Published in final edited form as:

Alzheimer Dis Assoc Disord. 2022 ; 36(2): 176–179. doi:10.1097/WAD.0000000000000474.

Comparing performance of different predictive models in estimating disease progression in Alzheimer's disease.

Ali Ezzati, MD[1,2], **Andrea R. Zammit, PhD**[1], **Richard B. Lipton, MD**[1,2]

[1]Department of Neurology, Albert Einstein College of Medicine, Bronx, NY, USA.

[2]Department of Neurology, Montefiore Medical Center, Bronx, NY, USA.

Abstract

Background: Automatic classification techniques provide tools to analyze complex data and predict disease progression.

Methods: A total of 305 cognitively normal (CN); 475 patients with amnestic mild cognitive impairment (aMCI); and 162 patients with dementia were included in this study. We compared performance of 3 different methods in predicting progression from aMCI to dementia: 1) Index-based model; 2) Logistic regression (LR); and 3) ensemble linear discriminant (ELD) machine learning (ML) models. LR and ELD models were trained using data from CN and dementia subgroups, and subsequently were applied to aMCI subgroup to predict their disease progression.

Results: Performance of ELD models were better than LR models in prediction of conversion from aMCI to AD at all time-frames. ELD models performed better when a larger number of features were used for prediction.

Conclusions: ML models have substantial potential to improve predictive ability for cognitive outcomes.

Keywords

Alzheimer's disease; MCI; dementia; Machine learning; predictive analytics

1. Introduction

Identifying individuals who will progress to Alzheimer's dementia (AD) over specified time intervals with high probability has been difficult. Quantitative risk prediction for AD using structured data sources and classical statistical methods have been available for many years[1]. Over the last decade and with the rapid growth of information on individuals' health and the increased availability of biomarkers, investigators are shifting toward using advanced

Corresponding Author: Ali Ezzati, Albert Einstein College of Medicine, 1225 Morris Park Avenue, Bronx, NY 10461, Tel: 718.430.3885, Fax: 718.430.3870, ali.ezzati@einsteinmed.org.

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

multivariate methods such as clustering, latent class analysis (LCA)[2] and machine learning (ML)[3] to improve classification and quantitative risk prediction of cognitive decline[4].

Despite the potential advantages of ML techniques, their use in clinical AD research has remained limited. This is in part due to doubts over efficiency of these models and lack of head to head comparison with conventional methods. In this study, we aimed to compare the performance of different models in predicting progression from amnestic mild cognitive impairment (aMCI) to AD using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI).

2. Methods

2.1. Study design and participants

Data used for this analysis were obtained from the ADNI database (www.adni.loni.usc.edu). The ADNI is an ongoing cohort, which was launched in 2003 as a public–private partnership. The individuals included in the current study were initially recruited as part of ADNI-1, ADNI-GO, and ADNI-2. This study was approved by the IRB of all participating institutions. Informed written consent was obtained from all participants at each site.

Eligible participants for this study had neuropsychiatric measures, CSF biomarkers and MRI measures at the baseline visit, and at least 6 months of follow up. Considering that missing data can affect performance of classifiers, making head-to-head comparison of them impractical, cases with missing data were deleted listwise and not considered for further analysis. Participants. A total of 942 participants met the criteria for inclusion in this study. Table, Supplemental Digital Content 1 summarizes characteristic of participant.

2.2. Study measures (features)

Study features included cognitive test scores, APOE4 gene status, volumetric MRIs and CSF biomarkers. For details about study measures see Table, Supplemental Digital Content 2.

2.3. Study Feature-sets

The following sets of features were used in the statistical analysis:

- Feature-set-1 (5 features) was adapted based on the study by Steeland et al[5]: ApoE4 status, memory summary score, Functional Activities Questionnaire (FAQ) score, hippocampal volume, and CSF Tau/Aβ ratio.
- Feature-set-2 (14 features) was an expansion of variables included in feature-set 1: Demographics (age, gender, education), ApoE4 status, composite scores for the memory and executive function domains, CSF biomarkers ($A\beta_{1-42}$, Tau, P-tau$_{181p}$), volumetric MRI measures (hippocampus, entorhinal cortex, middle temporal lobe, fusiform gyrus, and whole brain volume).

