Published in final edited form as:

Appl Artif Intell. 2014 January 1; 28(3): 243–257. doi:10.1080/08839514.2014.883903.

Robust Feature Selection Technique using Rank Aggregation

Chandrima Sarkar^{1,*}, Sarah Cooley², and Jaideep Srivastava¹

¹College of Science and Engineering University of Minnesota at Twin Cities

²Masonic Cancer Center, University of Minnesota at Twin Cities

Abstract

Although feature selection is a well-developed research area, there is an ongoing need to develop methods to make classifiers more efficient. One important challenge is the lack of a universal feature selection technique which produces similar outcomes with all types of classifiers. This is because all feature selection techniques have individual statistical biases while classifiers exploit different statistical properties of data for evaluation. In numerous situations this can put researchers into dilemma as to which feature selection method and a classifiers to choose from a vast range of choices. In this paper, we propose a technique that aggregates the consensus properties of various feature selection methods to develop a more optimal solution. The ensemble nature of our technique makes it more robust across various classifiers. In other words, it is stable towards achieving similar and ideally higher classification accuracy across a wide variety of classifiers. We quantify this concept of robustness with a measure known as the Robustness Index (RI). We perform an extensive empirical evaluation of our technique on eight data sets with different dimensions including Arrhythmia, Lung Cancer, Madelon, mfeat-fourier, internet-ads, Leukemia-3c and Embryonal Tumor and a real world data set namely Acute Myeloid Leukemia (AML). We demonstrate not only that our algorithm is more robust, but also that compared to other techniques our algorithm improves the classification accuracy by approximately 3–4% (in data set with less than 500 features) and by more than 5% (in data set with more than 500 features), across a wide range of classifiers.

1 Introduction

We live in an age of exploding information where accumulating and storing data is easy and inexpensive. In 1991 it was pointed out that the amount of stored information doubles every twenty months (13). Unfortunately, the ability to understand and utilize this information does not keep pace with its growth. Machine learning provide tools by which large quantities of data can be automatically analyzed. Feature selection is one of the fundamental steps of machine learning. Feature selection identifies the most salient features for learning and focuses a learning algorithm on those properties of the data that are most useful for analysis and future prediction. It has immense potential to enhance knowledge discovery by extracting useful information from high dimensional data as shown in previous studies in various important areas (15, 18, 20, 22). In this paper, we propose to develop an improved rank aggregation based feature selection method which will produce a feature set that is

*To whom correspondence should be addressed; sark0071@umn.edu.

robust across a wide range of classifiers than the traditional feature selection techniques. In our previous work (16) we developed the idea of rank aggregation based feature selection approach and we showed that feature selection for supervised classification tasks can be accomplished on the basis of ensemble of various statistical properties of data. In this paper, we extend our idea by developing the rank aggregation based feature selection algorithm with exclusive rank aggregation approaches such as Kemeny (10) and Borda (5). We evaluate our algorithm using five different classifiers over eight data set with varying dimensions.

Feature selection techniques can be classified into filter and wrapper approaches (7), (23). In this paper, we focus on Filter Feature Selection because it is faster and more scalable (15). Feature selection techniques using distinct statistical properties of data have some drawbacks; for example information gain is biased towards choosing attributes with a large number of values and Chi square is sensitive to sample size. This indicates a statistical bias towards achieving the most optimal solution for classification problem. In other words, there will be a variation in classification performance due to the partial ordering imposed by the evaluation measures (for example information gain and chi square statistics) over the space of hypotheses (19). It has been shown that ensemble approaches reduces the risk of choosing a wrong hypothesis from many existing hypothesis in the solution space (11) Ensemble technique has also been used in various applications showing notable improvement in the results such as (3,21) and (1). To the best of our knowledge, no other study has focused on an extensive performance evaluation of rank aggregation based feature selection technique using exclusive rank aggregation strategies such as Kemeny (10).

To summarize, this paper makes the following contributions:

1. Development of a novel rank aggregation based feature selection technique using exclusive rank aggregation strategies namely Borda (5) and Kemeny (10).
2. Extensive performance evaluation of the rank aggregation based feature selection method using five different classification algorithms over eight data set of varying dimensions. We perform pairwise statistical tests with 5% significance level to prove the statistical significance of the classification accuracy results.
3. We introduced the concept of robustness of a feature selection algorithm which can be defined as the property which characterizes the stability of a ranked feature set towards achieving similar classification accuracy across a wide range of classifiers. We quantify this concept with an evaluation measure namely Robustness Index (RI).
4. We propose to use the concept of inter-rater agreement for improving the quality of rank aggregation approach for feature selection.

