Interdiscip Sci Comput Life Sci
DOI 10.1007/s12539-016-0144-0

ORIGINAL RESEARCH ARTICLE

A New Decision Tree to Solve the Puzzle of Alzheimer's Disease Pathogenesis Through Standard Diagnosis Scoring System

Ashwani Kumar[1] · **Tiratha Raj Singh**[1]

Received: 26 June 2015 / Revised: 27 November 2015 / Accepted: 6 January 2016

Abstract Alzheimer's disease (AD) is a progressive, incurable and terminal neurodegenerative disorder of the brain and is associated with mutations in amyloid precursor protein, presenilin 1, presenilin 2 or apolipoprotein E, but its underlying mechanisms are still not fully understood. Healthcare sector is generating a large amount of information corresponding to diagnosis, disease identification and treatment of an individual. Mining knowledge and providing scientific decision-making for the diagnosis and treatment of disease from the clinical dataset are therefore increasingly becoming necessary. The current study deals with the construction of classifiers that can be human readable as well as robust in performance for gene dataset of AD using a decision tree. Models of classification for different AD genes were generated according to Mini-Mental State Examination scores and all other vital parameters to achieve the identification of the expression level of different proteins of disorder that may possibly determine the involvement of genes in various AD pathogenesis pathways. The effectiveness of decision tree in AD diagnosis is determined by information gain with confidence value (0.96), specificity (92 %), sensitivity (98 %) and accuracy (77 %). Besides this functional gene classification using different parameters and enrichment analysis, our finding indicates that the measures of all the gene assess in single cohorts are sufficient to diagnose AD and will help in the prediction of important parameters for other relevant assessments.

Keywords Dementia · Alzheimer's disease · Mini-Mental State Examination · Classification · Decision tree · Clustering · Validation

1 Introduction

Alzheimer's disease (AD) is the most common form of escalating dementia in the elderly. It is a neurodegenerative disorder marked by the neuropathologic hallmark of intracellular neurofibrillary tangles (NFT) and extracellular amyloid plaques that accumulate in susceptible brain regions [1]. Difficulties remembering recent events are often early symptoms. Later symptoms include impaired communication, disorientation, confusion, poor judgment, behavioral changes and, ultimately, difficulty in speaking, swallowing and walking [2, 3].

AD is sixth dominating cause of death in the USA. Current statistics indicates that about 25–30 million people are afflicted from AD, and the number of cases will triple by 2050 due to increasing life expectancy [4]. According to 2015 population tally, an estimated 5.3 million American of all ages are suffered from this disease [5]. In addition to this, AD puts psychological and economical burden on caregivers and exhibits a major public health problem being among the most expensive disease at global level [6].

The amyloid and the tau hypothesis recognized amyloid precursor protein (APP) and tau proteins as inducers and key players of the disease [7]. Genes encoding these two

Electronic supplementary material The online version of this article (doi:10.1007/s12539-016-0144-0) contains supplementary material, which is available to authorized users.

✉ Tiratha Raj Singh
tiratharaj@gmail.com

Ashwani Kumar
ashwani.kumar@mail.juit.ac.in

[1] Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan, H.P. 173234, India

Published online: 20 January 2016

proteins have a varied influence on developing AD, ranging from the autosomal dominant inheritance in the familial forms (1–5 % of cases) to the polygenic background in late-onset (>65 years of age) and sporadic AD (=95 % of cases). In addition to the genetic component, the risk of developing AD is influenced by several other factors which include socio-demographic, lifestyle, environment and medical conditions [8]. Age and female sex represent varied risk of developing AD (Fig. 1).

The genetic component, however, seems to be of major concern, since according to twin studies, a major part of the risk of sporadic AD is genetically determined [9]. The APP, presenilin 1 (PSEN1) and presenilin 2 (PSEN2) genes are currently known to be involved in the familial forms of AD [10]. Identification and characterization of dominant mutations of these genes were auxiliary for the understanding of the biological mechanisms which lead to enhanced Aß accumulation and senile plaques formation [11]. In contrast to the familial AD, the causing factors of the Aß accumulations and other pathological mechanisms remain mostly unclear in the sporadic form. The complex genetic model of sporadic form suggests that several heterogeneous susceptibility sets of genes may converge on the pathological processes that underlie the disease [12]. However, so far only the apolipoprotein E (APOE) gene has been definitively associated with the risk of AD [13].

