Indian Journal of Science and Technology, Vol 10(12), DOI: 10.17485/ijst/2017/v10i12/104975, March 2017
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

An Ensemble Classifier Adopting Random Subspace Method based on Fuzzy Partial Mining

P. Kayal[1]* and S. Kannan[2]

[1]Research and Development Centre, Bharathiar University, Coimbatore – 641046, Tamil Nadu, India; kayalpaddu@gmail.com
[2]Department of Computer Applications, M. K. University, Madurai - 625021, Tamil Nadu, India; skannanmku@gmail.com

Abstract

Objectives: Ensemble classification with fuzzy partial mining is a novel approach. The random subspace ensemble classifier contains several classifiers working on original attribute space. The aim of this paper is to examine the appropriateness of the random subspace ensemble method for fuzzy partial mining classification and thereby develop an algorithm Ensemble Classification on Fuzzy Utility Mining (ECFUM) by using a skill utility measure in addition to Support and Confidence. **Methods/Statistical Analysis**: The algorithm show high accuracy with ensemble classifier than solitary classifiers. The classifier is trained on random subspace method which is suitable when there is more number of attributes for the classification, where in many of the fuzzy rule based classification systems suffer increase in dimensionality. **Findings:** The unique integration of ensemble classification with fuzzy partial weighted mining generates Fuzzy Association Rules and Class Association Rules. Fuzzy association rules have been generated which holds the attributes association. Class association rules have been generated which holds the target class for the attribute association. The resultant classifier produced, shows credible results with better accuracy. ECFUM generates more number of hidden interesting rules compared to traditional associative classifiers. These hidden rules play a major role in later prediction of the algorithm. **Improvements:** Future work concentrates on the role of infinite sampling on class association rule with higher order confidence precedence to standardize the predictive power of the algorithm ECFUM.

Keywords: Ensemble Classification, Fuzzy Mining, Partial Weight, Random Subspace, Weighted Utility

1. Introduction

Associative classification[1,2] is a technique of fusing association rule mining and classification. Association rule mining does not have predetermined target class whereas classifications do. Such classification learns from instances whose classes are predestined. It is not sure that all the instances are learned well since it has varied data distribution strategy. Research says that the solution to such problem is Ensemble classification. Ensemble classification is based on many base classifiers. Not a single classifier is trusted rather it learns from set of classifiers and combines the prediction of those multiple classifiers to obtain maximum accuracy. It is a supervised learning algorithm which can be trained and used for predictions. Most common base classifiers are neural network, support vector machine, and k-NN classifiers[3]. The intension of using Ensemble classification is to reduce variance and bias by forming an ensemble of diverse classifiers. The ensemble methodology is well explained in Figure 1. The fuzziness of the data helps to classify associative relations between attributes of uncertainty. The role of fuzzy membership functions and weights over the attributes can be used for enhancing the prediction capability of the ensemble classifier model. These adjustment factors can be supportive while training the classifier of uncertain tuples for better prediction results.

There is a general credence about the higher accuracy of ensemble classifiers compared to solitary classifiers. Bagging was first introduced by Breiman[4]. Breiman's

*Author for correspondence

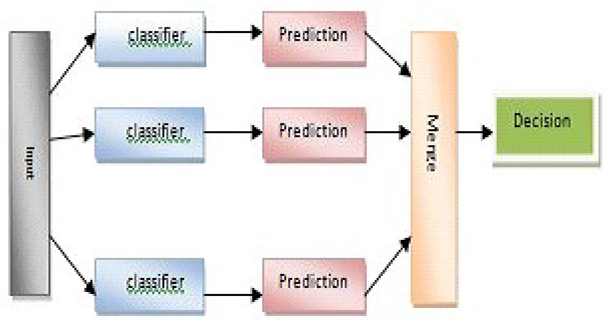


Figure 1. Ensemble model.