The remainder of the paper is organized as follows. Section 2 is the methodology. Section 3 describes the experimental results and discussion. In section 4, we present our conclusion.

*In this paper we use the terms variables, features and attributes with the same meaning.

2 Methodology

The process of rank aggregation based feature selection technique consists of the following steps: A non-ranked feature set is evaluated with n feature selection/evaluation techniques. This gives rise to n sets of ranked feature sets which differ in their rank ordering. The following step consists of executing rank aggregation on the feature sets using either Borda (5) or Kemeny Young (10) strategy to generate a final ranked feature set. The entire process of rank aggregation is documented inside the dotted box in figure 1.

2.1 Rank Aggregation

Rank aggregation is the process that combines ranking results of a fixed set of candidates from multiple ranking functions to generate a single better ranking. Rank aggregation can be done in different ways namely Borda (5) and Kemeny (6). Rank aggregation step has been described in Algorithm 1

2.1.1 Borda Method—In this paper we use rank aggregation based on Borda (5) ranking. For this we use a position based scoring mechanism to calculate the score of a feature. A pre-determined score is dedicated to each position in a list generated from each feature selection technique (this score is same for all the lists). For a distinct feature, the final score is the sum of all the positional scores from all the lists as given in equation 1. The final rank of a feature is determined from the final score.

$$score_{final} = \sum_{i=1}^n score_{pos(i,j)} \quad (1)$$

Where n is the total number of features selection techniques (or ranker) used. $pos(i, j)$ is the j^{th} position of a feature ranked by the ranker i . $score_{p(i,j)}$ is the score of a feature in list i generated by ranker i at j^{th} position. $score_{final}$ is the sum of all the positional score from all the lists. In this paper, we consider a single feature selection technique as a ranker and the candidates as the features.

2.1.2 Kemeny Method—The Kemeny rule is sometimes interpreted as a maximum likelihood estimator of the “correct” ranking (4) and for every pair of candidates, any voter ranks the better candidate higher, with probability $p > 1/2$, independently. The Kemeny rule is given as follows - Let x_1 and x_2 be two candidates, r be a ranking and v be the vote, let $\delta_{x_1, x_2}(r, v) = 1$ if there exists an agreement between r and v on the relative ranking of x_1 and x_2 that is either both rank of x_1 is higher, or both rank of x_2 is higher, or else 0 if they disagree. Let T' be the total number of pairwise agreements i.e. the agreement of a ranking r with a vote v which is given by $\sum_{x_1, x_2} \delta_{x_1, x_2}(r, v)$. Then a Kemeny ranking r maximizes the sum of the agreements with the votes given by $\sum_v \sum_{x_1, x_2} \delta_{x_1, x_2}(r, v)$ (4). Since, computing optimal Kemeny aggregation is NP-Hard for $r \geq 4$ (6), in this paper we use the 2-approximation of Kemeny optimal aggregation (17) which has been shown to run in time $O(R * c \log c)$, where R denotes total number of rankers and c denotes the number of candidates.

2.1.3 Analysis of rankers with Inter Rater Agreement—Inter Rater Agreement (IRA) can be used as a pre-processing step prior to the rank aggregation step. The main motivation for this step is the analysis of the homogeneity or consensus among the rank ordering generated by each ranker. Each ranker uses different measures for evaluating the candidates and hence generates a different rank ordering. There can be possibility that the rank ordering generated by one of the ranker is highly inconsistent with the other rankers. This might cause the final aggregated rank ordering to be far away from the ground truth (optimal) ordering. In this paper, we make the assumption that rank ordering which is in consensus with the majority of the rankers are closest to the ground truth. Hence, for improving the rank ordering generated by the ranker, we propose to use the concept of Inter-Rater Agreement (IRA) which analyses the degree of agreement among rankers. We use Intraclass Correlation (ICC) (2) approach for calculating IRA. The ICC assesses rating reliability by comparing the variability of different ratings of the same subject to the total variation across all ratings and all subjects, as formulated in Equation 2.

$$ICC = \frac{V(b)^2}{(V(w)^2 + V(b)^2)} \quad (2)$$

where $V(w)^2$ is the pooled variance within subjects, and $V(b)^2$ is the variance of the trait between subjects. The IRA lies between 0 to 1 where 0 is the least reliable rating and 1 is the most reliable rating for a group of rankers. We use a heuristically determined threshold T which is used for eliminating rankers who tends to disrupts the homogeneity in ranking from the group of rankers.

