Machine Learning Models Reveal the Importance of Clinical Biomarkers for the Diagnosis of Alzheimer's Disease

Article · June 2020

DOI: 10.3233/SHTI200599

9 authors, including:

Amal Ali
Hamad bin Khalifa University
1 PUBLICATION **0** CITATIONS
SEE PROFILE

Mohammad Islam
Southern Connecticut State University
7 PUBLICATIONS **22** CITATIONS
SEE PROFILE

Mowafa Househ
Hamad bin Khalifa University
215 PUBLICATIONS **1,934** CITATIONS
SEE PROFILE

Tanvir Alam
Coventry University
18 PUBLICATIONS **8** CITATIONS
SEE PROFILE

The Importance of Health Informatics in Public Health during a Pandemic
J. Mantas et al. (Eds.)

doi:10.3233/SHTI200599

Machine Learning Models Reveal the Importance of Clinical Biomarkers for the Diagnosis of Alzheimer's Disease

Mahmoud Ahmed REFAEE [a], Amal Awadalla Mohamed ALI [a], Asma Hamid ELFADL [a], Maha F. A. ABUJAZAR [a], Mohammad Tariqul ISLAM [b], Ferdaus Ahmed KAWSAR [c], Mowafa HOUSEH [a], Zubair SHAH [a] and Tanvir ALAM [a,1]
[a] *College of Science and Engineering, Hamad Bin Khalifa University (HBKU), Doha, Qatar*
[b] *Computer Science Department Southern Connecticut State University, USA*
[c] *Department of Computing, East Tennessee State University, USA*

Abstract. Alzheimer's Disease (AD) is a neurodegenerative disease that causes complications with thinking capability, memory and behavior. AD is a major public health problem among the elderly in developed and developing countries. With the growth of AD around the world, there is a need to further expand our understanding of the roles different clinical measurements can have in the diagnosis of AD. In this work, we propose a machine learning-based technique to distinguish control subjects with no cognitive impairments, AD subjects, and subjects with mild cognitive impairment (MCI), often seen as precursors of AD. We utilized several machine learning (ML) techniques and found that Gradient Boosting Decision Trees achieved the highest performance above 84% classification accuracy. Also, we determined the importance of the features (clinical biomarkers) contributing to the proposed multi-class classification system. Further investigation on the biomarkers will pave the way to introduce better treatment plan for AD patients.

Keywords: Dementia; Alzheimer's Disease; Mild Cognitive Impairment; ADNI

1. Introduction

Age-related cognitive decline and impairment is a major public health problem among the elderly in developed and developing countries [1]. When age-related cognitive decline has mild effects in few activities of daily life, it is termed as mild cognitive impairment (MCI) and when this reaches a more aggressive form of impairment affecting almost all activities of life including memory, communication, planning, preparing, etc. then it may be termed as Dementia [1]. Alzheimer's disease (AD) is the most common form of Dementia, which accounts for 50% to 60% percent of all cases [1]. Usually, Dementia is preceded by MCI, and there exists a grey boundary between them which makes it difficult to identify the early onset of Dementia. Identifying novel risk factors can help determine the early onset of MCI and Dementia, especially during middle age

[1] Corresponding Author. Tanvir Alam, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar; Email: talam@hbku.edu.qa.

(45 to 65 years) which has the potential to delay or mitigate the risk of developing one third of Dementia cases [1]. There is no single test that is prescribed to determine dementia due to AD. Physicians mostly rely on a variety of approaches and tools to make a diagnosis which includes blood tests, brain imaging in addition to memory and cognitive tests. Developing tools that can automatically differentiate the status of a person as being cognitively normal (CN), MCI or AD will help address this important problem. In this work, we propose a machine learning-based technique to distinguish control subjects with no cognitive impairments, AD subjects, and subjects with mild cognitive impairment (MCI), often seen as precursors of AD. To further validate the outcomes from our model, we investigate the relative importance of clinical biomarkers that were used in the model development based on the SHapley Additive exPlanations (SHAP) [2].

