

A MACHINE LEARNING FRAMEWORK FOR ASSESSMENT OF COGNITIVE AND FUNCTIONAL IMPAIRMENTS IN ALZHEIMER'S DISEASE: DATA PREPROCESSING AND ANALYSIS

N. Vinutha¹, S. Pattar¹, S. Sharma¹, P. D. Shenoy¹, K.R. Venugopal²

1. Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, K. R. Circle, Bangalore, India; 2. Bangalore University, Bangalore, India

Corresponding Author: Vinutha N, Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, K. R. Circle, Bangalore 560 001, India. Email: vinutha1v@gmail.com, Tel.: +91-9916-663-693

J Prev Alz Dis 2020;
Published online February 7, 2020, <http://dx.doi.org/10.14283/jpad.2020.7>

Abstract

The neuropsychological scores and Functional Activities Questionnaire (FAQ) are significant to measure the cognitive and functional domain of the patients affected by the Alzheimer's Disease. Further, there are standardized dataset available today that are curated from several centers across the globe that aid in development of Computer Aided Diagnosis tools. However, there are numerous clinical tests to measure these scores that lead to a challenging task for their assessment in diagnosis. Also, the datasets suffer from common missing and imbalanced data issues. In this paper, we propose a machine learning based framework to overcome these issues. Empirical results demonstrate that improved performance of Genetic Algorithm is obtained for the neuropsychological scores after Miss Forest Imputation and for FAQ scores is obtained after subjecting it to the Synthetic Minority Oversampling Technique.

Key words: Alzheimer's disease, functional activities questionnaire, genetic algorithm, logistic regression, imputation, missforest, neuropsychological scores, synthetic minority oversampling technique.

Introduction

Alzheimer's Disease (AD) is the most common form of dementia that occurs among older adults in their mid-60s. It is an irreversible and progressive disorder that is caused due to the deposition of neurofibrillary tangles and amyloid plaques. This protein deposition damages the brain and impacts on its size, which further reduces the functional ability of neurons (1). AD is characterized by the development of various symptoms that are assessed through the several techniques such as neuropsychiatric inventory questionnaire, neuropsychological scores, Functional Activities Questionnaire, medical history, physical and neurological examination, etc. Among these neuropsychological scores are significant markers that aid in diagnosis of the AD because they help to identify the cognitive impairment in a subject. Along with the

neuropsychological scores, the study of functional impairment using FAQ scores is also important because cognitive declinement in a person may also lead to functional impairment; hence, many research studies focus on utilizing these scores to detect the AD by developing Computer-Aided Diagnosis (CAD) methods using machine learning and pattern recognition techniques.

Alzheimer Disease Centers (ADCs) digitizes the clinical test results of the subjects and curates them to the standardized datasets. These datasets are freely available to the researchers and scientists to develop CAD tools for the AD diagnosis using clinical scores and imaging biomarkers which comprises of both cross-sectional and longitudinal data. However, the longitudinal data collected from the ADCs have some common problem such as missing and imbalanced data. In this work, we address these challenges by developing a machine learning based framework for CAD of the AD through prediction of missing values by imputation method and sampling of imbalanced data by Synthetic Minority Oversampling Technique. After computing the missing values and creating the balanced data, feature selection technique is then employed to select the most relevant features so as to obtain an improved classification results for the assessment and diagnosis of the AD.

Contributions

- To this end, the paper makes following contributions.
- Calculation of missing data: We employ imputation method to predict the missing data in the study dataset to improve the classification results.
 - Data balancing: In the proposed framework, sampling technique is utilized to balance the study data that outperform the imbalanced training model.
 - Feature selection: Genetic Algorithm (GA) based elitism technique is used to retain the fittest individual features that aids in effective assessment and diagnosis of the AD from cognitive and functional impairment perspectives.

Organization

The rest of the paper is organized as follows. Literature survey is presented in Section 2. Problem definitions of the work and its objectives along with the proposed machine learning based framework for CAD of the AD are described in Section 3. Section 4 details on the feature selection and evaluation methodologies. In Section 5, experiments and results are presented. Finally, the paper concludes in Section 6.

Literature Survey

In this section, we review the recent papers that use machine learning techniques, feature selection and statistical analysis techniques for CAD of the AD. In recent years, the CAD is considered to be the second opinion in the diagnosis of the patients' disease and also to assist the doctors to make correct decisions. Thus, the researchers and scientists to show the focus towards the development of the CAD using machine learning and pattern recognition techniques. In this regards, several standard datasets for CAD of AD are present viz. National Alzheimer's Coordinating Center (NACC) - <http://alz.washington.edu>; Mayo Clinic Study of Aging (MCSA) - <https://www.mayo.edu/>; and, Alzheimer's Disease Neuroimaging Initiative (ADNI) - <http://adni.loni.usc.edu>.