APOE is involved in lipid transport and metabolism. Furthermore, it plays a specific role in the central nervous system, including neuronal development, regeneration and certain neurodegenerative processes [14]. The polymorphism of the APOE gene determines three isoforms of APOE protein (e2, e3 and e4) with different conformation and lipid-binding properties [15]. A proportional relationship was found between the number of the inherited e4 alleles and the risk of developing AD and the age at onset. The APOE e4 isoform prefers very low-density lipoprotein, and it is less effective in cholesterol transport as compared

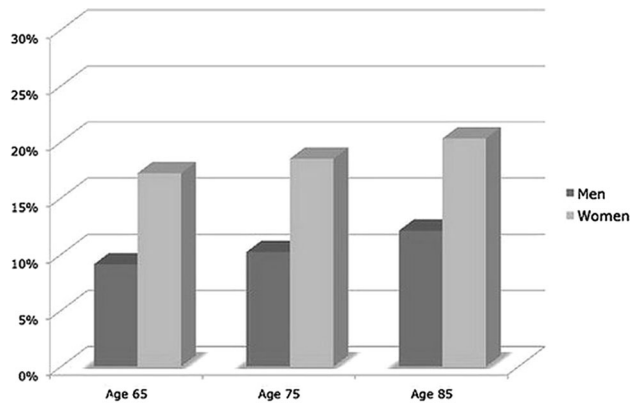

Fig. 1 Estimated lifetime risk of AD by age and sex ratio depicted in *bar chart*

to the other APOE isoforms. Membrane cholesterol modulates the cleavage of the APP protein, and in the presence of the e4 isoform, the balance is shifted to the production of Aß [16]. The amyloid cascade hypothesis has been the predominant model of molecular mechanisms underlying the pathogenesis of AD. According to this model, the genetic epidemiology of sporadic as well as familial AD remains a very active area of research, since a large part of the genetic etiology is still poorly understood and remains unresolved [17]. The aim of our work was to investigate direct entities and related attributes which are presumably involved in AD pathogenesis.

2 Materials and Methods

Classification of geneset data was carried out through RapidMiner Studio 6.2.0, Weka 3.6.9 and enrichment analysis for functional gene classification through David tool. Gene ontology (GO) study was done through web-based GEne SeT AnaLysis Toolkit (WebGestalt). All these analyses were performed on Window 7.0 platform running on a Lenovo PC with an Intel Core i5-2100 CPU processor and 4 GB of RAM.

2.1 Decision Tree

The decision tree represents tree-like structure to classify the data. Decision tree generates rules for classification [18]. A tree is represented by the set of nodes, leaves or branches. The root node is the attribute from which classification process starts, and the internal node corresponds to each of the questions about the particular attribute of the problem [19]. The branches coming out of the each node are labeled with possible values of attributes [20]. Each leaf node corresponds to a decision. The algorithm for generation of decision tree is partitioned into two parts: Top-down approach for induction of decision tree algorithm to choose features that partition the training data according to some evaluation function. Partitions are recursively split until some convergence criteria are reached. Secondly, the decision tree is pruned in order to avoid problem of over fitting [21] (Fig. 2).

The success of decision tree learning algorithm depends on evaluation criterion used to select the feature for splitting. Decision tree learning algorithm uses heuristic for estimating the best feature [22]. In our view, the measure yields a real positive number where the larger value indicates a set where there is more likelihood of all class values being present. Specifically, we use gain ratio (GR) as criteria for selection of features [23]. C4.5 algorithm implemented in RapidMiner tool treats missing value differently from normal values [24]. Association rules were generated

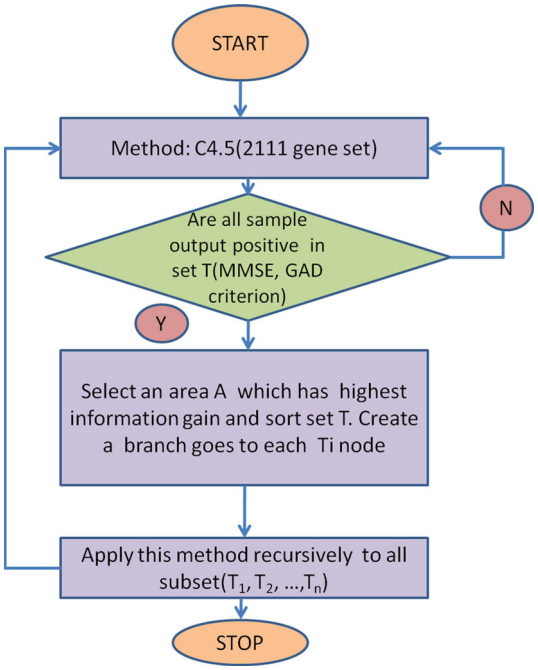

Fig. 2 Pipeline for decision tree building used for supervised classification of large gene dataset

which were determined by using Weka Tool, through various features.

