DOI: 10.1002/alz.040990

BIOMARKERS

POSTER PRESENTATIONS

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

Biomarkers (non-neuroimaging) / Novel biomarkers

Is machine learning prediction of Aβ positivity consistent? An assessment of multiple datasets

Elisabetta Grecchi[1] | Enrico Grisan[2,3] | Christopher Buckley[1] | Jan Wolber[1]

[1] GE Healthcare, Amersham, United Kingdom

[2] University of Padova, Padova, Italy

[3] London South Bank University, London, United Kingdom

Correspondence
Elisabetta Grecchi, GE Healthcare, Amersham, United Kingdom.
Email: elisabetta.grecchi@ge.com

Abstract

Background: Recent applications of machine learning methodologies in AD suggest that synergistic utilisation of imaging and non-imaging biomarkers may improve the ability to predict subject's likelihood to present with amyloid pathology prior to performing a PET scan. In this work we developed an algorithm for prediction of patient amyloid positivity that shows robust performance across different databases, thus being more likely to be suitable for real-world application.

Method: Machine learning (ML) algorithms that combine imaging (MRI volumes), genetic (ApoE status), psychometric (MMSE and CDR), and demographic (age/gender) data were developed to predict a patient's probability of being Aβ positive. The patient's [^{18}F]flutemetamol PET scan served as the standard of truth (SoT). Two ML methodologies (LASSO and RUS-BOOST) were tested to tackle both the unbalance between Aβ positive/negative patients and selection of volumetric brain regions. The algorithms were trained and tested using combination of 5 different databases, excluding all healthy controls: MCI Progression Phase III trial (204 MCI), AIBL (52 MCI, 16 pAD), Biofinder (117 SCD, 147 AD), and a subset of ADNI data selected for the ADNI evolution prediction data (564 MCI, 147 AD).

Result: The cross-validation results and the performance of the algorithms with different test sets are reported in Tables 1 and 2. We report AUC, accuracy, sensitivity, specificity, PPV and NPV for a probability threshold of 0.5 (Table1). PPV and NPV are also reported for threshold optimized and selected by the algorithm to provide a sensitivity of approximately 0.75 (Table 2), which results in a lower rate of misclassified amyloid positive.

Conclusion: We show that the developed algorithms can be confidently used across several independently acquired datasets. Our results suggest that for optimal usage as screening tool in a specific clinical trial, ML techniques should be adjusted for the characteristic of the specific population under analysis. Despite fairly consistent performance of the two methods within cross validation, results can vary when applying the learned models to different datasets, suggesting that cohort selection criteria, composition, and geographical origin may additionally influence outcomes.

TABLE 1

		Accuracy				Sensitivity				Specificity			
		LASSO		RUSBost		LASSO		RUSBoost		LASSO		RUSBoost	
		Internal 5-fold cross validation	External cross validation	Internal 5-fold cross validation	External cross validation	Internal 5-fold cross validation	External cross validation	Internal 5-fold cross validation	External cross validation	Internal 5-fold cross validation	External cross validation	Internal 5-fold cross validation	External cross validation
TRAINING DATASET	PhaseIII trial A	0.69 (0.08)	0.69 (0.09)	0.63 (0.09)	0.65 (0.03)	0.27 (0.22)	0.58 (0.09)	0.55 (0.21)	0.61 (0.07)	0.90 (0.07)	0.78 (0.26)	0.67 (0.09)	0.68 (0.08)
	AIBL	0.76 (0.09)	0.55 (0.1)	0.72 (0.08)	0.65 (0.06)	0.94 (0.08)	0.97 (0.03)	0.80 (0.13)	0.68 (0.23)	0.30 (0.30)	0.15 (0.13)	0.47 (0.38)	0.6 (0.24)
	Biofinder	0.80 (0.04)	0.76 (0.13)	0.75 (0.06)	0.70 (0.12)	0.51 (0.08)	0.73 (0.15)	0.60 (0.16)	0.63 (0.16)	0.92 (0.03)	0.74 (0.27)	0.81 (0.06)	0.75 (0.13)
	ADNI	0.68 (0.05)	0.50 (0.21)	0.68 (0.03)	0.67 (0.07)	0.74 (0.03)	0.99 (0.01)	0.68 (0.03)	0.79 (0.08)	0.58 (0.11)	0.01 (0.02)	0.68 (0.05)	0.55 (0.08)
	PhaseIII trial A + AIBL	0.68 (0.04)	0.75 (0.14)	0.66 (0.05)	0.70 (0.13)	0.57 (0.04)	0.67 (0.21)	0.59 (0.07)	0.56 (0.17)	0.76 (0.05)	0.81 (0.06)	0.71 (0.08)	0.86 (0.12)

TABLE 2

					Accuracy				Sensitivity				Specificity			
					LASSO		RUSBost		LASSO		RUSBoost		LASSO		RUSBoost	
			LASSO	RUSBost	Internal 5-fold cross validation	External cross validation	Internal 5-fold cross validation	External cross validation	Internal 5-fold cross validation	External cross validation	Internal 5-fold cross validation	External cross validation	Internal 5-fold cross validation	External cross validation	Internal 5-fold cross validation	External cross validation
TRAINING DATASET	PhaseIII trial A	TRESHOLD	0.22	0.40	0.50 (0.11)	0.63 (0.14)	0.55 (0.10)	0.67 (0.11)	0.74 (0.01)	0.94 (0.03)	0.74 (0.01)	0.81 (0.1)	0.39 (0.16)	0.22 (0.19)	0.45 (0.16)	0.51 (0.13)
	AIBL		0.66	0.47	0.64 (0.05)	0.61 (0.05)	0.66 (0.05)	0.64 (0.05)	0.72 (0.02)	0.80 (0.12)	0.72 (0.02)	0.72 (0.2)	0.43 (0.15)	0.37 (0.23)	0.47 (0.30)	0.56 (0.22)
	Biofinder		0.2	0.40	0.69 (0.06)	0.68 (0.09)	0.63 (0.10)	0.65 (0.07)	0.79 (0.01)	0.89 (0.04)	0.79 (0.01)	0.88 (0.1)	0.65 (0.08)	0.39 (0.2)	0.57 (0.15)	0.32 (0.23)
	ADNI		0.45	0.43	0.67 (0.03)	0.49 (0.21)	0.66 (0.05)	0.61 (0.19)	0.78 (0.02)	0.99 (0.01)	0.80 (0.00)	0.93 (0.02)	0.49 (0.08)	0.01 (0.01)	0.46 (0.12)	0.34 (0.12)
	PhaseIII trial A + AIBL		0.32	0.33	0.66 (0.09)	0.7 (0.09)	0.60 (0.08)	0.62 (0.08)	0.78 (0.01)	0.83 (0.06)	0.78 (0.01)	0.88 (0.06)	0.56 (0.15)	0.55 (0.13)	0.46 (0.14)	0.32 (0.09)