Machine Learning Approaches for CAD of AD

McCutcheon et al., (2) proved that the depression is independent of AD pathology. Gavett et al., (3) have identified the education, age and baseline scores as the best predictors of AD progression by utilizing the longitudinal data from NACC. Hassenstab et al., (4) suggest that the new scores with high correlation value can be utilized as a substitution with the old score in the longitudinal analysis. In paper (5), Besser et al., shows that different tools are required to diagnose AD and PD patients by subjecting the neuropsychological and clinical scores from NACC data to multivariable linear regression analysis. Barrens et al., (6) shows that both the cognitive domain and behavioral impairment are observed at the initial stages of AD, and it gradually increases with the increasing age groups.

Battista et al., (7) obtained Logical Memory, FAQ and Alzheimer's Disease Assessment Scale-Cognitive Behavior (ADAS-Cog) as the best predictors by optimizing the scores using machine learning techniques. Machulda et al., (8) proposed the framework to categorize the incident MCI into four different clusters: Amnesic, Dysexecutive, Dysnomia and Subtle Cognitive Impairment (SCI). Zhu et al., (9) proposed the personalized model for AD diagnosis. The personalized model utilizes the reweighting of training

subjects, therefore it aids in the distribution of testing data and also to refine the classifier based on training weights. Moradi et al., (10) prove that the structural brain atrophy is highly correlated to the Rey's Auditory Verbal Learning Test (RAVLT) Immediate and Percent Forgetting scores and the RAVLT Immediate score also shows high predictive accuracy for the AD conversion.

Statistical Analysis of the Clinical Scores of AD

Green et al., (11) performed the statistical analysis on NACC data, to find the transition probabilities of AD subjects. John et al., (12) shows that impairment under episodic memory, language, executive function and attention domain are significant to the individuals with the symptomatic AD. Kielb et al., (13) determined that the scores related to the episodic memory are significant in the baseline visit. Thus, it is considered to be a sensitive biomarker to determine MCI. Musa et al., (14) conducted a study on the AD patients in Chile to assess the behavioral changes during the progression of AD by using NPI-Q scores. In paper (15), the objective of Chen et al., is to identify the predictors that lead to the progression to MCI from normal.

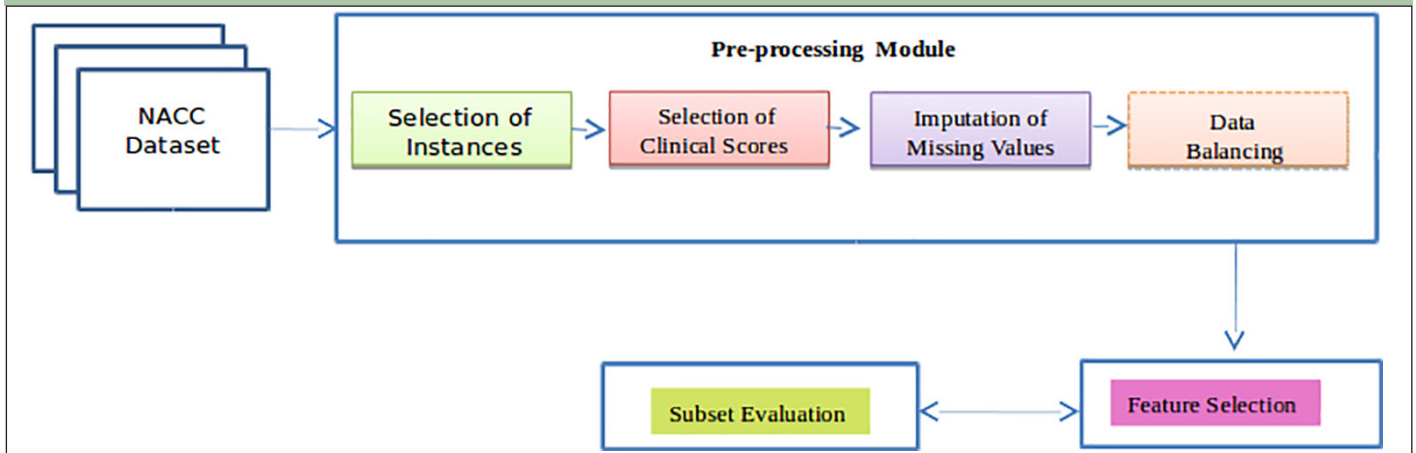
Feature Selection Techniques in the CAD of AD

Zhu et al., (16) applied Temporally Structured Support Vector Machine (TS-SVM) on MCI-NC to detect AD before its onset and also to obtain the better classification. Sorensen et al., (17) shows that the feature selection technique have an important role in the multiclass classification of AD. Beheshti et al., (18) proposed the framework for CAD of AD that comprises the feature extraction by Voxel based Morphometry (VBM), In the second stage, voxel values are extracted from the reduced Volume of Interest. Then feature selection is performed by t-test scores and GA. Finally, the obtained optimal subsets of features are classified by the linear Support Vector Machine (SVM) classifier.