2.2 Alzheimer's Gene Dataset

In this study, we have used Alzheimer's gene data available from various standard online resources such as ensemble gene, AlzGene, GenCard and NCBI. We selected 2111 raw genes that are known to relevant with Alzheimer's disease. Supplementary Table 1 contains list of genes used in our study. The dataset in our study consists of 14 attributes, which are describing different features of AD genes [25, 26]. Geneset was selected on the basis of the literature, and these genes were not only related to AD, in general, but also involved in other diseases. The most prioritizing attributes were selected based on the attribute selection techniques [27].

For selecting the attributes, the ranker technique was used. For the given dataset, the five different methods were applied. Chi-squared attribute Evaluation is the most widely used qualitative feature selection method. In order to reduce the effect of the bias resulting from the use of information gain, a variant known as gain ratio was utilized [28]. The gain ratio adjusts the information gain for each attribute to allow for the breadth and uniformity of the attribute values. Gain ratio is defined by the formula:

$$\text{Gain ratio} = \text{information gain}/\text{intrinsic information}.$$

The information gain attribute evaluation method evaluates the worth of an attribute by measuring the information gain with respect to the class.

$$\text{Gain}\,(\text{Class}, \text{Attribute}) = \text{H}\,(\text{Class}) - \text{H}\,(\text{Class}/\text{Attribute}),$$

where H is the information entropy.

2.3 Architecture and Model

The decision tree model is shown in Fig. 3. It begins with the collection of AD genes from different online resources, which is followed by preprocessing of the dataset including converting data into numeric values. Using geneset of 2111 gene, we use RapidMiner statistical tool and Weka Tool for data mining task. Further, we validated our model by using tenfold cross-validation application operator. Gene data include the descriptions such as gene name, association score, chromosomal position and MMSE score [29]. Enrichment analysis was also performed for this geneset to group genes on the basis of similarity using clustering algorithm. For more tightly associated gene in each group, stringency was maintained by keeping the kappa threshold 0.3 as anything below this threshold have great chance to be a noise. Finally, GO annotation was described through biological processes and molecular functions for the same.

3 Results and Discussion

Analysis was performed using classification of data by a J48 algorithm implemented in Weka, and C4.5 was used in RapidMiner. This analysis results in the correct classification of 950 out of 2111 genes. MMSE and Huge navigator were the most informative variable in this geneset obtained from information gain criterion. The classification algorithm provides MMSE score cutoff value for a different stage of the disease. We generated different rules through decision tree by selecting some specific parameter at each iteration [30].

The error rates and other features with respective classifier test for all decision trees are presented in Table 1.

After preprocessing the data which involve cleaning, integration and data reduction, the data mining was done. All analyses in this study involve tenfold cross-validation method to test the measures without pruning the tree. A single dataset of 2111 genes was used to build decision tree (DT). The dataset contains information about important attributes associated with genes. In order to improve the classifier, a supervised resample classifier was applied on data to assess the performance of the classifier. The number

