

Evaluation of Neuropsychological Tests in Classification of Alzheimer's Disease



N. Vinutha, R. Jayasudha, K. S. Inchara, Hajira Khan, Sonu Sharma,
P. Deepa Shenoy and K. R. Venugopal

Abstract Many neuropsychological tests are available to measure cognitive declinement in a person affected by Alzheimer's disease. To evaluate his/her current stage in dementia and also to find the disease progression, it is necessary to perform a serial assessment of tests. As a result, the huge amount of data gets collected which depends on the number of neuropsychological tests performed to examine the patient and also with the number of visits to the clinic. From the previous correlation studies, it is observed that high computational time is required to process many neuropsychological tests. Therefore, the scores obtained from these tests are subjected to attribute selection algorithms. The six different attribute selection algorithms are used to rank the attributes, but the top four ranked attributes are consistent with InfoGain and OneR attribute evaluators. So, we subject the ordered attributes from these two attribute evaluators to different classifiers with 10-fold cross-validation. The random forest classifier performed better with InfoGain and OneR attribute evaluators. Therefore, an accuracy of 99.1% and ROC area of 0.999 is obtained from the set of top four attributes, and similar results are obtained from the set of top six and seven attributes with the combination of Infogain and OneR with BayesNet classifier.

Keywords Alzheimer's disease · Attribute selection · Cognitive domains Classification · Neuropsychological test

N. Vinutha (✉) · R. Jayasudha · K. S. Inchara · H. Khan · S. Sharma
P. D. Shenoy · K. R. Venugopal
Department of Computer Science and Engineering, University Visvesvaraya College
of Engineering, Bangalore University, Bangalore, India
e-mail: vinutha1v@gmail.com

© Springer Nature Singapore Pte Ltd. 2019
C. R. Panigrahi et al. (eds.), *Progress in Advanced Computing and Intelligent
Engineering*, Advances in Intelligent Systems and Computing 714,
https://doi.org/10.1007/978-981-13-0224-4_14

1 Introduction

Alzheimer's Disease (AD) [1] is an irreversible form of dementia that occurs among adults who fall in the age group of 40–90, but most commonly seen after 65 years. It is caused by the deposition of amyloid plaques and neurofibrillary tangles in different regions of the brain. As a result of deposition, it shows the impact on the brain size and further, the functional ability of neurons is reduced and gradually gets destroyed. These changes are noticeable and can be measured before the development of symptoms. The symptoms begin with short-term memory loss which continues further to long-term memory loss and change in behaviors and language (aphasia) that become severe day by day as the disease progresses. There are several factors like age, family history, smoking, obesity, diabetes, and high blood pressure that also increase the risk to AD. Hence, a wide range of techniques such as medical imaging, neuropsychological tests, medical history, physical, and neurological examination are performed to assess the clinical diagnosis of the patients.

Neuropsychological testing [2] is a measure of cognitive decline in a person. It is utilized to identify the ability of an individual to perform day-to-day activities in the diseased state. It is necessary to follow the serial assessments of the tests to evaluate the performance of a person after subjecting to medication. From the previous studies, it is inferred that a specific pattern is developed in an affected patient of the similar age group that can be differentiated from the normal aging. Further, better discrimination can also be done by the fusion of neuroimaging data with neuropsychological scores or by the combination of genetic risk factors with neuropsychological scores. Table 1 shows the list of neuropsychological tests with their associated domains.

Each test follows a standardized protocol and is conducted with the help of pencil, paper, visual aids, and computer. Due to multiple cognitive deficits, it is

Table 1 List of neuropsychological test

Cognitive domains	Neuropsychological test
Episodic memory	Logical memory (WMS-III)
Semantic memory	Wechsler Adult Intelligence Scale (WAIS)-II
Attention/working memory	WAIS-III, Number span
Executive function	Trail B, Digit symbol, Trail A
Depression	Geriatric Depression Scale (GDS)
Dementia severity	Mini Mental State Examination (MMSE)
Language (verbal fluency)	Animal list, Vegetable list
Language (naming)	Boston Naming Test (BNT)

always better to utilize the combination of tests from various domains that are helpful to characterize the pattern developed due to cognitive impairments and also to make a better decision in clinical diagnosis.

