UNDERGRADUATE RESEARCH IN NATURAL AND CLINICAL SCIENCE AND TECHNOLOGY (URNCST) JOURNAL
Read more URNCST Journal articles and submit your own today at: https://www.urncst.com

PRIMARY RESEARCH OPEN ACCESS

Feasibility Study: Machine Learning in Neurodegenerative Disorders, Alzheimer's Disease

Xiangxuan Kong, HBSc Student [1]*

[1] Department of Computer Science, Faculty of Arts and Science, University of Toronto, Toronto, Ontario, Canada M5S 1A4

*Corresponding Author: noah.kong@mail.utoronto.ca

Abstract

Introduction: Clinical decision support systems (CDSSs), powered by machine learning and artificial intelligence, have demonstrated potential in clinical diagnosis and intervention for neurological and psychiatric disorders. Considering the importance of early detection and intervention of Alzheimer's disease (AD), this study aims to explore the potential of a data-driven non-knowledge-based machine learning CDSS for predicting AD diagnoses in individuals. In non-knowledge-based CDSSs, no prior knowledge about AD or any other disorder impacts the decision-making of classification models

Method: In this study, publicly available data of 14037 data points collected by the Alzheimer's Disease Neuroimaging Initiative were used for model training and testing. Binary classification and multiclassification machine learning were applied, and results from six mainstream classification models were analyzed.

Results: The binary classification models (AD diagnosis present or absent) gave accuracies around 0.92-0.93, and the multiclassification models gave accuracies around 0.85-0.87. Logistic regression model (binary classification) had the highest overall hit rate (0.93). This model maintained this hit rate when only features with over 90% non-empty data are available.

Discussion: Binary classification models are more reliable for diagnosing AD than multiclassification models. The high hit rates of the logistic regression model (binary classification) on generally available data implicate its feasibility.

Conclusion: There is strong potential for a complete machine learning-based CDSS to aid in AD diagnoses in the future.

Keywords: clinical decision support system; Alzheimer's disease; artificial intelligence; machine learning

Introduction

As a dominant technique of artificial intelligence (AI), machine learning stimulates the development of new patents and involves more than a third of the invented AI [1]. As the name suggests, AI is intelligence made by humans, usually demonstrated by machines. Daily applications of AI include the recommendation systems used by mainstream social media and search engines like Google. In addition, AI may improve medical diagnosis as a type of machine learning that has shown great potential in precision medicine [2].

Machine learning (ML) is a computational method which analyzes and learns patterns from data and makes predictions using the new input data [3]. Instead of producing programmed or designed outcomes, a machine learning algorithm gives answers based on what it has learned from the data it had been previously trained on. The three main approaches to machine learning are unsupervised, supervised, and reinforcement learning [4]. In supervised learning, the output values are provided in training, while in unsupervised learning, data are unlabeled. Usually, supervised learning is used to make predictions about outcomes, and unsupervised learning is used to reveal relationships in the data. Machine learning is widely applied across various fields, including economics, bioinformatics, and medical diagnosis [2]. This study focuses on utilizing supervised machine learning to aid in diagnoses of neurodegenerative disorders.

A non-knowledge-based clinical decision support system (CDSS) uses machine learning to study the pattern of a given dataset and gives case-specific suggestions based on this training data to apply to new data [5]. A CDSS can provide clinicians with recommendations about clinical treatments or early warnings, which helps increase the efficiency and reliability of medical interventions and diagnoses. Evidence suggests an added benefit to utilizing a CDSS in psychiatric settings to help front-line health care. For example, research has shown that CDSSs can help predict the effect of pharmaceutical medications for individuals with psychiatric disorders and inform early intervention for those engaging in suicidal thoughts [6,7].

CDSSs have also been created using clinical data to aid in diagnoses of neurodegenerative diseases such as Alzheimer's disease (AD). AD is a chronic neurodegenerative disease that is relatively common among older adults over 65 years old, causing 60% to 70% of dementia cases [8]. The gold standard for diagnosis is based on medical imaging and mainstream cognitive tests. These tests include Mini-Mental State Examination (MMSE) and some sub-tests of the Preclinical Alzheimer's Cognitive Composite (PACC). PACC is a composite score that combines the results of different neuropsychological assessments and tests the potential cognitive decline of a cognitively normal population to inform early detection of AD [9]. Other auxiliary examinations include routine blood examinations, clinical urine samples, and genomic tests. Difficulty in memory, behavioural symptoms, loss of motivation, and language problems are symptoms of AD [10]. As the disease advances, functional independence is expected to deteriorate, and life expectancy is expected to shorten [11]. The syndrome known as mild cognitive impairment (MCI) is of particular importance, which represents a prodromal stage of dementia between average aging-related cognitive decline and severe dementia. Participants with MCI show deficits in memory and executive functioning, but the deficiencies are not as severe as to influence independent functioning. However, the cause of AD is still poorly understood [12]. Therefore, early detection and intervention of AD are crucial for prevention efforts, which CDSSs can aid. For example, a knowledge-based CDSS was applied to aid in diagnosing AD and validated in three hospitals in Spain [13]. In addition to the patient's clinical data, such as results from MRI scans, this system also involves clinical judgement and reasoning based on prior knowledge about AD diagnosis. Whether a non-knowledge-based CDSS can aid in AD diagnoses is unclear.

