



## A HYBRID CLASSIFICATION MODEL EMPLOYING GENETIC ALGORITHM AND ROOT GUIDED DECISION TREE FOR IMPROVED CATEGORIZATION OF DATA

R. Geetha Ramani, Lakshmi Balasubramanian and Alaghu Meenal A.

Department of Information Science and Technology, College of Engineering, Guindy, Anna University, Chennai, India

E-Mail: [rgeetha@yahoo.com](mailto:rgeetha@yahoo.com)

### ABSTRACT

Data mining algorithms play a major role in analyzing the vast data available in many fields like multimedia, medicine, business, education etc. Classification techniques have been extensively adopted for the purpose of pattern analysis. Several classification algorithms have been proposed in the literature. Yet demand exists for classification algorithms that yield higher accuracies. Hybrid classification procedures were also attempted in the literature. In this paper, the concept of Genetic Algorithm and Decision Tree is employed collectively for achieving better accuracies. The proposed methodology adopts genetic search to generate subsets of the attributes of the data and these subsets are evaluated using the Root Guided Decision Tree. This process results in a final decision tree with relevant set of attributes and yielding higher accuracy. The algorithm is validated on the datasets obtained from UCI repository and retinal dataset acquired from a publicly available High Resolution Fundus image Dataset.

**Keywords:** data mining, classification, decision tree, genetic algorithm, UCI dataset.

### INTRODUCTION

The huge availability of data and the necessity to retrieve useful information from it has increased the demand of efficient data mining algorithms [1-3]. Data mining is a branch of computational intelligence which aims at deriving useful and hidden patterns in the available data. Data mining constitutes of supervised and unsupervised learning techniques. Supervised learning techniques require class label of the data for the learning process while the unsupervised learning group data based on some similarity measure. Classification techniques fall under the supervised learning technique and has been widely used for the purpose of data analysis. Decision trees are one of the most effective ways for representing the rules built by the classification model. Several decision trees were proposed in the literature in aim to classify the data and form rules. Some of them include C4.5 [4], Best First Tree (BFT) [5], Classification and Regression Tree (CART) [6] etc. Yet, the demand for new classification algorithms that yield higher accuracies exists. Attempts have also been made to design hybrid classification models that combine either two classification algorithms or combine the technique of supervised and unsupervised methods or combine some other concept of computational intelligence with that of the classification techniques. These hybrid techniques yielded better accuracies than the individual classification models. In this paper, the concept of Genetic Algorithm and Decision Trees has been employed collectively in the view to achieve increased classification accuracies. Genetic Algorithms (GA) [7] are a part of evolutionary computing, inspired by Darwin's theory about evolution. Solution to a problem solved by Genetic Algorithms is evolved [8].

The proposed model utilizes Genetic Algorithm for generating subsets of attributes of the available data. These subsets of attributes are evaluated through Root Guided Decision Tree. The Root Guided Decision Tree (RGDT) [9] is built as a forest of trees where the number of trees built is based on the number of features in the training data. Every attribute is given as a root node for a tree and the tree with best accuracy is used for learning the rules. The subsets generated by Genetic Algorithms evolve based on their ability to generate best RGDT hence resulting in the relevant set of attributes and its corresponding best RGDT. The proposed classification model is validated on datasets from UCI machine learning repository [10] and a publicly available retinal image dataset namely High Resolution Fundus Image Database [11].

The paper is organized as follows: Section 2 presents the related work. Section 3 explains the proposed classification model employing Genetic Algorithms and Root Guided Decision Tree. Section 4 highlights the experimental results. Finally Section 5 concludes the paper.

### Related work

Classification through Decision Trees [12] offers a rapid and an effective method for analyzing datasets. Decision Tree is where a tree is constructed to model the classification process. Different decision trees exist in the literature. Hybrid variations of decision trees were also analyzed to achieve better performance. This section provides a brief discussion on the various decision trees and hybrid models available in the literature.



