JGG 2021;69:137-146
doi: 10.36150/2499-6564-N311

Comparison of neural network and logistic regression for dementia prediction: results from the PREADViSE trial

Xiuhua Ding [1], Frederick Schmitt [3,4], Richard Kryscio [2,3,5], Richard Charnigo [2,5]

[1] Western Kentucky University, Department of Public Health, Bowling Green, Kentucky; [2] University of Kentucky, Department of Statistics, Lexington, Kentucky; [3] University of Kentucky, Sanders-Brown Center on Aging, Lexington, Kentucky; [4] University of Kentucky, Department of Neurology, Lexington, Kentucky; [5] University of Kentucky, Department of Biostatistics, Lexington, Kentucky

Objective. Two systematic reviews suggest that current parametric predictive models are not recommended for use in population dementia diagnostic screening. This study was to compare predictive performance between logistic regression (conventional method) and neural network (non-conventional method).
Method. Neural network analysis was performed through the R package "Neuralnet" by using the same covariates as the logistic regression model. Results. Results show that neural network had a slightly apparently better predictive performance (area under curve (AUC): 0.732 neural network vs. 0.725 logistic regression). Neural network performed similarly as logistic regression. Furthermore, logistic regression confirmed that the interaction effect among covariates, which elucidated from neural network.
Conclusions. Neural network performed slightly apparently better than logistic regression, and it is able to elucidate complicated relationships among covariates.

Key words: prediction, neural network, logistic regression, dementia

Received: December 24, 2020
Accepted: February 15, 2021

Correspondence
Xiuhua Ding
Department of Public Health, Western Kentucky University, 1906 College Heights Blvd, Bowling Green, KY 42101. Tel. 270-745-3618. Fax 270-745-4437. E-mail: xiuhua.ding@wku.edu

Conflict of interest
The Authors declare no conflict of interest

How to cite this article: Ding X, Schmitt F, Kryscio R, et al. Comparison of neural network and logistic regression for dementia prediction: results from the PREADViSE trial. Journal of Gerontology and Geriatrics 2021;69:137-146. https://doi.org/10.36150/2499-6564-N311

INTRODUCTION

Rising prevalence of dementia has become a major public health concern as disability associated with dementia, especially at the late stage, leads to high personal, social and economic costs. Early identification of individuals with high risk for dementia is of great importance for dementia prevention, especially when interventions are developed. To identify these high-risk individuals as early as possible, effective predictive or prognostic models with risk factors become a research priority. So far, many studies have attempted to find a useful prediction model.
Conventional prediction models have been predominantly developed from logistic regression [1-4] or proportional hazards regression analysis [5-10]. For non- conventional models, the classification tree is the most often used method [11,12]. Alternative approaches, also include non-conventional statistical learning methods such as random forest [13] and neural network analyses [14]. Covariates used in the majority of predictive modeling studies include demographic variables, such as age, education, body mass index (BMI), medical comorbidity (e.g., history of cardiovascular disease)

or cognitive tests. Recently, studies have incorporated genetic risk factors and imaging data into predictive models [7,15]. However, some have argued that genetic risk factors and neuroimaging variables have not significantly increased discriminative accuracy [15]. Further, these data are often difficult and expensive to obtain [16], while evidence shows that a third of Alzheimer's disease (AD) cases worldwide may be due to modifiable risk factors [17].

Tang et al. conducted an updated systematic review based on their 2010 review to evaluate latest development on methods for prediction of dementia risk [18,19]. The updated systematic review concluded that despite the significant increase in the number of risk modeling studies, the predictive accuracy of these parametric models has not changed to a significant degree after 5 years (range 0.49-0.91 in 2010 review, and 0.49-0.89 in the 2015 review), and none of the methods are recommended for dementia risk prediction in the population setting due to insufficient consideration of sample selection, model diagnostics, and model validation [18,19]. Most of risk models included in the reviews were developed from conventional methods including logistic regression or COX proportional hazard model, and none of risk models was created by using neural network. In this study, we sought to compare the predictive performance between neural network and logistic regression using mainly mental status and self-reported data from the Prevention of Alzheimer's Disease with Vitamin E and Selenium (PREADViSE) trial with a known AD genetic risk (APOE genotype) and clinical diagnosis of dementia to compare predictive models.

METHODS

Study sample and data sources

The PREADViSE trial was an ancillary study to the Selenium and Vitamin E Cancer Prevention Trial (SELECT) (a large prostate cancer prevention randomized controlled trial (RCT)) [20] and was designed to evaluate the effectiveness of antioxidant supplements vitamin E and selenium in preventing incident AD and other forms of dementia. During the recruiting period (2002 to 2009) PREADViSE enrolled 7,547 non-demented male participants age 62 years and older (60+ if African American) from 128 participating SELECT clinical sites in the US, Canada, and Puerto Rico. The eligibility criteria for participating in PREADViSE included active SELECT enrollment at a participating site, and absence of dementia and other active neurologic conditions that affect cognition such as major psychiatric disorder, including depression.