Therefore, this preliminary study aimed to assess the feasibility of a supervised non-knowledge-based machine learning algorithm specific to AD. As the name suggests, no prior knowledge about AD influences the model's learning and decisions. Instead, this machine learning model uses pre-existing datasets as training data and provides classification responses based on new, untrained data. In this paper, classification models from non-knowledge-based CDSSs were trained to predict from a given dataset which individuals were cognitively normal (CN) or had MCI or AD. Six models were examined: XGBoost, Random Forest, K-Nearest Neighbors, Support Vector Machine, logistic regression and ensemble methods. In addition to examining the feasibility of a non-knowledge-based machine learning algorithm in AD diagnoses, some potential correlations between the other demographic or clinical variables and AD diagnosis were analyzed. Training the non-knowledge-based CDSS with collected data may, in the long term, offer implications for future studies on prevention efforts most effective for AD.

Methods

Data Description

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org.

Starting in 2004, ADNI collects neuroimaging, clinical, and cognitive test data from older adult research participants aged 55 to 90 years from Canada and the US and combines expertise and funding from both private and public sectors to improve the prevention and treatment of AD. As part of this initiative, neuropsychological tests such as the Clinical Dementia Rating Scale (CDR) and a modified version of the PACC (mPACC) are administered to participants longitudinally.

The dataset involves 113 variables and 14037 data rows, with each row representing data from a single timepoint from a participant in ADNI. In addition, the variables include the results of the five primary cognitive tests for *AD*, including the ADAS, MMSE, Rey Auditory Verbal Learning Test, Montreal Cognitive Assessment, and Everyday Cognition Scale. In addition, some neuroimaging measures like brain ventricle volumes and other demographic variables, including sex, gender, education level, and race, were also included.

The output variable in this study was the presence or absence of AD. The algorithm was trained to output three possible diagnostic outcomes for this variable: CN, MCI, or AD. The dataset's dependent variable (diagnosis) consists of 3310 CN outcomes, 2230 AD outcomes and 4410 MCI outcomes.

Data Analysis

Before model training, the data underwent imputation, i.e., the substitution of missing data using statistical measures of central tendency (mean, mode or median) of available data, as well as normalization, i.e., organizing data to reduce data redundancy and increase integrity. Data without a value for the output variable were removed (n = 4087). To balance the distribution of the three possible outcome levels, some data rows with CN or MCI outcomes were removed from the remaining data, and the number of deleted rows depended on the results of the train test split, which are mentioned in the Results section below. In the binary and multi-classification models, all data columns were accepted. The mean of the column replaced missing numeric data, and the mode of the column replaced missing categorical data. Afterwards, two datasets were created: a

training and testing datasets. Seventy percent of the remaining data were randomly chosen to be placed in the training dataset, and the other 30% were automatically put into the testing dataset.

Six binary classification models (XGBoost, Random Forest, Support Vector Machine, K-Nearest Neighbors, logistic regression and ensemble methods that combine the previous five models) were built through Python software [14]. In these binary classification models focusing on the presence or absence of AD specifically, CN/MCI results were labelled as 0, and AD outcomes were marked as 1. Six multiclassification models were also made to account for the MCI, where CN outcomes were labelled as 0, MCI were labelled as 1, and AD outcomes were labelled as 2.

Because of the differences in binary versus multiclassification, the training data and testing data for each type of classification underwent an undersampling phase, resulting in different sample sizes for binary and multiclassification models. Applying the Boruta feature selection algorithm to the training data eliminated irrelevant variables from the training and testing data through feature selection to reduce the model's complexity. In addition, features unrelated to the output variable, for example, the participant ID number, were eliminated.

First, these six models studied the training data. Since the model is supervised, the values of the output variable in training data were accessible by the models. Finally, these models were then tested by the testing data. The predictions made by the models according to the selected input variables were compared with the actual output. Confusion matrices of accuracy, negative predictive value (the ratio of rejections predicted by the model to correct rejections), positive predictive value (the ratio of predicted hits to correct hits), areas under the curve (AUCs) from a receiver operating curve, and SHAP value summary plots were demonstrated.

The machine learning method that gave the highest accuracy was trained and tested solely by the features with over 90% non-missing data, and these results were examined. These features were also compared with the important features marked by Boruta algorithm before the binary classification and multiclassification model training.

Results

After the undersampling phase, the training data size was 4659, which included 1553 data points for each class, and the testing data size was 2031 data points, which had 677 data points for each class category.

