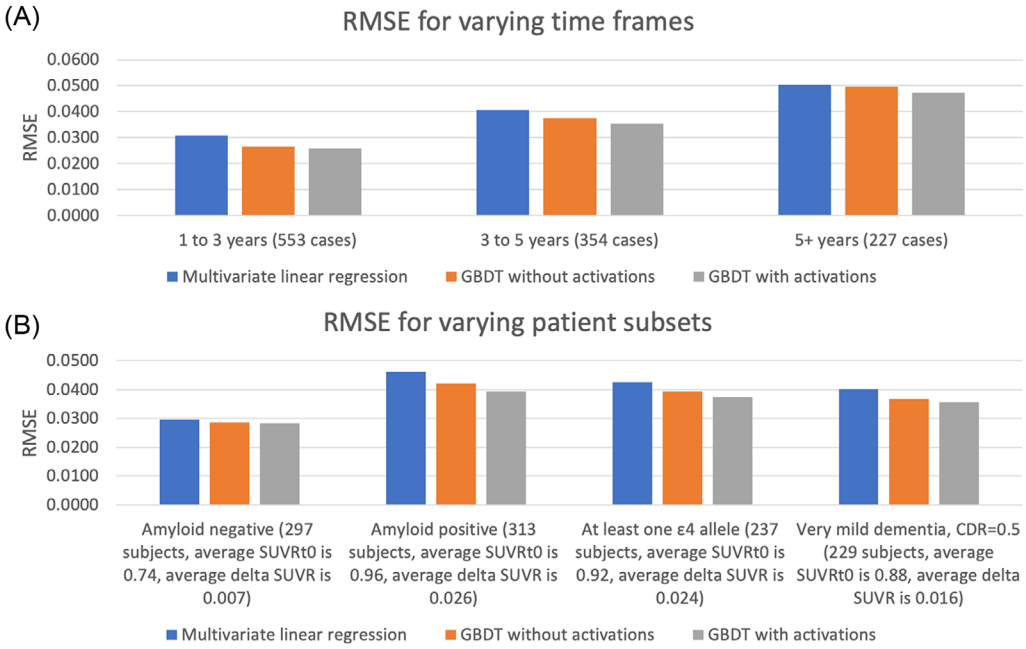

FIGURE 3 A, Root mean-squared error (RMSE) for predicting ΔSUVR (standardized uptake value ratio) at different time periods after baseline. Overall performance decreases for predictions farther in the future. The gradient-boosted decision tree (GBDT) with activations was always the best-performing model. B, Test set RMSE for predicting ΔSUVR in different patient subsets. Again, the GBDT with activations always performed best.

best performance (0.0090 ± 0.0004 train, 0.0339 ± 0.0003 test). While the performance difference between train and test for the GBDT with imaging activations suggests residual overfitting, applying regularization did not improve test set performance, and of course, the ultimate proof of their superiority is their better performance on the held-out test set. There was significantly better performance of the GBDT model with activations compared to the GBDT model that used dummy activations ($P < 2.35e\text{-}9$) or the GBDT model that only used the eight clinical features ($P < 0.00388$), measured using the Wilcoxon rank sum test. We additionally compared the correlation between ΔSUVR predictions and ground truth changes, shown in Figure S3 in supporting information. Similar to the findings of RMSE, the worst performing method is linear regression (correlation coefficient $R = 0.21$), while the best performing method is GBDT with deep activations ($R = 0.47$).

We analyzed model performance for the various time horizons of prediction (Figure 3). Two follow-up scans were excluded, as they were performed less than a year from baseline. In the shortest timeframe (1–3 years), accuracy was highest for all ML algorithms. This performance decreases slightly in the 3- to 5-year and 5+-year time frames. In all cases, GBDT with activations performed best.

GDBT with activations also performed better in many different subsets of the full cohort used in clinical trials. Results for initially amyloid-negative patients, amyloid-positive patients, patients with at least one *APOE* ε4 allele, and patients with mild cognitive impairment (CDR 0.5) are shown in Figure 3. In initially amyloid-negative patients, performance was similar between the different models, but for the other groups, the GBDT model with activations performed better.