The SELECT study supplements were discontinued by its Data Safety Monitoring Committee in 2008 following a futility analysis on its primary endpoint of prostate cancer incidence [21], and then participants in PREADViSE and SELECT were invited to continue as participants in observational cohort studies, and 4,271 of 7,547 original PREADViSE volunteers consented to continue participation. In order to maximize the consistency and completeness of follow-up, only participants who were screened in both the RCT and exposure phases of PREADViSE are included in the current study (n = 3784). PREADViSE was approved by the University of Kentucky Institutional Review Board (IRB) as well as the IRBs at each SELECT study site. Each participant provided written informed consent.

Mental status screening

The Memory Impairment Screen (MIS) [22] was used as the primary screening instrument for memory impairment in both the RCT and observational portions of PREADViSE. The MIS was given annually. If participants failed the MIS (that is, the participant scored 5 or less out of 8 on either the immediate or delayed recall portion of the MIS), a second tier screen was administered. An expanded Consortium to Establish a Registry in AD battery (CERAD-e) [23] was used during the RCT period and the modified Telephone Interview for Cognitive Status (TICS-m) [24], was used during the observational study. Both the CERAD-e and the TICS-m assess participants' global cognitive function. Failure on the secondary screen (T score $\leq$ 35 on CERAD-e battery or total score $\leq$ 35 on TICS-m) would lead to a recommendation for a clinic visit with their local physician. Records from the clinic visit were reviewed by 3-5 expert clinicians, including two neurologists and at least one neuropsychologist, for a consensus diagnosis. In cases where the neurologists disagreed in their diagnoses, the study PI made the final determination. Annual screenings were completed in May 2014, and a small number of participants were followed for medical records through August 2015.

Covariates

APOE genotype was obtained for 3,681 participants (ε2/2: 26 (0.71%); ε2/3: 459 (12.47%); ε2/4: 86 (2.34%); ε3/3: 2,240 (60.85%); ε3/4: 808(21.95%); ε4/4: 62(1.68%)). These genotypes were converted to a dummy indicator for at least one ε4 allele, where the presence of at least one ε4 allele was considered a carrier. For 103 subjects without APOE information, SAS 9.4® procedure PROC MI was used to impute missing values for the indicator variable based on family history of dementia. Four imputed data sets were generated; participants with two or more positive imputations for

APOE ε4 were coded as *APOE* ε4 positive. *APOE* ε4 positivity is a major risk factor for AD-type dementia [25]. Other data collected included age at baseline, race, BMI, years of education, as well as self-reported indicators of cardiovascular disease (i.e., diabetes, hypertension, and smoking), coronary artery bypass graft (CABG), congestive heart failure, hypertensive medication, and memory change at the baseline. These data were obtained at enrollment and annually thereafter as recognized risk factors for dementia [26]. History of significant cognitive or motor impairment due to stroke was an exclusion criterion, thus stroke was not considered in the models.

Case ascertainment

To create a predictive model, we used clinical dementia status (dementia vs. non- dementia) at end of follow-up as the outcome in this study. Dementia cases were identified through two methods. First, as described above, a medical records-based consensus diagnosis was used. Date of diagnosis was assigned as the date of the failed screen. Second, because many participants were reluctant to obtain medical workups for their memory, additional longitudinal measures including the AD8 Dementia Screening Interview [27], self-reported medical history, self-reported diagnosis of dementia, use of memory enhancing prescription drug, and cognitive scores including the MIS, CERAD-e Score, NYU Paragraph Delayed Recall, and TICS-m were used to identify cases. The diagnostic criteria for the second method were AD8 total of ≥ 1 (at any time during follow-up) to indicate functional impairment [27] plus one of the following: a self-reported diagnosis of dementia, use of a memory enhancing prescription drug (donepezil, rivastigmine, galantamine, or memantine), or cognitive score below cutoffs for intact cognition on any test (for example: 1.5 SDs below expected performance based on age and education normative data [28]. The date of diagnosis was assigned to the earliest event.

Data analysis

Chi-square and t-test statistics were used to examine differences in categorical and continuous variables between dementia groups except for congestive heart disease (Fisher's exact test was used). Univariate logistic regressions were performed first, and only those variables significantly associated with probability of dementia at univariate analysis were included in the multivariable logistic regression. Covariates included in the initial multivariable logistic regression model were age, education, smoking, *APOE*-ε4 allele status (any vs. none), history of hypertension, diabetes, coronary artery bypass graft (CABG), antihypertensive medication use, and memory change. In the model, age and education were used as continuous variables, and the rest were binary variables (yes vs. no). Logistic regression with backward elimination method was performed to compare with the neural network. Covariates age, education, *APOE*-ε4, and self-reported memory change were left in the final logistic regression model without interaction terms. Then neural network elucidated an interaction effect among years of education, age at baseline, status of memory change, and *APOE*-ε4 allele status. Logistic regression was conducted again to confirm the interaction effects.