Feature Selection with Boruta

Table 1. Features marked as "important" by the Boruta algorithm with the XGBoost classifier

Feature Name	Meaning
CDR-SB	Clinical Dementia Rating Scale Sum of Boxes score, an indicator of global cognitive dysfunction.
MMSE	Score on the Mini-Mental State Examination, a test of global cognition (i.e., participants' orientation, attention and calculation, recall, and language ability)
mPACCdigit	Score on the ADNI modified Preclinical Alzheimer's Cognitive Composite (PACC) with Digit Symbol Substitution. It tests the potential cognitive decline of cognitively normal population
mPACCtrailsB	ADNI modified PACC with Trails B, a test of cognitive flexibility
Ethnicity_Hisp/Latino	Ethnic identity as Hispanic or Latino
Race_Black	Whether the participant identifies as Black
Race_White	Whether the participant identified as White
Marital_Widowed	Whether the participant is widowed
Marital_Never married	Whether the participant has never been married
FLDSTRENG_1.5 Tesla MRI	Whether the field strength of MRI used was 1.5 Tesla or 3.0 Tesla

Multiclassification Accuracy

Table 2. Results from multiclassification models (created in Python)

Model	Accuracy	NPV	PPV	AUC-ROC	Hit rate AD	Hit rate MCI	fHit rate CN
XGBoost model	0.863	[0.938, 0.904, 0.951]	[0.908, 0.788, 0.896]	0.965	0.90	0.79	0.91
Random Forest	0.873	[0.938, 0.920, 0.95]	[0.929, 0.788, 0.912]	N/A	0.91	0.79	0.9
Logistic regression	0.853	[0.962, 0.859 , 0.967]	[0.836, 0.839, 0.881]	0.947	0.88	0.84	0.84
Support Vector Machine	0.877	[0.962, 0.902, 0.952]	[0.888, 0.825, 0.918]	not applicable	0.92	0.83	0.89
K-Nearest Neighbors	0.856	[0.946 0.888 0.950]	[0.872, 0.790 0.904]	0.948	0.9	0.79	0.87
Ensemble Method	0.874	[0.947, 0.908, 0.956]	[0.907, 0.807, 0.909]	0.969	0.91	0.81	0.91

Confusion Matrices of Models

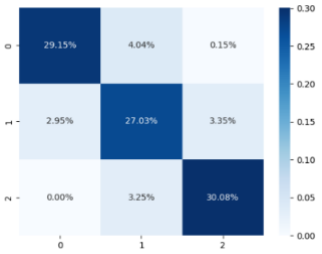

Figure 1A

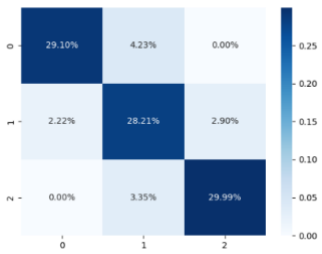

Figure 1B

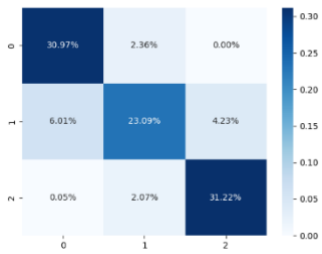

Figure 1C

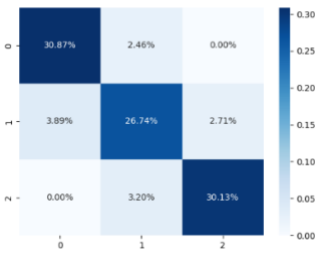

Figure 1D

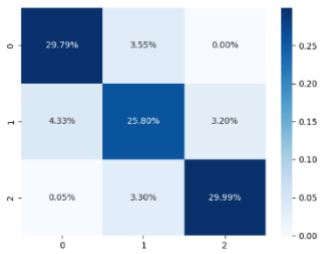

Figure 1E

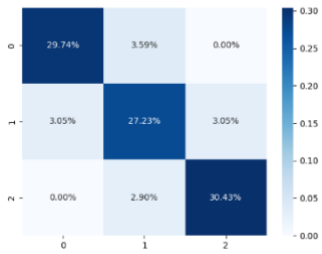

Figure 1F

Figure 1. Confusion matrices of multiclassification models for the A) XGBoost model; B) Random Forest model; C) logistic regression model; D) support vector machine model; E) K-Nearest Neighbors model; and F) ensemble method. Figure created using Python.org.