Since data normalization is considered to be essential for improving performance of ML models, continuous measures were normalized using standardized scores.

2.4. Data analysis

Three different methods were used for prediction of outcomes:

Method 1: Prediction using an established index score. *Using feature set 1 and b*ased on the algorithm suggested by Steeland et al[5], we created an index score and assigned aMCI participants into 2 groups: Dementia-like, those who had all the harmful risk factors (APOE4 positive, FAQ>0.5, memory summary score <0.26, hippocampal volume<6696mm^3, Tau/Aβ ratio≥0.43 (Dementia-like), and CN-like, those who were negative for at least one of these risk factors. CN-like, were those individuals who did not meet the criteria mentioned above.

Method 2: Conventional multivariate statistical modeling. The traditional approach to develop clinical risk prediction models involves the use of regression models, such as logistic regression (LR) to predict outcomes.

Method 3: Ensemble Linear Discriminant (ELD) models. While many ML models have proven to be effective tools for predictions of outcomes, in a previous study[4] we showed that ELD perform exceptionally well in ADNI data. Therefore, ELD model was selected for this study.

ELD is among the family of classification methods known as ensemble learning, in which the output of an ensemble of simple and low-accuracy classifiers trained on subsets of features are combined (e.g., by weighted average of the individual decisions), so that the resulting ensemble decision rule has a higher accuracy than that obtained by each of the individual classifiers. Linear Discriminant Analysis (LDA) was proposed by R. Fischer in 1936.[6] It consists of finding the projection hyperplane that minimizes the interclass variance and maximizes the distance between the projected means of the classes. Similar to PCA, these two objectives can be solved by solving an eigenvalue problem with the corresponding eigenvector defining the hyperplane of interest. This hyperplane can be used for classification. In this work, we combined linear discriminant functions (i.e., hyperplanes that dichotomize the samples based on subsets of features) to construct the ensemble classifier.

For both method data-driven methods (LR, and ELD), baseline data was used to train models to learn classifying cognitively normal individuals from individuals with dementia. Subsequently, for validation, baseline data from MCI participants were used to classify them into those more similar to cognitively normal individuals (CN-like), or those who were more similar to individuals with dementia (Dementia-like). The accuracy of the predicted outcomes (CN-like or Dementia-like) for aMCI population for each of the methods described above was evaluated using the available clinical outcomes from the longitudinal follow-up data. Considering change in proportion of MCI subgroups over time (due to drop outs, death, etc.), the accuracy is reported separately for each wave of follow-up at 6, 12, 24, 36, and 48 months. Furthermore, we computed sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the models for predicting conversion to Dementia at each follow-up time-point.

For simplicity, we refer to LR and ELD model applied to feature-set 1 and feature-set 2 as LR1/ELD1 and LR2/ELD2, respectively.

Comparison of classification performance.—We used the McNemar test to compare the performance of classification models.[7]

RESULTS

2.5. Demographics and baseline characteristics

Participants' characteristics are summarized in Table, Supplemental Digital Content 1. Follow-up data was available for 475 aMCI cases at 6 months, 463 cases at 1y, 403 cases at 2y, 345 cases at 3y, and 238 cases at 4y. The cumulative proportions of individuals who progressed from aMCI to dementia at 6m, 1y, 2y, 3y, and 4y were 6.3%, 14.5%, 25.8%, 28.1%, and 31.5%, respectively.

2.6. Performance of different methods in predicting conversion from aMCI to AD (table 1)

Method 1 (index-based): Using the developed index score yielded predictions with poor sensitivity, declining from 46.6% at 6m to 26.6% at 48m, but high specificity, increasing from 87.6 at 6m to 99.4 at 48m.

Method 2 (Logistic Regression): Relative importance of predictors for LR1 and LR2 models are summarized in Table, Supplemental Digital Content 3. LR1 showed high sensitivity, declining from 93.3 at 6m to 77.3 at 48m, while specificity increased from 60.0% to 85.2%. LR2, had lower sensitivity (ranging from 53.3% at 6m to 34.6%) and higher specificity (ranging from 83.1% at 6m to 94.5% at 48m) in comparison with LR1.