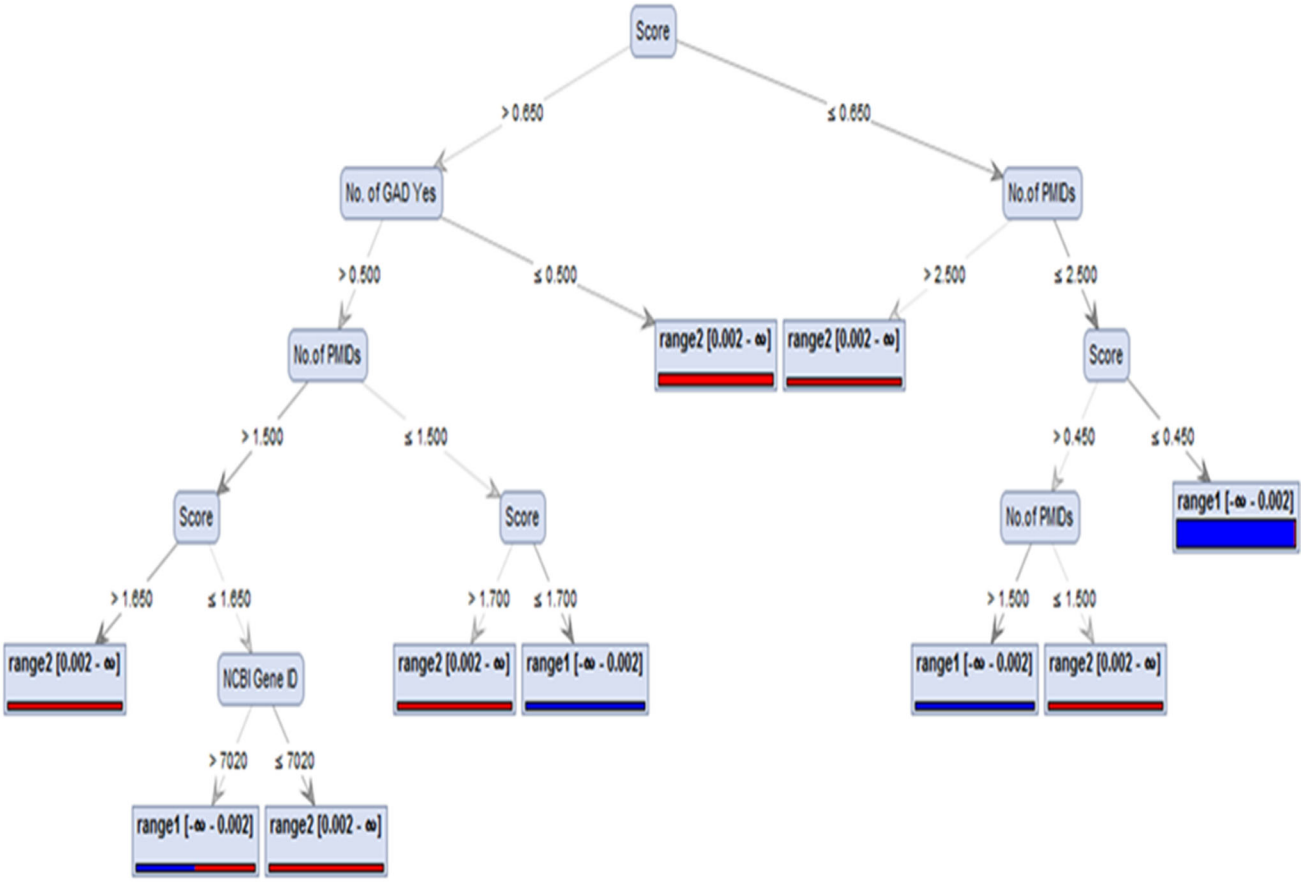

Fig. 3 Decision tree for the AD related genes. MMSE score is the root node at which classification of gene data has been done

Table 1 Statistical measure of performance of association rule (AR) based feature selection for sample group (AD genes dataset) using multiple setups

Data	Classifier	DT = 1	Features → error (%)				DT = 2	Feature → error (%)			
			F7	F11	F18	F32		F7	F11	F18	F32
2111 genes	C4.5	50	25.7	25.3	25.3	21.01	50	11.9	11.4	11.4	9.8
		100	23.9	23.2	23.2	20.4	100	9.8	9.5	9.5	7.4
		150	16.5	16.2	16.2	13.0	150	3.8	3.5	3.5	1.6

of tree was held constant at 100 while the number of features was kept varied at various points. Authenticity of built decision tree depends on accuracy which was measured 77 % with sensitivity 86 and specificity of the classifier was 81. On applying the priori algorithm, classifier generates ten best rules and was also verified (Tables 2, 3).

3.1 Decision Tree Induction

C4.5 algorithm implemented in the RapidMiner tool was used to make decision tree.

For making the decision tree, those feature were selected for classification whose values were not constant. Also we discretized the data on the basis of frequency. Different color node at the bottom of the tree reveals the class definition of leaf nodes.

3.2 Extracting Rule from RapidMiner

Based on the general process of data classification, we first identify a validation method to be used in modeling phase. The next step was to apply learning algorithm to training data and to generate rules. Decision trees were preferred since they are easy to understand. If the depth of tree increases, then IF–THEN rules can be extracted from decision tree.

To extract rule, every rule is drawn for each path from root node to the leaf node using some logical operators such as AND or IF. The summary table of decision tree (DT-1) is given below.

Table 2 Association between different classes based on a priori algorithm and best rule generated from classifier

Minimum support 0.2		
Minimum confidence 0.9		
Number of cycle performed 16		
Best rules found		
a1 = false a5 = false 24	class = c0 24	conf:(1)
a5 = false a8 = false 24	class = c0 24	conf:(1)
a5 = false a6 = false 23	class = c0 23	conf:(1)
a8 = false class = c1 22	a5 = true 22	conf:(1)
a5 = false a7 = true 21	class = c0 21	conf:(1)
a5 = false a9 = false 21	class = c0 21	conf:(1)
a3 = false a5 = false 20	class = c0 20	conf:(1)
a6 = false class = c1 20	a5 = true 20	conf:(1)
a2 = false a5 = false 27	class = c0 26	conf:(0.96)
a4 = false a5 = false 23	class = c0 22	conf:(0.96)

3.3 Generation of Another Decision Tree Using CHAID Algorithm to Determine How Variable Best Combined Using Chi-Square Interaction Method

The performance of classifiers for both decision trees is calculated from the confusion matrix (Fig. 4; Table 4).