Neural network

As an extension of generalized linear models (GLM), artificial neural network (ANN) was applied to explore the complex relationship between covariates and response [29]. Unlike GLM, ANN does not need to specify the form of the relationship between covariates and response variables. In this study, ANN was performed in R package "Neuralnet" under R (version 3.1.2), details about Neuralnet can be found in Gunther's paper [30]. Multilayer perceptron (MLP) [29,31] is the main model for neural network, which consists of vertices and directed edges called neurons and synapses respectively. Neurons are organized as layers and connected by synapses. Our ANN model had three neuron layers: input, hidden, output (See Figure 1). The input layer included all covariates in separate neurons, and the output layer consisted of the response variable (output). The layers between input and output layers are referred as hidden layers because they are not observed. For each synapse, a weight is assigned to indicate the effect of the corresponding neuron. Based on Fristsch, then "all data will pass through the neural network as signals, and these incoming signals will be first processed by the integration function, and then by activation function to transform as output of the neuron" [29]. According to Hornik[32], one hidden layer is sufficient to fit any real-valued continuous function [29,32].

Then supervised learning is applied in which desired output (or given output) is defined and is compared to the predicted output. Weights are also chosen at this stage [29,33]. The starting weights are usually assigned randomly from the standard normal distribution. To fit the neural network, the following steps are repeated [29]:

1 neural network calculates an output for given inputs and starting weights;
2 an error function such as sum of squared errors (SSE) or the cross-entropy will be applied to measure the difference between the actual output and predicted output;
3 then all weights are adapted based on the rule of a learning algorithm;

4 the process will stop if the pre-specified criterion (rule of a learning algorithm) is reached, for example, all absolute partial derivatives of the error function with respect to the weights are smaller than a given threshold or a specified maximum step is reached [29]. The resilient backpropagation algorithm (rprop+) is the most commonly used learning algorithm [34]. Weights are modified by searching in the opposite direction of the partial derivatives until a local minimum is found [29]. Additional technical details about ANN can be found in Gunther's technical report and Quintana's paper [29,35].

Our ANN input layer included four covariates including age, education, *APOE*-ε4, and self-reported memory change, in order to be directly comparable to the logistic regression model. We decided to have 10 hidden units based on the consideration of the sample size. The output layer had one neuron, which was dementia status at end of follow-up. Logistic regression and identity function were used as the activation function and integration function, respectively. The "rprop+" algorithm was used to determine the weights. AUC was calculated to compare the performance between logistic regression and ANN on classification of dementia status.

Descriptive analysis and logistic regression were conducted by using SAS 9.4® (SAS Institute, Inc., Cary, NC). Statistical significance was set at p = 0.05.

RESULTS

Table I presents the general characteristics of participants in both RCT and central follow up. Of 3,784 subjects, 227 had been diagnosed with dementia at the end of follow-up. Compared to subjects who did not develop dementia, subjects who developed dementia were older at baseline, less educated, were more often smokers, more likely carried the *APOE*-ε4 allele, used antihypertensive medication, and reported experiencing a memory change at baseline (Tab. I).

Based on preliminary analysis (data not shown), the prediction error in the neural network did not change dramatically as the threshold of the partial derivatives of the error function changed; we chose 0.1 as the threshold. Fig. 1 depicts the neural network structure for the current study and shows the final weights of the corresponding synapses. These weights were used to calculate the estimated probability of incidence of dementia. To interepret the association found in the neural network, the estimated probabilities of having dementia for 36 hypothetical subjects are presented in Table II. The measure of association for having dementia given a certain covariate in the neural network depends on the covariate and other covariates in the model.

Keeping other covariates in the model constant, as age increased, the estimated probability of having dementia is increased (Tab. II). For example, subjects 1, 2, and 3, who represent persons who are without an *APOE*-ε4 allele, absence of memory change, lower education(13.2 years), are aged at 62.2 years, at 67.2 years, and at 72.2 years, respectively. The estimated probabilities of developing dementia for subjects 1, 2, 3 are 0.029, 0.053, and 0.068, respectively. Similarly, we can conclude that the presence of an *APOE*-ε4 allele is associated with increased estimated probability of having dementia (subject 3 *vs* subject 12 in Table II).

As illustrated in Table 2, the effect of education on risk of dementia depended on age, *APOE*-ε4 allele status, and perception of memory change. Higher education was associated with lower risk of dementia only in younger subjects (62.2 years) with the exception of

Table I. General Characteristics of participants in PREADViSE.