Method 3 (Ensemble Linear Discriminant): Prediction sensitivities for conversion from aMCI to AD for ELD1at 6, 12, 24, 36, and 48 months were 83.3, 76.1, 68.2, 65.9, and 58.6, respectively. Specificities for this model at 6, 12, 24, 36, and 48 months were 72.3, 77.3, 84.9, 90.3, and 93.8, respectively. ELD2 had slightly higher sensitivities, ranging from 86.6 at 6m to 68.0 at 48m, and lower specificity, ranging from 64.4 at 6m to 90.1 at 48m, in comparison with ELD2.

2.7. Comparison of performance of all three methods.

Next, we compared head-to-head performance of different models based on correct classification using McNemar test (see Table, Supplemental Digital Content 4). Both LR1 and ELD1 performed better than method 1 (index score) in prediction of disease progression. LR models performed better when using feature-set 1 in comparison with feature-set 2 ($p<0.001$ for prediction at all follow-up periods). However, ELD models using feature-set 2 outperformed ELD models using feature-set 1 at all time-frames ($p<0.001$). ELD models persistently and across all time frames performed better than LR models.

Discussion

In this study, we showed that classifiers such as logistic regression and ELD models are effective tools for prediction of disease progression in patients with aMCI. These multivariate models outperform index scores created based on individual variable cutoffs. Owing to its flexibility and easier handling of a larger number of potential predictors, ML is claimed to outperform traditional statistical modeling[8], however, a recent systemic review of literature[9] showed that logistic regression models often perform remarkably well[9]. To compare these two approaches, we ran all models in the same data set using two feature set: a smaller, previously selected feature set, and a larger feature set and endeavored to make the comparison as fair as possible. While performance of ELD models were better than LR models based on some performance metrics, these differences were not always consistent or significant. The observed differences in performance might be due to several factors such as model characteristics, signal-to-noise ratios, ratio of features to sample size and how models handle it. Testing this hypothesis rigorously may require independent sample validation.

We showed that performance of LR models decreased when a larger feature-set was used. This is likely to be due to a few factors: The key to developing a high performance LR model is to choose the correct variables to enter into the model. While it is tempting to include as many features as possible, this can negatively impact effect of true predictors on prediction performance and lead to large standard errors with wide and imprecise confidence intervals, or, conversely, lead to identifying false predictors. Prior studies indicate that if input variables are highly correlated with one another (also known as multicollinearity), then the effect of each on the regression model becomes less precise.[10] it is recommended that when a pair of variables are highly correlated (i.e., correlation coefficient of >0.8), one of them should be removed from the model. None of the variables used in our models correlated at that level, however some of our input variables such as volumetric MRI measures had moderate correlations with each other (correlation coefficient rang: 0.3–0.5), which might negatively affect performance of LR models. Furthermore, we showed that performance of ML models improved with increased number of predictors. This is in line with prior studies that have suggested that ML models might perform better than LR models when complexity and number of predictors are increased.[11]

One strength of this study was having separate training and test samples. Another strength was using and comparing three different methods for prediction of disease progress. However, a few limitations should be noted. Variables included in feature-set 1 were selected based on knowledge from another study, but they were not specifically selected and tuned to be used as classifiers. No formal feature-selection was used for feature-set 2. This approach can negatively affect performance of predictive models, specially LR models. For the purpose of simplicity, we only compared performance of LR models with one ML models, while other models might have incremental value in progression-prediction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This work was supported by grants from the National Institute of Health (NIA K23 AG063993; Ezzati; NIA K01AG054700, Zammit, NIA P01 AG03949, Lipton), the Alzheimer’s Association (Ezzati, 2019-AACSF-641329), the Cure Alzheimer’s Fund (Ezzati, Lipton), and the Leonard and Sylvia Marx Foundation (Lipton).

