Featured Article

Forecasting the progression of Alzheimer's disease using neural networks and a novel preprocessing algorithm

Jack Albright[1,*], for the Alzheimer's Disease Neuroimaging Initiative[2]

The Nueva School, San Mateo, CA

Abstract

Introduction: There is a 99.6% failure rate of clinical trials for drugs to treat Alzheimer's disease, likely because Alzheimer's disease (AD) patients cannot be easily identified at early stages. This study investigated machine learning approaches to use clinical data to predict the progression of AD in future years.

Methods: Data from 1737 patients were processed using the "All-Pairs" technique, a novel methodology created for this study involving the comparison of all possible pairs of temporal data points for each patient. Machine learning models were trained on these processed data and evaluated using a separate testing data set (110 patients).

Results: A neural network model was effective (mAUC = 0.866) at predicting the progression of AD, both in patients who were initially cognitively normal and in patients suffering from mild cognitive impairment.

Discussion: Such a model could be used to identify patients at early stages of AD and who are therefore good candidates for clinical trials for AD therapeutics.

© 2019 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Machine learning; Neural networks; Disease progression; Alzheimer's disease; Mild cognitive impairment; Dementia; Longitudinal studies

1. Introduction

Alzheimer's disease (AD) is the most common neurodegenerative disease in older people [1,2]. AD takes a significant toll on patients' daily lives, causing a progressive decline in their cognitive abilities, including memory, language, behavior, and problem solving [3–6].

Changes to AD patients' cognitive abilities often start slowly and become more rapid over time [7,8]. Doctors and other caregivers monitor the progression of AD in patients by evaluating the degree of decline in the patients' cognitive abilities [9], which are often divided into 3 general categories: cognitively normal (NL), mild cognitive impairment (MCI), and dementia [5,10]. Patients with MCI and dementia both suffer from reduced cognitive abilities, but MCI has a less severe effect on everyday activities, and patients suffering from dementia often have additional symptoms such as trouble with reasoning or impaired judgment [11,12].

Unfortunately, there is no cure for AD at this time, and progress on identifying a cure has been slow [10]. None of the five medications currently approved by the FDA to treat AD have been shown to delay or halt its progression [9]. Instead, they only temporarily improve patients' symptoms [9,13,14]. And, the most recently approved medication is

just the combination of two existing drugs for treating AD, donepezil and memantine, which were approved by the FDA 23 and 16 years ago, respectively [14]. Despite considerable efforts to find a cure for AD, there is a 99.6% failure rate of clinical trials for AD drugs [13,15]. In early 2018 alone, two groups ended their AD clinical trials because their drugs failed to prevent the progression of AD [16,17]. The difficulty in finding treatments for AD is most likely a combination of uncertainty over the cause of AD and the fact that AD patients cannot be easily identified at early stages [10,13,18]. Even the FDA recognizes the importance of identifying patients who are at risk of developing AD but who do not have any noticeable cognitive impairment [19]. For this reason, AD research would benefit from the ability to use current medical data to predict the mental state of patients in future years to identify patients who are good candidates for clinical trials before they become symptomatic [2,10,13,20,21].

Several different types of data that are relevant to assessing the mental state of AD patients and the progression of AD in general have been identified. One of the largest genetic risk factors for AD is the presence of 1 or 2 copies of the ε4 allele of the *APOE* gene, which encodes a particular variant of the enzyme Apolipoprotein E [5]. Physical changes to the brain have also been shown to be correlated with the progression of AD. For example, a decline in neurogenesis in the hippocampus is one of the earliest changes to brain physiology seen in AD patients and is thought to underlie cognitive impairments associated with AD [6]. The progression of AD also accelerates the normal atrophy of brain tissue caused by aging, as evidenced by increased enlargement of the ventricles of the brain over time [22]. One study demonstrated a 4-fold difference in the rate of ventricle enlargement in AD patients and normal controls over a six-month interval [22]. Cognitive tests have also been widely used for early detection of AD [8]. Several commonly used tests, such as ADAS11 and ADAS13, are based on the Alzheimer's Disease Assessment Scale (ADAS), which is a brief cognitive test battery that assesses learning and memory, language production, language comprehension, constructional praxis, ideational praxis, and orientation [23,24]. ADAS11 scores range from 0 to 70, and ADAS13 scores range from 0 to 85, with higher scores indicating more advanced stages of AD [24]. Similar cognitive tests, such as the Mini–Mental State Examination (MMSE), the Rey Auditory Verbal Learning Test (RAVLT), and the Functional Activities Questionnaire (FAQ) have also been used to assess the progression of AD in individual patients [25,26]. ADAS has been found to be more precise than the MMSE [27], and the RAVLT only addresses verbal recall [28], thus providing less diagnostic information than either of the other two. Similarly, the FAQ only assesses patients' ability to perform certain tasks [29] and therefore is more limited in scope than the MMSE and ADAS.

In recent years, machine learning techniques have been applied to the diagnosis of AD patients with great success. For example, Esmaeilzadeh et al. achieved an accuracy of 94.1% using 3D convolutional neural networks to diagnose AD on a data set with 841 patients [30]. Similar results were obtained by Long et al., who used a support vector machine to diagnose AD based on a magnetic resonance imaging (MRI) scan data set (n = 427 patients; mean best accuracy = 96.5%) [31] and Zhang et al., who used MRI scans, FDG–positron emission tomography scans, and CSF biomarkers to diagnose AD (n = 202 patients; AD vs. NL accuracy = 93.2%, MCI vs. NL accuracy = 76.4%) [32].