But there is a quick rise in the cost of medical and healthcare system. This is due to accumulation of a large amount of data and lot of time requirement by an expert to process the collected data and to make a decision in diagnosis and treatment of patients. All the problems mentioned above can be handled by machine learning approach [3, 4] as it plays a significant role in feature reduction and also retains only those features that lead to high performance.

1.1 Motivation

Neuropsychological scores have tremendous scope to integrate and validate under various domains. However, consideration of all the clinical scores requires more computational time. Thus, identifying a small subset of scores is very crucial for the correlation studies with either neuroimaging data or genetic risk factors.

1.2 Contribution

- To identify the visit with more number of demented cases.
- Identify suitable attribute selection algorithms based on ranking method.
- Evaluate different machine learning algorithms for classification of AD.
- To identify a minimum set of attributes with better performance.

1.3 Organization

The paper is organized as follows: Literature survey is presented in Sect. 2. Proposed system architecture for Alzheimer's disease classification is explained in Sect. 3. Experiments and results are presented in Sect. 4. The paper concludes in Sect. 5.

2 Literature Survey

Enormous techniques are developed by researchers to focus toward the prediction of AD. The papers [5–9] provide state-of-the-art survey on clinical scores to measure the progression of disease using longitudinal data, correlation studies, replacement of the existing neuropsychological tests with equivalent new tests, and the contribution of single and multiple predictors toward prediction of AD.

McCutcheon et al. [5] have evaluated whether AD pathology and depression are related to each other in MCI and Mild Dementia. The study requires clinical and neuropathological data. The GDS is obtained as a result, by subjecting the covariates Neuritic Plaque (NP) score and Braak stages of Neurofibrillary (NF) to the regression model. The outcome showed that GDS is not related to NP score or cognitive decline or their combination. Hence, it can be said that depression in early AD is evident to be independent of NP and NF pathology.

Authors in [6] have suggested four new non-proprietary tests in NACC's UDS neuropsychological battery. The suggested tests can be used as a replacement for the existing tests by measuring the correlation factor between them. To assess the correlation between each of the previous and new tests, a crosswalk study is conducted. Tests having good correlation are said to have high prediction accuracy. These equivalent scores can be considered for the longitudinal analysis.

The authors of the paper [7] have proposed the development of a multi-domain model to predict the progression of dementia in Alzheimer's disease. The data obtained from NACC are used in the evaluation of transition probabilities between the health states based on the behavioral, functional abilities, cognitive function, and also to analyze the status of symptoms. From the above analysis, it is inferred that there is a transition in the stages of AD within a time span of 12 months and the model helped in the assessment of AD.

Lee Gavett et al. [8] considered the longitudinal data of healthy older adults from NACC dataset. The longitudinal data between two and three annual visits were considered for each subject. The followup scores and baseline test scores of eleven neuropsychological tests are used in linear mixed effect regression to obtain Reliable Change Intervals (RCI) and also to calculate the cumulative frequency of the raw scores. It is inferred that age, education, and baseline test scores are good predictors. Tests related to attention and executive functioning are significant to healthy aging, and tests related to episodic and semantic memory are effective with relevance to practice effects.

According to John et al. [9], the cognitive performance of neuropsychological tests from UDS dataset has been interpreted in two approaches, namely shared variance and unique variance. In the first approach, the latent factor is used as a

single predictor for measures of severity, whereas the second approach utilizes 12 raw scores from the neuropsychological tests as the predictors of dementia diagnosis. A logistic regression analysis is performed on single and multiple predictors to obtain a log-odd ratio, model fit statistics, and classification accuracy. The results thus obtained from logistic regression revealed the significance of each test in the diagnosis of dementia.

3 Proposed Work

Figure 1 illustrates the architecture of the proposed work. It consists of four modules:

- (1) Data collection,
- (2) Preprocessing,
- (3) Attribute selection and
- (4) Classification.