The first decision tree inputs only representative genes test samples accurately at 98.94 % using tenfold cross-validation strategy (Tables 5, 6).

Based on gene data, ten rules were generated for prediction of disease in DT-I and six rules in DT-II. On the basis of standard scoring system of MMSE score, performance of the model was improved (Tables 7, 8).

3.4 Data Analysis by Weka Tool

The AD gene dataset contains 499 relevant instances and 14 attributes. The data were analyzed using a WEKA software utilizing decision tree J4.8 classification algorithm and Bayesian network, and a Naïve Bayes algorithm. The classifiers were directly applied without any feature (gene) selection. The number of top-ranked genes selected using feature selection techniques and then classifiers technique was applied, on the data. The Relief Feature Attribute Evaluator is used in WEKA Explorer with a default parameter setting (Table 9).

Table 4 Summary of binary decision tree (DT) II

Learning algorithm	C4.5
Attribute selection criterion	Gain ratio
Input	2111 gene dataset
Minimal gain	0.01
Maximal depth	12
Validation	Tenfold cross-validation
Minimal split size	5
Minimal leaf size	1
Number of pre-pruning	3

Decision trees were generated by RapidMiner Studio and Weka Tool. Key attribute selection is a very critical step for the impactful decision tree. There are a bunch of methods available, but ranker search method is very powerful technique among these.

The performance of the classification model is measured by its predictive accuracy from the independent dataset. First, the classification model was built from a subset of the data called the training set, where the algorithm knows the values of both predictor attributes and classes for the data instances. After the model was built, its predictive accuracy was then measured in a separate subset of the data, called the test set, where the algorithm knows only the values of the predictor attributes (not classes) for data instances. So, this measure of predictive accuracy measures the generalization ability of the classification model and gave it better applicability.

3.5 Enrichment Analysis

3.5.1 GO Analysis

Enrichment analysis of the gene highlights the most relevant terms associated with GO with given gene list. Annotation includes GO which includes biological process,

Table 3 Summary table of decision tree (DT-I) with all essential parameters

Learning algorithm	C4.5
Attribute selection criterion	Specifying the used method for selecting attributes, we choose gain ratio for this criterion
Minimal size for split	4
Minimal leaf size	1
Minimal gain	0.1
Maximal depth	20
Confidence value	0.25
Number of pre-pruning	3

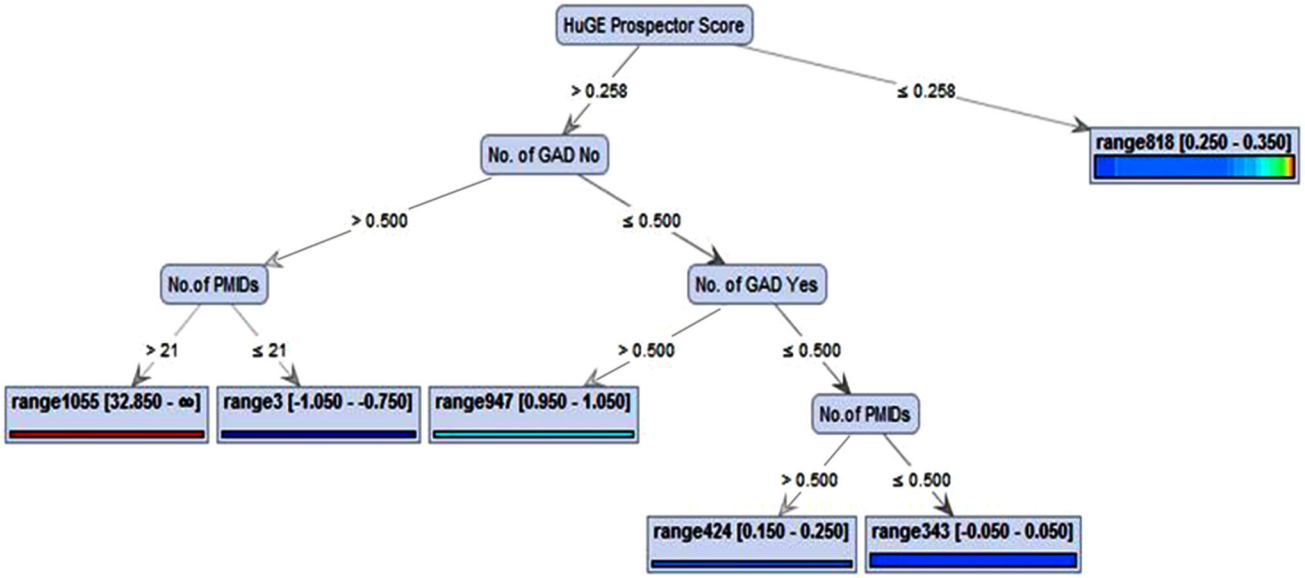

Fig. 4 Decision Tree of gene data based on Huge Prospector score as root node and number of GAD are important parameter for classification