Table 1
Features from the ADNI data set used to train the machine learning models

Variable	Meaning	8-Feature training group	11-Feature training group	15-Feature training group
DX	Earlier diagnosis	X	X	X
ADAS13	13-item Alzheimer's Disease Assessment Scale	X	X	X
Ventricles	Ventricular volume	X	X	X
AGE	Age	X	X	X
FAQ	Functional Activities Questionnaire		X	X
PTRACCAT	Race	X	X	X
Hippocampus	Hippocampal volume	X	X	X
APOE4	# of ε4 alleles of *APOE*	X	X	X
MMSE	Mini-Mental State Examination		X	X
ADAS11	11-item Alzheimer's Disease Assessment Scale		X	X
RAVLT_immed	Rey Auditory Verbal Learning Test— total number of words memorized over 5 trials			X
RAVLT_learn	Rey Auditory Verbal Learning Test— number of words learned between trial 1 and trial 5			X
RAVLT_forg	Rey Auditory Verbal Learning Test— number of words forgotten between trial 5 and trial 6			X
RAVLT_perc_forg	Rey Auditory Verbal Learning Test— percentage of words forgotten between trial 5 and trial 6			X
N/A	Time difference	X	X	X

Abbreviation: ADNI, Alzheimer's Disease Neuroimaging Initiative.

However, the focus of these earlier studies was to use current medical data to diagnose a patient's present cognitive state, in effect demonstrating that a computer can replicate a doctor's clinical decision-making. What is needed is a way to use machine learning to predict future diagnoses of AD patients [10].

2. Methods

2.1. Alzheimer's Disease Neuroimaging Initiative data

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (https://adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by the principal investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD.

The ADNI study has been divided into several phases, including ADNI-1, ADNI-GO, and ADNI-2, which started in 2004, 2009, and 2011, respectively. ADNI-1 studied 800 patients, and each subsequent phase included a mixture of new patients and patients from the prior phase who elected to continue to participate in the study. The ADNI patient data were preprocessed to flag missing entries and to convert nonnumeric categories (such as race) into numeric data. Data were sorted into three data sets (LB1, LB2, and LB4) based on criteria established by The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) Challenge (https://tadpole.grand-challenge.org/) [10].

The LB2 and LB4 data sets consist of data from 110 patients who participated in ADNI-1, who continued to participate in ADNI-GO/ADNI-2, and who were not diagnosed with AD as of the last ADNI-1 time point. Specifically, LB2 contains all observations of these patients from ADNI-1, and LB4 contains all observations of these patients from ADNI-GO/ADNI-2. The LB1 data set consists of ADNI data for all remaining patients (n = 1737). Generally speaking, LB1 was used as a training and validation data set, whereas LB2 and LB4 were later used to test the ability of machine learning models to predict the progression of AD on an independent patient population.

2.2. All-pairs technique

Data from LB1 were then further processed using a novel methodology developed for this project called the all-pairs technique, which can be summarized as follows: Let R be the number of patients in the data set and B be the number of biomarkers (or other clinical data) being evaluated as features. For each patient P_i ($1 \leq i \leq R$), the ADNI database in-
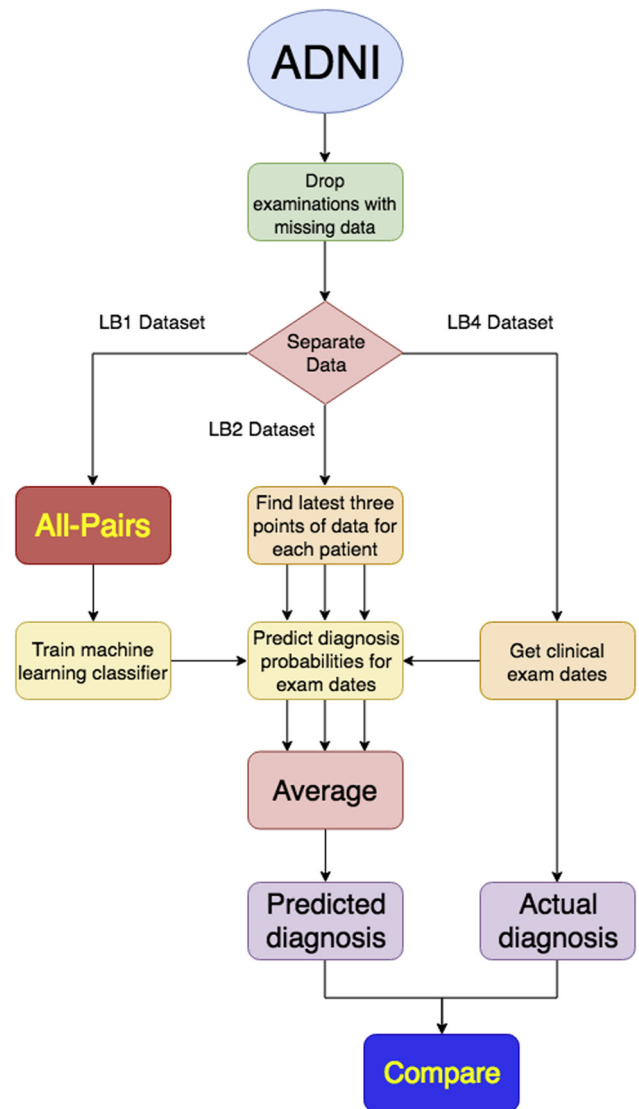

Fig. 1. Flowchart showing methodology for training and evaluating machine learning models.