SHAP Value Summary Plots for XGBoost and Random Forest Models

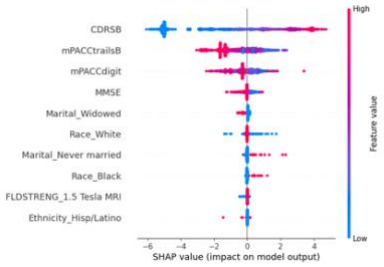

Figure 2A

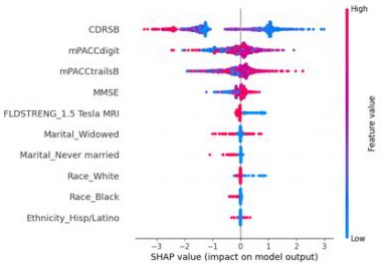

Figure 2B

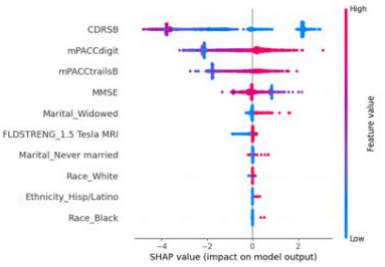

Figure 2C

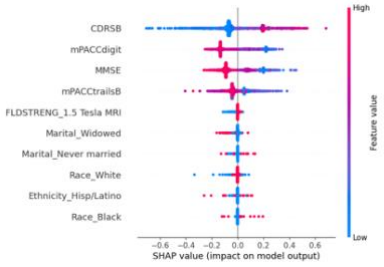

Figure 2D

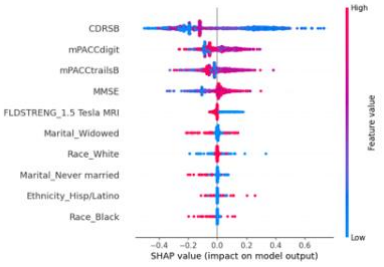

Figure 2E

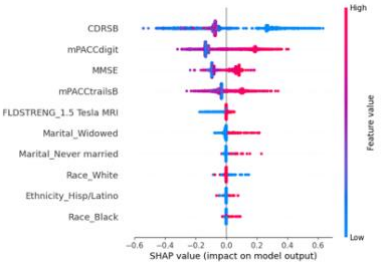

Figure 2F

Figure 2. SHAP value summary plots (created in Python) of models (if applicable) for three classes. Figure 2A–C show the plots of XGBoost models, for AD, MCI and CN outcomes respectively. Figure 2D–F show the plots of Random Forest models, for AD, MCI and CN outcomes respectively. Figure created using Python.org.

Binary Classification

After undersampling, the training data size was 3106 data points, with 1553 AD outcomes and 1553 non-AD outcomes. For testing data, this was withs 1354 data points with 677 AD and 677 non-AD outcomes.

Feature Selection With Boruta

Table 3. Features that are marked as "important" by the Boruta algorithm with XGBoost classifier

Feature Name	**Meaning**
CDR-SB	Clinical Dementia Rating Scale Sum of Boxes score, an indidcator of cognitive dysfunction.
MMSE	Mini-Mental State Examination, measuring the participants' orientation, registration, attention and calculation, recall, and language ability.
mPACCdigit	ADNI modified Preclinical Alzheimer's Cognitive Composite (PACC) with Digit Symbol Substitution. It tests the potential cognitive decline of cognitively normal population
mPACCtrailsB	ADNI modified Preclinical Alzheimer's Cognitive Composite (PACC) with Trails B. It tests the potential cognitive decline of cognitively normal population.
FLDSTRENG_1.5 Tesla MRI	Whether the field strength of MRI used is 1.5 Tesla

Binary Classification Accuracy

Table 4. Binary classification models and their results

Model	Accuracy	NPV	PPV	Hit rate AD	Hit rate non-AD (CN or MCI)
XGBoost model	0.919	0.914	0.925	0.91	0.93
Random Forest	0.915	0.912	0.918	0.91	0.92
Logistic regression	0.933	0.931	0.935	0.93	0.93
Support Vector Machine	0.932	0.922	0.943	0.92	0.94
K-Nearest Neighbors	0.921	0.915	0.927	0.92	0.93
Ensemble method	0.931	0.926	0.936	0.93	0.94

Confusion Matrices of Models (XGBoost, Random Forest, Logistic Regression, Support Vector Machine, K-Nearest Neighbors, Ensemble Method)

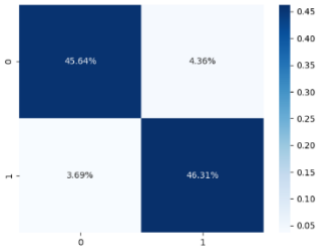

Figure 3A

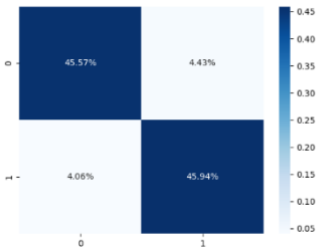

Figure 3B

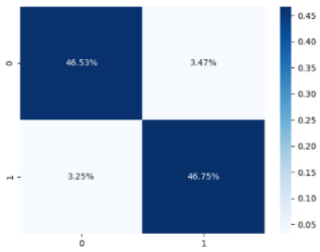

Figure 3C

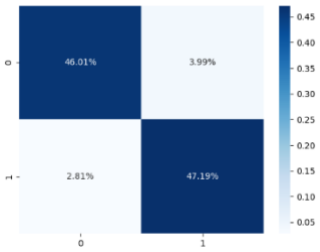

Figure 3D

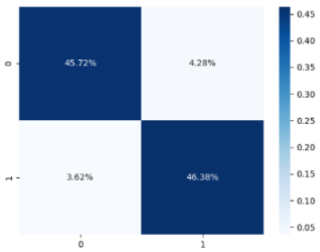

Figure 3E

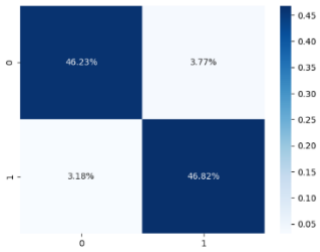

Figure 3F

Figure 3. Confusion matrices of binary classification models created in Python (Figure 3A shows the confusion matrix of XGBoost model; Figure 3B shows the confusion matrix of Random Forest model; Figure 3C shows the confusion matrix of logistic regression model; Figure 3D shows the confusion matrix of Support Vector Machine model; Figure 3E shows the confusion matrix of K-Nearest Neighbors model; Figure 3F shows the confusion matrix of ensemble method). Figure created using Python.org.