implementation of bagging on classification trees used medium as well as huge sized data set and it was applied to linear regression, Regression trees and Nearest Neighbour classifiers[5]. Bagging has given enhanced performance on unstable classifiers. A classifier is announced stable when relatively small changes in the training set do not react in the classifiers[6]. Boosting focuses on examples that were misclassified by earlier classifiers. They increase the weight of incorrectly classified examples which ensures they will become more important in the next iterations. Misclassification errors for these examples count more heavily than the right one. Adaboost is the general algorithm of boosting which is well explained[7]. Adaptive Neuro-Fuzzy Inference Systems (ANFIS) discuss how to build ensembles by using different type's offuzzy membership functions[8]. General Fuzzy Min-Max (GFMM) neural network is described[9]. It is used for classification, clustering or both. Their results can be crisp or fuzzy. This classification accuracy was better compared to many conventional classifiers.

2. Ensemble Classification

2.1 Random Subspace Classifier

Random Subspace [RS] method proposed by[10] is a kind of ensemble classifiers which consist of various classifiers in a subspace of data feature space. Classification results are based on these individual classifiers output by majority voting. This method can be used with many classifiers like linear classifier[11], nearest-neighbour classification[12], support vector machine[13] and many others too. The benefit of RS is, since the algorithm subspace the original data size, the training objects looks smaller for the original data, where as it is larger for the subspace data. The original data size is reduced, but training object size remains the same which gives more training sample size, helps for better classification[11]. This algorithm is a suitable choice where in there are high number of features. The algorithm is attracted by researchers for many reasons like simplified model, easy interpretation, training times are shorter compared to other models, enhanced generalization with reduced over fitting.

2.2 RS Algorithm

"Random Subspace samples data from the original feature set and builds one base classifier on each subset. The ensemble assigns a class label by either majority voting or averaging of output probabilities. Let $f=\{x_1,..x_n\}$ be the set of n features. To construct an RS ensemble with L classifiers, collect L samples, each of size M, drawn without replacement from a uniform distribution over X. Each feature subset defines a subspace of X of cardinality M, and a classifier is trained"[14] by base classifiers like Support Vector machine, K-nearest neighbour, Discriminant analysis. The concluding ensemble decision is made by majority vote. This shows how Random Subspace ensembles propose a response to the difficulty of huge dimensionality. Usually the classifiers can be trained effortlessly in smaller subspaces, and the feature-to-instance ratio enhances significantly. The accuracy of classification is not disturbed by replacing a single classifier with an ensemble. The RS ensemble requires two parameters, the size of the ensemble and the feature subsetcardinality[14].

3. Problem Definition

3.1 Proposed Classification

Classification based on fuzzy utility mining is atypical work in field. Here we introduce a new ensemble algorithm on classification called Ensemble Classification on Fuzzy Utility Mining (ECFUM) which is based on our previous work Fuzzy Partial Weighted Utility Mining (FPWUM) algorithm[14]. The FPWUM algorithm explains well about the use of special measure called SUF (skill utility factor) and generating hidden interesting Fuzzy association rules. The ensemble classification is developed here with the SUF measure, extracting all the integrity of the same. The idea of implementing SUF in Ensemble classification works extremely well and the results reveals the same. The classifier algorithm is based on Random subspace method used when there is high dimension of predictors, and it is persuasive too.

3.1.1 Problem Statement

Definition 1: A Database D contains several tupleswith values in the form of whole numbers.

Definition 2: A tuple in D contains attributes A_i and values a_{ij}, and a class denoted by C_j .

Definition 3: An ItemSet(IS) can be described as a set of disjoint attribute values contained in a training case, denoted < $(A_{i1}, a_{i1}), \ldots, (A_{ij}, a_{ij})$>.

Definition 4: A Frequent Item Set (FIS) is the ItemSet (IS) which has the support value >min_support.

Definition 5: A rule item r_i is of the form <FrequentItemSet, c>, where $c \in C$ is a class.

Definition 6: A rule item r_i passes min_support if (FIS(r)/ D)≥min_support.

Definition 7: A rule item r_i passes the min_confidence threshold if (r_iFreq(r)/ FIS (r)) ≥ min_confidence.