cludes L_i separate examinations by a physician. Then, $E_{i,j}$, the jth examination of the ith patient, can be defined as a multidimensional vector as follows:

$$E_{i,j} = \left[d_{i,j}, b_{i,j,1}, b_{i,j,2}, \ldots, b_{i,j,B}, c_{i,j}\right]$$

where $d_{i,j}$ is the date of the examination, $b_{i,j,k}$ ($1 \leq k \leq B$) are different biomarkers (or other clinical data), and $c_{i,j}$ is the clinical state of the patient (normal, MCI, or dementia) as measured during that examination. The all-pairs technique transforms these examination data to generate a feature array X and target array Y that are used to train the machine learning models. Specifically, for every i,j_a,j_b, where $1 \leq i \leq R$ and $1 \leq j_a < j_b \leq L_i$, a row of X and a corresponding cell of Y are calculated:
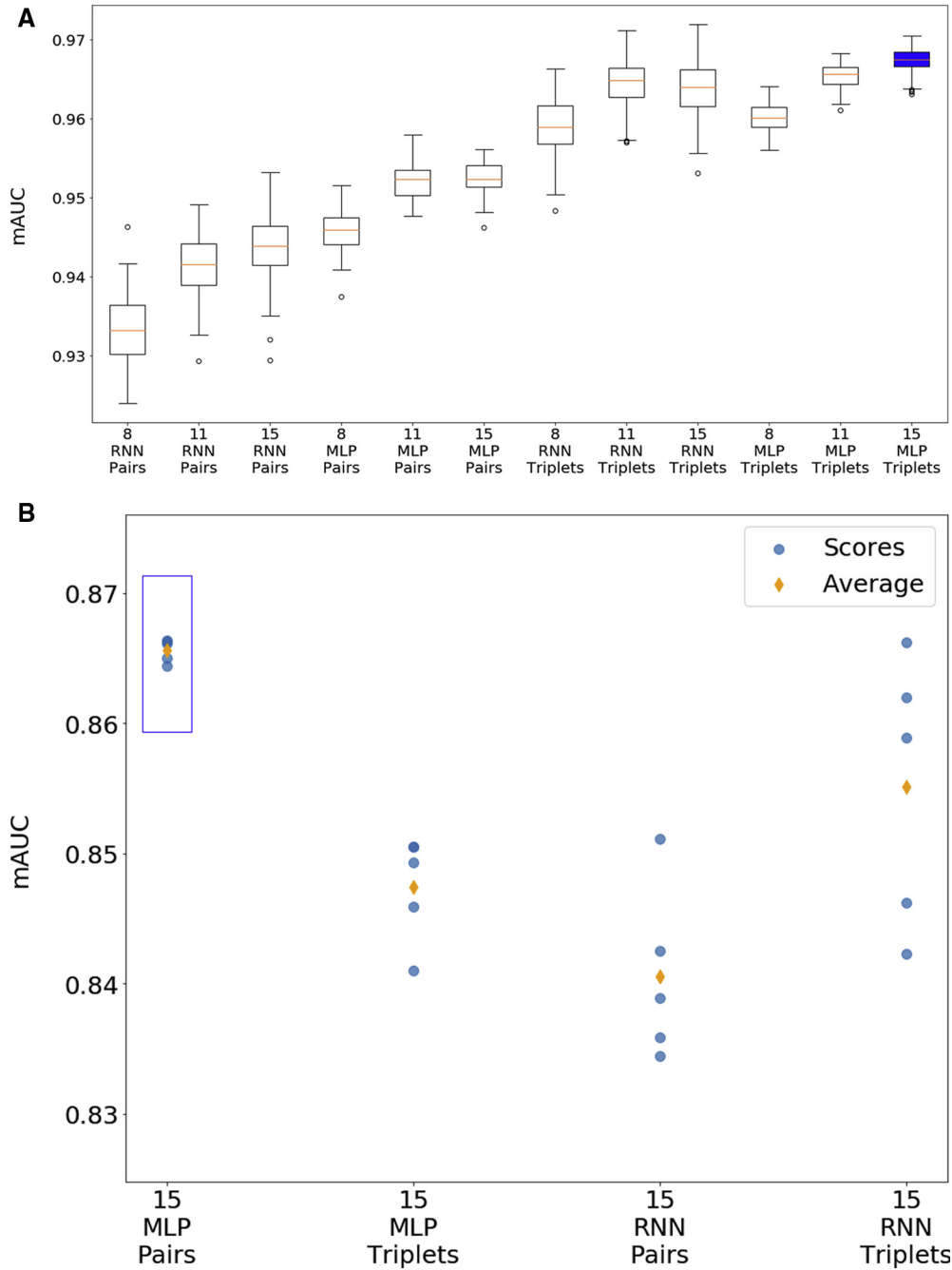
Fig. 2. Performance of various machine learning models. (A) Box-and-whisker plots showing performance of models on 100 random splits of 1737-patient training data set (LB1). (B) Performance of models on 110-patient testing data set (LB2/LB4) (scores from 5 iterations plus average score).

$$X_{row} = \left[d_{i,j_b} - d_{i,j_a}, b_{i,j_a,1}, b_{i,j_a,2}, ..., b_{i,j_a,B}, c_{i,j_a}\right]$$

$$Y_{cell} = c_{i,j_b}$$

This approach can be extended from pairs of examinations (j_a, j_b) to triplets (j_a, j_b, j_c) as follows:

$$X_{row} = \left[d_{i,j_c} - d_{i,j_b}, d_{i,j_b} - d_{i,j_a}, b_{i,j_b,1}, ..., b_{i,j_b,B}, \right.$$
$$\left. c_{i,j_b}, b_{i,j_a,1}, ..., b_{i,j_a,B}, c_{i,j_a}\right]$$

$$Y_{cell} = c_{i,j_c}$$

Individual rows and cells are then assembled to create the X and Y arrays that are used for training of the machine learning model as well as cross-validation studies.