Problem Definition and Proposed Approach

The problem that we consider in this paper is described as follows. "Diagnosis of the AD by identification of Cognitive and Functional impairment at the baseline and follow up visits using clinical scores through machine learning approach". Further, our objective is to determine the (i) Effect of missing data on the classification; (ii) Role of imbalanced data on the classification and (iii) Which are the best (optimal) and minimal features for AD diagnosis.

Figure 1. The Proposed Framework for CAD of the AD



Proposed Framework for CAD of the AD

The flowchart of the proposed framework is represented in Figure 1 and described below. The longitudinal data collected from the NACC dataset are subjected to the preprocessing module which comprises the selection of instances, and clinical scores, Imputation of missing values and data balancing. The preprocessed data is then subjected to GA to select the best subset of features. Further, the obtained subsets of features are evaluated by Logistic Regression (LR) model. Next, the obtained results are analyzed to find the significance of the selected subset of neuropsychological and FAQ scores.

To attain these objectives, we have to overcome numerous challenges existing with the study data. For the collected study data, we have the following objectives:

- Missing Data Prediction: To effectively balance all the attributes of the dataset. Thus, it leads to the unbiased training phase for machine learning techniques.
- Imbalanced Data Handling: To marginally distribute the data samples between different classes. Therefore, the prediction model is effective in learning the features from all the classes.
- Feature Selection: To select the relevant scores from the neuropsychological test and FAQ scores.

Data Collection

The NACC database used in our proposed work is developed through the contribution of standardized clinical data by several ADCs across united states (19). The collected longitudinal data of AD, MCI and normal subjects comprises of 11,735 unique instances and 467 features and these instances have their visit range varying from 1-12 for the period of September 2005 to June 2016. So out of 467 features, 16 neuropsychological scores, 10 FAQ scores are significant because neuropsychological scores provide information about the current levels of cognitive performance in a person and the loss of cognitive function may also affect the change in the functional and behavioral state of the person. Thus, in addition to the neuropsychological scores, FAQ is also used by a neuropsychologist to assess the severity of the disease and also to distinguish between the different stages of the AD. Hence, these scores are selected in our study for further analysis and are shown in Table 1 and Table 2 respectively.

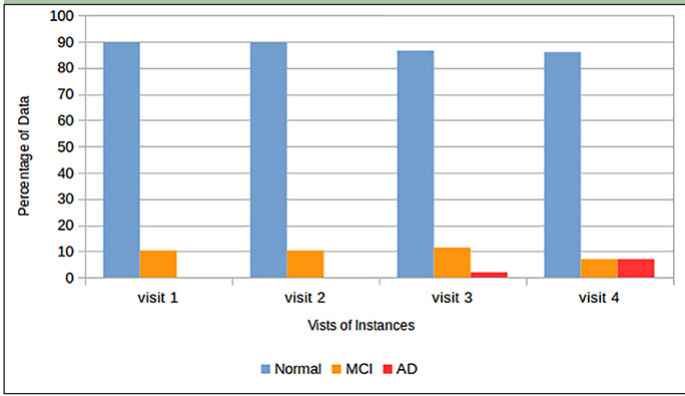
Table 1. The neuropsychological scores

FN	Neuropsychological Scores	Cognitive Domain	FN	Neuropsychological Scores	Cognitive Domain
1	CDRSUM	Dementia Severity	9	DIGIB	Working Memory
2	CDRGLOB	Episodic Memory	10	DIGIBLEN	Working Memory
3	NACCGDS	Depression	11	ANIMALS	Language
4	NACCMME	Dementia Severity	12	VEG	Language
5	LOGIMEM	Episodic Memory	13	TRAILA	Executive Function
6	MEMUNITS	Episodic Memory	14	TRAILB	Executive Function
7	DIGIF	Working Memory	15	WAIS	Working Memory
8	DIGIFLEN	Working Memory	16	BOSTON	Language

*FN- Feature Number

Table 2. The functional activities questionnaire scores

FAQ	Feature Number
BILLS	1
TAXES	2
SHOPPING	3
GAMES	4
STOVE	5
MEAL PREP	6
EVENTS	7
PAYATTN	8
REMDATES	9
TRAVEL	10

Figure 2. The Percentages of Data for Different Visits


Preprocessing

Selection of Patient's Visits

The collected data from NACC have the instances with their visit range varying from one to twelve. From the clinician diagnosis, we observe 50% of MCI patients are converted to AD within the time duration of 24 to 36 months. So, we preprocess the NACC data to have the instances with the minimum of 24 to 36 months between baseline and follow up visits. The selected period is significant because the conversion of MCI patients to AD are more in visit-4, which is as shown in Figure 2.