Definition 8: Any rule item r_i that passes the min_support threshold is said to be a frequent rule item(FreqRules).

Definition 9: Any rule item r_i that passes the SUF threshold is said to be Special ruleitem(SplRules).

Definition 10: Sensitivity of a rule is defined as the number of true positives divided by the addition of number of true positives and number of false negatives represented as

$$P(T^+|D^+) = TP / (TP + FN) \tag{1}$$

Definition 11: Specificity of a rule is defined as the number of true negatives divided by the sum of number of true negatives and number of false positives represented as

$$P(T^-|D^-) = TN / (TN + FP) \tag{2}$$

3.2 The Proposed ECFUM Algorithm

The ECFUM algorithm uses the Subspace type of ensemble for classification. It is a supervised learning algorithm where the database contains the attributes and the class label. In our sphere, there are 12 base attributes exclusive of a class attribute. Each base attribute is sub-divided into four sub-attributes holding values on a grade of four membership functions shown in Figure 2. And there are four possible class labels for the class attribute shown in Table 1 and well explained[15].

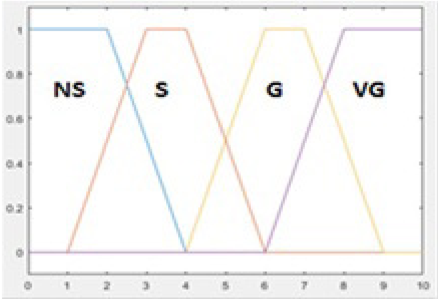


Figure 2. Fuzzy membership.

Table 1. Class label

NS	Not satisfactory
S	Satisfactory
G	Good
VG	Very good

3.2.1 Fuzzy Associative Classification Rules

The algorithm 1 explains about the fuzzy association rule generation. Algorithm 2 explains about the calculations of support and Skill Utility Factor (SUF).Initially scan the database for the attributes fuzzy membership degree and the weight of the corresponding attributes. Find the support, confidence and the special measure SUF which has proven to uncover hidden interesting itemset .

Algorithm 1: FAR Generation
Input: Dataset D, Weight w, Fuzzy support, confidence and suf threshold as min_supp,min_conf, min_suf respectively.
Output: Fuzzy association rule set carrying attributes associations.

Step 1: Scan database D, find 1-item frequent set L_1.
Step 2: Using L_{k-1} generate candidate item set c_k.
Step3: For each item set c in c_k calculate fuzzy support, if it is greater than min_supp, add to L_k, else step 4.
Step 4: If it is lesser than min_supp, find min_suf of the item set C. If it satisfies min_suf threshold add to L_k else prune the item set.
Step 5: For each item set in L_k, calculate Fuzzy confidence, if is greater than min_conf, then add it to Fuzzy association rules FARs.
Step 6: Continue step 2 until L_k is greater than 1, else go to step 7.
Step 7: Prune un-interesting rules.

Algorithm 2: Calculate min_support and min_suf
Input: Itemset I,f_{ij}-Fuzzymembership of j^{th} value of ith attribute,w_{ij}-weight of j^{th} value of i^{th} attribute.
output: Itemset support and Fuzzy partial weighted SUF.

Step 1: For each transaction t in D,Find the fuzzy value associated with item(i) in I,calculate the fuzzy value of I using min operator for each transactions.
Step 2: Sum up all the fuzzy transaction values of each calculated in step 1 and divide by the size of the dataset which is respresented as supp(I).
Step 3: After step 1 ads 2, if the supp(x) is less than min_supp for the respective transaction ,find the weight(W_i)of each item in itemset(I_i) and multiply withcorresponding fuzzy value as I_i X W_i(IW).
Step 4: Sum up all(IW) and divide by the number of occurences of the itemset(I_i) as given in equation 3.