2.3. Evaluation using LB2 and LB4 data sets

In the case of LB2, each examination can be characterized by the same vector $E_{i,j}$ as used for the LB1 data set.
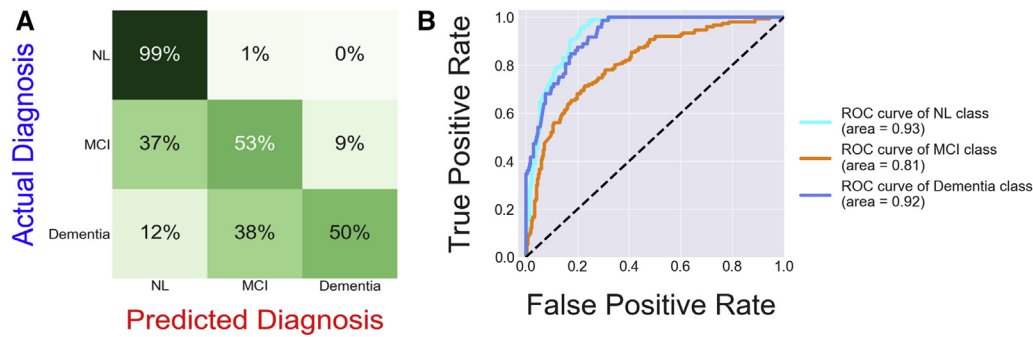
Fig. 3. (A) Confusion matrix comparing diagnoses predicted by the best-performing neural network with actual diagnoses. (B) Receiver operating characteristic curves based on output of best-performing neural network. Abbreviations: MCI, mild cognitive impairment; NL, cognitively normal.

However, because the LB2 data are only used to test the machine learning models and not for training, no comparison between examinations is performed. Instead, $E_{i,j}$ is transformed into the input vector by replacing the $d_{i,j}$ term with a time variable that represents the number of months into the future that the machine learning model should make a prediction. Specifically, for each patient P_i in LB2, the machine learning algorithm is applied to a series of input vectors of the form:

$$[t, b_{i,j,1}, b_{i,j,2}, \ldots, b_{i,j,B}, c_{i,j}]$$

where t is the time variable. Input vectors are generated based on each patient's last three examinations in LB2, or for all the examinations if the patient has less than three. The probabilities calculated by the machine learning algorithm based on these examinations are averaged to generate predicted probabilities for patient P_i at time t. When comparing the model's predictions against the actual diagnoses in LB4, t is set to the time difference between the applicable examination in LB2 and the applicable examination in LB4. For the time courses shown in Figs. 4 and 5, t is set to an integer between 1 and 84, inclusive.

2.4. Features

The features analyzed by the machine learning models consisted of 13 biomarkers or other types of clinical data present in the ADNI data set, all of which have been cited in published papers as correlating with AD progression. These 13 features are summarized in Table 1 and include genetic biomarkers (APOE4), physical biomarkers (ventricular volume/ICV ratio, hippocampal volume), the results of behavioral tests (ADAS13 and ADAS11 scores, FAQ score, MMSE score, and 4 types of RAVLT scores), and basic demographic information (age and race). As described previously, two additional features that were generated during preprocessing of the data using the all-pairs technique were also included in the models, namely the clinical diagnosis at the earlier of two examinations and the time difference between examinations.

3. Results

3.1. Model performance

A flowchart summarizing the overall methodology for training and evaluating the machine learning models is shown in Fig. 1. Various machine-learning classifiers, including support vector machines, logistic regression, gradient boosting classifiers, random forests, multilayer perceptron neural networks, and recurrent neural networks, were implemented using the Python libraries *Scikit-learn* and *Keras* (backed by *TensorFlow*). Each classifier was then evaluated on the processed data derived from LB1 using 7-fold cross-validation.

The effectiveness of each classifier was measured using a specialized version of a receiver operating characteristic area under the curve (ROC-AUC) score for multiclass classification ("mAUC score"), as previously described by Hand and Till [33]. The ROC-AUC score is a balanced metric for classifiers that considers both the true-positive rate (percentage of actual positives that are called correctly) and the false-positive rate (percentage of actual negatives that are called incorrectly). The mAUC variant of this score takes all ordered pairs of categories (i, j), measures the probability that a randomly selected element from category i would have a higher estimated chance of being classified as category i than a randomly selected element from category j, and averages all these probabilities. A classifier that works perfectly would have an mAUC score of 1; a classifier that guessed randomly would result in an mAUC score of 0.5.