2.Materials and Methods

2.1 Dataset Description

The data used in this research was collected from The Alzheimer's Disease Neuroimaging Initiative (ADNI) [3]. The ADNI dataset includes 20 features known to be important for the diagnosis of AD. It contains 819 Control Subjects, 399 cases of Dementia, and 1050 cases of MCI. The features include demographic information such as gender, marital status, and age as well as genetic information such as APOE4, where presence or absence of this gene is the strongest risk factor for AD [4]. The clinical assessment considered for this study were the volumes of hippocampus, ventricles, brain, fusiform gyrus, entorhinal and middle temporal cortex, and average fluorodeoxyglucose (FDG)-positron emission tomography (PET) (FDG-PET) of angular, temporal, and posterior cingulate. Except for APOE4, all others are volumes of different regions of the brain that were calculated in mm3. We considered all the above mentioned measurements as a feature vector to develop the proposed classification model to distinguish AD, MCI and the CN group. Recently, Shabaz et al. [5] worked on a similar problem and the authors considered six different machine learning (ML) algorithms which reported an 88% accuracy in distinguishing AD subjects from others. However, Shabaz et al. considered the Clinical Dementia Rating Scale Sum of Boxes (CDRSB) as one of the features. CDRSB is a clinical tool used to differentiate CN vs. MCI vs. AD and it was designed to be, primarily, used by clinicians, to support in diagnosing and determining the severity level of dementia. Unfortunately, its reliability is suboptimal in very mild dementia cases [6]. Researchers developed a modified version of CDRSB (mCDR) to make it easier to use [7]. CDRSB depends on the clinicians' experiences, and its performance in early-stages of dementia has a lower reliability. As a result, we excluded CDRSB from the feature vector for our proposed model.

2.2 Machine Learning Model Development and Evaluation

Five different ML models: Gradient Boosting Decision Trees (XGBoost), Neural Network (NN), Random Forest (RF), Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA) were trained and evaluated against the dataset using a nested ten-fold cross-validation for model selection and assessment.

3. Results & Discussions

Among the five different ML algorithms we explored, XGBoost showed the best results. From Table 1, we can see that the XGBoost-based model achieved the highest performance with 95.20% accuracy when we included CDRSB as a feature and 84.44% accuracy when we left CDRSB out of our feature list. We also observed high values of multiclass Matthews Correlation Coefficient (MCC), which demonstrates a high degree of confidence in our reported accuracies. Table II shows the normalized confusion matrix for the three categories for our best performer (XGBoost) algorithm.

Table 1. Summary of the performance of the models

Performance Metric	**XGBoost**	**NN**	**RF**	**SVM**	**LDA**
Considering CDRSB					
Accuracy (%)	95.20	92.00	94.03	88.38	83.84
MCC (%)	92.85	88.09	91.12	82.88	75.92
Without Considering CDRSB					
Accuracy (%)	84.44	78.22	81.46	76.00	74.63
MCC (%)	76.822	67.88	72.78	64.42	64.42

Table 2. Normalized confusion matrix for XGBoost; Rows (columns) denote true (predicted) labels

	Accuracy (%) Considering CDRSB			**Accuracy (%) without Considering CDRSB**		
	CN	**AD**	**MCI**	**CN**	**AD**	**MCI**
CN	96.29	0.00	3.71	84.48	0	15.52
AD	00.10	98.29	1.62	0.00	97.62	2.38
MCI	2.86	6.10	91.05	19.05	9.71	71.24

SHAP values for each feature, highlighted in the X-axis of Figure 1, represent the relative importance of corresponding features in the classification task of distinguishing AD, MCI and the CN group. From Figure 1, based on the model without considering CDRSB, it is obvious that highly contributing features in the classification task were the scores from the clinical assessment tools Mini-Mental State Examination (MMSE) [8] and Rey Auditory Verbal Learning Test (RAVLT) [9]. Values on the X-axis of Figure 1 represent the relative importance of features in this classification task. The average score of the AD group was less than the MCI and CN group's average score (Table 3), which clearly indicates the effectiveness of these tests. Age was a contributing factor, and the average age of AD (74.87) group was higher than the MCI (72.88) and CN (72.84) group.