$$suf = \sum Ii \times Wi \div N \quad (3)$$

3.2.2 Class Association Rule Generation

The Class Association Rules (CARs) are generated warily after many necessary pruning steps explained in 3.2.4. While generating the CARs from fuzzy association rules we have overcome many attribute selection problems. For instance, attribute 1,2,3,4 should not occur together in a Fuzzy association rules. 1 out of 4 continuously for the 48 attributes should occur in the association rules as seen in Table 2. The FARs carries the association (Item set) of the attributes. The CARs carry the class for the associations. Assigning class for the association is done on class majority voting technique explained in 3.2.3. Once the CARs are generated they are fed as input for the subspace ensemble classifier which is used for classification. The CARs are converted into a classifier readable format. The ECFUM classifier is a class association rule based classifier model built with Freq. Rules and Spl. Rules only.

The algorithm works on reasonable passes. For any pass the algorithm performs 5 operations. First the n-frequent items are found. Secondly with the n-frequent item set, frequent rule items are found (antecedent).Third, the frequent rule items are pruned. Fourth the corresponding COV table of frequent rule items are calculated. Fifth, the rule items are assigned with class variable (consequent). The other way, the antecedent and consequent of a rule is determined by the above steps.

Table 2. FARs/CARs

FARs	CARs
1, 5 → 9, 14	1, 5, 9, 14 → classlabel
2, 11, 15 → 19, 44	2, 11, 15, 19, 44 → classlabel
18, 23, 28, 31 → 34, 39	18, 23, 28, 31, 34, 39 → classslabel

3.2.3 Voting Technique

The database is scanned for known class attribute for the frequent item set/frequent associations. It will find out all possible classes for the frequent item set and takes the count of Class Occurrence Variable (COV). The COV has the count of classes the association belongs to, as shown in Table 3. The class with higher COV is assigned as class for the association. When a frequent item set is having negligible difference in the count between classes in the COV, the association is pruned since it may create chaos in the prediction.

3.2.4 Ranking and Pruning

Rule ranking in associative classifiers are based on confidence, support and anti-size [16]. R1 and R2 are rules with support and confidence as supp (R1), conf (R1), and supp (R2), conf (R2) respectively. R1 is said to be higher ranked, if

- The confidence of R1 is greater than that of R2
- The confidence values of R1 and R2 are same but support of R1 greater than R2.
- Confidence and support are same but anti-size of R1 is less than R2.

Pruning is done of necessity since the accessibility of huge rule database may enhance the accuracy of the classifier without upsetting its performance. Usually such large database gives enhanced predictive power. Since all attribute combinations are considered for rule's condition, the possibilities of redundant rules are common.1. Rule redundancy pruning discards specific rules with

Table 3. Sample COV table

Rule items	COV for four class label
1, 5, 7	4, 2, 0, 0
3, 6, 10	0, 3, 1, 0
6, 11	0, 0, 16, 6
1, 5, 9, 13, 18	0, 2, 11, 19

lesser confidence values than general rules. Such pruning minimizes the occurrence of redundant rules in classifier. 2. Basic pruning on rules which do not satisfy min_confidence is pruned. Rules that do not satisfy min_suf are pruned.3. Pruning after voting is performed for the algorithm which prunes cases where the COV of an item set is not clear. That is when the COV of the item set falls equal in more than one class label.
Anun-interesting rule R1 is said to be pruned if

- The suf of R1 is lesser than min_suf.
- The confidence of R1 is lesser than min_conf.
- R1 &R2 are redundantwith same confidence and suf.

Algorithm 3: Building the ECFUM Classifier
Input: Rule item set*Ri* and dataset *D*
Output:ECFUM Classifier.

Step 1: For a transaction t in dataset D, organize the rule-item set$r_i$using random subspace method to classify t.
Step 2: Traverse the data set D and follow Step 1 for all transactions, and take count of COV.
Step 3: Prune un-interesting rules from the ruleset and-calculate classificationerror- rate.
Step 4: Iterate Step 1 to 3 until the error rate is minimized.