The importance of various individual features was explored by removing individual features and measuring the effect on RMSE. In general, for models without activations, removing baseline SUVR and delta time made the biggest difference. When deep activations were used, only delta time omission led to a significant degradation in prediction performance (Figure S4 in supporting information).

3.3 | Implications for study selection

We evaluated the ability of the various models to identify the fastest 10% of true amyloid accumulators (Figure 4). The top 10% progressors had an average change in amyloid SUVR of 0.105 while the other subjects had an average SUVR change of 0.0010. When selecting using linear regression, we find 19.7% of this group are also in the top 10% of ground truth patients (sensitivity 0.20, specificity 0.91). This is already twice the number of subjects compared to the expected number of subjects selected via random pick. Selection based on GBDT without activations led to 29.5% identification (sensitivity 0.30, specificity 0.92). Highest performance was obtained using GBDT with activations, with 37.7% of fast progressors identified, an almost 4x increase in yield compared to a random selection (sensitivity 0.38, specificity 0.93). Further information about sensitivity and specificity along with positive and negative likelihood ratios are included in the Table S2 in supporting information. We see similar results when we limit progressors to certain time frames. When limited to 1- to 3-year or 3- to 5-year follow-up scan timeframes, we find an even better relative performance of GBDT

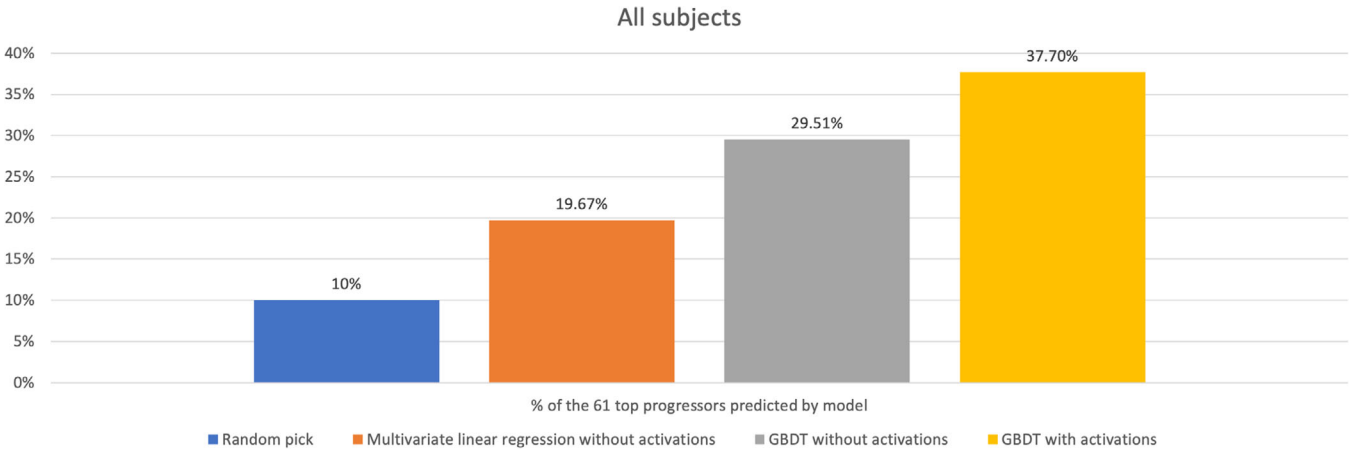

FIGURE 4 Percentage of ground truth top 61 (top 10%) progressors who are also found in top 61 highest predicted ΔSUVR (standardized uptake value ratio)

TABLE 2 Percentage of fastest amyloid progressors predicted via various "simple" selection methods compared to ML methods (bolded)

Selection method	% of the 61 top progressors (top 10%) predicted by method
Highest FAQtotal score with at least one ε4 allele	8.2%
Random pick	10.0%
Subjects with at least one ε4 allele	13.5%
Amyloid positive cases with at least one ε4 allele	15.7%
Amyloid positive cases	16.0%
Mildly amyloid positive cases	19.2%
Multivariate linear regression without activations	***19.7%***
Mildly amyloid positive cases with at least one ε4 allele	25.7%
GBDT without activations	***29.5%***
GBDT with activations	***37.7%***

Abbreviations: FAQ, Functional Activities Questionnaire; GBDT, gradient-boosted decision tree; IQR, interquartile range; ML, machine learning. Italic boldface text represents the machine learning methods tested in this study.

with activations compared to GBDT without activations. More detail can be found in Figure S7 in supporting information.