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

Conflict of interest:

R.B.L. receives research support from the following sources unrelated to this manuscript: NIH: 2PO1 AG003949 (mPI), 5U10 NS077308 (PI), R21 AG056920 (Investigator), 1RF1 AG057531 (Site PI), RF1 AG054548 (Investigator), 1RO1 AG048642 (Investigator), R56 AG057548 (Investigator), U01062370 (Investigator), RO1 AG060933 (Investigator), RO1 AG062622 (Investigator), 1UG3FD006795 (mPI), 1U24NS113847 (Investigator), K23 NS09610 (Mentor), K23AG049466 (Mentor), K23 NS107643 (Mentor). He also receives support from the Migraine Research Foundation and the National Headache Foundation. He serves on the editorial board of Neurology, senior advisor to Headache, and associate editor to Cephalalgia. He has reviewed for the NIA and NINDS, holds stock options in eNeura Therapeutics and Biohaven Holdings; serves as consultant, advisory board member, or has received honoraria from: Abbvie (Allergan), American Academy of Neurology, American Headache Society, Amgen, Avanir, Biohaven, Biovision, Boston Scientific, Dr. Reddy's (Promius), Electrocore, Eli Lilly, eNeura Therapeutics, Equinox, GlaxoSmithKline, Grifols, Lundbeck (Alder), Merck, Pernix, Pfizer, Supernus, Teva, Trigemina, Vector, Vedanta. He receives royalties from Wolff's Headache 7th and 8th Edition, Oxford Press University, 2009, Wiley and Informa.

References

1. Weiner MW, Veitch DP, Aisen PS, et al. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimer's & Dementia 2013;9:e111–e194.
2. Zammit AR, Muniz-Terrera G, Katz MJ, et al. Subtypes based on neuropsychological performance predict incident dementia: Findings from the Rush Memory and Aging Project. Journal of Alzheimer's Disease 2018:1–11.
3. Ezzati A, Lipton RB. Machine Learning Predictive Models Can Improve Efficacy of Clinical Trials for Alzheimer's Disease 1, 2. Journal of Alzheimer's Disease 2020:1–9.
4. Ezzati A, Zammit AR, Harvey DJ, et al. Optimizing machine learning methods to improve predictive models of Alzheimer's disease. Journal of Alzheimer's Disease 2019;71:1027–1036.
5. Steenland K, Zhao L, John SE, et al. A 'Framingham-like'Algorithm for Predicting 4-Year Risk of Progression to Amnestic Mild Cognitive Impairment or Alzheimer's Disease Using Multidomain Information. Journal of Alzheimer's Disease 2018;63:1383–1393.
6. Xanthopoulos P, Pardalos PM, Trafalis TB. Linear discriminant analysis. Robust data mining: Springer, 2013: 27–33.
7. Dietterich TG. Approximate statistical tests for comparing supervised classification learning algorithms. Neural computation 1998;10:1895–1923. [PubMed: 9744903]
8. Beam AL, Kohane IS. Big data and machine learning in health care. Jama 2018;319:1317–1318. [PubMed: 29532063]
9. Jie M, Collins GS, Steyerberg EW, Verbakel JY, van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. Journal of clinical epidemiology 2019.
10. Ranganathan P, Pramesh C, Aggarwal R. Common pitfalls in statistical analysis: logistic regression. Perspectives in clinical research 2017;8:148. [PubMed: 28828311]

11. Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. Journal of medical Internet research 2016;18:e323. [PubMed: 27986644]

Table 1.

Performance of different models in prediction of progression from aMCI to dementia.