---

4. Discussion

4.1 Data Source

We have used real-time data than synthetic data. We have collected more than 1000 samples from various HR personnel, HR Team and their feedback using questionnaire. The dataset contains12 base attributes and each base attribute have four sub-attributes. We have used 60% of the data for training the classifier and 40% of the data used for validation and testing. The performance of the ECFUM classifier is compared against customary classifiers and various ensemble techniques. The experimental figures and tables given are based on the part of database transactions and are not given completely.

Table 4. ECFUM frequent item set and CAR generation

Confidence (%)	No. of Freq Item set	Time taken to build classifier (s)
50	7102	20.086
60	8046	24.139
70	9871	57.989
80	13400	96.597
90	16240	151

4.2 Experimental Observations

The number of Frequent Item Set and the time taken to build the classifiers are presented in Table 4. The results

Table 5. Accuracy of ECFUM in % on various support and confidence thresholds

Algorithm/Threshold	0.1, 0.6	0.2, 0.5	0.2, 0.7
ECFUM	96.4	89	83.2
Boosting	50.4	54.4	58
Bagging	93.2	82	81
KNN	92.8	85.2	80
RusBoosted trees	50.4	45.6	56.4
Svm	81.6	87	78.8

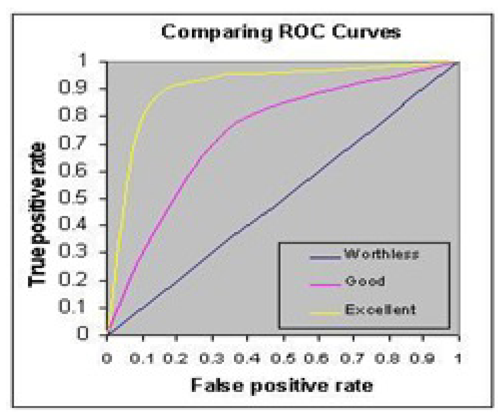
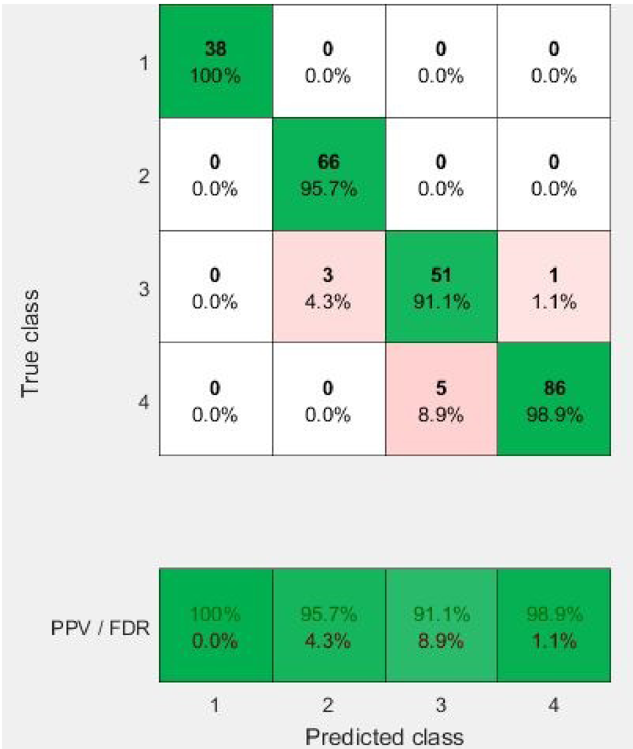


Figure 3. Sample ROC curve.

(a)



(b)

Figure 4. Confusion matrix at particular time **(a)** Per true class **(b)** Per predicted class.

Figure 5. Overall confusion matrix.