Characteristic	All subjects (n = 3784)	No dementia (n = 3557)	Dementia (n = 227)	P-value
Baseline age[b], y	67.2 ± 5.0	67.0 ± 5.0	70.1 ± 5.2	< 0.001
Education[b], y	15.5 ± 2.3	15.5 ± 2.3	15.0 ± 2.5	0.002
Black race[c]	318 (8.4)	294 (8.3)	24 (10.6)	0.22
Baseline smoking[c]	2018 (53.4)	1879 (52.9)	139 (61.2)	0.01
APOE-ε4 (≥ 1 ε4)c	956 (25.3)	871 (24.5)	85 (37.4)	< 0.001
Baseline hypertension[c]	2998 (39.7)	2703 (38.6)	295 (53.4)	<0.001
Baseline diabetes[c]	354 (9.4)	322 (9.1)	32 (14.1)	0.01
Baseline BMI[ab], kg/m^2	28.4 ± 4.3	28.4 ± 4.3	28.4 ± 4.5	0.93
Baseline CABG[ac]	135 (3.6)	119 (3.4)	16 (7.1)	0.004
Baseline congestive heart disease[c]	18(0.5)	16 (0.5)	2 (0.9)	0.29
Baseline antihypertensive medication[c]	1413 (37.3)	1308 (36.8)	105 (46.3)	0.004
Memory change[c]	852 (22.5)	762 (21.4)	90 (39.7)	< 0.001

[a]BMI: Body Mass Index; CABG: Coronary artery bypass graft; [b]mean ± standard deviation; [c]count (%)

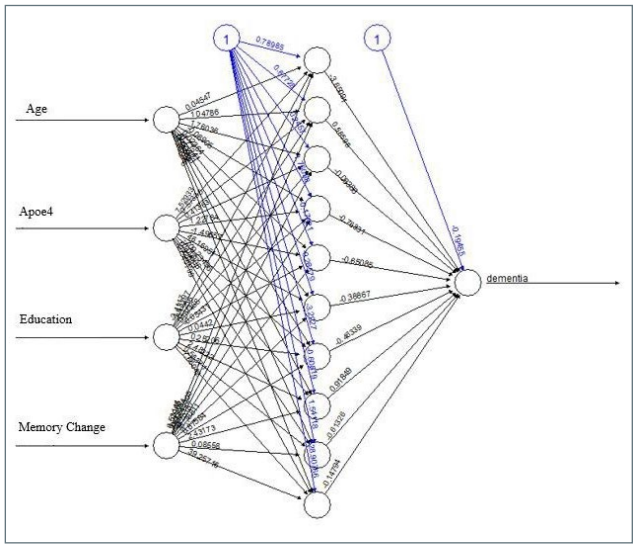

Figure 1. Neural network model for incidence of dementia in PREADViSE trial. Age, *APOE*- ε4, Education, Memory Change represent the 4 input neurons on the left side of the diagram. Each input neuron was connected with 10 hidden neurons (second column of empty circles from the left of figure) by 10 corresponding synaptic weights. The 10 hidden units and the output neuron – dementia were connected by the synapses starting from the hidden units and ending at the output layer. The first "1" in the circle from the left of the figure represents the intercepts of each hidden neuron, and the second "1" in the circle stands for the intercept of output neuron. These weights and intercepts were adapted to calculate the estimated probability of dementia. The model was stopped after 203,313 steps, and prediction error is 99.02907.

younger subjects with *APOE*-ε4 and self-reported memory change. For example, in the younger age group, the estimated probability of having dementia for subject 7 (= 0.003) with higher education (17.8 years) is much lower than subject 1 (= 0.029) with lower education (13.2 years). Similar comparisons can be made for subjects 10 (= 0.054) and 16 (= 0.002), but not the comparison between subject 28 (= 0.070) and subject 34 (= 0.098), who represent subjects with presence of both *APOE*-ε4 allele and memory change. Education did not show a protective effect on risk of dementia in older subjects (72.2 years), especially for older subjects with presence of both *APOE*-ε4 allele and memory change. No matter the education level, older subjects who were *APOE* ε4 carriers and had memory changes had the highest risk of dementia, such as subjects 30 (= 0.354), 33 (= 0.364), and 36 (= 0.372). In contrast, well-educated younger subjects who did not have either *APOE*-ε4 or memory changes had the lowest risk of dementia, such as subject 7 (= 0.003), subject 16 (= 0.002), and subject 25 (= 0.001).

According to the results in neural network in which the effect of education was modified by age, status of *APOE*-ε4, and status of memory change, logistic regression were performed to confirm the interaction effects. Table III shows parameter estimates and p values for one 3-way interaction regression model from logistic regression final model, and the four-way interaction term and other three-way interaction terms were removed from final model due to the insignificance. The 3-way interaction among education in years, *APOE*-ε4 allele carrier, and self- reported memory changes is significant (p = 0.01). To demonstrate the effect modification identified in the logistic regression model, same 36 hypothetical subjects are presented in table 4. Subjects with presence of both *APOE*-ε4 allele and memory change had highest estimated probability of having dementia, such as subject 30 ($\tilde{p}$ = 0.664), and subject 33 ($\tilde{p}$ = 0.594). Age modified the effect of education; however, comparing subjects 28, 29, 30, or subje cts 19, 20, 21 from Table IV, the interaction effect among age, education, *APOE*-ε4 and memory change is not significant.