Statistical Significance for Choosing Fourth Visit

The paired t-test is used as a statistical procedure to compare CDRSUM score of the instances at two different visits. We choose CDRSUM in the statistical analysis because it is a measure of cognitive dysfunction and dementia severity. To perform statistical analysis, we assume the null hypothesis and calculate the paired t-test scores between two different visits using Equation 1.

Null Hypothesis: The mean difference of CDRSUM score between two visit times is zero and p-value ≥ 0.05

$$t = \frac{m}{s/\sqrt{n}} \quad (1)$$

Where, t is the t-statistic; m is mean of the group; s is standard deviation, and n is the sample/group size.

The obtained p-value is $2.2e-16$. Thus, we reject the null hypothesis at a 95% confidence interval and conclude that there is a significant difference between the mean of CDRSUM scores at two different visits and the obtained results of t-test scores and mean differences are tabulated in Table 3.

From Table 3, we observe the magnitude of t-test score is high in concern to the visit 1 and visit 4 clinical scores of the selected instances and it also shows that there is a larger mean difference between these visits. Thus we choose visit1 and visit4 for further analysis.

Table 3. Paired T-Test scores between different visits

Visits	t-test score	Mean Difference
1 and 2	6.368	0.2158
1 and 3	12.258	0.7341
1 and 4	19.304	1.8024
2 and 3	8.7816	0.5182
2 and 4	17.028	1.5865
3 and 4	11.86	1.0682

Computation of Missing Values

The missing values observed in the obtained longitudinal data from NACC are computed using Miss Forest Imputation method (20). Let the considered data be in the form of a matrix with columns corresponding to clinical scores P_j , where $[j=1,2,\dots,m]$ and rows corresponding to instances V_i , where $[i=1,2,\dots,n]$ of normal and demented subjects. Miss forest begins with the initial calculation of missing values by mean imputation method. In the next step, the number of missing values for P_j 's are determined and sorted them based on its increasing order. Where NAP_j represents the count of missing values (NA) for clinical score P_j . Based on the number of missing values, rearrange P_j with $NAP_1 < NAP_2 < NAP_3 \dots < NAP_m$. Fit a random forest for the observed values of P_x and P_j as the target and predictor variable respectively. Then predict one of the missing values of P_x based on trained random forest and repeat it for every attribute P_j . This process continues for several iterations until the difference between previous imputed matrix and newly imputed matrix i_{+1} increases for the first time. The Function 1 represent the step by step procedure followed for Miss Forest Imputation method.

Data Balance by SMOTE

Synthetic Minority Oversampling Technique (SMOTE) is a hybrid based method, where the samples of minority class are synthesized and samples of majority class are

removed randomly until the balanced ratio of both the classes is generated. Function 2 describes the synthesis of new minority samples by considering each sample M_i from the minority class and its nearest neighbors. The nearest neighbors are selected based on the minimum Euclidean distance concerning to each minority sample considered. In the next step, we calculate the difference between the feature vector of each minority sample and one of the randomly selected neighbors. Then, the calculated difference is multiplied by a random number between 0 and 1. By adding this result to the M_i , generates a new sample. It is followed by the under-sampling of samples from majority class.

Function 1. Miss Forest

Input : Original data with N*M matrix with missing values.
Output : Imputed data with N*M matrix without missing values
Step 1: The initial calculation of missing values is performed by using mean imputation method.
Step 2: Let data is of N * M where N is the number of records and M is the number of attributes.
Step 3: Count the number of missing values of each attribute A_j where $j = 1,2,3 \dots m$.
Step 4: Rearrange the attributes in ascending order based on the missing values.
Step 5: A random forest is fit by considering the observed values of both the target variable A_x and independent variables A_j .
Step 6: Then the missing value of the target variable A_x is predicted by using the trained random forest.
Step 7: If the difference between the matrix and newly created matrix x_{i+1} meets the stopping criteria, then imputation stops.

Function 2. Smote algorithm

Input: Data, with M number of Samples from Minority Classes, Amount of synthetic samples N%.
Output: N% of M Synthetic Minority Samples.
Initialize: Number of nearest neighbor k, $n = N\%$ of M.
Step 1: K nearest neighbors are computed for each instance M_i from Minority Class M.
Step 2: Randomly choose one of the K nearest neighbors of Minority Sample M_i .
Step 3: Diff: Calculate the difference between the nearest neighbor and M_i .
Step 4: Alpha: A random number between 0 and 1.
Step 5: New Sample: $M_i + (Diff * Alpha)$
Step 6: Repeat the step 2 - step 5 until n new samples are synthesized.

Feature selection and evaluation

Genetic Algorithm (GA)

GA (21) is a population-based heuristic search technique, which is inspired by the biological process of natural selection. It evolves all the possible candidate solutions from the population and the best candidate is selected through the evolutionary approach. This approach is composed of three operators such as selection, crossover and mutation. Function 3 lists the steps carried by a GA technique.