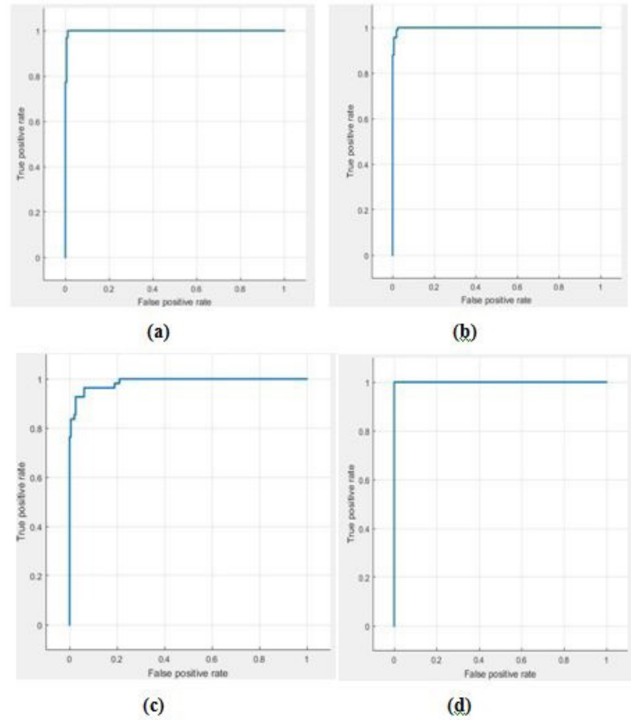


Figure 6. ROC curve for the four class labels.

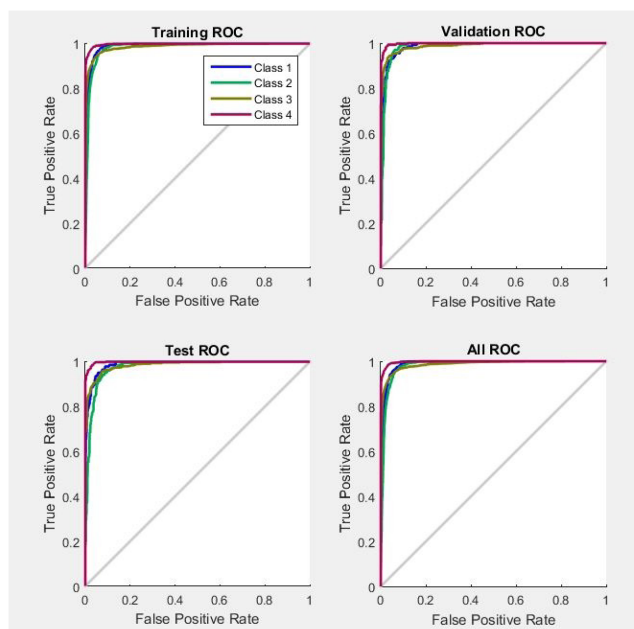


Figure 7. ROC of EPFUM at different phases.

of many classifiers are compared and presented in Table 5 which explains result values taken for three different thresholds. The Receiver Operating Characteristic (ROC) curve and the confusion matrix are presented to analyse the accuracy and performance of ECFUM

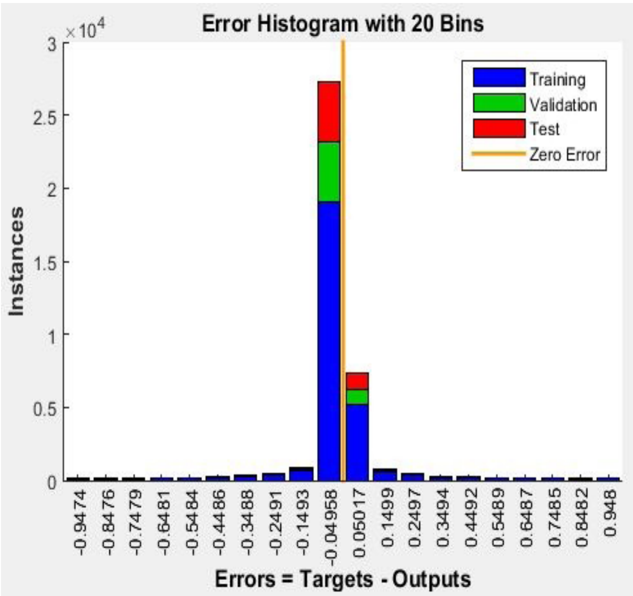


Figure 8. Error histogram of the ECFUM.