Comparison of overall performance between logistic regression and ANN for predicting incident dementia was recorded in Table V. For calculating sensitivity, etc., a model-estimated probability greater than 0.05 was regarded as a positive classification. ANN had similar predictive accuracy as logistic regression (AUC in neural network = 0.732 *vs* AUC in logistic regression 0.725). Overall, neural network has similar performance in sensitivity (83.2%) and negative predictive value (98.0%) than sensitivity (83.3%) and negative predictive value (98.0%) in logistic regression, but worse in the positive predictive value (9.9% in neural network *vs* 9.8% in logistic regression).

DISCUSSION

The purpose of this study was to compare predictive accuracy for incident dementia between neural network and logistic regression in the PREADViSE trial. Neural network analysis showed slightly apparently improved predictive accuracy (AUC = 0.732) compared to logistic regression (AUC = 0.725). The model obtained from the neural network had similar performance as the logistic regression model. Similar association between covariates and the dementia outcome were found in neural network and logistic regression, but the model in neural network is more difficult to interpret. Furthermore, neural network may easily elucidate more complex relationships between model variables, such as education, age, APOE, and self-reported memory change, and the outcome. While higher education is usually considered universally protective against dementia [36,37], the effect of

Table II. Illustration effect of education by age, APOE-ε4 allele and memory change status for 36 hypothetical subjects from neural network.

	Age[a]					
	Old		Average		Young	
Education[b]	Subject ID	$\hat{p}$ [c]	Subject ID	$\hat{p}$ [c]	Subjects ID	$\hat{p}$ [c]
Absence of APOE-ε4 allele and absence memory change						
Low	3	0.068	2	0.053	1	0.029
Average	6	0.066	5	0.046	4	0.018
High	9	0.062	8	0.037	7	0.003
Presence of APOE-ε4 allele and absence of memory change						
Low	12	0.148	11	0.103	10	0.054
Average	15	0.128	14	0.079	13	0.028
High	18	0.104	17	0.053	16	0.002
Absence of APOE-ε4 allele and presence of memory change						
Low	21	0.145	20	0.126	19	0.091
Average	24	0.113	23	0.096	22	0.048
High	27	0.100	26	0.053	25	0.001
Presence of APOE-ε4 allele and presence of memory change						
Low	30	0.354	29	0.056	28	0.070
Average	33	0.364	32	0.071	31	0.084
High	36	0.372	35	0.058	34	0.098

Note: [a]Young: 62.2 years; Average: 67.2 years; Old: 72.2 year; [b]Low: 13.2 years of education; Average: 15.5 years of education; High: 17.8 years of education; [c]: estimated probability of having dementia

Table III. Parameter estimates from logistic regression model with interaction.

Variables	Estimate (SE)	P-value
Intercept	4.78 (5.59)	0.39
Age at baseline	-0.10 (0.08)	0.22
Education in years	-0.91 (0.37)	0.01
Presence of *APOE-ε4 allele*	-1.91 (2.11)	0.37
Presence of memory change	2.45 (2.07)	0.24
Age at baseline * presence of *APOE-ε4 allele*	0.05 (0.03)	0.04
Age at baseline * education in years	0.01 (0.005)	0.02
Age at baseline * presence of memory change	-0.002 (0.03)	0.91
Education in years * presence of *APOE-ε4 allele*	-0.08 (0.07)	0.30
Presence of *APOE-ε4 allele* * presence of memory change	-4.85 (1.99)	0.01
Education in years * presence of memory change	-0.10 (0.07)	0.18
Education in years * presence of *APOE-ε4 allele* * presence of memory change	0.32 (0.13)	0.01

education on dementia in the neural network depended on age, *APOE*-ε4 allele status, and self-reported memory change.

Stephan et al. [18] evaluated predictive accuracy of dementia prediction models and found that poor predictive accuracy is associated with single-factor models, long follow-up intervals, and all-cause dementia for outcome ascertainment, which assumes all dementias share risk factors. The Canadian Study of Health and Aging (CSHA) [38] showed lower predictive accuracy (AUC = 0.77 in 10-year follow-up than 5–year follow-up (AUC = 0.83). The predictive accuracy (AUC = 0.732) in our neural network model is slightly lower than the CSHA 10-year study, but is comparable to the Gothenburg H-70 1901-02 birth cohort for 20-years of follow-up (AUC = 0.74) [39]. In contrast to the CSHA study, Exalto et al did not find any significantly different results on predictive accuracy in two Cardiovascular Risk Factors,

Table IV. Estimated probability of having dementia from multivariable logistic regression to illustrate interaction effects among age at baseline, education, APOE-ε4 allele and memory change.