Various decision trees such as ID3 [13], C4.5 [4], Best First Tree [5], CART [6] etc. are briefly presented here. ID3 [13] algorithm chooses the best attribute based on entropy and information gain for constructing the tree. Then, C4.5 [4] algorithm was proposed which utilized the basic concept of ID3 but computes Gain Ratio for evaluation of attributes. Further, Grafted C4.5 [14] was introduced which generates a grafted decision tree from a C4.5 tree algorithm. It is an inductive process that adds nodes to inferred decision trees.

Another decision tree, CART [6] gives the results as either classification or regression trees, depending on categorical or numeric data set. It is a binary decision tree as it generates only two branches at each node. In another attempt, Best First Tree was introduced which works on the principle of maximum reduction of impurity. Further, REP Tree is a fast decision tree learner which builds a decision tree or regression tree using information gain as the splitting criterion, and prunes it using reduced error pruning. Another classification procedure, Naive Bayes (NB) [15] is also widely used for classification of real life problems. Various works have analyzed the performance of these classifiers which are briefed below.

In 2010, Karegowda *et al.* [16] used wrapper approach with Genetic Algorithms as random search technique for subset generation, with different classifiers namely C4.5, Naive Bayes, Bayes networks and Radial basis function as subset evaluating mechanism on four datasets namely Pima Indians Diabetes Dataset, Breast Cancer, Heart Stat log and Wisconsin Breast Cancer. In 2011, Aman Kumar Sharma *et al.* [17] investigated four decision trees namely Alternating Decision Tree, C4.5, ID3 and CART algorithms for classification of spam e-mail dataset and it was observed that C4.5 performed the best with an accuracy of 92.76%. Aruna *et al.* [18] provided an empirical comparison of accuracy, precision and recall of C4.5 and CART trees on different datasets from the UCI repository. GeethaRamani *et al.* [19] investigated the performance of various classifiers on a fundus image dataset to identify images that are normal, affected by Retinopathy and affected by Glaucoma. It was observed that C4.5 and Random Tree achieved the highest training accuracy. Shomona Gracia Jacob *et al.* [3] demonstrated that C4.5 achieved 100% classification accuracy on the various medical datasets available in UCI repository.

Hybrid models were also analyzed in regard to yield high performance. Some of these hybrid models are discussed here. Polat and Gunes [20] proposed a hybrid classification system based on a C4.5 classifier and a one-against-all method to enhance the classification accuracy for multi-class classification problems. Their one-against-all method constructed  $M$  number of binary C4.5 decision tree classifiers, each of which separated one class from all of the rest. Another approach was introduced for building classification model based on adjusted cluster analysis classification called classification by clustering [21]. There

existed similarities between instances clustered in a cluster and the target class assigned to it. So, in each cluster, the target class distribution was calculated. When a threshold for the number of instances stored in a cluster was attained, all the instances in each cluster were classified pertaining to the appropriate value of the target class. Subsequently, Aitkenhead [22] introduced a co-evolving decision tree method, where a large number of attributes in datasets were considered. They proposed a novel combination of Decision Trees and evolutionary methods, such as the bagging approach and back propagation neural network approach to enhance the classification accuracy.

Then, in 2014, an integration of supervised and unsupervised learning method was presented [23]. K-Means clustering was combined with decision tree, Bayesian network, logistic regression, multilayer perceptron, radial basis function, and support vector machine algorithms to enhance the accuracy results. Subsequently, Farid *et al.* proposed two hybrid models based on the concept of removal of misclassified instances [24]. Firstly, Naive Bayes Classifier was employed on the data followed by the application of C4.5 classifier to the correctly classified instances of the Naive Bayes Classifier. Another hybrid model, in which C4.5 was applied first on the data, followed by the use of naive bayes classifier on the correctly classified instances from the C4.5 classifiers.

Though there exists a variety of decision trees, there still exist demand for new classification models yielding high accuracy. The proposed classification model is described in the next section.

### Proposed hybrid classification model

Decision trees have been widely used for the purpose of analyzing huge data and deriving hidden patterns from it. The proposed hybrid classification model employs Genetic Algorithm (GA) [7] and Root Guided Decision Tree (RGDT) [9] in view to achieve higher accuracy.