Logistic Regression

Logistic Regression is the type of regression analysis suitable for the data with binomial response (dependent) variable and one or more explanatory (independent) variables (22). The response variables considered for our study is normal and MCI or AD. While the neuropsychological or FAQ scores are considered as explanatory variables. The prediction of the response variable is obtained by the combination of explanatory variables and its coefficient which is as represented in Equation (2).

$$Q = \frac{e^{\beta_0 + P_1\beta_1 + P_2\beta_2, \dots, P_n\beta_n}}{1 + e^{\beta_0 + P_1\beta_1 + P_2\beta_2, \dots, P_n\beta_n}} \quad (2)$$

Where Q is the response variable, β_0 is the intercept term, $P = [P_1, P_2, P_3, \dots, P_n]$ is a set of explanatory variables and $\beta = [\beta_1, \beta_2, \beta_3, \dots, \beta_n]$ is a set of the regression coefficient.

Function 3. To separate the instances for four visit times

Input: Neuropsychological or FAQ Data
Output: Best features at the end of iterations
Initialize: <i>max_iterations, pop_size, cross over probability(P_c), mutation probability(P_m)</i>
Step 1: Initialize <i>population</i> of <i>pop_size</i>
Step 2: while <i>itr</i> \leq <i>max_iteration</i> do
evaluate <i>population</i> using eq (2)
while <i>size(new_population)</i> $<$ <i>pop_size</i> do
select parents from population using Tournament Selection
perform cross-over on parents with cross-over probability(P _c)
perform mutation with probability(P _m)
append new individuals to <i>new_population</i>
end while
<i>population</i> \Leftarrow <i>new_population</i>
<i>Itr</i> \Leftarrow <i>itr</i> + 1
end while

Experiments and results

The proposed framework is implemented in R tool and we have carried out the different experiments on the selected neuropsychological and FAQ scores at two different visit times.

Evaluation Parameters

To measure the performance of the logistic regression classifier, we considered the metrics such as Accuracy, MSE and AUC.

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP} + \text{FP}+\text{TN}+\text{FN}} \quad (3)$$

$$\text{MSE} = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (4)$$

AUC = Area Under Curve with respect to Sensitivity and 1-Specificity.

Where, True Positive (TP)-No of instances correctly classified as the demented subjects; True Negative (TN)-No of instances incorrectly classified as the non-demented subjects; False Positive (FP)-No of instances correctly classified as the non-demented subjects; False Negative (FN)-No of instances incorrectly classified as the non-demented subjects, and Predicted Value by LR model (\hat{y}_i)-Observed Value of the Scores in the dataset (y_i).

Parameters used for GA

In all the experiments, the parameters considered for GA are as follows: Initial population N of size 50, crossover rate Pc of 0.9, mutation rate Pm of 0.1, number of elitism is set to 2 and 20 numbers of generations is set as termination criteria.

Experiments on Neuropsychological Scores

In first experiment, MCI patients from normal subjects are discriminated by considering the neuropsychological scores from clinical visit-1 and these have 565 MCI and 4716 normal subjects (23). Next, we determine the percentage of missing values for all the scores from visit-1 and obtained 0.83%. Since the percentages of missing value are less, it did not show an impact on the classifier model. Thus, these scores are handled by SMOTE technique and optimized by GA.

The results obtained show that GA selects the same subset of features consistently across 20 generations and the performance evaluation for this selected subset of scores is shown in Table 4. The frequently selected scores of visit-1 are 2 (CDR GLOB), 3 (NACC GDS), 6 (MEMUNITS), and 8 (DIGIFLEN) respectively. From the obtained results, we infer the cognitive domains

effective to discriminate MCI patients from normal are episodic memory, working memory and depression. The combination of these features also shows better performance. Later, we perform cross-validation by dividing 4716 normal subjects into 10 equal sizes and thus we obtain 471 normal subjects in each subset and we consider 565 MCI subjects. The common features obtained with the highest frequency in both the approaches are as follows: 2 (CDR GLOB), 3 (NACC GDS), 6 (MEMUNITS), but the results of the proposed framework achieves better than the cross-validation results.

Table 4. Performance evaluation of subset of (visit-1) neuropsychological scores

Feature Number	AUC	Accuracy	MSE
1	0.947	0.9265	0.057
2,6	0.967	0.9403	0.042
2,4,6	0.968	0.9393	0.043
2,3,6,8	0.971	0.9427	0.041
8,3,2,6,10	0.971	0.9427	0.042

Next, the neuropsychological scores having 381 AD and 4398 normal subjects from clinical visit-4 are considered for the second experiment. We determined the percentage of missing data and it ranges between 4.54% - 16.18% and 8 out of 16 have the missing data equal to and above 14.27%. Hence, we subject the visit-4 data to two different multiple imputation approaches such as MICE and Miss Forest to compute the missing values. Then the imputed dataset are subjected to GA and the selected features are evaluated by using MSE as the parameter. The obtained results in Table 5. shows that the imputed dataset yield smaller MSE than the original dataset and by the comparison of two imputation methods, we observe that Miss Forest outperforms MICE. Thus, we consider Miss Forest in our proposed framework.