SHAP Value Summary Plots of Models (if applicable) for AD Predictions and Non-AD Predictions

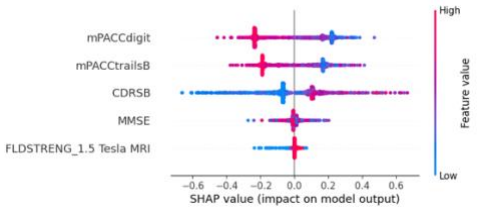

Figure 4A

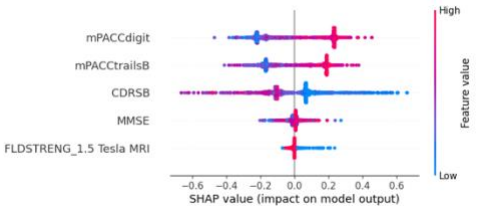

Figure 4B

Figure 4. SHAP value summary plots of models for two classes. Figure 4A and 4B show the SHAP value summary plots of the Random Forest model for AD outcomes and non-AD outcomes, respectively. Figure created using Python.org.

Binary classification Using Logistic Regression, Using Only Features with Over 90% Non-Missing Data

After undersampling, the training data size is 3106, with 1553 AD and 1553 non-AD and that of testing data is 1354 with 677 AD and 677 non-AD.

Features Selection

Note that the Boruta Algorithm does not apply to the data; only features with over 90% non-missing data are kept.

Selected features (variables in Python) include age, sex, education, RAVLT, ethnicity, race, martial status, ApoE4 genotype (a genetic risk factor for AD), CDRSB, ADAS11, ADAS13, ADASQ4, MMSE, RAVLT immediate, RAVLT_learning, RAVLT_forgetting, RAVLT_perc_ forgetting, TrailsB, FAQ, mPACCdigit, and mPACCtrailsB.

RAVLT_learning and RAVLT _forgetting stand for the scores in RAVLT learning section and forgetting section. RAVLT_perc_forgetting stands for the RAVLT Percent Forgetting. RAVLT immediate stands for the sum of the 5 trials in RAVLT.

Logistic Regression Model (Binary) Using Only Features with Over 90% Non-Missing Data

The overall accuracy was 0.932. The NPV and PPV values are 0.933 and 0.931, respectively. The hit rates for AD and non-AD (CN or MCI) are both 0.93. Below is the confusiton matrix for the model.

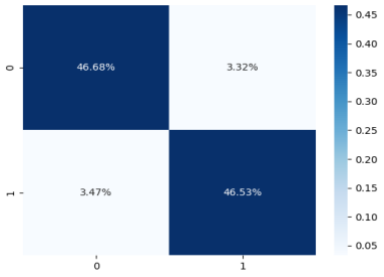

Figure 5. Confusion matrix for logistic regression model using only features with over 90% non-missing data. Figure created using Python.org.

Discussion

Interpretation of Results

This study aimed to show the potential of a non-knowledge-based CDSS in neurodegenerative disorders like AD. Notably, the scores on standardized cognitive tests, including the MMSE, CDR-SB, and ADNI-modified PACC, were strongly related to AD diagnosis. The impact of these features was more critical than the other demographic variables, such as race and marital status. According to the SHAP value summary plots (Figure 2, Figure 4), the CDR-SB had the highest absolute SHAP value, meaning that CDR-SB is a significant factor in determining AD outcomes. The next highest were mPACCdigit, mPACCtrailsB and MMSE. The CDR is a brief assessment of cognitive dysfunction [15]. Aspects of memory, orientation, judgement and problem solving, as well as community affairs such as home life, hobbies, and personal care, were evaluated, with a 0-3 scale where 0 refers to no impairment, and 3 refers to severe impairment. CDR-SB refers to the sum of all boxes (i.e., total score) and is previously validated to assess the stages the dementia severity and distinguish MCI from dementia [16].

According to the SHAP value summary plots (Figure 2A, Figure 2C, Figure 2D and Figure 2F), Individuals with higher scores on the CDR-SB are more likely to be predicted to have AD.

This study's other three essential features that were important to AD diagnoses were the MMSE, mPACC digital symbol substitution test and Trail Making Test (TMT) part B [17]. The Digit Symbol Substitution Test (DSST) is an assessment of executive functioning. Participants are provided with nine digit-symbol pairs and are then asked to write down the corresponding symbol as fast as possible under each digit [18]. DSST can help indicate brain injury, dementia, age, and depression as a neuropsychological test [19]. The MMSE measures the participants' orientation, attention recall, and language ability, with a final score out of 30. Cognitive impairment is indicated by a score lower than 23 [20].

The Trail Making Test (TMT) is one of the most popular neuropsychological tests of executive functioning (as well as visual search and processing speed). In TMT part B, participants are given numbers and letters randomly distributed on paper and are instructed to sequentially draw lines connecting the numbers and letters in an alternating way (e.g., 1, A, 2, B, 3, C, 4, D etc.) [19].

DSST and TMT part B scores are positively associated with CN predictions and negatively associated with AD predictions. In other words, the machine learning models are more likely to make CN predictions on individuals with higher DSST and TMT part B scores. MMSE results were not as impactful on AD predictions as on the three previous cognitive tests. Still, it is noticeable that the MMSE score positively correlates with MCI predictions, according to the Random Forest and XGBoost SHAP (Figure 2B and Figure 2E) value plot.