	Age[a]					
	Old		Average		Young	
Education[b]	Subject ID	$\hat{p}$ [c]	Subject ID	$\hat{p}$ [c]	Subjects ID	$\hat{p}$ [c]
Absence of APOE-ε4 allele and absence memory change						
Low	3	0.059	2	0.044	1	0.032
Average	6	0.055	5	0.036	4	0.023
High	9	0.051	8	0.029	7	0.016
Presence of APOE-ε4 allele and absence of memory change						
Low	12	0.315	11	0.205	10	0.126
Average	15	0.299	14	0.172	13	0.091
High	18	0.284	17	0.143	16	0.066
Absence of APOE-ε4 allele and presence of memory change						
Low	21	0.164	20	0.125	19	0.095
Average	24	0.126	23	0.084	22	0.055
High	27	0.097	26	0.056	25	0.032
Presence of APOE-ε4 allele and presence of memory change						
Low	30	0.664	29	0.525	28	0.382
Average	33	0.594	32	0.415	31	0.256
High	36	0.519	35	0.313	34	0.161

Note: [a]Young: 62.2 years; Average: 67.2 years; Old: 72.2 year; [b]Low: 13.2 years of education; Average: 15.5 years of education; High: 17.8 years of education; [c]: estimated probability of having dementia

Table V. Comparison of predictive performance of logistic regression and neural network.

	Logistic regression[a]	Neural network[a]
Sensitivity	83.3%	83.2%
Specificity	51.3%	51.4%
Positive predictive value	9.8%	9.9%
Negative predictive value	98.0%	98.0%

Note: [a]Area under curve: 0.725 in logistic regression and 0.732 in neural network

Aging, and Incidence of Dementia (CAIDE) studies based on follow-up time (one is 10 years follow-up and another one is 36 years) [40,41]. Follow-up time in the current study was over 10 years.

Based on covariates used to generate the predictive model, models generated in the previous papers can be summarized into the following categories: 1) demographic only models; 2) cognitive test based models with or without demographic data; 3) comorbidity data models; 4) genetic and biomarker models; and 5) models including demographics, comorbidities, genetics, and biomarkers. Our logistic regression model included age, education, APOE, and memory changes to predict incident dementia and had moderate predictive accuracy (AUC: 0.725, sensitivity 83.3% and specificity 51.3%). A similar study [40] in which the model was derived from demographic variables, health risk factors and *APOE*, obtained slightly better diagnostic accuracy (AUC = 0.78).This study also argued that diagnostic accuracy did not change significantly after removing *APOE* from the model (AUC: 0.77; sensitivity 77% and specificity 63%). Other models include neuroimaging information and/or neuropsychological tests. Tang et al. argued in their review that genetic information and/or imaging data do not improve diagnostic accuracy significantly [15,19]. Furthermore, predictive models using one or multiple neuropsychological tests as covariates seem to have higher predictive accuracy, but there is not direct comparison for these two approaches due to between-study variation, such as different criteria for outcome measurement [18]. Waite and colleagues argued that refining the subgroups of dementia types may improve diagnostic accuracy, but is unlikely to be cost effective because defining these subgroups of dementia can be expensive [42].

On the other hand, from machine learning and classical statistical methods, neural network in several studies has demonstrated superior ability to capture complex relationships in data compared to classical statistical methods [43,44]. Also, neural networks obtained higher predictive accuracy rate than linear discriminant analysis and successfully distinguished Alzheimer's patients from control

aged 80 years and older in the Nun study using neurofibrillary tangles and neurotic plaques counts (AUC was not reported) [45]. In contrast, Maroco et al. [46] suggested that random forest and linear discriminant analysis performed better than other statistical methods, such as neural network, support vector machines, and logistic regression based on the consideration of predictive accuracy, sensitivity, specificity. They also argued that neural networks and logistic regression are inappropriate for unbalanced data, which means small frequency vs. large frequency group in response variable. Furthermore, Song et al. [47] compared the machine learning methods with classic statistical methods for two biomedical datasets: one was from patient care records and another was from a population survey, and they did not find big differences in prediction between the two datasets, which indicates that the quality of the data, such as accuracy and completeness may be more important than leaning technique.

Strengths for this study include larger sample size and long follow-up. We were also able to consider most well-established risk factors for dementia, including demographic, genetic, and medical characteristics. This study also had some limitations. Our outcome diagnosis was based on two criteria due to lack of medical records from many participants, our case ascertainment may be less accurate. However, misclassification of diagnosis is independent of exposure measurement, so non-differential misclassification is unlikely. Thus, results are likely biased toward the null.

CONCLUSIONS

Neural network did not significantly improve predict accuracy over logistic regression and also increased difficulty of interpretation of the association between the outcome and covariates. The most important aspect to improve performance of a model, does not depend on statistical methods, or computational techniques, but depends on how much accurate information the dataset contains. In the future, similar studies should focus on refining the definition of outcome diagnosis, improving quality of assessments, and performing validation after generating a risk model. Neural network may not improve prediction, but it has potential to identify interaction effects with conducting multiple testing.

Acknowledgments

We thank Dr. Erin Abner for her assistance with helpful comments. Financial Disclosure: PREADViSE (NCT00040378) is supported by National Institute on Aging (NIA) R01 AG019421. Additional support for the current study comes from NIA R01 AG038651 and P30 AG028383. SELECT was supported by NCI grants CA37429 and UM1 CA182883. NCI was involved in the design of SELECT.