We compared these results to other ways of selecting fast progressors using the entire cohort. If we randomly select baseline amyloid-positive subjects, we would correctly predict the top 61 progressors in 16% (50/313 of baseline amyloid-positive subjects). Among the other methods of choosing fast progressors (at least one *APOE* ε4 allele, mildly positive amyloid patients, amyloid positive with at least one *APOE* ε4 allele, and mildly amyloid positive subjects with at least one *APOE* ε4 allele), best performance is the last group (25.7%), still significantly lower than the GBDT with activations model (Table 2).

We looked at model performance in clinically relevant subgroups. For mildly positive amyloid patients, of the top 20% fastest progressors, performance increases to 29.0% for linear regression, 41.9% for GBDT without activations, and 45.2% for GBDT with activations. For subjects with mild dementia at baseline (CDR 0.5), performance increases from 41.3% for linear regression, 43.5% for GBDT without activations, and 60.9% for GBDT with activations. These subjects had a baseline SUVR of 0.88 (interquartile range [IQR] 0.75, 1.00), similar to that of the subjects identified by GBDT with activations (0.92 [IQR 0.85, 0.97]). These findings are shown in Figure S5 in supporting information.

4 | DISCUSSION

In this study, we have extended prior work using neural networks to predict current SUVR for amyloid PET studies in the ADNI cohort [19] to predict SUVR in the future. We accomplished this by training a network on longitudinal studies and by including clinical and genetic features. We found that a GBDT that includes clinical, demographic, and genetic features combined with deep activations created from ResNet-50 had the best performance. We showed the value of this quantitatively, measuring the mean error in the prediction of the SUVR change, as well as on a practical basis, showing that using this approach can identify the fastest amyloid accumulators in both the entire test dataset as well as in clinically relevant subpopulations at a 2- to 4x higher rate than random selection or other commonly used selection methods. This latter capability might be useful to enrich research studies that target this biomarker, such as an amyloid-clearing pharmacological agent, reducing costs and speeding up clinical trials. Fundamentally, the idea of using deep learning to combine imaging and clinical information with the goal of predicting future imaging biomarkers, including the possible use in patients receiving different treatments, could be a fruitful pathway toward more personalized medicine.

Prior work in this area with the ADNI dataset has focused on predicting clinical outcomes.[3,25–28] In this study, we take a different

approach, focusing on predicting a quantitative biomarker which may be less subject to error compared to clinical evaluations. For example, the IDEAS trial showed that the assessments of clinicians diagnosing AD was often inaccurate.[28] Predicting the development of a relevant quantitative feature such as SUVR is objective and could find use in data mining, clinical trial assessments, and longitudinal analysis. We also believe that using regression (i.e., predicting the continuous value of an important biomarker) rather than classification allowed the algorithm to learn more from each individual subject's information.

The current study shows the importance of the deep features. In particular, we show that adding deep features to the GBDT improves performance for both RMSE and selection of fast progressors. We also showed that performance improves by obtaining better deep features by training on a larger number of PET scans (i.e., 1441 vs. 831 subjects), strongly suggesting that the model is learning relevant features for this prediction task. It also highlights the value of large shared datasets such as ADNI for ML methods using deep learning feature identification. We also found that the use of deep features make the model less dependent on missing clinical, demographic, or genetic data as shown by the studies in which individual features are selectively removed. One advantage of combining deep activations with the GBDT structure enabled the evaluation of the role of specific features and how sensitive the models are to missing data.