In the end, two classifiers, a multilayer perceptron implemented in *Scikit-learn* ("MLP") and a recurrent neural network implemented in *Keras* ("RNN") were found to have the best performance in the cross-validation studies. Both these classifiers are types of neural networks. An MLP consists of a layer of input nodes, a layer of output nodes, and one or more hidden layers between the input and output layers. Each input vector is fed into the input nodes, and the value of each node in every other layer is dependent on the values of the nodes in the previous layer. Like an MLP, an RNN consists of multiple nodes organized into layers, but the outputs of some of the hidden layers are fed back into the same layer so that earlier input vectors can influence the
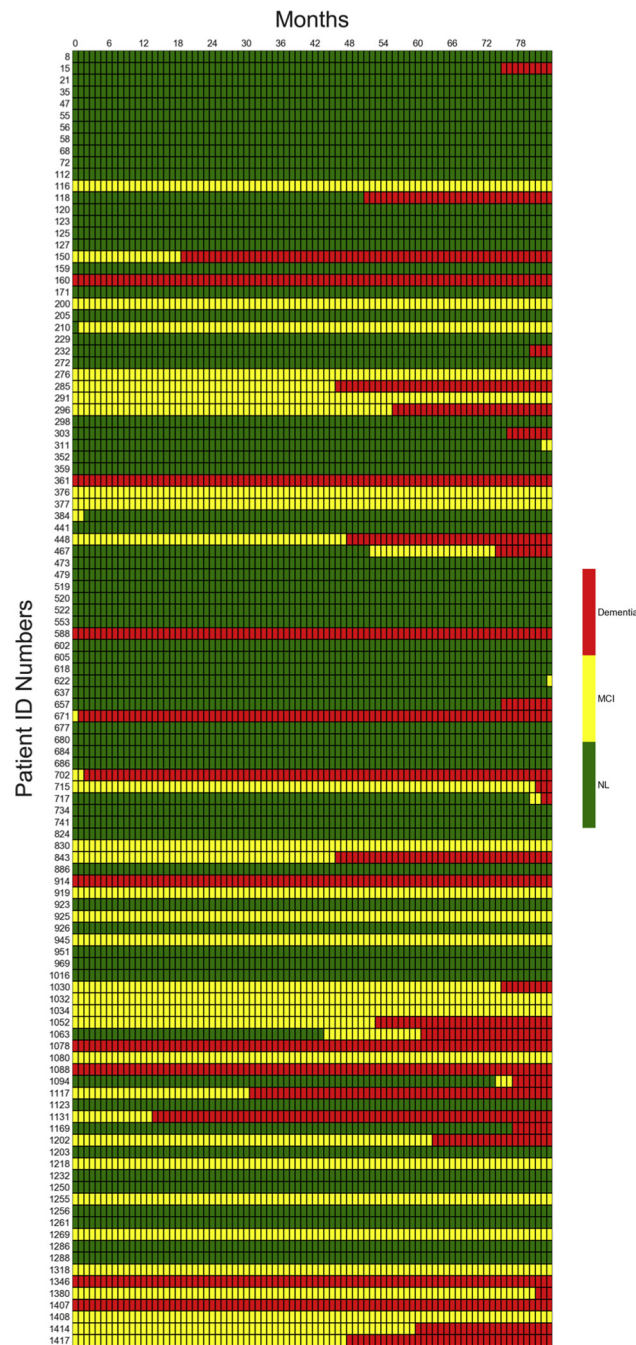
Fig. 4. Month-by-month predicted diagnoses for all 110 patients in LB2/LB4 data set over 7 years.

was applied to pairs of patients' doctors' visits or triplets of visits. Table 1 details which features were included in the 8-, 11-, and 15-feature training groups. The columns in Table 1 labeled "Variable" and "Meaning" provide the name of each feature as it appears in the ADNI data set as well as the corresponding description for each feature, respectively.

Each of the 12 models were trained and tested on LB1-derived data using 7-fold cross-validation. The best performing model was the MLP trained on 8 features and triplets of time points, which achieved an mAUC score of 0.9631. However, the differences among the mAUC scores for all the models were quite small and in many cases well within 1 standard deviation (SD), suggesting that some of the variability between the scores might be due to random chance rather than actual differences in predictive performance. In addition, because the training and testing of neural networks, as well as the cross-validation process, relies to some extent on algorithms that use random numbers, the scores will change slightly each time that the models are subjected to cross-validation. Consequently, a single set of cross-validation results may not provide a conclusive answer.

To assess the performance of these models in a more rigorous manner, each of the 12 models was also evaluated on a series of random splits of the preprocessed LB1 data set. For each model, the preprocessed LB1 data set was randomly separated into a training data set and testing data set using a 70:30 ratio, and this was repeated to create 100 pairs of training and testing data sets that were then used to train and test the model. The use of a large number of randomly generated splits produces a distribution of mAUC scores that better reflect the overall performance of the model and minimizes the effects of outliers. The results are shown as box-and-whisker plots in Fig. 2. Using this more rigorous approach, the model with the highest average mAUC score was an MLP trained on data with 15 features and with triplets of patients' examinations, which had an average mAUC score of 0.967 and a standard deviation of 0.0016. This model is highlighted in blue in Fig. 2.

3.2. Prediction of Alzheimer's disease progression

To assess real-world performance, select models were also trained on the entire LB1 data set (after processing with the all-pairs technique) and then evaluated on data derived from LB2 and compared to actual diagnoses in LB4, asking whether early biomarkers and other early clinical data for the 110 patients in LB2 can predict their later diagnoses. The actual examination dates in LB4 vary from patient to patient, but they generally cover the 7-year period of ADNI-GO/ADNI-2. LB4 contains a total of 417 examinations, or an average of 3.79 examinations per patient. As in the cross-validation studies, the performance of each model was assessed based on mAUC scores. Out of the 12 previously discussed models, the 4 models that used 15 features were selected for testing against LB2/LB4, namely

outputs for later input vectors. This allows the RNN to "remember" earlier inputs, which has been shown to be particularly useful when analyzing data consisting of multiple observations taken at different time points [34,35].