Table 5 Confusion matrix for only representative 2111 gene for DT-I (without missing values)

	True range	True range	Class precision
Pred. range	1058	13	98.79 %
Pred. range	2	348	99.43
Class recall	99.81 %	96.40 %	

Table 6 Confusion matrix for only representative 2111 gene for DT-II (with missing values using CHAID method)

	True range	True range	Class precision
Prediction range	946	13	75.7 %
Pred. range	2	460	99.43 %
Class recall	75.7 %	83.5 %	

cellular component and molecular function [31]. The annotation coverage provides investigators with much more power to analyze their genes using different biological aspects in a single space. Annotation result for biological process and molecular function are displayed in the form of bar chart as shown in Figs. 5 and 6, respectively.

Members of independent groups fall into one of two mutually exclusive categories. Fisher's exact test was used to determine whether the proportions of those falling into each category differ by group [32]. To avoid over counting duplicated genes, the Fisher exact statistics is calculated based on

Table 8 Class for building and using a 0–R classifier

Decision table summary	
Number of training instances	100
Number of rules	31
Start set	No attributes
Search direction	Forward
State search after five node expansions	
Total number of subsets evaluated	68
Merit of best subset found	85
Evaluation (for feature selection)	CV (leave one out)
Feature set: 1, 6, 7, 9, 10, 11	
Correctly classified instances	77 %
Incorrectly classified instances	23 %
Kappa statistic	0.4912
Mean absolute error	0.3316
Root mean squared error	0.4002
Total number of instances	100

Predicts the mean (for a numeric class) or the mode (for a nominal class)

Table 7 Decision table classifier output

	TP rate	FP rate	Precision	Recall	F-measure	ROC area	Class
	0.818	0.324	0.831	0.818	0.824	0.83	C0
	0.676	0.182	0.657	0.676	0.667	0.83	C1
Weighted avg.	0.77	0.275	0.772	0.77	0.771	0.83	

Table 9 Attribute–criterion–feature selection from the data on the basis of rank generated from the score

Attribute Evaluator	Supervised Filtered Attribute Evaluator	
Score	Attribute	Rank
0.255714	a5	6
0.038926	a3	4
0.024319	a8	9
0.009714	a2	3
0.005152	a7	8
0.003551	a9	10
0.003551	a1	2
0.002202	a6	7
0.000531	a4	5
0.000168	a0	1

Feature selection is the technique of removing irrelevant features and to reduce dimensionality. The ranker search method is used to select attribute, and each selected attribute is ranked

corresponding ensemble gene IDs by which all redundancies in original IDs were removed. All results of chart report were generated after passing through the thresholds (by default, Max. Prob. $\leq$0.1 and Min. Count $\geq$2) [33].

The threshold of EASE score, a modified Fisher's exact *P* value, ranges from 0 to 1 and describes a statistically significant number of genes in the list with respect to the number in the population of genes from which the list derives. Fisher's exact *P* value = 0 represents perfect enrichment. Usually, *P* value is equal to or smaller than 0.05 to be considered strongly enriched in the respective annotation categories.

3.5.2 *Gene Functional Classification*

For this purpose, hypergeometric statistical method was applied using Bonferroni method at significant level (0.05). From 2111 geneset, David tool found 958 valid IDs and 320 genes were irrelevant (supplementary file II). This classification of gene was based on important parameter as *enrichment score* which rank the biological significance of gene group based on overall EASE score of all enriched annotation terms. From 958 genes in DAVID tool, only 74 genes passed the filter and only this number of gene shows the common annotation term profile of functional group based on frequency [34]. Finally, 39 functional gene groups came as output on the basis of enrichment score range from 39.92 to 1.207 in decreasing order by applying certain parameter like kappa threshold of 0.3 and multiple linkage threshold of 0.50 as default setting (Supplementary file II).