Feature-set	Model	Follow-up, years	Sensitivity, %, 95% CI	Specificity, %, 95% CI	PPV, %, 95% CI	NPV, %, 95% CI	Accuracy, %, 95% CI	AUC	Progression rate, %[a]	N
1	Index	0.5	46.6	87.6	20.2	96.0	85.0	0.67	6.3	475
		1	41.7	90.4	42.4	90.1	83.3	0.66	14.5	463
		2	32.6	94.9	69.3	80.2	78.9	0.64	25.8	403
		3	28.8	97.5	82.3	77.8	78.2	0.63	28.1	345
		4	26.6	99.3	95.2	74.6	76.4	0.63	31.5	238
	LR	0.5	93.3 (77.9–99.1)	60.0 (55.3–64.5)	13.5 (11.9–15.4)	99.2 (97.2–99.8)	62.1 (57.6–66.4)	0.77	6.3	475
		1	88.0 (77.8–94.7)	65.1 (60.2–69.8)	29.9 (26.7–33.4)	96.9 (94.4–98.4)	68.4 (64.0–72.7)	0.77	14.5	463
		2	80.7 (71.8–87.8)	73.9 (68.5–78.8)	51.8 (46.5–57.1)	91.7 (88.1–94.3)	75.6 (71.2–79.8)	0.77	25.8	403
		3	76.2 (66.6–84.3)	79.0 (73.4–83.9)	58.7 (52.1–65.0)	89.5 (85.5–92.4)	78.2 (73.5–82.5)	0.78	28.1	345
		4	77.3 (66.2–86.2)	85.2 (78.9–90.3)	70.7 (62.1–78.1)	89.1 (84.3–92.6)	82.7 (77.4–87.3)	0.81	31.5	238
	ELD	0.5	83.3 (65.3–94.4)	72.3 (68.0–76.4)	16.8 (14.0–20.2)	98.4 (96.6–99.3)	73.0 (68.8–77.0)	0.78	6.3	475
		1	76.1 (64.1–85.7)	77.2 (72.8–81.3)	36.1 (31.1–41.5)	95.0 (92.5–96.7)	77.1 (73.0–80.9)	0.77	14.5	463
		2	68.2 (58.4–77.0)	84.9 (80.4–88.8)	61.2 (53.9–68.0)	88.5 (85.3–91.1)	80.6 (76.4–84.4)	0.77	25.8	403
		3	65.9 (55.6–75.3)	90.3 (85.9–93.7)	72.7 (64.0–80.0)	87.1 (83.7–90.0)	83.4 (79.1–87.2)	0.78	28.1	345
		4	58.6 (46.7–69.9)	93.8 (89.0–97.0)	81.4 (70.0–89.2)	83.1 (79.0–86.6)	82.7 (77.3–87.3)	0.76	31.5	238
2	LR	0.5	53.3 (34.3–71.6)	83.1 (79.3–86.5)	17.5 (12.5–24.0)	96.3 (94.7–97.4)	81.2 (77.4–84.7)	0.68	6.3	475
		1	35.8 (24.5–48.5)	83.5 (79.6–87.1)	26.9 (20.0–35.3)	88.5 (86.5–90.2)	76.6 (72.5–80.4)	0.60	14.5	463
		2	38.4 (29.0–48.5)	90.3 (86.4–93.4)	57.9 (47.5–67.8)	80.8 (78.3–83.1)	76.9 (72.5–81.0)	0.64	25.8	403
		3	38.1 (28.5–48.5)	93.5 (89.7–96.2)	69.8 (57.5–79.8)	79.4 (76.7–81.9)	77.9 (73.2–82.2)	0.66	28.1	345
		4	34.6 (24.0–46.5)	94.4 (89.8–97.4)	74.2 (25.6–37.8)	75.8 (58.8–85.4)	75.6 (72.6–78.8)	0.65	31.5	238

Feature-set	Model	Follow-up, years	Sensitivity, %, 95% CI	Specificity, %, 95% CI	PPV, %, 95% CI	NPV, %, 95% CI	Accuracy, %, 95% CI	AUC	Progression rate, %[a]	N
	ELD	0.5	86.6 (69.3–96.2)	64.4 (59.8–68.9)	14.1 (12.0–16.6)	98.6 (96.6–99.4)	65.8 (61.4–70.1)	0.76	6.3	475
		1	85.0 (74.2–92.6)	69.1 (64.4–73.7)	31.8 (28.1–35.8)	96.4 (93.9–98.0)	71.4 (67.1–75.6)	0.77	14.5	463
		2	80.7 (71.9–87.8)	78.6 (73.5–83.1)	56.7 (21.6+30.4)	92.1 (88.7–94.6)	79.1 (74.9–83.0)	0.80	25.8	403
		3	79.3 (70.0–86.9)	86.2 (81.4–90.3)	69.3 (62.0–75.9)	91.4 (87.8–94.0)	84.3 (80.1–88.0)	0.83	28.1	345
		4	68.0 (56.2–78.3)	90.1 (84.5–94.3)	76.1 (66.1–83.9)	85.9 (81.4–89.5)	83.1 (77.8–87.7)	0.79	31.5	238

Note.

[a] Represents progression rate from aMCI to AD at each follow up timeframe based on longitudinal data, which is included in the table for the purpose of comparison with PPV/NPV derived from the models. See texts for details on variables included in each feature-set.

Abbreviations: N= Number of participants at each follow up; LR= logistic regression; ELD= ensemble linear discriminant; CI= confidence interval; PPV= positive predictive value; NPV = negative predictive value.