2.1.4 K-step feature subset Selection—The K -step feature subset selection is a post processing step to the rank aggregation step with a focus on generating a feature subset from the final rank aggregated feature set. In this process, firstly, for each classification algorithm, we determine the classification accuracy of each top i feature subset where $1 \leq i \leq k$ where k is the total number of features in the feature subset. Next, we select the feature subset with the maximum classification accuracy across all the classification algorithms used, as our final feature subset as given in Algorithm 2.

2.1.5 Evaluation Measure—We evaluate our algorithm based on three evaluation measures as discussed below -

1. **Classification accuracy** - accuracy is calculated as the percentage of correctly classified instances by a given classifier. At first we obtain a feature subset of size K using a K -step feature subset selection approach as described in the previous section. Classification accuracy of this feature subset was determined and recorded using five different classifiers for evaluation purpose.
2. **F-measure** - Weighted (by class size) average F-measure was obtained from the classification using feature subset with the same five classifiers as above.
3. **Robustness Index (RI)** - We define robustness as the property that characterizes the stability of a feature subset towards achieving similar classification accuracy across

a wide range of classifiers. In order to quantify this concept we introduce a measure called robustness index (RI) which can be utilized for evaluating the robustness of a feature selection algorithm across a variety of classifiers. Intuitively, RI measures the consistency with which a feature subset generates similar (ideally higher) classification accuracy (or lower classification error) across a variety of classification algorithms when compared with feature subsets generated using other methods. The step-by-step process of robustness index is described in algorithm 3

The motivation behind the concept of robustness is as follows: it is not an easy task to determine the best classifier to use for a classification task prior to actually using that model. A robust technique helps one to choose a classification model with the minimum risk in choosing an inappropriate model.

2.2 Experimental Setup

2.2.1 Data sets used—We used eight different types of data sets shown in Table 3. Acute Myeloid Leukemia or AML is a real world data set that contains 69 demographic, genetic, and clinical variables from 927 patients who received myeloablative, T-cell replete, unrelated donor (URD) stem cell transplants (16). We also used data sets including Embryonal Tumours of the Central Nervous System (14), madelon and Internet-ads (12), Leukemia-3c, Arrhythmia, Lung Cancer and mfeat-fourier (9) from UCI KDD as listed in Table 3.

2.2.2 Statistical test of significance—We performed pairwise t-test with a 5% significance level in order to measure the statistical significance of our classification accuracy result. The null hypothesis is that the difference in between the classification accuracy result obtained from the two algorithms considered in the pairwise test, comes from a normal distribution with mean equal to zero and unknown variance. We reject the null hypothesis if p – value is less than 5% significance level. The result is given in Table 4.

3 Experimental Results and Discussion

We evaluated the performance of our feature selection algorithm in terms of classification accuracy, F-measure and robustness by comparing with three feature selection techniques namely information gain attribute evaluation, symmetric uncertainty attribute evaluation and chi square attribute evaluation (8). We selected feature selection techniques with the help of IRA method assuming an IRA threshold of 0.75 (heuristically determined). We refer our rank aggregation based feature selection algorithm as *Kemeny* and *Borda* (using Kemeny and Borda method respectively) in the figures shown in this paper.

The results of classification accuracy over eight data sets are given in Figure 2. Using Algorithm 1 and 2 we generate feature subsets for each data set indicated in a bracket in the tables 3 and 5. Next, we perform classification using five different classifiers (shown in Table 1) whose results are given in figure 2a, 2b, 2c, 2d, 2e, 2f, 2g, 2h. Figures 2a, 2c 2h and 2e shows classification accuracy for data sets with over 500 variables. In these four data sets, the accuracy with Kemeny and Borda is more than 5% higher as compared to those with the three single feature selection methods. In the four other data sets shown in figure

2b, 2d, 2f and 2g, the classification accuracy is higher by approximately 3–4 % across all the classifiers.

Next, we performed pairwise statistical significance test with a 5% significance level to prove the statistical significance of our accuracy results. We calculated the p – values for every data set, comparing *Kemeny* and *Borda* with three feature selection techniques as depicted in Table 4.