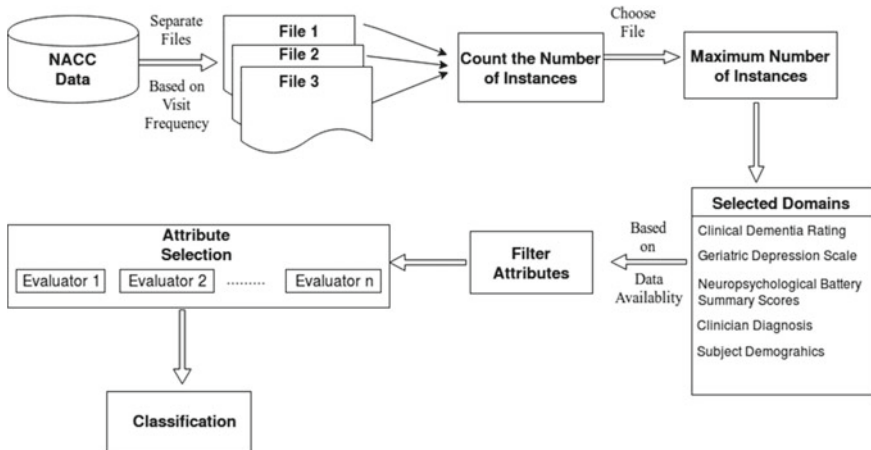


Fig. 1 System architecture for classification of Alzheimer’s disease

3.1 Data Collection

Data for the research work are collected from National Alzheimer's Coordinating Center (NACC), as it comprises the data of various Alzheimer's Disease Centers (ADCs). The collected NACC data constitute of subject demographics, health history, global staging, clinical dementia rating, neuropsychiatric inventory questionnaire, geriatric depression scale, functional activities questionnaire, clinician judgment of symptoms, clinical diagnosis, and neuropsychological battery summary scores for 11,735 unique instances.

3.2 Preprocessing

Patients' visits are available from 1 to 12. The preprocessing step begins by mapping unique IDs of the instances to the set of consecutive integers and to identify the number of visits available for each patient as shown in Algorithm I.

Algorithm I: To count the number of visits by each ID

Input: NACC Data

Output: Array containing number of visits by each ID

Initialization:

Map Unique ID's To 0, 1, 2.....N

No_of_Visits [0 to N] \leftarrow 0

Function

for $i \leftarrow [1 \text{ to Total ID's}]$

No_of_Visits [ID] \leftarrow No_Visits [ID] + 1
 $i \leftarrow i+1$

end for

The identification of number of visits for each instance is followed with the determination of demented cases from each visit time. After the determination, it is observed that the number of demented patients is increased with the higher visit times. Therefore, we group the instances with visits three to twelve into separate files as shown in Algorithm II.

Algorithm II: To group, the ID's with 3 to 12 visits

Input: *No_of_Visits***Output:** *2D Array Group, Size of Group***Initialize:***Group [3 to 12][0 to N] ← 0**Size_of_Group [0 to 12] ← 0***Function:****for** *i* ← [3 to 12] **for** *j* ← [0 to N] **if** *No_of_Visits[j] == i* *push j into Group[i]* *Size_of_Group[i]++* **end if** *j ← j+1* **end for** *i ← i+1***end for**

After the separation of files based on visit times, we count the number of patients for each selected visit and chose the file with the largest number of instances as shown in Algorithm III. From analysis of all the visits, we infer that the three times visited data are the largest with 1345 unique instances as shown in Fig. 2. Therefore, we consider three times visit data in our study.