This study has several limitations. We did not test our performance in an external dataset, as significantly sized cohorts with similar information and scope to ADNI are not publicly available. It is possible that a cohort with different baseline distributions of disease or with a different mix of PET scanners might lead to poorer performance. However, given that ADNI is the largest multicenter cohort of its kind, at minimum, the weights trained in this model could be used as pretraining using a transfer learning approach if current performance in a new cohort is not sufficient. Next, our choice of using the models to identify the 10% of the cohort with the fastest amyloid accumulation is necessarily arbitrary and meant to demonstrate the value of the GBDT method in a practical task. Other thresholds will yield different performance. Another limitation is that amyloid levels are not directly related to cognition, as patients can be amyloid positive without cognitive deficits. Other biomarkers may be more relevant for prediction, including tau and fluorodeoxyglucose PET and MRI-based cortical thickness in relevant subregions. Amyloid was chosen due to the fact that more than 2500 cases were available in ADNI, but we suspect that the current framework would enable the prediction of other biomarkers, given datasets of appropriate size. Finally, we recognize that the selected eight clinical/genetic features are somewhat arbitrary and were chosen because they were available for all or most cases and were felt by the investigators to be relevant to disease progression. Using these data along with the deep features from the baseline images, we found good performance for predicting amyloid accumulation. It is certainly possible that using different subsets or other metrics and models could perform even better, and this would be a fruitful path for future investigations.

5 | CONCLUSION

We trained a ML algorithm to combine deep image features with clinical, demographic, and genetic information to predict future changes in amyloid deposition. Practically, it was shown to be superior to several other methods of identifying patients to recognize fast progressors. This method is adaptable to study other important imaging biomarkers and to assess the effects of different treatments and may have advantages over models trained to predict clinical endpoints.

ACKNOWLEDGMENTS

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.

CONFLICTS OF INTEREST

F. H. R. discloses funding received from his PhD scholarship at CompCancer, paid through a working contract at Charité Berlin; funding received from his PhD advisor, Dagmar Kainmüller, at the Kainmüller lab for technical equipment; a patent application submitted on the methods described in this paper; and consulting fees from Subtle Medical, Inc. E. C. M. discloses grant funding received from the US National Institutes of Health, consulting fees from Eli Lilly and Hoffmann-La Roche. G.Z. discloses grant funding received from the US National Institutes of Health, Bracco Imaging, and GE Healthcare paid to his institution; royalties from Cambridge University Press paid to him; licenses on MRI and AI patents paid to his institution and to him; consulting fees from Biogen as part of an advisory board paid to him; honoraria for scientific lectures from multiple societies paid to him; expert testimony

from various law firms paid to him; and equity in Subtle Medical, Inc. He is a board member of ISMRM and president-elect of the ASFNR.

AUTHOR CONTRIBUTIONS

Fabian Reith: study conceptualization, study design, data collection, data curation, data analysis, data interpretation, deep learning conceptualization, deep learning design, deep learning analysis, statistical analysis, manuscript drafting and editing. Elisabeth Mormino: study design, data interpretation. Greg Zaharchuk: funding, study conceptualization, study design, data collection, data interpretation, manuscript drafting, and editing.

ORCID

Fabian H. Reith https://orcid.org/0000-0002-0745-2726
Greg Zaharchuk https://orcid.org/0000-0001-5781-8848