The dataset is composed of attributes and instances. The relevance of the attributes in deriving the patterns is very important. Some attributes which do not contribute useful information, may deviate the rules and hence decrease the classification accuracy. Hence the process of choosing relevant attributes from the entire set of attributes gains more importance. Also when the number of attributes is very high, the dimensionality of the data increases, increasing the complexity of the process. In the proposed methodology, to attain a relevant set of attributes from the entire set of attributes, the concept of Genetic Algorithm is adopted. It is a random search method, capable of effectively exploring large search spaces [25]. Genetic Algorithms performs a global search unlike many search algorithms, which perform a local, greedy search. The basic idea is to evolve a population of



individuals, where each individual is a candidate solution to a given problem.

Initially, a set of random individuals (an individual represents a set of attributes in this case) is selected. The fitness of these individuals is computed through its ability to generate the best RGDT. Hence fitness is the accuracy obtained by the RGDT with the set of attributes in the considered individual. Further the algorithm proceeds with its genetic operators namely reproduction, crossover, and mutation. Reproduction passes the best individual to the next generation without applying any change to it. Crossover operation combines individuals with high fitness to generate better individuals and mutation alters an individual locally to attempt to create a better individual. Mutation also helps in overcoming the local maxima issue. This process of evolution in Genetic Algorithm continues till the termination criterion is reached (either the required fitness or the number of generations). In each generation, the population is evaluated and tested for termination of the algorithm. If the termination criterion is not satisfied, the population is operated upon by the Genetic Algorithm operators and then re-evaluated. This procedure is continued until the termination criterion is met. Once the termination criterion for the genetic search is reached, the best subset of attributes is returned by the Genetic Search for which the best tree is produced.

In the proposed model, Genetic Algorithm uses the Root Guided Decision Tree [26] for evaluation of the individuals. The Root Guided Decision Tree evaluates the subsets of attributes ( $m$ ) in each individual, where in each individual, every attribute of the subset is given as a root node and  $m$  trees are generated for each subset containing  $m$  attributes. Once all the trees for the subset are produced, the tree which produced the best merit is said to be assigned for that subset and the fitness for the subset is calculated. The algorithm for the proposed methodology is presented in Figure-1 while the RGDT algorithm is presented in Figure-2 [9].

Fitness - function evaluates how good a hypothesis is  
 Fitness\_threshold - minimum acceptable hypothesis  
 $p$  - size of the population  
 $r$  - fraction of population to be replaced  
 $m$  - mutation;  $P$ - population;  $D$ - Data  $A$ - Attributes  
 $M$ -Number of attributes.

**GA (Fitness, Fitness\_threshold,  $p$ ,  $r$ ,  $m$ )**

Step 1: Initialize:  $P \leftarrow p$  random subset of attributes  
 Step 2: Evaluate: for each  $h$  in  $P$ , where  $h$  contains  $\{D, A, M\}$ , compute FOREST\_OF\_RGDT ( $D, A, M$ )  
 Step 3: Compute fitness for every tree.  
 Step 3: while  $[\max_h \text{Fitness}(h)] < \text{Fitness\_threshold}$   
 Step 3.1: The Tree with maximum fitness is retained for next generation  $P_s$   
 Step 3.2: Select  $(1 - r)$  members of  $P$  to add to  $P_s$  based on fitness  
 Step 3.2: Crossover: Probabilistically select pairs of hypotheses from  $P$ . For each pair,  $\langle h_1, h_2 \rangle$ , produce two offspring by applying the Crossover operator. Add all offspring to  $P_s$ .  
 Step 3.3: Mutate: Invert a randomly selected bit in  $m$ .  
 Step 3.4: Reproduction: The tree with the maximum fitness is retained and sent to  $P_s$   
 Step 3.4: Update:  $P \leftarrow P_s$   
 Step 3.5: Evaluate: for each  $h$  in  $P$ , compute Fitness ( $h$ )  
 Step 4: Return the subset from  $P$  that has the highest fitness  
**Output:** The tree with the best set of features.