Reference

1 Chary E, Amieva H, Pérès K, et al. Short-versus long-term prediction of dementia among subjects with low and high educational levels. Alzheimers Dement 2013;9:562-71. https://doi.org/10.1016/j.jalz.2012.05.2188

2 Wolfsgruber S, Jessen F, Wiese B, et al. The CERAD neuropsychological assessment battery total score detects and predicts Alzheimer disease dementia with high diagnostic accuracy. Am J Geriatr Psychiatry 2014;22:1017-28. https://doi.org/10.1016/j.jagp.2012.08.021

3 Song X, Mitnitski A, Rockwood K. Nontraditional risk factors combine to predict Alzheimer disease and dementia. Neurology 2011;77:227-34. https://doi.org/10.1212/WNL.0b013e318225c6bc

4 Tierney MC, Moineddin R, McDowell I. Prediction of all-cause dementia using neuropsychological tests within 10 and 5 years of diagnosis in a community-based sample. J Alzheimers Dis 2010;22:1231-40. https://doi.org/10.3233/JAD-2010-100516

5 Grober E, Sanders AE, Hall C, et al. Free and cued selective reminding identifies very mild dementia in primary care. Alzheimer Dis Assoc Disord 2010;24:284-90. https://doi.org/10.1097/WAD.0b013e3181cfc78b

6 Barnes DE, Beiser AS, Lee A, et al. Development and validation of a brief dementia screening indicator for primary care. Alzheimers Dement 2014;10:656-65.e1. https://doi.org/10.1016/j.jalz.2013.11.006

7 Verhaaren BF, Vernooij MW, Koudstaal PJ, et al. Alzheimer's disease genes and cognition in the nondemented general population. Biol Psychiatry 2013;73:429-34. https://doi.org/10.1016/j.biopsych.2012.04.009

8 Exalto LG, Biessels GJ, Karter AJ, et al. Risk score for prediction of 10 year dementia risk in individuals with type 2 diabetes: a cohort study. Lancet Diabetes Endocrinol 2013;1:183-90. https://doi.org/10.1016/S2213-8587(13)70048-2

9 Derby CA, Burns LC, Wang C, et al. Screening for predementia AD: time-dependent operating characteristics of episodic memory tests. Neurology 2013;80:1307-14. https://doi.org/10.1212/WNL.0b013e31828ab2c9

10 Jessen F, Wiese B, Bickel H, et al. Prediction of dementia in primary care patients. PLoS One 2011;6:e16852. https://doi.org/10.1371/journal.pone.0016852

11 Weakley A, Williams JA, Schmitter-Edgecombe M, et al. Neuropsychological test selection for cognitive impairment classification: a machine learning approach. J Clin Exp Neuropsychol 2015;37:899-916. https://doi.org/10.1080/13803395.2015.1067290

12 Haghighi M, Smith A, Morgan D, et al. Identifying cost-effective predictive rules of amyloid-beta level by integrating neuropsychological tests and plasma-based markers. J Alzheimers Dis 2015;43:1261-70. https://doi.org/10.3233/JAD-140705

[13] Mattsson N, Insel PS, Donohue M, et al. Predicting reduction of cerebrospinal fluid beta-amyloid 42 in cognitively healthy controls. JAMA Neurol 2015;72:554-60. https://doi.org/10.1001/jamaneurol.2014.4530

[14] Li L, Huang J, Sun S, et al. Selecting pre-screening items for early intervention trials of dementia – a case study. Stat Med 2004;23:271-83. https://doi.org/10.1002/sim.1715

[15] Seshadri S, Fitzpatrick AL, Ikram MA, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. Jama 2010;303:1832-40. https://doi.org/10.1001/jama.2010.574

[16] Stephan BC, Tzourio C, Auriacombe S, et al. Usefulness of data from magnetic resonance imaging to improve prediction of dementia: population based cohort study. Bmj 2015;350:h2863. https://doi.org/10.1136/bmj.h2863

[17] Norton S, Matthews FE, Barnes DE, et al. Potential for primary prevention of Alzheimer's disease: an analysis of population-based data. Lancet Neurol 2014;13:788-94. https://doi.org/10.1016/S1474-4422(14)70136-X

[18] Stephan BC, Kurth T, Matthews FE, et al. Dementia risk prediction in the population: are screening models accurate? Nat Rev Neurol 2010;6:318-26. https://doi.org/10.1038/nrneurol.2010.54

[19] Tang EY, Harrison SL, Errington L, et al. Current developments in dementia risk prediction modelling: an updated systematic review. PLoS One 2015;10:e0136181. https://doi.org/10.1371/journal.pone.0136181

[20] Lippman SM, Klein EA, Goodman PJ, et al. Effect of selenium and vitamin E on risk of prostate cancer and other cancers: the Selenium and Vitamin E Cancer Prevention Trial (SELECT). Jama 2009;301:39-51. https://doi.org/10.1001/jama.2008.864