Table 2 shows the comparison of robustness index as calculated using Algorithm 3. Lower the value of RI, more robust is the technique, i.e. Robustness index equals 1 is more robust than an RI equals 3. We can see from Table 2 that both *Kemeny* and *Borda* has robustness index of either 1 or 2 with every data set. This shows that *Kemeny* and *Borda* are more robust than the other traditional feature selection techniques. The motivation behind this analysis is that, when one is unable to decide on the best classification algorithm to use on a given data set, our feature selection algorithm will help with a technique that will ensure a lower classification error over a variety of classification algorithms. The number in parentheses beside the data set names in Table 2 indicates the size of the feature subset.

In Table 5 we compare the weighted (by class size) average F-measures (defined as the harmonic means of precision and recall), generated using *Kemeny* and *Borda* with three feature selection methods using five different classifiers as given in Table 1. The number in parentheses beside the data set names in every figure indicates the size of the feature subset used for classification. We can see that F-measure with *Kemeny* and *Borda* is higher in almost all the cases. This shows that apart from accuracy, the sensitivity and specificity generated with different classifiers using our rank aggregation based feature selection method can be improved.

The results of these analysis show that our rank aggregation based feature selection algorithm is an efficient technique suited for various kinds of data sets including the ones with features greater than 1000. Our method gives a higher classification accuracy, f-measure and greater robustness than the other traditional methods over a wide range of classifiers. This method is advantageous especially in cases where it is difficult to determine the best statistical property for evaluation of a given data set. The greatest advantage in having a robust technique is that, there will be fewer dilemmas in deciding on the most appropriate classifier to use from the vast range of choices.

4 Conclusion

In this paper, we propose a novel rank aggregation based feature selection technique beneficial for classification tasks. The results of our algorithm suggest that our proposed ensemble technique yields higher classification accuracy, higher f-measure and greater robustness than single feature selection techniques on data sets with different range of dimensions. The eight different data set that we used has dimensions from as low as 57 (lung cancer data set) to as high as 7130 (Leukemia-3c data set). We found that our algorithm improved accuracy, F-measure and robustness of classification in all the data sets. We proved the statistical significance of our classification accuracy results using a pairwise t-test with a 5% significance level. This shows that our feature selection technique is suited

for high dimensional data applications, especially in situations where it is difficult to determine the best statistical property to use for evaluation of a feature set. We conclude by stating that our robust feature selection technique is an appropriate approach to be utilized in situation where one faces the dilemma of choosing the most suitable classifiers and the best statistical property to use for an improved result on a given data set. Our experiments and the results provide initial evidence for the success of our feature selection framework. We believe that this framework has the potential to bring about improvement in the accuracy and robustness of various classification tasks in many different applications.

Acknowledgments

AML data resource in this work was supported by National Institutes of Health/NCI grant P01 111412, PI Jeffrey S. Miller, M.D, utilizing the Masonic Cancer Center, University of Minnesota Oncology Medical Informatics shared resources. We would like to thank Atanu Roy for his critical reviews and technical feedback during developing this research.

References and Notes

1. Abeel T, Helleputte T, Van de Peer Y, Dupont P, Saeys Y. Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. *Bioinformatics*. 2010; 26(3):392–398. [PubMed: 19942583]
2. Bartko J. On various intraclass correlation reliability coefficients. *Psychological bulletin*. 1976; 83(5):762.
3. Budnik M, Krawczyk B. On optimal settings of classification tree ensembles for medical decision support. *Health informatics journal*. 2013; 19(1):3–15. [PubMed: 23486822]
4. Conitzer, V. PhD thesis. IBM; 2006. Computational aspects of preference aggregation.
5. de Borda, JC. Memoire sur les Elections au Scrutin. *Histoire de l'Academie Royale des Sciences*; Paris: 1781.
6. Dwork, C.; Kumar, R.; Naor, M.; Sivakumar, D. Rank aggregation methods for the web; *Proceedings of the 10th international conference on World Wide Web*; ACM; 2001. p. 613-622.
7. Guyon I, Elisseeff A. An introduction to variable and feature selection. *The Journal of Machine Learning Research*. 2003; 3:1157–1182.
8. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I. The weka data mining software: an update. *ACM SIGKDD Explorations Newsletter*. 2009; 11(1):10–18.
9. Hettich, S.; Bay, S. The uci kdd archive. 1999.
10. Kemeny J. Mathematics without numbers. *Daedalus*. 1959; 88(4):577–591.
11. Khossainov, R.; Hess, A.; Kushmerick, N. Ensembles of biased classifiers; *Proceedings of the 22nd international conference on Machine learning*; ACM; 2005. p. 425-432.
12. Pathical, SP. PhD thesis. University of Toledo; 2010. Classification in High Dimensional Feature Spaces through Random Subspace Ensembles.
13. Piateski, G.; Frawley, W. Knowledge discovery in databases. MIT press; 1991.
14. Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME, Kim JY, Goumnerova LC, Black PM, Lau C, et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature*. 2002; 415(6870):436–442. [PubMed: 11807556]
15. Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. *Bioinformatics*. 2007; 23(19):2507–2517. [PubMed: 17720704]
16. Sarkar, C.; Cooley, S.; Srivastava, J. Improved feature selection for hematopoietic cell transplantation outcome prediction using rank aggregation; *Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on*; IEEE; 2012. p. 221-226.
17. Subbian K, Melville P. Supervised rank aggregation for predicting influence in networks. 2011 arXiv preprint arXiv:1108.4801.