To further investigate the MLP and the RNN and see if the performance of the MLP and RNN could be further improved by optimizing the training protocol, six variants of each of the neural networks were generated to examine the effects of changing the number of features being examined (8, 11, and 15) and whether the all-pairs technique
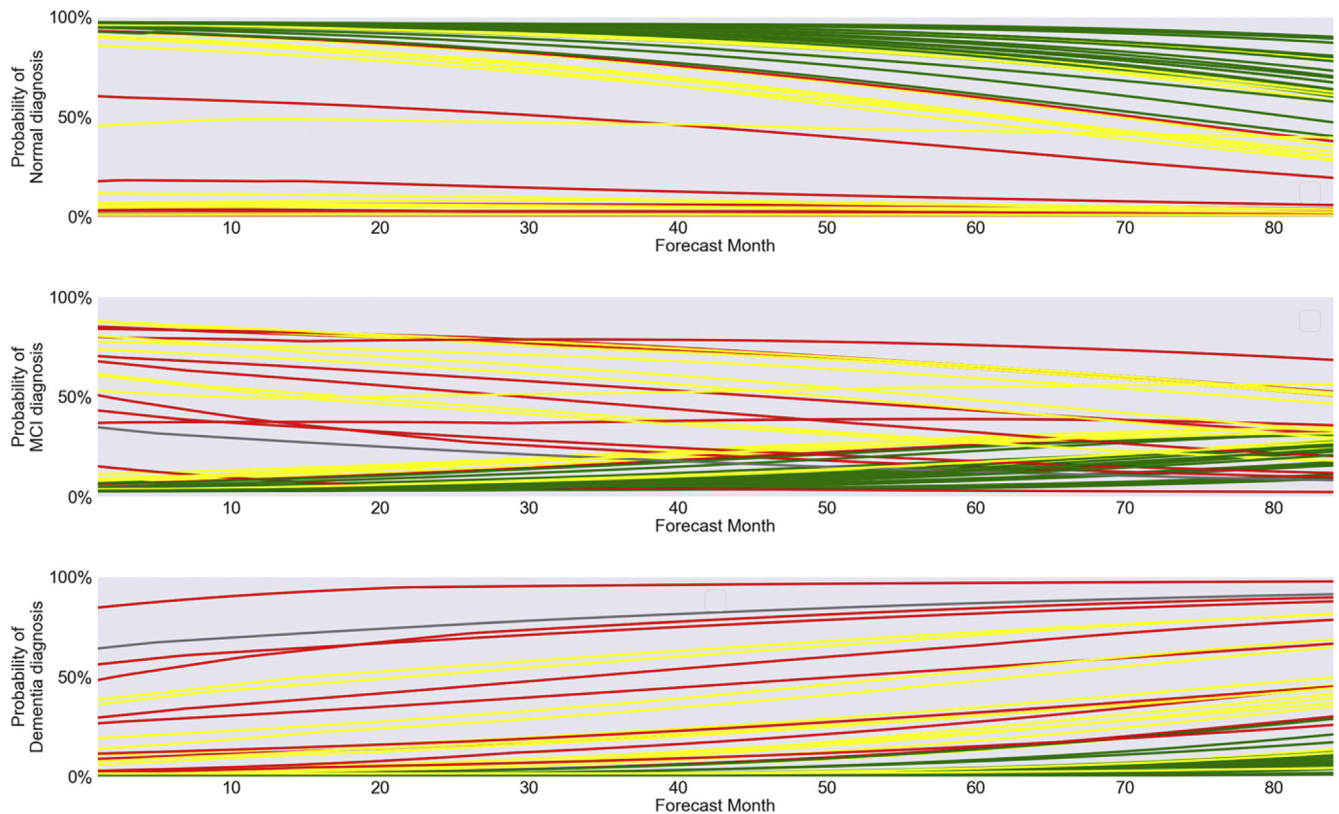
Fig. 5. Predicted likelihood of normal, MCI, and dementia diagnoses for 50 random patients in LB2/LB4 data set over 7 years. Abbreviation: MCI, mild cognitive impairment.

1. MLP trained using pairs of examinations,
2. MLP trained using triplets of examinations,
3. RNN trained using pairs of examinations, and
4. RNN trained using triplets of examinations.

Hyper-parameter optimization was conducted for each of these models, and each model was tested 5 times for each set of parameters to minimize the impact of random variation inherent in the neural network training process. In total, 27 different sets of parameters were tested for each model, consisting of 3 possible values for the alpha parameter, 3 possible values for the learning rate, and 3 possible values for the size of the hidden layers. Fig. 2 shows the results from the highest performing version of each of the 4 models listed previously.

The model and parameters with the best average mAUC score was the MLP trained on 15 features using pairs of examinations, which achieved an average score of 0.866 on the 110-patient test data set. This model is shown with a blue box in Fig. 2. This result represents an improvement over previously published work using the ADNI data set. In particular, Moore et al. [8] achieved an mAUC score of 0.82 with a random forest classifier; Ghazi et al. [34] achieved an mAUC score of 0.7596 with an RNN; and Nguyen et al. [35] achieved an average mAUC score of 0.86 with an RNN together with forward-filling data imputation.

Fig. 3 depicts a confusion matrix for this best performing model, which provides a visual indication of how well the diagnoses predicted by the model line up with the actual diagnoses. The confusion matrix reveals two types of mistakes occasionally made by the machine learning model: predicting a cognitively normal diagnosis when a patient is actually diagnosed with MCI and predicting an MCI diagnosis when a patient is actually diagnosed with dementia. These mistakes may both simply be the result of a small error in how the time variable is applied by the algorithm, which creates a lag in the diagnosis predictions.