Figure-1. Proposed algorithm employing GA and RGDT.

Let

$D$ : Dataset containing  $N$  instances along with their class label  
 $A$ : set of attributes  
 $M$ : Number of attributes

FOREST\_OF\_RGDT( $D, A, M$ )

Step 1: For  $i = 1$  to  $M$  do  
 Step 2: Call ROOT\_RGDT( $D, A, i$ )  
 Step 3: end  
 Step 4:  $\text{Tree}_{\text{best}} = \text{Tree}$  yielding the highest training accuracy  
 Step 5: Return  $\text{Tree}_{\text{best}}$

ROOT\_RGDT( $D, A, i$ )

Step 1: Create a root node  $RN$ .  
 Step 2: If all instances in  $D$  belong to the same class  $C$ , then  
     Return  $RN$  as the leaf node labeled with class  $C$ .  
 Step 3: Let  $a_i = i^{\text{th}}$  attribute in  $A$   
 Step 4: Label node  $RN$  with  $a_i$  and let it test the splitting criterion.  
 Step 5: For each outcome  $j$  of the splitting criterion,  
      $D_j =$  data instances in  $D$  satisfying outcome  $j$ .  
     If  $D_j$  is empty, then  
         Attach a leaf labeled with the majority class in  $D$  to node  $RN$ .  
     Else, attach the node returned by recursively calling  
         NONROOT\_RGDT( $D, A$ ).  
 Step 6: Return  $RN$ .

NONROOT\_RGDT( $D, A$ )

Step 1: Create a node  $N$ .  
 Step 2: If all instances in  $D$  belong to the same class  $C$ , then  
     Return  $N$  as the leaf node labeled with class  $C$ .  
 Step 3: If  $A$  is empty, then  
     Return  $N$  as a leaf node labeled with the majority class in  $D$ .  
 Step 4: For all attributes  $a$  in  $A$ , compute gain ratio as follows:  

$$\text{GainRatio}(a) = \frac{\text{Gain}(a)}{\text{SplitInfo}(a)}$$

$$\text{Where } \text{splitinfo}_a(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} * \log_2\left(\frac{|D_j|}{|D|}\right)$$
 Step 5: Assign  $a_{\text{best}} =$  attribute with maximum gain ratio  
 Step 6: Label node  $N$  with  $a_{\text{best}}$  and let it test the splitting criterion.  
 Step 7: For each outcome  $j$  of the splitting criterion,  
      $D_j =$  data instances in  $D$  satisfying outcome  $j$ .  
     If  $D_j$  is empty, then  
         Attach a leaf labeled with the majority class in  $D$  to node  $N$ .  
     Else, attach the node returned by recursively calling  
         NONROOT( $D_j, A$ ).  
 Step 8: return  $N$ .

Figure-2. Algorithm for generation of RGDT.



Various experiments were performed to evaluate the performance of the proposed classification model. The experimental results are discussed in the following section.

### Experimental results

Various experiments were conducted to assess the performance of the proposed algorithm. The proposed classification model was implemented in Weka 3.6.2, an open source data mining tool [27]. Different datasets were obtained from the UCI Machine Learning Repository [10] and public retinal image repository [11] to validate the ability of the proposed classification model in categorizing the data. The datasets acquired from UCI repository include Contact lenses, Diabetes, Soybean, Vote, Breast

Cancer, Weather, Zoo, Labor, Vowel, Primary Tumor, Hepatitis, Ionosphere, Vehicle, Lymph and Autos datasets. Another clinical dataset was obtained from publicly available database namely High Resolution Fundus image database (HRF) [11, 28]. The dataset consists of sample images containing healthy, Diabetic Retinopathy affected and Glaucoma affected images. In this work, it is attempted to categorize the images as either belonging to healthy, diabetic retinopathy or glaucoma affected (HRF-HGDR) from the texture features of the entire images. The details of the datasets highlighting the number of attributes, number of instances and number of classes are tabulated in Table-1.