In the next step, following domains such as subject demographics, global staging, clinical dementia rating, geriatric depression scale, clinician diagnosis, and neuropsychological battery scores are selected for the third visit data. From the above domains, we select the attributes that have data availability greater than 50%. So attributes such as *CDRSUM* (*Clinical Dementia Rating Sum Of Boxes*) (100%), *CDRGLOB* (*Global Clinical Dementia Rating*) (100%), *MEMORY* (100%), *COMPORT* (98.88%), *CDRLANG* (*Language*) (98.88%), *NACCGDS* (*Geriatric Depression Scale*) (91.59%), *NACCMSE* (*Mini Mental State Examination*) (58.73%), *LOGIMEM* (*Logical Memory*) (58.43%), *MEMUNIT* (*Logical Memory IIA-Delayed*) (58.36%), *DIGIF* (*Digit Span Forward*) (58.28%), *DIGIB* (*Digit Span Backward*) (58.21%), *ANIMALS* (*Animals List*) (91.07%), *VEG* (*Vegetables List*)

Algorithm III: To find the group with maximum number of ID's

Input: *Group, Size_of_Group*

Output: *Group having maximum no of ID's*

Initialize:

Max←0, *Group_No*←0

Function:

for *i*←[3 to 12]

if *Size_Of_Group*[*i*]≥*Max*

Max←*Size_Of_Group*[*i*]

Group_No←*i*

end if

i←*i*+1

end for

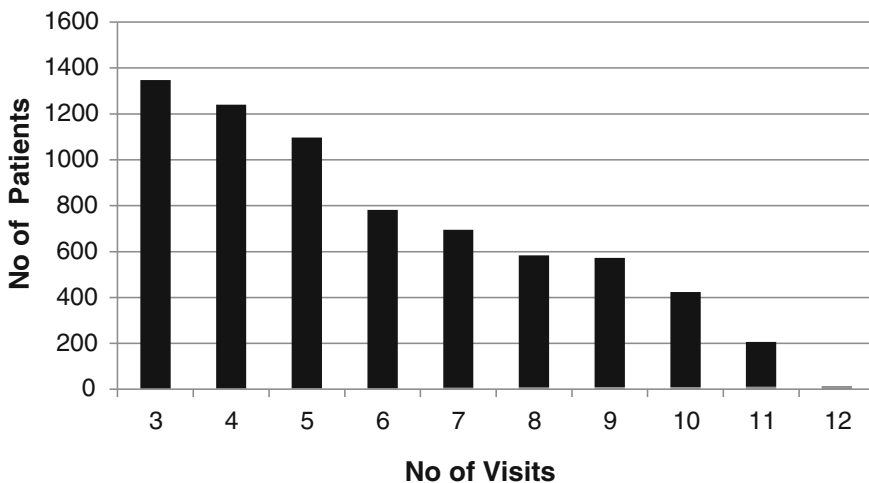


Fig. 2 Total number of instances based on number of visits

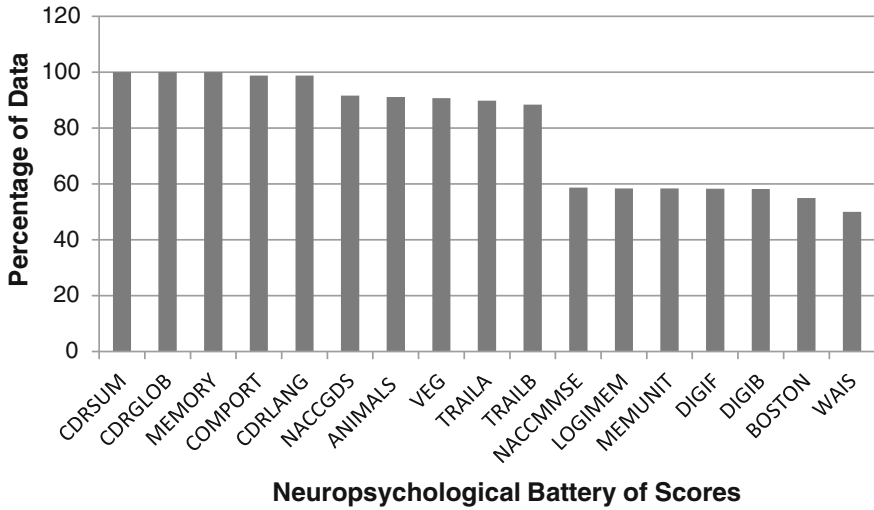


Fig. 3 The neuropsychological scores with data availability greater than 50%

(90.70%), *TRAILA* (*Trail Making Test Part A*) (89.73%), *TRAILB* (*Trail Making Test Part B*) (88.40%), *WAIS* (*Wechsler Adult Intelligence Scale*) (50.96%), and *BOSTON* (*Boston Naming Test*) (57.91%) are considered due to sufficient data availability as shown in Fig. 3.