In addition to the cognitive test results, participants' ethnicity and marital status were also influential for the XGBoost and Random Forest multiclassification model predictions. According to their SHAP value summary plots (Figure 2A and Figure 2D), AD predictions were positively associated with people who identify as Black. Whether participants identified as White did not have a noticeable effect on the model output. The XGBoost model implies that non-White participants are more likely to receive MCI and AD predictions. This disparity may result from the fact that people from other ethnic minorities and diverse communities have less access to health care than people who identify as White or Caucasian [21]. Recent perspectives suggest more studies about the potential relationship between race and AD for scientific and ethical justifications [22]. Previous research has suggested an association between race and AD; for example, people with AD who identify as African American and Latino may have greater postdiagnosis survival rates than people with AD who identify as White [23].

As for marital status, both XGBoost and Random Forest multiclassification models clearly illustrate that widowed participants were more likely to be cognitively normal, and the widowed marital status seems to be negatively correlated with AD predictions. Furthermore, according to the Random Forest and XGBoost models, participants who have never married are more likely to be predicted to have healthy cognition. The two models suggest that loneliness or isolation reduces the probability of giving MCI prediction. In other words, people with no marriage experience are less likely to have MCI. In addition, it is shown by the two models, especially the XGBoost model, that the lack of marriage experience is a mild positive influencer for AD prediction. However, the implication about the relationship between marriage and AD, suggested by this study, remains questionable. The model output contradicts with the conclusions of a previous cohort study of 257 old adults in 2020 which indicates that widowhood may be a potential risk factor for AD-related cognitive decline [24]. Also, it is claimed by one study, with a sample of over 800 000 participants in the UK, that people who have married have a lower risk of dementia than widowed and lifelong single people [25].

The last important feature selected is the field strength of MRI used, measured in Tesla (T). In contrast to race and marital status variables mentioned above, this feature was also selected by the Boruta algorithm in binary classification. In the dataset, field strength is a categorical variable, with 1.5T and 3.0T as two possible values for this variable. The SHAP value summary plots (Figure 2B-2C, and Figure 2E-2F) in multiclassification suggest that the participants using MRI of field strength 1.5T tend to be CN or MCI. But in binary classification, using 1.5T MRI correlates positively with the AD predictions and negatively with the non-AD predictions. Research suggests that the scanning results of the brain with this 2-field strength are mostly the same, except for some regional differences in white matter and grey matter in deep brain structures, cerebellum, and brainstem [26].

Binary classification models tend to have higher accuracy for AD predictions, usually between 0.91 and 0.93, than multi-classification models (around 0.85-0.88), which may result from variability by including those with MCI. Most of the false positives for AD and CN predictions fall into the MCI category. However, because of under-sampling, multiclassification models can use a larger sample size (6690) than binary classification models (4460). In multiclassification, all models except for the logistic regression classifier provide around accuracies of 0.9 for predicting AD, and the Support Vector Machine model and ensemble method offer the most accurate predictions. In binary classification, the logistic regression classifier provides the highest overall accuracy (0.933) among the machine learning models in binary classifications. Support Vector Machine and the ensemble method perform well under both situations. The logistic regression binary classifier using features with 90% or above non-missing data only was also built to simulate the

real-world problem. Without using the Boruta algorithm for feature selection, the model gave an overall accuracy of 0.932, similar to the logistic regression classifier in Result 3.2. This result indicates the potential feasibility of using machine learning non-knowledge-based CDSS for diagnosing AD solely on the relatively accessible data of participants. According to section 3.3, except for MRI field strength, all features marked as "important" in the multiclassification or binary classification selection have over 90% of non-missing data, which implies the high accessibility of the required features in the real world. In other words, the data collection of the important features identified by the Boruta Algorithm is feasible.

Comparison Between Studies

As the introduction outlined, a knowledge-based CDSS has been built and applied in hospitals for validation [8,13]. Another example includes AD diagnosis CDSS study using biomarkers; these models give high accuracies of predicting AD (around 85%) and agree with the vital relevance of CDR and MMSE in diagnosing AD. In this study, Lawton Instrumental Activities of Daily Living Scale, a self-report questionnaire measuring independent functioning and ability to live in community settings, also demonstrates high relevance with AD predictions [27]. In another study, machine learning models were built using novel biomarkers about synaptic dysfunctions, such as tau and amyloid beta protein aggregates common in AD pathology. The AUC score and accuracy for such models reached over 0.8 [28]. Results of this study showed the contrary, as tau and amyloid beta proteins as unimportant by the Boruta Algorithm. Nevertheless, training the machine learning models with data involving biomarkers and more tests of independent functioning, for example, scores on the Lawton may create more precise AD predictions in future studies.

Limitations and Implications

A complete CDSS should contain several stages before the final decision. And this study needs a larger sample size to be more convincing.

An essential ethical issue associated with CDSS is whether it is appropriate to use the data collected from the participants [29]. The data used in this study does not involve any information that reveal participants' identities. At the time of data collection, participants provided consent for experimenters to collect and distribute their data. In the dataset, each of the participants were assigned a roster ID. The building of a CDSS in clinical settings should always respect the participant's privacy, even though the construction of models does not necessarily involve the identity of participants.