**Table-1.** Details of the datasets used for experimentation.

Dataset	Number of attributes	Number of instances	Number of classes
Contact Lenses	4	24	3
Diabetes	8	768	2
Soybean	35	683	19
Vote	16	435	2
Breast Cancer	9	286	2
Weather	4	14	2
Zoo	17	101	7
Labor	16	57	2
Vowel	13	990	11
Primary Tumor	17	339	22
Hepatitis	19	155	2
Ionosphere	34	351	2
Vehicle	18	846	4
Lymph	18	148	4
Autos	25	205	6
HRF-HGDR	11	45	3

The experimental data is carefully chosen so that the algorithm is evaluated on all type of data with varying cardinalities of attributes, instances and classes. Performance of the decision trees are compared using the classification accuracy. Accuracy [29] is defined as the ratio of number of correctly classified instances to the total number of instances. Evaluation techniques used for assessing a classification model include cross validation, leave one out cross validation, bootstrapping and train-test techniques etc. In this paper, the results obtained through cross validation are demonstrated as classification accuracy.

### Performance comparison of different decision tree classifiers

Experiments were performed to evaluate the different decision trees. Five existing decision trees namely C4.5, Best First Tree (BFT), Classification and Regression Trees (CART), Reduced Error Pruning Tree (REP) and RGDT were tested on the dataset to exhibit the outstanding performance of the proposed classifier model. Ten fold cross validation was set for experimental trials. Table-2 exhibits the classification accuracy (%) of the different decision trees. The results reported are the classification accuracy obtained from the unpruned trees.

**Table-2.** Performance comparison of different classifiers based on classification accuracy (%).

Dataset	C4.5	BFT	CART	REP	RGDT
HRF-HGDR	60.00	57.77	57.77	57.77	<b>71.11</b>
Contact Lenses	70.83	75.00	75.00	70.83	<b>83.33</b>
Diabetes	72.65	71.75	71.75	70.31	<b>73.57</b>
Soybean	91.36	91.80	91.80	89.60	<b>92.39</b>
Vote	96.32	94.94	94.94	95.86	<b>96.32</b>
Breast Cancer	69.58	60.48	60.48	66.78	<b>71.68</b>
Weather	57.14	64.28	64.28	64.28	<b>71.43</b>
Zoo	92.07	19.80	19.80	40.59	<b>95.05</b>
Labor	78.94	78.94	77.19	77.19	<b>87.72</b>
Vowel	83.53	81.21	81.31	84.04	<b>87.98</b>
Primary Tumor	40.41	39.23	39.23	35.39	<b>41.89</b>
Hepatitis	80.64	80.00	80.00	78.06	<b>83.22</b>
Ionosphere	91.45	88.88	88.88	89.74	<b>92.88</b>
Vehicle	72.81	70.33	70.21	73.28	<b>73.99</b>
Lymph	77.07	77.02	77.02	72.29	<b>81.08</b>
Autos	84.39	75.60	76.58	81.46	<b>85.37</b>

From Table-2, it is seen that RGDT performs the highest for all the datasets and C4.5 performs the second highest. Hence experimental trials were conducted employing Genetic Algorithm with RGDT and Genetic Algorithm with C4.5.

#### **Performance comparison of hybrid classification model employing GA and decision trees**

Investigation to assess the performance of the hybrid algorithms employing GA and decision trees was performed. The parameter settings for Genetic Algorithms include initial population size of 10, maximum number of generations of 50, Single point crossover with crossover probability of 0.6 and mutation probability of 0.33. Table-3 presents the results of the experimental trials employing GA with RGDT and GA with C4.5.





**Table-3.** Performance of hybrid classification model employing GA and decision tree based on classification accuracy (%).