18. Termenon M, Graña M, Besga A, Echeveste J, Gonzalez-Pinto A. Lattice independent component analysis feature selection on diffusion weighted imaging for alzheimer's disease classification. *Neurocomputing*. 2012
19. Vilalta, R.; Oblinger, D. ICML. Citeseer: 2000. A quantification of distance bias between evaluation metrics in classification; p. 1087-1094.
20. Wang G, Song Q. Selecting feature subset for high dimensional data via the propositional foil rules. *Pattern Recognition*. 2012
21. WoŹniak M, Graña M, Corchado E. A survey of multiple classifier systems as hybrid systems. *Information Fusion*. 2013
22. Yang, Y.; Pedersen, J. MACHINE LEARNING-INTERNATIONAL WORKSHOP THEN CONFERENCE. MORGAN KAUF-MANN PUBLISHERS, INC; 1997. A comparative study on feature selection in text categorization; p. 412-420.
23. Yu L, Liu H. Feature selection for high-dimensional data: A fast correlation-based filter solution. *MACHINE LEARNING-INTERNATIONAL WORKSHOP THEN CONFERENCE*. 2003; 20:856.

Figure 1.

Flow diagram of the rank aggregation based feature selection algorithm. n is the number of feature evaluation technique (using different statistical properties of data).

Figure 2.

Comparison of Classification Accuracies in different datasets using *Kemeny*, *Borda* and 3 single feature selection technique.

Table 1

Classification Algorithms Used

Classifiers	Settings
Naive Bayes	estimator classes
J48	pruned C4.5 decision tree
KNN	k=3; brute force search algorithm; Euclidean distance function
AdaboostM1	base classifier: Decision Stump, 10 boost iterations, % of weight mass for training was 100
Bagging	weak classifier: fast decision tree learner, bag size as 100%, 10 bagging iterations

Table 2
Robustness Index IG- Information gain, SU - Symmetric Uncertainty, CS- ChiSquare

	Feature Selection Techniques				
	Kemeny	Borda	InfoGain	SymmetricU	ChiSquare
Data Sets	1	1	3	2	3
	2	1	3	3	4
	1	2	4	5	3
	2	1	4	3	3
	1	1	2	3	2
	1	1	2	2	2
	1	1	2	3	3
	2	1	3	4	5
	2	1	3	4	5

Table 3

Data Sets with attributes and instances

Data sets		
Data Set	Attributes	Instances
Lung Cancer	57	32
AML	69	927
Mfeat-fourier	77	2000
Arrhythmia	280	452
Madelon	501	2600
Internet-Ads	1559	3279
Leukemia-3c	7130	72
Embryonal Tumor	7130	60

Table 4
Results of paired ttest for different data sets. IG- Information gain, SU - Symmetric Uncertainty, CS-ChiSquare

Lung Cancer			AML			mfeat-fourier			Embryonal Tumor		
	Borda	Kemeny		Borda	Kemeny		Borda	Kemeny		Borda	Kemeny
IG	0.0327	0.0401	IG	0.0509	0.0404	IG	0.0212	0.0227	IG	0.0155	0.0197
SU	0.009	0.0074	SU	0.0571	0.0451	SU	0.0227	0.0243	SU	2.7E-04	6.0E-06
CS	0.0046	0.0066	CS	0.0509	0.0404	CS	0.0227	0.0243	CS	0.0225	0.0255