Table 5. Comparison of MSE of the features from visit-4 before and after imputation

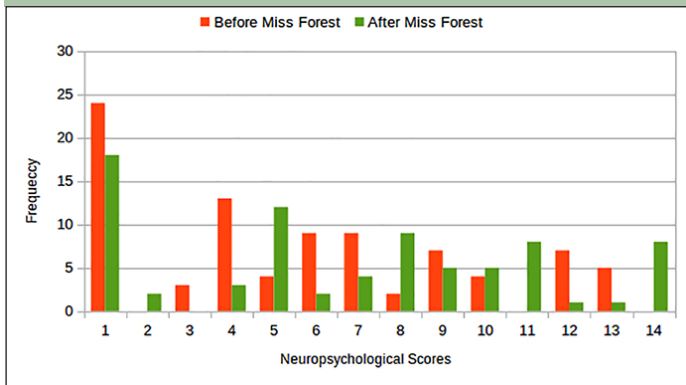
No of Features	Average MSE	Average MSE (with Miss Forest)	Average MSE (with MICE)
1	0.0191	0.0093	0.012
2	0.0191	0.0091	0.01026
3	0.01895	0.00909	0.01035
4	0.01863	0.00888	0.00972
5	0.01875	0.00889	0.00998

The significant features obtained with higher frequency after the miss forest imputation and SMOTE are 1(CDR SUM), 5 (LOGIMEM), 8 (DIGIFLEN), 14 (TRAILB), 11 (ANIMALS) respectively as shown in

Figure 3. From the obtained results, we infer the cognitive domains effective to discriminate AD patients are dementia severity, episodic memory, working memory and attention, executive function and language. Also, we observe that the feature selection for the considered neuropsychological scores of clinical visit-4 is improved by incorporating two preprocessing techniques: Miss Forest Imputation and SMOTE.

Next, we perform cross-validation for the neuropsychological scores of visit-4 similar to the visit-1 scores. We observe some of the common features obtained with highest frequency in both the approaches are as follows:1 (CDRSUM), 14 (TRAIL B) and 11 (ANIMALS) but the result of the proposed approach outperforms the cross-validation.

Figure 3. The Frequencies of Neuropsychological Scores Selected by GA



Experiments on FAQ Scores

We considered FAQ scores of visit-1 having 4716 instances from majority class and 565 instances from minority class in the ratio of 0.893:0.1069 and visit-4 have 4398 normal and 381 AD subjects in the ratio of 0.9202:0.079 for the third and fourth experiment. As these FAQs are scored based on the informant’s response, the missing values are less, therefore the data is not subjected to the imputation technique, but the scores from visit-1 and visit-4 are subjected to SMOTE technique because the imbalanced data subjected to the LR model shows AUC value in the range of 0.6955 to 0.7793, thus there is a requirement to improve the performance of the model. Then the GA is run for 20 times on imbalanced FAQ scores for the instances from both the visits and we obtain best subsets of features with minimum size. This is followed with the measurement of prediction performance of these selected feature subsets by LR model, by choosing AUC as the metric. The same procedure is followed for the balanced data, which is formed by applying SMOTE technique on the original dataset by choosing the percentage of oversampling equal to 600 and under sampling equal to 100. Then the performance of original and the balanced data are measured for the GA selected subset of features with the

size varying from 1-5. It is followed with the calculation of average AUC value for each selected subset size with and without SMOTE. The results of average AUC for both visit-1 and visit-4 are shown in Table 6. The improved performance shows the significance of SMOTE to handle the imbalanced data.

Table 6. Comparison of average AUC of the FAQ scores for visit-1 and visit-4 with and without smote technique

No of Features	Average AUC for Visit 1 (without SMOTE)	Average AUC for Visit 1 (with SMOTE)	Average AUC for Visit 4 (without SMOTE)	Average AUC for Visit 4 (with SMOTE)
1	0.6955	0.8138	0.9162	0.9187
2	0.7027	0.8699	0.91565	0.9587
3	0.7411	0.9029	0.9349	0.96835
4	0.7253	0.9328	0.9589	0.9784
5	0.7793	0.9342	0.94084	0.9749

The frequently selected features for visit-1 after SMOTE technique are 1 (BILLS), 6 (MEAL PREP), 9 (REMDATES) as shown in Table 7 and for the visit-4, we obtain 2 (TAXES), 8 (PAY ATTN), 9 (REMDATES) as the functional scores which aids to differentiate AD from normal subjects as shown in Table 8. Then we perform the cross-validation for both the visit-1 and visit-4 FAQ scores and it is observed that the same features are selected in both the approaches but the proposed approach outperforms the cross-validation results.