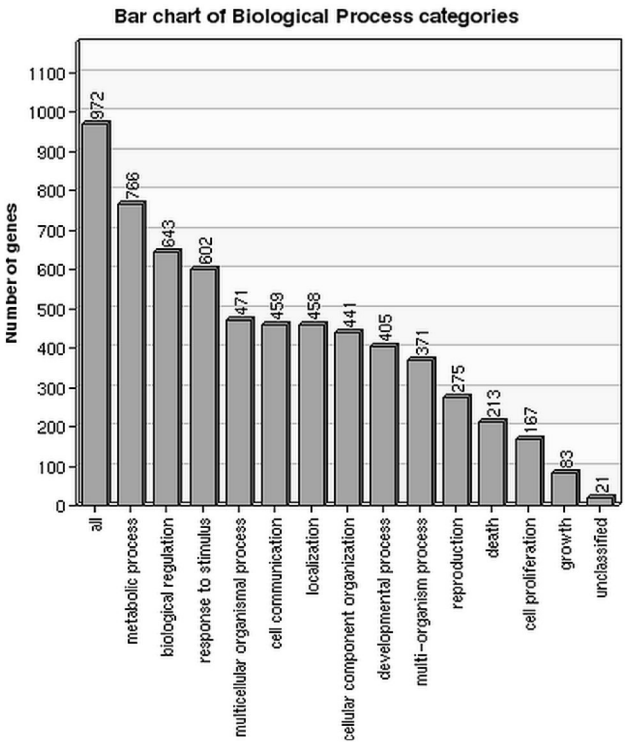

Fig. 5 Biological process under GO term; some biological phenomenon, commonly recognized series of events affecting state of an organism

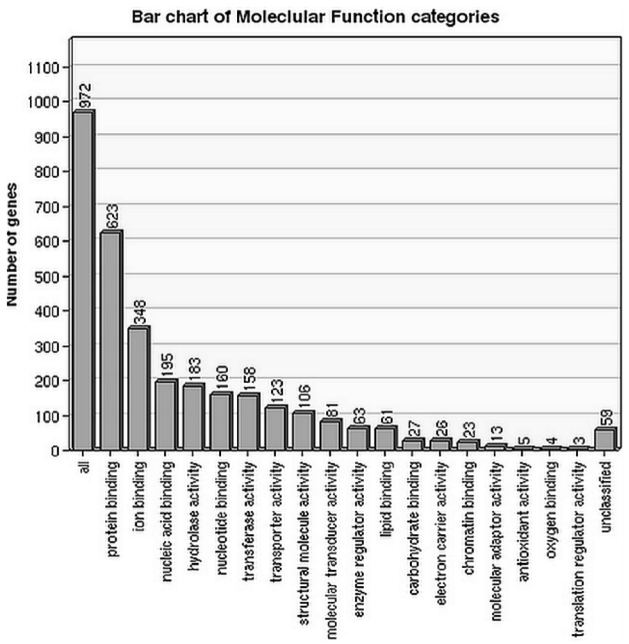

Fig. 6 Molecular function of involved gene under GO categories. The function carried out by a gene product; one product may carry out many functions; a set of functions together makes up a biological process

4 Conclusion

For human understanding, it is important to generate simple logic based classifier, i.e., in the form of a simple decision tree which describe the target concept. However, these simple decision trees may be of lower predictive value than other complex classifiers, but precision or accuracy generated from this classification model for gene dataset of 2111 genes is good (>80 %) while applying a C4.5 algorithm which follow branch and bound method for classification. The absolute error rate is very low. Specificity and sensitivity are also calculated to determine the suitability of designed model and are significant. Finally, from this study it is concluded that MMSE and relevance association score are important attributes for classification of genes and labeled them to a particular class. This type of testing and analysis has been used for selection of gene for expression arrays, automated protein data annotation, automatic cancer diagnosis, plant genotype discrimination, classifying gene expression profiles and computational model for mutational sites and then to extract best rules from designed models. Enrichment analysis alters us about the actual role of genes in term of GO, genes functional classification, pathway analysis, disease association, drug association and phenotype analysis. From gene functional classification analysis, we found APOE, PSEN1, GRN, ACE, BCHE, PRNP, IL1A are key genes that are strongly associated with AD whose association score ranges from 526.8 to 19.1 (Supplementary Table 1). Our proposed decision tree models and enrichment analysis of target genes will serve as a standard for computing biological phenomenon related to disease and exhibit their relevance toward AD conditions and its early diagnosis.

Acknowledgments Authors would like to acknowledge financial support from ICMR (BIC/12(33)/2012) to TRS.