REFERENCES

1. Grill JD, Karlawish J. Addressing the challenges to successful recruitment and retention in Alzheimer's disease clinical trials. *Alzheimers Res Ther*. 2010;2:34.
2. Clement C, Selman LE, Kehoe PG, Howden B, Lane JA, Horwood J. Challenges to and facilitators of recruitment to an Alzheimer's disease clinical trial: a qualitative interview study. *J Alzheimers Dis*. 2019;69:1067-1075.
3. Ding Y, Sohn JH, Kawczynski MG, et al. A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brain. *Radiology*. 2019;290:456-464.
4. Hekler A, Utikal JS, Enk AH, et al. Pathologist-level classification of histopathological melanoma images with deep neural networks. *Eur J Cancer*. 2019;115:79-83.
5. Miotto R, Li L, Dudley JT. Deep Learning to Predict Patient Future Diseases from the Electronic Health Records. *Advances in Information Retrieval*. Springer International Publishing; 2016:768-774.
6. Yoo Y, Tang LW, Brosch T, et al. Deep Learning of Brain Lesion Patterns for Predicting Future Disease Activity in Patients with Early Symptoms of Multiple Sclerosis. *Deep Learning and Data Labeling for Medical Applications*. Springer International Publishing; 2016:86-94.
7. Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med*. 2018;79:3055-3071.
8. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. *Nature*. 2018;555:487-492.
9. Liu F, Jang H, Kijowski R, Bradshaw T, McMillan AB. Deep learning MR imaging-based attenuation correction for PET/MR imaging. *Radiology*. 2018;286:676-684.
10. Landau SM, Fero A, Baker SL, et al. Measurement of longitudinal β-amyloid change with 18F-florbetapir PET and standardized uptake value ratios. *J Nucl Med*. 2015;56:567-574.
11. Brendel M, Högenauer M, Delker A, et al. Improved longitudinal [(18)F]-AV45 amyloid PET by white matter reference and VOI-based partial volume effect correction. *Neuroimage*. 2015;108:450-459.
12. Chen K, Roontiva A, Thiyyagura P, et al. Improved power for characterizing longitudinal amyloid-β PET changes and evaluating amyloid-modifying treatments with a cerebral white matter reference region. *J Nucl Med*. 2015;56:560-566.
13. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189-198.
14. Pfeffer RI, Kurosaki TT, Harrah CH Jr, Chance JM, Filos S. Measurement of functional activities in older adults in the community. *J Gerontol*. 1982;37:323-329.
15. Liu C-C, Liu C-C, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol*. 2013;9:106-118.
16. Ellis RJ, Olichney JM, Thal LJ, et al. Cerebral amyloid angiopathy in the brains of patients with Alzheimer's disease: the CERAD experience, Part XV. *Neurology*. 1996;46:1592-1596.
17. Seabold S & Statsmodels PerktoldJ: Econometric and statistical modeling with Python. 9th Python in Science Conference. 2010.
18. Paszke A, Gross S, Chintala S, et al. Automatic differentiation in PyTorch. 2017 [cited 15 Jan 2019]. https://openreview.net/forum?id=BJJsrmfCZ
19. Reith F, Koran ME, Davidzon G, Zaharchuk G, & Alzheimer's Disease Neuroimaging Initiative. Application of deep learning to predict standardized uptake value ratio and amyloid status on 18F-Florbetapir PET using ADNI data. *AJNR Am J Neuroradiol*. 2020;41:980-986.
20. He K, Zhang X, Ren S, Sun J. *Deep Residual Learning for Image Recognition*. 2015. http://arxiv.org/abs/1512.03385. arXiv [cs.CV].
21. Kingma DP, Adam BaJ. *A Method for Stochastic Optimization*. 2014. http://arxiv.org/abs/1412.6980. arXiv [cs.LG].
22. Deng J, Dong W, Socher R, Li L-J, Li K, ImageNet Fei-FeiL, & ImageNet (Fei-Fei L). *A Large-Scale Hierarchical Image Database*. http://www.image-net.org/papers/imagenet_cvpr09.pdf.
23. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29:1189-1232.
24. Ke G, Meng Q, Finley T. LightGBM: a Highly Efficient Gradient Boosting Decision Tree. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al. eds. *Advances in Neural Information Processing Systems*. Curran Associates, Inc.; 2017:3146-3154.
25. Lu D, Popuri K, Ding GW, Balachandar R, Beg MF, Alzheimer's Disease Neuroimaging Initiative. Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease. *Med Image Anal*. 2018;46:26-34.
26. Singh S, Srivastava A, Mi L, et al. Deep learning based classification of FDG-PET data for Alzheimers disease categories. *Proc SPIE Int Soc Opt Eng*. 2017:10572.
27. Choi H, Jin KH, Alzheimer's Disease Neuroimaging Initiative. Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. *Behav Brain Res*. 2018;344:103-109.
28. Rabinovici GD, Gatsonis C, Apgar C, et al. Association of amyloid positron emission tomography with subsequent change in clinical management among medicare beneficiaries with mild cognitive impairment or dementia. *JAMA*. 2019;321:1286-1294.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Reith FH, Mormino EC, Zaharchuk G. Predicting future amyloid biomarkers in dementia patients with machine learning to improve clinical trial patient selection. *Alzheimer's Dement*. 2021;7:e12212. https://doi.org/10.1002/trc2.12212