Dataset	C4.5	RGDT	GA+C4.5	GA+RGDT
HRF-HGDR	60.00	71.11	<b>75.55</b>	<b>80.00</b>
Contact Lenses	70.83	83.33	87.5	87.5
Diabetes	72.65	73.57	74.21	76.43
Soybean	91.36	92.39	93.41	93.41
Vote	96.32	96.32	96.78	97.01
Breast Cancer	69.58	71.68	75.87	75.87
Weather	57.14	71.43	71.42	78.5
Zoo	92.07	95.05	98.01	98.02
Labor	78.94	87.72	87.71	92.98
Vowel	83.53	87.98	84.34	87.98
Primary Tumor	40.41	41.89	44.24	47.19
Hepatitis	80.64	83.22	85.16	87.09
Ionosphere	91.45	92.88	93.73	93.73
Vehicle	72.81	73.99	75.17	76.29
Lymph	77.07	81.08	85.81	85.81
Autos	84.39	85.37	85.85	86.82

Form Table-3, it is evident that the performance of the proposed classifier model based on Genetic Algorithm and Root Guided Decision Tree outperforms the existing classification models.

The proposed hybrid classification model can thus be utilized for the purpose of efficient categorization of real time problems.

## CONCLUSIONS

Many application areas utilises data mining algorithms to derive useful information from raw data. There have been many decision trees in the literature to solve numerous real world problems. C4.5, Best First Tree, Classification and Regression Trees and Reduced Error Pruning Tree are some of the most widely used decision trees. Root Guided Decision tree is a decision tree in which the root control is obtained. In this paper, a hybrid model employing Genetic Algorithm and Root Guided Decision tree is proposed. Genetic Algorithm is used to evolve the relevant subset of attributes while Root Guided Decision tree is utilized to assess the merit of the subset of the attributes. The final relevant set of attributes and hence the best decision tree is obtained achieving high accuracy results. The performance of the proposed model was evaluated on UCI Machine Learning repository and publicly available retinal image datasets. Experimental results affirm the fact that the hybrid Genetic Algorithm

and RGDT combination exhibits outstanding performance when compared to the other classification models.

## REFERENCES

- [1] Jiawei Han, Micheline Kamber and Jian Pei. 2011. "Data mining: Concepts and techniques", The Morgan Kaufmann Series in Data Management Systems, Third Edition.
- [2] Shanthi A. and Geetha Ramani R. 2011. "Classification of vehicle collision patterns in road accidents using data mining algorithms", International Journal of Computer Applications, vol. 35, no. 12, pp. 30-37.
- [3] Shomona Gracia Jacob and R. GeethaRamani. 2010. "Data mining in clinical data sets: a review", International Journal of Applied Information Systems, vol. 4, no. 6, pp. 15-26.
- [4] Steven L. Salzberg. 1993. "C4.5: Programs for machine learning by J. Ross Quinlan", Morgan Kaufmann Publishers, Machine Learning, vol. 16, no. 3, pp. 235-240.