[21] Goodman PJ, Hartline JA, Tangen CM, et al. Moving a randomized clinical trial into an observational cohort. Clin Trials 2013;10:131-42. https://doi.org/10.1177/1740774512460345

[22] Buschke H, Kuslansky G, Katz M, et al. Screening for dementia with the memory impairment screen. Neurology 1999;52:231-8. https://doi.org/10.1212/wnl.52.2.231

[23] Morris JC, Heyman A, Mohs RC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. Neurology 1989;39:1159-65. https://doi.org/10.1212/wnl.39.9.1159

[24] de Jager CA, Budge MM, Clarke R. Utility of TICS-M for the assessment of cognitive function in older adults. Int J Geriatr Psychiatry 2003;18:318-24. https://doi.org/10.1002/gps.830

[25] Lopez OL. The growing burden of Alzheimer's disease. Am J Manag Care 2011;17(Suppl 13):S339-45.

[26] Barnes DE, Yaffe K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. Lancet Neurol 2011;10:819-28. https://doi.org/10.1016/S1474-4422(11)70072-2

[27] Galvin JE, Roe CM, Powlishta KK, et al. The AD8: a brief informant interview to detect dementia. Neurology 2005;65:559-64. https://doi.org/10.1212/01

[28] Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med 2004;256:183-94. https://doi.org/10.1111/j.1365-2796.2004.01388.x

[29] Fristsch FGS.Neuralnet: training of neural networks. The R Journal 2010;2.

[30] Stefan Fritsch FG. Package'neuralnet'. CRAN 2012.

[31] Bishop CM. Neural networks for patter recognition. Oxford: Clarendon press 1995.

[32] Hornik K. Multiplayer feedforward networks are universal approximators, 1989;2:7.

[33] Rojas RU. Neural networks – a systematic introduction. Berlin: Springer-Verlag 1996.

[34] Riedmiller M. Rprop – description and implementation. Details Technical Rep 1994.

[35] Quintana M, Guàrdia J, Sánchez-Benavides G, et al. Using artificial neural networks in clinical neuropsychology: high performance in mild cognitive impairment and Alzheimer's disease. J Clin Exp Neuropsychol 2012;34:195-208. https://doi.org/10.1080/13803395.2011.630651

[36] EClipSE Collaborative Members, Brayne C, Ince PG, et al. Education, the brain and dementia: neuroprotection or compensation? Brain 2010;133:2210-6. https://doi.org/10.1093/brain/awq185

[37] Wilson RS, Hebert LE, Scherr PA, et al. Educational attainment and cognitive decline in old age. Neurology 2009;72:460-5. https://doi.org/10.1212/01.wnl.0000341782.71418.6c

[38] Tierney MC, Yao C, Kiss A, et al. Neuropsychological tests accurately predict incident Alzheimer disease after 5 and 10 years. Neurology 2005;64:1853-9. https://doi.org/10.1212/01.WNL.0000163773.21794.0B

[39] Mitnitski A, Skoog I, Song X, et al. A vascular risk factor index in relation to mortality and incident dementia. Eur J Neurol 2006;13:514-21. https://doi.org/10.1111/j.1468-1331.2006.01297.x

[40] Kivipelto M, Ngandu T, Laatikainen T, et al. Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. Lancet Neurol 2006;5:735-41. https://doi.org/10.1016/S1474-4422(06)70537-3

[41] Exalto LG, Quesenberry CP, Barnes D, et al. Midlife risk score for the prediction of dementia four decades later. Alzheimers Dement 2014;10:562-70. https://doi.org/10.1016/j.jalz.2013.05.1772

[42] Waite LM, Broe GA, Grayson DA, et al. Preclinical syndromes predict dementia: the Sydney older persons study. J Neurol Neurosurg Psychiatry 2001;71:296-302. https://doi.org/10.1136/jnnp.71.3.296

[43] Parsons TD, Rizzo AA, Buckwalter JG. Backpropagation and regression: comparative utility for neuropsychologists. J Clin Exp Neuropsychol 2004;26:95-104. https://doi.org/10.1076/jcen.26.1.95.23932

[44] French BM, Dawson MR, Dobbs AR. Classification and staging of dementia of the Alzheimer type: a comparison between neural networks and linear discriminant analysis. Arch Neurol 1997;54:1001-9.

[45] Grossi E, Buscema MP, Snowdon D, et al. Neuropathological findings processed by artificial neural networks (ANNs) can perfectly distinguish Alzheimer's patients from controls in the Nun Study BMC Neurol 2007;7:15. https://doi.org/10.1186/1471-2377-7-15

[46] Maroco J, Silva D, Rodrigues A, et al. Data mining methods in the prediction of dementia: a real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. BMC Res Notes 2011;4:299. https://doi.org/10.1186/1756-0500-4-299

[47] Song X, Mitnitski A, Cox J, et al. Comparison of machine learning techniques with classical statistical models in predicting health outcomes. Stud Health Technol Inform 2004;107:736-40.