3.3 Attribute Selection

Attribute selection is a process of searching the best subset of attributes from a given dataset. Various measures considered for attribute selection are correlation, distance, information, dependence, and consistency. The two different approaches to attribute selection are wrapper and filter method. In wrapper method, the subset selection is based on the learning algorithm, so the computational time increases for every subset that is evaluated in the context of the learning model, whereas, in filter method, the relevance of attribute is measured by using their correlation with the dependent variable and it is computationally faster since it does not involve training of the model. In our study, filter-based attribute evaluators are used to order the attributes based on the obtained rank.

3.4 Classification

All the ordered attributes obtained from filter-based attribute evaluators are subjected to supervised classifiers. Each classifier is evaluated based on the performance measures such as sensitivity, 1-specificity, ROC area, and accuracy. Some of the supervised classifiers used for our study are random forest, BayesNet, Random Committee, AdaBoost, and Naive Bayes.

4 Experiments and Results

In the proposed system, preprocessing is the first step performed to obtain the data required for our study. The preprocessed data are then subjected to attribute selection algorithms such as *OneRAttributeEval*, *InfoGainAttributeEval*, *GainRatioAttributeEval*, *ReliefAttributeEval*, *SymmetricalUncertAttributeEval*, and *CorrelationAttributeEval*. The ranked attributes obtained from these algorithms are further subjected to classifiers with 10-fold cross-validation. The classifiers, random forest, and BayesNet performed better with an accuracy of 99.4 and 99.1% for all the 22 attributes, that were ordered based on ranks obtained for *Infogain* and *oneR* attribute evaluators. However, our aim is to predict AD with a minimum number of attributes. Hence, the least-ranked attribute is removed each time and subjected to above classifiers until the minimal subset with the highest accuracy and ROC area is obtained.

The top-ranked attributes from *InfoGainAttributeEval* and *OneRAttributeEval* are *{CDRSUM, MEMORY, CDRGLOB, NACMMSE, ANIMALS, MEMUNITS, VEG, LOGIMEM...}* and *{CDRSUM, MEMORY, CDRGLOB, NACMMSE, LOGIMEM, CDRLANG, TRAILB...}*, respectively. It is observed that top four ranked attributes are common in these two attribute evaluators, so the performance is measured by the classifier with a minimal set of attributes.

An accuracy of 99.1% and ROC area of 0.999 is obtained from the top six attributes for the combination of *InfoGainAttributeEval* with BayesNet classifier, and same results are obtained from top seven attributes for the combination of *OneRAttributeEval* with BayesNet classifier. Figures 4 and 5 show the plot of ROC area versus number of attributes for BayesNet classifier.

However, the combination of *InfoGainAttributeEval* and *OneRAttributeEval* with random forest classifier results with an accuracy of 99.1% and ROC area of 0.999 from the top four attributes. Figures 6 and 7 show the plot of ROC area versus number of attributes for random forest classifier. The comparison of performance measures is shown in Table 2.