Additionally, part of the public's concern about machine learning models and machine learning-based CDSS may result from the lack of a prior knowledge and theoretical basis in terms of medical science. Physicians are able to give the logical deduction for their diagnoses and interventions using medical training, knowledge, and previous cases of their patients with AD. However, non-knowledge-based CDSS can not explain clinical decisions from the perspective of medical sciences. For example, MRI field strength does not (and should not) seem influential in diagnosing AD. Still, the machine learning models from binary and multi-classifications suggest the association but can not explain it.

The data include results from the same patients over time, with the longitudinal scans separated by at least one year for each individual. Each data point is considered independent during the data training because of the removal of identity in the data set. Therefore, the training and testing data would have a lower variability than expected.

Furthermore, there exists various diagnoses aside from AD, which are not involved in this study. Failing to account or detect these disorders may result in false alarms of AD or MCI. In future studies, it would be more rigorous to examine and, if necessary, exclude the interference of other related disorders while exploiting the datasets.

However, the exploitation of the machine learning model in clinical research can sometimes give significant implications about associations between factors in data science and provide additional support for clinicians' decisions. For example, using the real-world dataset collected by the ADNI, the models in this study verify the importance of some particular cognitive tests like the MMSE, DSST, and TMT, as well as marital status, race, CDR score and MRI field strength in predicting MCI and AD.

Conclusions

Based on the accuracy of machine learning models in this study, there is potential to create a non-knowledge-based CDSS with machine learning algorithms for the diagnosis of neurodegenerative disorders in the future. Non-knowledge-based machine learning models, especially ones using support vector machine or ensemble methods, may provide informative and supplementary methods for clinical diagnosis of AD in the future.

List of Abbreviations Used

AD: Alzheimer's disease
CN: cognitively normal
MCI: mild cognitive impairment
CDSS: clinical decision support system
NPV: negative predictive value
PPV: positive predictive value
AUC-ROC: area under the receiver operating characteristic curve
CDR-SB: clinical dementia rating scale sum of boxes
MMSE: mini-mental state examination
mPACC: modified preclinical Alzheimer's cognitive composite
PACC: preclinical Alzheimer's cognitive composite

Conflicts of Interest
There are no conflicts of interest regarding publication of this manuscript.

Ethics Approval and/or Participant Consent
Data used in this study are acquired from the *Alzheimer's Disease* Neuroimaging Initiative. The original data is collected from participants worldwide with their consent.

Authors' Contributions
XK: made substantial contributions to the design of the study, collected and analysed data, drafted the manuscript, and gave final approval of the version to be published.

Acknowledgements
The author expresses great thanks to Andy Tai. He provided important articles for references, gave valuable suggestion and instructions to the analysis of the data and the visualization of results, and helped revise the final paper. The author also thanks The URNCST Journal for pairing them with Andy Tai and providing a valuable opportunity. For the Alzheimer's Disease Neuroimaging Initiative 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: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

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

References
[1] Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthcare Journal. 2019;6(2):94. https://doi.org/10.7861%2Ffuturehosp.6-2-94
[2] Khosla V. Do we need doctors or algorithms? TechCrunch [Internet]. 2012 Jan 10; Available from: https://techcrunch.com/2012/01/10/doctors-or-algorithms/
[3] Mitchell T. Machine Learning. New York, NY: McGraw-Hill Professional; 1997.
[4] Bishop CM, Nasrabadi NM. Pattern recognition and machine learning. New York: Springer; 2006 Aug 17.
[5] Syeda-Mahmood TF. Role of machine learning in clinical decision support (Presentation Recording). In: Hadjiiski LM, Tourassi GD, editors. Medical Imaging 2015: Computer-Aided Diagnosis. SPIE; 2015. http://dx.doi.org/10.1117/12.2084708
[6] Hu Z, Jing Y, Xue Y, Fan P, Wang L, Vanyukov M, et al. Analysis of substance use and its outcomes by machine learning: II. Derivation and prediction of the trajectory of substance use severity. Drug Alcohol Depend. 2020;206(107604):107604. http://dx.doi.org/10.1016/j.drugalcdep.2019.107604
[7] Zheng L, Wang O, Hao S, Ye C, Liu M, Xia M, et al. Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records. Translational Psychiatry. 2020 Feb 20;10(1):72. https://doi.org/10.1038/s41398-020-0684-2
[8] Simon RP, Aminoff MJ, Greenberg DA. Clinical Neurology. Lange Medical Books/McGraw-Hill; 2009 Mar 9.
[9] Insel PS, Weiner M, Mackin RS, Mormino E, Lim YY, Stomrud E, et al. Determining clinically meaningful decline in preclinical Alzheimer disease. Neurology. 2019 Jul 23;93(4):e322-33. https://doi.org/10.1212/wnl.0000000000007831
[10] Burns A. Iliffe S. Alzheimer's disease. BMJ. 2009;338:b158. https://doi.org/10.1136/bmj.b158
[11] Alzheimer's disease fact sheet [Internet]. National Institute on Aging. Available from: https://www.nia.nih.gov/health/alzheimers-disease-fact-sheet
[12] Knopman DS, Amieva H, Petersen RC, Chételat G, Holtzman DM, Hyman BT, et al. Alzheimer disease. Nat Rev Dis Primers. 2021 May 13;7(1):33. https://doi.org/10.1038/s41572-021-00269-y