References

1. Honjo K, Black SE et al (2015) Alzheimer's disease, cerebrovascular disease, and the β-amyloid cascade. Can J Neurol Sci 39(06):712–728
2. Braak H, Del Tredici K (2012) Where, when, and in what form does sporadic Alzheimer's disease begin? Curr Opin Neurol 25(6):708–714
3. Katzman R, Saitoh T (1991) Advances in Alzheimer's disease. FASEB J 5(3):278–286
4. Dartigues JF, Letenneur L (2000) Genetic epidemiology of Alzheimer's disease. Curr Opin Neurol 13(4):385–389
5. Williams-DeVane ClarLynda R et al (2013) Decision tree-based method for integrating gene expression, demographic, and clinical data to determine disease endotypes. BMC Syst Biol 7(1):119
6. Hardy J, Selkoe DJ (2002) The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. Science 297(5580):353–356
7. Hoenicka J (2005) Genes in Alzheimer's disease. Rev Neurol 42(5):302–305
8. Panigrahi PP, Singh TR (2013) Computational studies on Alzheimer's disease associated pathways and regulatory patterns using microarray gene expression and network data: revealed association with aging and other diseases. J Theor Biol 334:109–121
9. Scheuner D, Eckman C et al (1996) Secreted amyloid β-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. Nat Med 2(8):864–870
10. Wortmann M (2012) Dementia: a global health priority-highlights from an ADI and World Health Organization report. Alzheimers Res Ther 4(5):40
11. Strittmatter WJ, Saunders AM et al (1993) Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci 90(5):1977–1981
12. Levy E, Carman MD et al (1990) Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrhage, Dutch type. Science 248(4959):1124–1126
13. Corder EH, Saunders AM et al (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 261(5123):921–923
14. Hingorani AD, Liang CF et al (1999) A common variant of the endothelial nitric oxide synthase (Glu298→ Asp) is a major risk factor for coronary artery disease in the UK. Circulation 100(14):1515–1520
15. Heyman A, Wilkinson WE et al (1984) Alzheimer's disease: a study of epidemiological aspects. Ann Neurol 15(4):335–341
16. De Mãntaras RL (1991) A distance-based attribute selection measure for decision tree induction. Mach Learn 6(1):81–92
17. Fayyad UM, Irani KB (1992) On the handling of continuous-valued attributes in decision tree generation. Mach Learn 8(1):87–102
18. Maccioni RB, FarÃas G et al (2010) The revitalized tau hypothesis on Alzheimer's disease. Arch Med Res 41(3):226–231
19. Hastie T, Tibshirani R et al (2005) The elements of statistical learning: data mining, inference and prediction. Math Intell 27(2):83–86
20. Jensen R, Shen Q (2007) Fuzzy-rough sets assisted attribute selection. Fuzzy Syst IEEE Trans 15(1):73–89
21. Cuevas A, Febrero M et al (2004) An anova test for functional data. Comput Stat Data Anal 47(1):111–112
22. Tombaugh TN, McIntyre NJ (1992) The mini-mental state examination: a comprehensive review. J Am Geriatr Soc 40(9):922–935
23. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. Ijcai 14(2):1137–1145
24. Quinlan JR (2014) C4. 5: programs for machine learning. Elsevier, Philadelphia
25. Murphy C (1998) Induced decision trees for temporal medical data. In: AMCIS 1998 proceedings, p 66
26. Zhou Xiao Jia, Dillon Tharam S (1991) A statistical-heuristic feature selection criterion for decision tree induction. IEEE Trans Pattern Anal Mach Intell 8:834–841
27. Erdoğan O, Aydin SY (2013) Predicting the disease of Alzheimer with SNP biomarkers and clinical data using data mining classification approach: decision tree. Stud Health Technol Inform 205:511–515
28. Gutiérrez SLM, Rivero MH, Ramírez NC, Hernández E, Aranda-Abreu GE (2014) Decision trees for the analysis of genes

involved in Alzheimer's disease pathology. J Theor Biol 357:21–25
29. Yaneli AAM, Nicandro CR, Efrén MM, Nancy PC, Gabriel AMH (2013) Assessment of Bayesian network classifiers as tools for discriminating breast cancer pre-diagnosis based on three diagnostic methods. In: Batyrshin I, González Mendoza M (eds) Advances in artificial intelligence. Springer, Berlin, pp 419–431
30. Benuskova L, Kasabov N (2008) Modeling brain dynamics using computational neurogenetic approach. Cogn Neurodyn 2(4):319–334
31. Sehgal M, Singh TR (2014) Systems biology approach for mutational and site-specific structural investigation of DNA repair genes for xeroderma pigmentosum. Gene 543(1):108–117
32. Zhang CB, Zhu P, Yang P, Cai JQ, Wang ZL, Li QB, Bao ZS, Zhang W, Jiang T (2015) Identification of high risk anaplastic gliomas by a diagnostic and prognostic signature derived from mRNA expression profiling. Oncotarget 6(34):36643–36651
33. Sehgal M, Gupta R, Moussa A, Singh TR (2015) An integrative approach for mapping differentially expressed genes and network components using novel parameters to elucidate key regulatory genes in colorectal cancer. PLoS One 10(7):e0133901
34. Piovesan D, Giollo M, Ferrari C, Tosatto SC (2015) Protein function prediction using guilty by association from interaction networks. Amino Acids 47(12):2583–2592