Table 7. Performance evaluation of subset of (visit-1) FAQ scores

Feature Number	AUC	Accuracy	MSE
9	0.8138	0.8002	0.1478
1, 9	0.8829	0.8675	0.1040
9,4	0.8569	0.8652	0.1088
2,6,9	0.9276	0.8985	0.0839
4,1,9	0.8946	0.8975	0.0848
6,2,9,3	0.9332	0.9128	0.0717
6,1,9,8	0.9290	0.9138	0.0708
2,6,7,8,9	0.9342	0.921	0.0650

Since the proposed approach shows better performance than the cross-validation results. So, the features selected by our proposed approach are considered to be significant for CAD of AD.

Table 8. Performance evaluation of subset of (visit-4) FAQ scores

Feature Number	AUC	Accuracy	MSE
2	0.9187	0.8980	0.0870
2,9	0.9587	0.9283	0.0577
8,2,9	0.9672	0.9414	0.0450
9,6,2	0.9615	0.9419	0.0442
2,9,8,5	0.9841	0.9530	0.0368
9,2,1,8	0.9727	0.9510	0.0383
1,8,2,4,9	0.9658	0.9606	0.0335
4,5,8,2,9	0.9845	0.9535	0.0339

Conclusions

In this paper, the proposed framework is significant to improve the performance of feature selection by GA in our paper. The tabulated results show the frequently selected neuropsychological and FAQ scores. The neuropsychological scores are required to identify the impaired cognitive domain and FAQ are required to identify an impaired functional domain. From the obtained results we infer that the cognitive domains impaired for visit-1 are memory and depression. In visit-4 it identifies the language and executive function along with the memory domain. The impaired functional domain for visit-1 are paying bills, remembering appointments, meal preparation. In visit-4 it identifies assembling tax records, remembering appointments and paying attention. In future, we perform the feature selection on the selected scores by using different optimization techniques.

Acknowledgement: We thank the curators of NACC database for providing the data for our research work.

Conflict of interest: All the authors declare no funding received from any source and there is no conflict of interest.

Ethical standards: This work was conducted in accordance with the guidelines provided by the Bangalore University.

References

- G. Emilien, C. Durlach, K. L. Minaker, B. Winblad, S. Gauthier, and J.M. Maloteaux, «Alzheimer Disease,» *Neuropsychology and Pharmacology*, 2012.
- S. T. McCutcheon, D. Han, J. Troncoso, V. E. Koliatsos, M. Albert, C. G. Lyketsos, and J.-M. S. Leoutsakos, «Clinicopathological correlates of depression in early Alzheimer's disease in the NACC,» *International Journal of Geriatric Psychiatry*, 2016.
- B. E. Gavett, L. Ashendorf, and A. S. Gurnani, «Reliable Change on Neuropsychological Tests in the Uniform Data Set,» *Journal of the International Neuropsychological Society*, vol. 21, no. 07, pp. 558–567, 2015.
- J. Hassenstab, S. E. Monsell, C. Mock, C. M. Roe, N. J. Cairns, J. C. Morris, and W. Kukull, «Neuropsychological Markers of Cognitive Decline in Persons with Alzheimer Disease Neuropathology,» *Journal of Neuropathology & Experimental Neurology*, vol. 74, no. 11, pp. 1086–1092, 2015.
- L. M. Besser, I. Litvan, S. E. Monsell, C. Mock, S. Weintraub, X.-H. Zhou, and W. Kukull, «Mild Cognitive Impairment in Parkinson's Disease versus Alzheimer's Disease,» *Parkinsonism Related Disorders*, vol. 27, pp. 54–60, 2016.