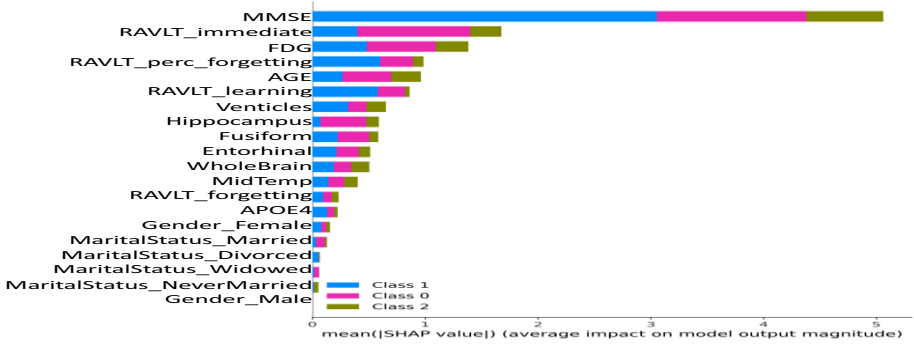

Figure 1. Summary plot showing variable importance based on SHAP values. Values on X-axis represent the relative importance of features in the classification task; Class 0: CN, Class 1:AD, Class 2: MCI

Table 3. The average value for different clinical measurements for AD, MCI and CN group

	Scores from Assessment				Region of Brain Volume (mm^3)		
	CN	**MCI**	**AD**		**CN**	**MCI**	**AD**
MMSE	29.09	27.62	23.17	Wholebrain	1028182	1028720	984537
RAVLT_learning	6.02	4.12	1.92	Fusiform	17652	17468	16153
RAVLT_forgetting	3.69	4.56	4.54	Entorhinal	3687	3492	3055
RAVLT_immediate	45.49	34.58	22.99	Hippocampus	7145	6791	6079
FDG-PET	1.27	1.24	1.11	Ventricles	36347	40443	47888

The current study demonstrates that the average volume of the whole brain, and specifically Hippocampus, Fusiform, and Entorhinal areas are decreasing in contrast to ventricle size that are relatively increasing (Table 3). FDG-PET, an advanced model of imaging, contributed heavily in differentiating AD vs. MCI vs. CN. Our findings show that the classification relied upon the glucose metabolism within specific areas of the brain reflecting the function of the different regions [10]. The results demonstrate that patients with lower values for AD (1.11) compared to MCI (1.24) and CN (1.27) have a less active metabolism in the brain related area which impacts cognitive function.

4.Conclusions

Based on our analysis, we showed the effectiveness of the cognitive test for AD in the clinical setup. We also showed that combining other clinical measurements from imaging data improves the performance of the model to distinguish AD from MCI and the CN group.

References

[1] Livingston G., et al., Dementia prevention, intervention, and care, The Lancet 390(10113) (2017), 2673-2734.
[2] Lundberg SM, Lee SI, A unified approach to interpreting model predictions, in Advances in neural information processing systems (2017).
[3] Petersen RC, et al., Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization, Neurology 74(3) 2010, 201-209.
[4] Sanan DA, et al., Apolipoprotein E associates with beta amyloid peptide of Alzheimer's disease to form novel monofibrils. Isoform apoE4 associates more efficiently than apoE3, The Journal of clinical investigation, 94(2) (1994), 860-869.
[5] Shahbaz M, et al., Classification of Alzheimer's Disease using Machine Learning Techniques, 2019.
[6] Schafer K, et al., Alzheimer's Disease Cooperative Study. Reliability of monitoring the clinical dementia rating in multicenter clinical trials, Alzheimer Dis Assoc Disord 18(4) (2004), 219-222.
[7] Duara R, et al., Diagnosis and staging of mild cognitive impairment, using a modification of the clinical dementia rating scale: the mCDR, International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences, 25(3) (2010), 282-289.
[8] Folstein MF, Folstein SE, McHugh PR, "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician, Journal of psychiatric research 12(3) (1975), 189-198.
[9] Schmidt M, Rey auditory verbal learning test: A handbook, Western Psychological Services Los Angeles, CA.,1996.
[10] Marcus C, Mena E, Subramaniam RM, Brain PET in the diagnosis of Alzheimer's disease, Clinical nuclear medicine, 39(10) (2014), e413.