[13] Sanchez E, Toro C, Carrasco E, Bonachela P, Parra C, Bueno G, et al. A knowledge-based clinical decision support system for the diagnosis of Alzheimer disease. In: 2011 IEEE 13th International Conference on e-Health Networking, Applications and Services. IEEE; 2011. https://doi.org/10.1109/HEALTH.2011.6026778
[14] vanRossum G. Python reference manual. Department of Computer Science [CS]. 1995 Jan 1(R 9525).
[15] Balsis S, Benge JF, Lowe DA, Geraci L, Doody RS. How do scores on the ADAS-Cog, MMSE, and CDR-SOB correspond? Clin Neuropsychol. 2015;29(7):1002-9. https://doi.org/10.1080/13854046.2015.1119312
[16] O'Bryant SE, Lacritz LH, Hall J, Waring SC, Chan W, Khodr ZG, et al. Validation of the new interpretive guidelines for the clinical dementia rating scale sum of boxes score in the national Alzheimer's coordinating center database. Arch Neurol. 2010 Jun;67(6):746-9. https://doi.org/10.1001/archneurol.2010.115
[17] Donohue MC, Sperling RA, Salmon DP, Rentz DM, Raman R, Thomas RG, et al. The preclinical Alzheimer cognitive composite: Measuring amyloid-related decline. JAMA Neurol. 2014 Aug;71(8):961-70. https://doi.org/10.1001/jamaneurol.2014.803
[18] Kron J, Brukner P, Khan K. The encyclopedia of exercise, sport and health. St Leonards, NSW, Australia: Allen & Unwin; 2004.
[19] Adam C, Craton N. Concussions: A head-on approach. Canadian Journal of Diagnosis. 2002;19(2):101-16.
[20] Kurlowicz L, Wallace M. The mini-mental state examination (MMSE). Journal of Gerontological Nursing. 1999 May 1;25(5):8-9. https://doi.org/10.3928/0098-9134-19990501-08
[21] 2021 Alzheimer's disease facts and figures. Alzheimers Dement. 2021 Mar;17(3):327-406. https://doi.org/10.1002/alz.12328
[22] Weiner MF. Perspective on race and ethnicity in Alzheimer's disease research. Alzheimers Dement. 2008 Jul; 4(4): 233–8. https://doi.org/10.1016%2Fj.jalz.2007.10.016
[23] Mehta KM, Yaffe K, Pérez-Stable EA, Stewart A, Barnes D, Kurland BF, et al. Race/ethnic differences in AD survival in US Alzheimer's Disease Centers. Neurology. 2008 Apr 1;70(14):1163-70. https://doi.org/10.1212/01.wnl.0000285287.99923.3c
[24] Biddle KD, Jacobs HI, Uquillas FD, Zide BS, Kirn DR, Properzi MR, et al. Associations of widowhood and β-amyloid with cognitive decline in cognitively unimpaired older adults. JAMA Netw Open. 2020 Feb; 3(2): e200121. https://doi.org/10.1001%2Fjamanetworkopen.2020.0121
[25] Sommerlad A, Ruegger J, Singh-Manoux A, Lewis G, Livingston G. Marriage and risk of dementia: systematic review and meta-analysis of observational studies. J Neurol Neurosurg Psychiatry. 2018 Mar;89(3):231-8. https://doi.org/10.1136/jnnp-2017-316274
[26] West J, Blystad I, Engström M, Warntjes JB, Lundberg P. Application of quantitative MRI for brain tissue segmentation at 1.5 T and 3.0 T field strengths. PLoS One. 2013 Sep 16;8(9):e74795. https://doi.org/10.1371/journal.pone.0074795
[27] Carvalho CM, Seixas FL, Conci A, Muchaluat-Saade DC, Laks J, Boechat Y. A dynamic decision model for diagnosis of dementia, Alzheimer's disease and mild cognitive impairment. Computers in Biology and Medicine. 2020 Nov 1;126:104010.
[28] Chang CH, Lin CH, Lane HY. Machine learning and novel biomarkers for the diagnosis of Alzheimer's disease. Int J Mol Sci. 2021 Mar 9;22(5):2761. https://doi.org/10.3390/ijms22052761
[29] Bonney W. Is it appropriate, or ethical, to use health data collected for the purpose of direct patient care to develop computerized predictive decision support tools. Stud Health Technol Inform. 2009;143:115-21. https://pubmed.ncbi.nlm.nih.gov/19380924/

Article Information
Managing Editor: Jeremy Y. Ng
Peer Reviewers: Andy Tai, Ricky Chow
Article Dates: Received Aug 13 22; Accepted Dec 15 22; Published Feb 03 23

Citation
Please cite this article as follows:
Kong X. Feasibility study: Machine learning in neurodegenerative disorders, Alzheimer's disease. URNCST Journal. 2023 Feb 03: 7(2). https://urncst.com/index.php/urncst/article/view/426
DOI Link: https://doi.org/10.26685/urncst.426