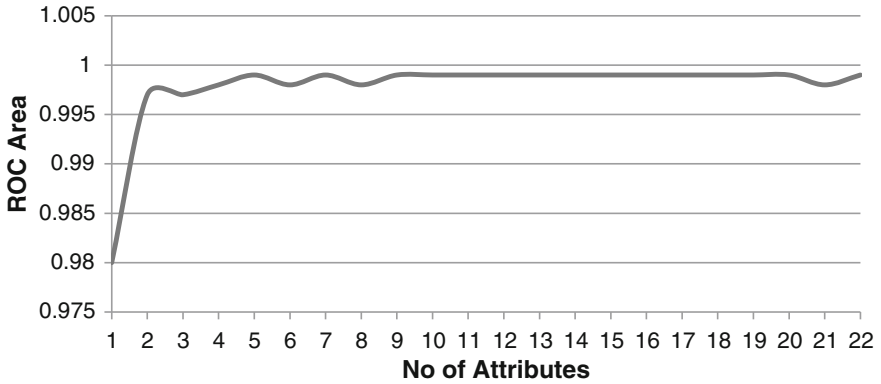


Fig. 4 The plot of ROC area versus number of attributes for OneR with BayesNet

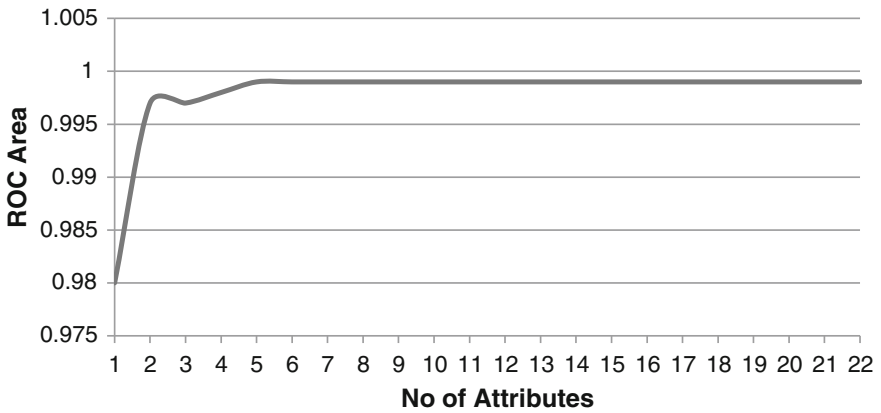


Fig. 5 The plot of ROC area versus number of attributes for Infogain with BayesNet

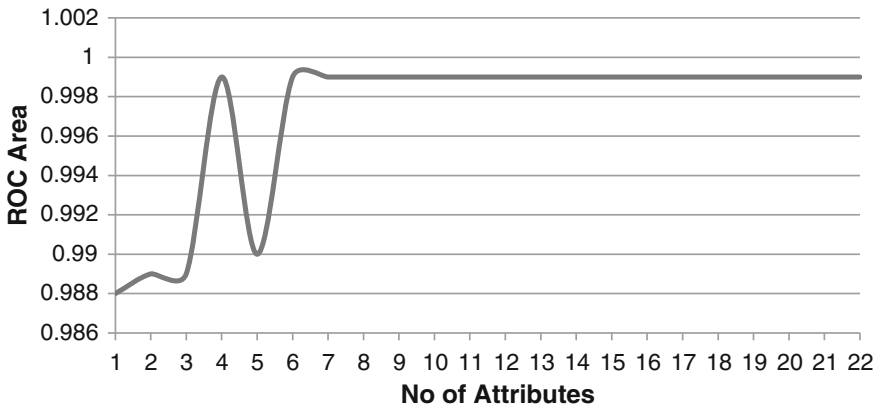


Fig. 6 The plot of ROC area versus number of attributes for OneR with Random forest

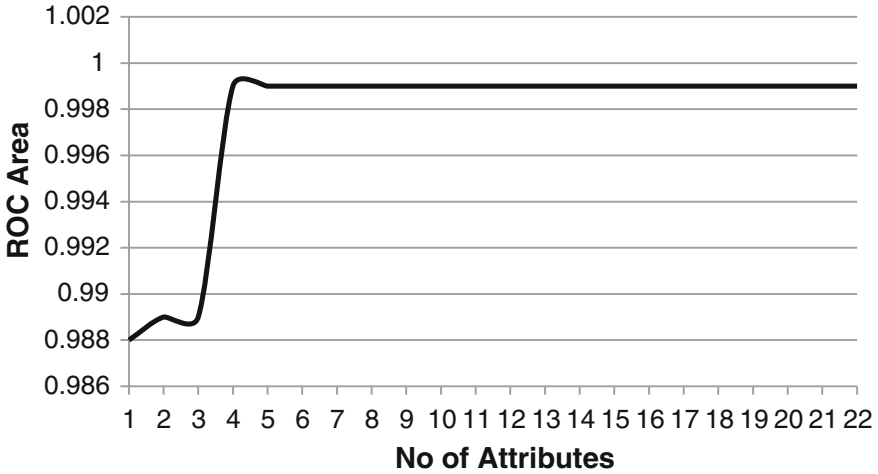


Fig. 7 The plot of ROC area versus number of attributes for Infogain with random forest