Fig. 3 also shows a group of receiver operating characteristic (ROC) curves based on the output of the best performing model, which measure how well the model can separate two groups: patients with a particular diagnosis and patients with one of the other diagnoses. Interestingly, the ROC score for the MCI class is lower than the other two ROC scores, suggesting that the model is having more difficulty separating patients with MCI from the other two groups. Based on these individual ROC scores, the model's average mAUC score (0.866) could be dramatically increased if the model's ability to separate MCI patients from non-MCI patients was improved.

The best performing model was also used to predict the future diagnosis of normal, MCI, or dementia for all 110 patients in the LB2/LB4 data sets on a month-to-month basis over an 84-month (7-year) period, as shown in Fig. 4. The

model predicted that some patients would remain normal over the entire 7-year period, while others would progress from normal to MCI to dementia.

Time courses were also generated for a random subset of these patients (n = 50) showing how the likelihood of normal, MCI, or dementia diagnoses is forecasted to vary over the 84 months, as shown in Fig. 5. The color of each curve (green, yellow, or red) indicates the most severe diagnosis actually received by the patient (normal, MCI, or dementia, respectively) during the 7-year period. As can be seen in Fig. 5, patients who remained normal over the entire 84-month period generally received very low predicted probabilities of MCI or dementia diagnoses from the model. Similarly, patients who were diagnosed with dementia at some point during this period generally received high predicted probabilities of dementia from the model.

4. Discussion

Previous work published by others has shown that machine learning algorithms can accurately classify a patient's current cognitive state (normal, MCI, or dementia) using contemporaneous clinical data [30–32]. This project has extended this previous work by looking at how past and present clinical data can be used to predict a patient's future cognitive state and by developing machine learning models that can correlate clinical data obtained from patients at one time point with the progression of AD in the future. Several of the machine-learning models used in this project were effective at predicting the progression of AD, both in cognitively normal patients and patients suffering from MCI. In addition, a novel all-pairs technique was developed to compare all possible pairs of temporal data points for each patient to generate the training data set. By comparing data points at different points in time, the all-pairs technique adds time as a variable and therefore does not require fixed time intervals, which are unlikely to occur in "real-life" data [36]. These techniques could be used to identify patients having high AD risk before they are diagnosed with MCI or dementia and who would therefore make good candidates for clinical trials for AD therapeutics. Because the inability to identify AD patients at early stages is believed to be one of the primary reasons for the frequent failure of AD clinical trials, these techniques may help increase the chances of finding a treatment for AD.

RESEARCH IN CONTEXT

1. Systematic review: Literature on machine learning techniques for the prediction or the diagnosis of Alzheimer's disease (AD) was reviewed, including those that also used the Alzheimer's Disease Neuroimaging Initiative data set as a source of AD patient data.

2. Interpretation: This work suggests that machine learning models, such as the neural networks used in this paper, can be used to make accurate predictions of the future clinical state of AD patients. This paper introduced a novel preprocessing algorithm (the all-pairs technique) that facilitates the effective training of such models.

3. Further directions: To further ascertain the ability of the machine learning model described in this paper to make accurate assessments of AD patients, future work may include evaluation of the model's performance on other data sets besides the Alzheimer's Disease Neuroimaging Initiative data set, and the model's performance using different biomarkers than those chosen for this paper.

References

[1] Selkoe DJ, Lansbury PJ Jr. Alzheimer's disease is the most common neurodegenerative disorder. In: Siegel GJ, Agranoff BW, Albers RW, Fisher SK, Uhler MD, eds. Basic Neurochemistry: Molecular, Cellular and Medical Aspects. Philadelphia: Lippincott-Raven; 1999.

[2] Mathotaarachchi S, Pascoal TA, Shin M, Benedet AL, Kang MS, Beaudry T, et al. Identifying incipient dementia individuals using machine learning and amyloid imaging. Neurobiol Aging 2017;59:80–90.

[3] Reisberg B, Gordon B, McCarthy M, Ferris SH. Clinical symptoms accompanying progressive cognitive decline and Alzheimer's disease. In: Alzheimer's dementia. Totowa, NJ: Humana Press; 1985. p. 19–39.

[4] Humpel C. Identifying and validating biomarkers for Alzheimer's disease. Trends Biotechnol 2011;29:26–32.

[5] Liu CC, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. Nat Rev Neurol 2013;9:106.

[6] Mu Y, Gage FH. Adult hippocampal neurogenesis and its role in Alzheimer's disease. Mol Neurodegeneration 2011;6:85.

[7] Teri L, McCurry SM, Edland SD, Kukull WA, Larson EB. Cognitive decline in Alzheimer's disease: a longitudinal investigation of risk factors for accelerated decline. J Gerontol Ser A Biol Sci Med Sci 1995; 50A:M49–55.

[8] Moore P, Lyons T, Gallacher J. Random forest prediction of Alzheimer's disease using pairwise selection from time series data. PLoS One 2019;14:e0211558.

[9] Neugroschl J, Wang S. Alzheimer's disease: diagnosis and treatment across the spectrum of disease severity. Mt Sinai J Med 2011; 78:596–612.

[10] Marinescu RV, Oxtoby NP, Young AL, Bron EE, Toga AW, Weiner MW, et al. TADPOLE Challenge: Prediction of Longitudinal Evolution in Alzheimer's Disease; 2018. arXiv: 1805.03909 [q-bio.PE].