classification algorithm. A perfect test will have a ROC plot that passes in the course of the upper left corner. The closer the curve follows the upper-left border of the ROC space, the more accurate the test. The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test[17] as seen in Figure 3. The confusion matrix is a visualized table that explains the performance of the classification model on test data for which the true values are known is shown in Figure 4 and 5. The ROC curve shows trade-off between sensitivity and specificity which are directly proportional shown in Figure 6 and 7. Sensitivity is the true positive rate and specificity is the false positive rate which is (1-specificity). The error histogram of the algorithm ECFUM is presented in Figure 8.

5. Conclusion

Initially the algorithm takes the entire positive item set considering the universal proven support measure and then it works on the remaining item set for algorithmic utility weighting. We have identified three main challenge of the algorithm 1. Appropriateness of the random subspace ensemble method for fuzzy partial mining classification 2. The use of skill utility measure helps for the algorithmic growth and there by generating hidden interesting fuzzy associations and CARs 3. The prediction capability of the algorithmic system is found to be high compared to the traditional classifiers methods. Since the ECFUM data mining model is based on human leadership skills prediction, the fuzzy partial weight approach with subspace learners suits well the bucket.

6. References

1. Fadi T, Suhel H, Hussein AJ. Parallel associative classification data mining frameworks based map reduce. Parallel Processing Letters. 2015 Jun; 25(2).
2. Abdelhamid N. MAC: A multiclass associative classification algorithm. Journal of Information and Knowledge Management. 2012; 11(2). Crossref
3. Tasnim ARS. Ensemble classifiers and their applications: A review. International Journal of Computer Trends and Technology. 2014; 10(1):31-5. Crossref
4. Breiman L. Bagging predictors. Machine Learning. 1996; 24(2):123–40. Crossref
5. Farhad SG, Seyyed RK, Isa M. A new approach in bloggers classification with hybrid of K-nearest neighbor and artificial neural network algorithms. Indian Journal of Science and Technology. 2015; 8(3):237–46. Crossref
6. Juana C, Larry S, Lawrence O. Ensembles of Fuzzy Classifiers. IEEE International Conference on Fuzzy Systems; 2007. p. 1-5.
7. Yoav F, Robert E, Schapire. A short introduction to boosting. Journal of Japanese Society for Artificial Intelligence. 1999; 14(5):771-80.
8. Lima CAM, Coelho ALV, Zuben VFJ. Fuzzy systems design via ensembles of anfis. Proceedings of the IEEE International Conference on Fuzzy Systems; 2002. p. 506–11. Crossref
9. Gabrys B, Bargiela A. General Fuzzy min-max neural networkfor clustering and classification. IEEE Transactions on Neural Networks. 2000; 11(3):769–83. Crossref
10. Ho, Tin K. The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1998; 20(8):832-44. Crossref
11. Skurichina Marina. Bagging, boosting and the random subspace method for linear classifiers. Pattern Analysis and Applications. 2002; 5(2):121–35. Crossref
12. Tremblay G. Optimizing nearest neighbour in random subspaces using a multi-objective genetic algorithm. 17th International Conference on Pattern Recognition; 2004. p. 208–11. Crossref
13. Dacheng T, Xiaoou T, Xuelong L, Xindong W. Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006; 28(7):1088-99 Crossref

14. Ludmila I, Juan J, Catrin O, David E, Stephen J. Random subspace ensembles for fMRI classification. IEEE Transactions On Medical Imaging. 2010; 29(2):531-42. Crossref
15. Kayal P, Kannan S. A partial weighted utility measure for fuzzy association rule mining. Indian Journal of Science and Technology. 2016; 9(10):1-6. Crossref
16. Kannan S, Bhaskaran R. Role of interestingness measures in CAR rule ordering for associative classifier: An empirical approach. Journal of Computing. 2010; 2(1).
17. Zweig H, Gregory C. Receiver-Operating Characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. Clinical Chemistry. 1993; 39(4):561-77.