- [5] Shi, Haijia. 2007. "Best First Decision Tree Learning", University of Waikato.
- [6] L. Breiman, J.Friedman, C.J Stone and R.A.Olshen. 1984. "Classification and regression trees", Chapman and Hall/CRC.
- [7] D. Goldberg. 1989. "Genetic algorithms in search, optimization and machine learning", Addison-Wesley, First Edition.
- [8] R. Geetharamani and Lakshmi Balasubramanian. 2011. "Genetic algorithm solution for cryptanalysis of knapsack cipher with knapsack sequence of size 16", International Journal of Computer Applications, vol. 35, no. 11, pp. 17-23.
- [9] Geetha ramani, Lakshmi Balasubramanian, Alaghu Meenal. A. 2015. "Decision tree variants (Absolute Random Decision Tree and Root Guided Decision Tree) for improved classification of data", International Journal of Applied Engineering Research, Vol. 10, No. 17, pp. 13190-13195. [in press].
- [10] A. Frank, and A. Asuncion, A. 2010. "UCI machine learning repository", [http:// archive.ics.uci.edu/ml](http://archive.ics.uci.edu/ml). Accessed 29.07.2013.
- [11] Budai Attila and Jan Odstreilik, "High-Resolution Fundus (HRF) Image Database". Available at: <http://www5.cs.fau.de/research/data/fundus-images/>
- [12] Ian H.Witten and Elbe Frank. 2005. "Data mining Practical Machine Learning Tools and Techniques", Morgan Kaufmann Publishers, Second Edition.
- [13] J.R.Quinlan. 1986. "Induction of Decision Trees", Machine Learning, vol. 1, pp. 81-106.
- [14] Geoffrey I. Webb. 1999. "Decision Tree Grafting from the all-tests-but-partition", IJCAI Proceedings of the 16<sup>th</sup> International Conference on Artificial Intelligence, vol. 2, pp. 702-707.
- [15] L. Koc, T.A. Mazzuchi and S.Sarkani. 2012. "A network intrusion detection system based on a hidden naive Bayes multiclass classifier", Expert Systems with Applications, vol. 39, pp. 3492-3500.
- [16] Karegowda, Jayaram, Manjunath. 2010. "Feature Subset Selection Problem using Wrapper Approach in Supervised Learning", International Journal of Computer Applications, Vol. 1, No. 7, pp. 0975-8887.
- [17] Aman Kumar Sharma and Suruchi Sahni. 2011. "A Comparative Study of Classification Algorithms for Spam Email Data Analysis", International Journal on Computer Science and Engineering, Vol. 3, No. 5, pp. 1891-1895.
- [18] S. Aruna, S.P. Rajagopalan, and L.V. Nandakishore. 2011. "An Empirical Comparison of Supervised Learning Algorithms on in Disease Detection", International Journal of Information Technology Convergence and Services, Vol. 1, No. 4, pp. 81-92.
- [19] R. GeethaRamani, Lakshmi Balasubramanian and Shomona Gracia Jacob. 2010. "Automatic prediction of diabetic retinopathy and glaucoma through image processing and data mining techniques", Proceedings of International Conference on Machine Vision and Image Processing, pp. 163-167.
- [20] K. Polat and S. Gunes. 2009. "A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems". Expert Systems with Applications, Vol. 36, pp. 1587-1592.
- [21] B. Aviad, G. Roy. 2011. "Classification by clustering decision tree-like classifier based on adjusted clusters", Expert Systems with Applications. Vol. 38, pp. 8220-8228.
- [22] M.J. Aitkenhead. 2008. "A co-evolving decision tree classification method", Expert Systems with Applications, Vol. 34, pp. 18-25.
- [23] R. Sharareh, Niakan Kalhori, Xiao-Jun Zeng. 2014. "Improvement the accuracy of six applied classification algorithms through Integrated Supervised and Unsupervised Learning Approach", Journal of Computer and Communications, Vol. 2, pp. 201-209.
- [24] D.M. Farid, Li Zhang, CM Rehman. 2014. "Hybrid Decision Tree and Naïve Bayes Classifiers for multiclass classification tasks", Expert Systems with Applications, Vol. 41, No. 4, pp. 1937-1946.
- [25] K.F. Man, K.S. Tang and S.Kwong. 1996. "Genetic algorithms: Concepts and Applications". IEEE



Transactions of Industrial Electronics, vol. 43, no. 5, pp. 519-534.

- [26] Geetha ramani, Lakshmi Balasubramanian, Alaghu Meenal. A,” Hybrid Decision Classifier Model Employing Naive Bayes and Root Guided Decision Tree for Improved Classification”, International Journal of Applied Engineering Research, Vol. 10, No. 17, pp. 13245-13249.
- [27] Eibe Frank, Mark Hall, Peter Reutemann, and Len Trigg, 'Weka' 3  
<http://www.cs.waikato.ac.nz/ml/weka/>, GNU General Public License.
- [28] R. Geetha Ramani, Dhanapackiam. C and Lakshmi Balasubramanian. 2013. “Automatic Detection of Glaucoma in Fundus Images through Image Features”, International Conference on Knowledge Modelling and Knowledge Management.
- [29] Geetha Ramani R., Lakshmi Balasubramanian and Shomona Gracia Jacob. 2012. “Data Mining Method of Evaluating Classifier Prediction Accuracy in Retinal Data”, in the Proceedings of IEEE International Conference on Computational Intelligence and Computing Research, pp. 426-429.