[11] Singanamalli A, Wang H, Madabhushi A. Cascaded multi-view canonical correlation (CaMCCo) for early diagnosis of Alzheimer's disease via fusion of clinical, imaging and omic features. Scientific Rep 2017; 7:8137.

[12] Gamberger D, Lavrač N, Srivatsa S, Tanzi RE, Doraiswamy PM. Identification of clusters of rapid and slow decliners among subjects at risk for Alzheimer's disease. Scientific Rep 2017;7:6763.

[13] The need for early detection and treatment in Alzheimer's disease. EBioMedicine 2016;9:1–2.

[14] Alzheimer's Association. What we know today about Alzheimer's disease. https://www.actionalz.org/research/science/alzheimers_disease_treatments.asp; 2010. Accessed September 10, 2019.

[15] Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. Alzheimer's Res Ther 2014;6:37.

[16] Mumal I. Poor results prompt Takeda and Zinfandel to end phase 3 Alzheimer's therapy trial early. https://alzheimersnewstoday.com/2018/01/31/takeda-and-zinfandel-bring-early-end-to-phase-3-trial-of-alzheimers-therapy-pioglitazone/; 2018. Accessed September 10, 2019.

[17] Cortez M. Merck will end Alzheimer's trial as alternative approach fails. https://www.bloomberg.com/news/articles/2018-02-13/merck-will-end-alzheimer-s-trial-as-alternative-approach-fails; 2018. Accessed September 10, 2019.

[18] U.S. Food and Drug Administration. https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM596728.pdf; 2018. Accessed September 10, 2019.

[19] Gottlieb S. Statement from FDA commissioner Scott Gottlieb, M.D. on advancing the development of novel treatments for neurological conditions; part of broader effort on modernizing FDA's new drug review programs. https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm596897.htm; 2018. Accessed September 10, 2019.

[20] Teipel S, Drzezga A, Grothe MJ, Barthel H, Chételat G, Schuff N, et al. Multimodal imaging in Alzheimer's disease: validity and usefulness for early detection. Lancet Neurol 2015; 14:1037–53.

[21] Oxtoby NP, Alexander DC. Imaging plus X: multimodal models of neurodegenerative disease. Curr Opin Neurol 2017; 30:371–9.

[22] Nestor SM, Rupsingh R, Borrie M, Smith M, Accomazzi V, Wells JL, et al. Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's Disease Neuroimaging Initiative database. Brain 2008;131:2443–54.

[23] ADNI Staff. ADNI general procedures manual. https://adni.loni.usc.edu/wp-content/uploads/2010/09/ADNI_GeneralProceduresManual.pdf; 2010. Accessed September 10, 2019.

[24] Kueper JK, Speechley M, Montero-Odasso M. The Alzheimer's disease assessment scale-cognitive subscale (ADAS-Cog): modifications and responsiveness in pre-dementia populations. a narrative review. J Alzheimer's Dis 2018;63:423–44.

[25] Wan J, Zhang Z, Rao BD, Fang S, Yan J, Saykin AJ, et al. Identifying the neuroanatomical basis of cognitive impairment in Alzheimer's disease by correlation-and nonlinearity-aware sparse Bayesian learning. IEEE Trans Med Imaging 2014;33:1475–87.

[26] Tekin S, Fairbanks LA, O'Connor S, Rosenberg S, Cummings JL. Activities of daily living in Alzheimer's disease: neuropsychiatric, cognitive, and medical illness influences. Am J Geriatr Psychiatry 2001; 9:81–6.

[27] Balsis S, Benge JF, Lowe DA, Geraci L, Doody RS. How do scores on the ADAS-Cog, MMSE, and CDR-SOB correspond? Clin Neuropsychologist 2015;29:1002–9.

[28] Boone KB, Lu P, Wen J. Comparison of various RAVLT scores in the detection of noncredible memory performance. Arch Clin Neuropsychol 2005;20:301–19.

[29] Pfeffer RI, Kurosaki TT, Harrah CH, Chance JM, Filos S. Measurement of functional activities in older adults in the community. J Gerontol 1982;37:323–9.

[30] Esmaeilzadeh S, Belivanis DI, Pohl KM, Adeli E. End-to-end Alzheimer's Disease Diagnosis and Biomarker Identification; 2018. arXiv: 1810.00523 [cs.CV].

[31] Long X, Chen L, Jiang C, Zhang L. Prediction and classification of Alzheimer disease based on quantification of MRI deformation. PLoS One 2017;12:1–19.

[32] Zhang D, Wang Y, Zhou L, Yuan H, Shen D. Multimodal classification of Alzheimer's disease and mild cognitive impairment. NeuroImage 2011;55:856–67.

[33] Hand DJ, Till RJ. A simple generalisation of the area under the ROC curve for multiple class classification problems. Machine Learn 2001;45:171–86.

[34] Ghazi MM, Nielsen M, Pai A, Cardoso MJ, Modat M, Ourselin S, et al. Robust Training of Recurrent Neural Networks to Handle Missing Data for Disease Progression Modeling; 2018. arXiv: 1808.05500 [cs.CV].

[35] Nguyen M, Sun N, Alexander DC, Feng J, Yeo BT. Modeling Alzheimer's disease progression using deep recurrent neural networks. In: 2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI). IEEE; 2018. p. 1–4. https://doi.org/10.1109/PRNI.2018.8423955.

[36] Bernal-Rusiel JL, Greve DN, Reuter M, Fischl B, Sabuncu MR, Martinos AA. Statistical analysis of longitudinal neuroimage data with linear mixed effects models. Neuroimage 2013;66:249–60.