Table 2 Comparison of classification accuracy

Attribute evaluator	Classifier	No of attributes	Accuracy	ROC area	Sensitivity	1-Specificity
InfoGain	Random Forest	4	99.1	0.999	0.991	0.101
	BayesNet	6	99.1	0.999	0.99	0.068
OneR	Random Forest	4	99.1	0.999	0.991	0.084
	BayesNet	7	99.1	0.999	0.99	0.068

5 Conclusions and Future Work

The neuropsychological data for the instances with same visit number are significant to classify AD patients. Hence, we consider various domains for the selected visit as a measure of the cognitive declinment in a person. For each domain considered, the attributes only with sufficient data availability are selected to further analysis. After the refinement of data, it is subjected to six attribute selection methods. The performance of the ranked attributes is evaluated based on the metrics: sensitivity, 1-specificity, ROC area, and accuracy. An accuracy of 99.1% and ROC area of 0.999 is obtained from top four attributes by using OneRAttributeEval and InfoGainAttributeEval in combination with random forest classifier. Thus, it is inferred that these top four attributes have a significant role in the classification of AD.

The neuropsychological scores with data availability less than 50% are excluded in our study. Therefore, our focus is to handle a missing data and to study their significance in the classification of AD and in future, we extend our study on fourth visit data.

Acknowledgements The authors would like to thank curators of the NACC database for providing the data to conduct the research.

References

1. Emilien, G., Durlach, C., Minaker, K.L., Winblad, B., Gauthier, S., Maloteaux, J.M.: Alzheimer Disease: Neuropsychology and Pharmacology. Birkhauser (2012).
2. Harvey, P.D.: Clinical Applications of Neuropsychological Assessment, Vol. 14. Dialogues in Clinical Neuroscience (2012).
3. O’Kelly, N.: Use of Machine Learning Technology in the Diagnosis of Alzheimer’s Disease (2016).
4. Joshi, S., Shenoy, P.D., Venugopal, K.R., Patnaik, L.M.: Evaluation of Different Stages of Dementia Employing Neuropsychological and Machine Learning Techniques. In First IEEE International Conference on Advanced Computing (2009) 154–160.
5. McCutcheon, S.T., Han, D., Troncoso, J., Koliatsos, V.E., Albert, M., Lyketsos, C.G., Leoutsakos, J.M.S.: Clinicopathological Correlates of Depression in Early Alzheimer’s Disease in the NACC. International Journal of Geriatric Psychiatry (2016).
6. Monsell, S.E., Dodge, H.H., Zhou, X.H., Bu, Y., Besser, L.M., Mock, C., Hawes, S.E., Kukull, W.A., Weintraub S.: Results from the NACC Uniform Data Set Neuropsychological Battery Crosswalk Study, Vol. 30. Alzheimer Disease & Associated Disorders (2016) 134–139.
7. Green, C., Zhang, S.: Predicting the Progression of Alzheimer’s disease Dementia: A Multidomain Health Policy Model, Vol. 12. Alzheimer’s and Dementia (2016) 776–785.
8. Gavett, B.E., Ashendorf, L., Gurnani, A.S.: Reliable Change on Neuropsychological Tests in the Uniform Data Set, Vol. 21. Journal of the International Neuropsychological Society (2015) 558–567.
9. John, S.E., Gurnani, A.S., Bussell, C., Saurman, J.L., Griffin, J.W., Gavett, B.E.: The Effectiveness and Unique Contribution of Neuropsychological Tests and the Latent Phenotype in the Differential Diagnosis of Dementia in the Uniform Data Set, Vol. 30. Neuropsychology (2016) 946.