- J. Barnes, B. C. Dickerson, C. Frost, L. C. Jiskoot, D. Wolk, and W. M. van der Flier, «Alzheimer's Disease First Symptoms are Age Dependent: Evidence from the NACC Dataset,» *Alzheimer's Dementia*, vol. 11, no. 11, pp. 1349–1357, 2015.
- Battista, Petronilla, Christian Salvatore, and Isabella Castiglioni, «Optimizing Neuropsychological Assessments for Cognitive, Behavioral, and Functional Impairment Classification: A Machine Learning Study,» *Behavioural Neurology*, 2017.
- Machulda, M.M., Lundt, E.S., Albertson, S.M., Kremers, W.K., Mielke, M.M. et al. «Neuropsychological Subtypes of Incident Mild Cognitive Impairment in the Mayo Clinic Study of Aging». *Alzheimer's & Dementia*, vol. 15, pp. 878-887, 2019.
- Zhu, Yingying, Minjeong Kim, Xiaofeng Zhu, Jin Yan, Daniel Kaufer, and Guorong Wu. «Personalized Diagnosis for Alzheimer's Disease.» In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 205-213, 2017.
- Moradi, Elaheh, Ilona Hallikainen, Tuomo Hänninen, Jussi Tohka, and Alzheimer's Disease Neuroimaging Initiative. «Rey's Auditory Verbal Learning Test scores can be predicted from whole brain MRI in Alzheimer's disease.» *NeuroImage: Clinical*, vol. 13, pp. 415-427, 2017.
- C. Green and S. Zhang, «Predicting the Progression of Alzheimer's Disease Dementia: A Multidomain Health Policy Model,» *Alzheimer's & Dementia*, vol. 12, no. 7, pp. 776–785, 2016.
- S. E. John, A. S. Gurnani, C. Bussell, J. L. Saurman, J. W. Griffin, and B. E. Gavett, «The Effectiveness and Unique Contribution of Neuropsychological Tests and the Latent Phenotype in the Differential Diagnosis of Dementia in the Uniform Data Set,» *Neuropsychology*, vol. 30, no. 8, 2016.
- Kielb, Stephanie, Emily Rogalski, Sandra Weintraub, and Alfred Rademaker, «Objective Features of Subjective Cognitive Decline in a United States National Database,» *Alzheimer's and Dementia* vol.13, no. 12, pp.1337-1344, 2017.
- Musa, Gada, Fernando Henríquez, Carlos Muñoz-Neira, Carolina Delgado, Patricia Lillo, and Andrea Slachevsky. «Utility of the Neuropsychiatric Inventory Questionnaire (NPI-Q) in the Assessment of a Sample of Patients with Alzheimer's Disease in Chile,» *Dementia and Neuropsychologia*, vol.11, no. 2, pp. 129-136, 2017.
- Chen, Yingjia, Katherine G. Denny, Danielle Harvey, Sarah Tomaszewski Farias, Dan Mungas, Charles DeCarli, and Laurel Beckett. «Progression from Normal Cognition to Mild Cognitive Impairment in a Diverse Clinic-based and Community-based Elderly Cohort,» *Alzheimer's and Dementia*, vol.13, no. 4, pp. 399-405, 2017.
- Zhu, Yingying, Xiaofeng Zhu, Minjeong Kim, Dinggang Shen, and Guorong Wu. «Early Diagnosis of Alzheimer's Disease by Joint Feature Selection and Classification on Temporally Structured Support Vector Machine.» In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 264-272, 2016.
- Sørensen, Lauge, Christian Igel, Akshay Pai, Ioana Balas, Cecilie Anker, Martin Lillholm, Mads Nielsen, and Alzheimer's Disease Neuroimaging Initiative. «Differential Diagnosis of Mild Cognitive Impairment and Alzheimer's Disease using Structural MRI Cortical Thickness, Hippocampal Shape, Hippocampal Texture, and Volumetry.» *NeuroImage: Clinical*, vol.13, pp. 470-482, 2017.
- Beheshti, Iman, Hasan Demirel, Hiroshi Matsuda, and Alzheimer's Disease Neuroimaging Initiative. «Classification of Alzheimer's Disease and Prediction of Mild Cognitive Impairment-to-Alzheimer's Conversion from Structural Magnetic Resonance Imaging using Feature Ranking and a Genetic Algorithm.» *Computers in Biology and Medicine*, vol. 83 pp. 109-119, 2017.
- K. Steenland, J. Macneil, S. Bartell, and J. Lah, «Analyses of Diagnostic Patterns at 30 Alzheimer's Disease Centers in the US,» *Neuroepidemiology*, vol. 35, no. 1, pp. 19–27, 2010.
- Stekhoven, Daniel J., and Peter Bühlmann. «MissForest Non-Parametric Missing Value Imputation for Mixed-Type Data.» *Bioinformatics*, vol.28, no. 1, 2011.
- Johnson, Piers, Luke Vandewater, William Wilson, Paul Maruff, Greg Savage, Petra Graham, Lance S. Macaulay et al. «Genetic algorithm with logistic regression for prediction of progression to Alzheimer's disease,» *BMC bioinformatics* vol.15, no. 16, 2014.
- Hosmer Jr, David W., Stanley Lemeshow, and Rodney X. Sturdivant. *Applied logistic regression*. Vol. 398. John Wiley & Sons, 2013.
- Vinutha N, Sonu Sharma, P Deepa Shenoy, Venugopal K R., «Optimization of Neuropsychological Scores at the Baseline Visit Using Evolutionary Technique.» In the *Proceedings of Third IEEE International Women in Engineering Conference on Electrical and Computer Engineering (WIECON-ECE-2017)*, e-ISBN : 978-1-5386-2621-4, p-ISBN : 978-1-5386-2622-1, pp.55-59, December 18-19, 2017.