Madelon			Internet-Ads			Leukemia-3c			Arrhythmia		
	Borda	Kemeny		Borda	Kemeny		Borda	Kemeny		Borda	Kemeny
IG	0.0164	0.0164	IG	0.0119	0.0119	IG	0.0015	0.0015	IG	0.0157	0.0283
SU	0.0139	0.0139	SU	0.0119	0.0119	SU	5.3E-04	5.3E-04	SU	0.0049	0.013
CS	0.0164	0.0164	CS	0.0119	0.0119	CS	3.8E-04	3.8E-04	CS	0.0034	0.0084

Table 5

Results for F-measure

		Naive Bayes Classifier				
		Kemeny	Borda	IG	SU	CS
Data Sets	AML (25)	0.67	0.67	0.64	0.65	0.64
	mfeat-fourier (9)	0.79	0.79	0.77	0.77	0.77
	Arrhythmia (36)	0.67	0.67	0.53	0.56	0.54
	madelon (40)	0.60	0.60	0.53	0.53	0.53
	Internet-Ads (60)	0.95	0.95	0.94	0.94	0.94
	Leukemia-3c (60)	0.99	0.99	0.81	0.73	0.79
	Embryonal Tumor (60)	0.74	0.72	0.62	0.59	0.61
	Lung Cancer (8)	0.45	0.45	0.46	0.51	0.40

		J48 Decision Tree				
		Kemeny	Borda	IG	SU	CS
Data Sets	AML (25)	0.69	0.69	0.64	0.64	0.64
	mfeat-fourier (9)	0.76	0.76	0.73	0.73	0.73
	Arrhythmia (36)	0.65	0.65	0.46	0.50	0.48
	madelon (40)	0.75	0.75	0.51	0.53	0.51
	Internet-Ads (60)	0.97	0.97	0.95	0.95	0.95
	Leukemia-3c (60)	0.89	0.89	0.75	0.51	0.68
	Embryonal Tumor (60)	0.68	0.68	0.73	0.72	0.73
	Lung Cancer (8)	0.48	0.48	0.47	0.43	0.42

		k Nearest Neighbor				
		Kemeny	Borda	IG	SU	CS
Data Sets	AML (25)	0.63	0.63	0.60	0.62	0.60
	mfeat-fourier (9)	0.83	0.83	0.81	0.81	0.81
	Arrhythmia (36)	0.62	0.60	0.50	0.48	0.48
	madelon (40)	0.67	0.67	0.50	0.51	0.50
	Internet-Ads (60)	0.97	0.97	0.95	0.95	0.95

		k Nearest Neighbor			
		Kemeny	Borda	IG	CS
	Leukemia-3c (60)	0.96	0.96	0.63	0.61
	Embryonal Tumor (60)	0.77	0.79	0.64	0.64
	Lung Cancer (8)	0.38	0.38	0.41	0.52

		AdaBoost			
		Kemeny	Borda	IG	CS
Data Sets	AML (25)	0.69	0.69	0.63	0.63
	mfeat-fourier (9)	0.83	0.83	0.81	0.81
	Arrhythmia (36)	0.45	0.45	0.42	0.42
	maelon (40)	0.61	0.61	0.54	0.54
	Internet-Ads (60)	0.92	0.92	0.91	0.91
	Leukemia-3c (60)	0.90	0.90	0.71	0.59
	Embryonal Tumor (60)	0.73	0.73	0.57	0.56
	Lung Cancer (8)	0.47	0.47	0.47	0.47

		Bagging			
		Kemeny	Borda	IG	CS
Data Sets	AML (25)	0.69	0.69	0.63	0.63
	mfeat-fourier (9)	0.79	0.79	0.78	0.78
	Arrhythmia (36)	0.71	0.71	0.54	0.53
	maelon (40)	0.79	0.79	0.54	0.54
	Internet-Ads (60)	0.96	0.96	0.95	0.95
	Leukemia-3c (60)	0.93	0.93	0.73	0.70
	Embryonal Tumor (60)	0.76	0.76	0.62	0.57
	Lung Cancer (8)	0.44	0.44	0.35	0.33

Algorithm 1

Rank Aggregation

Input:

1) Feature Evaluation Technique (FET) set Q' and feature set S

Output:

Ranked Feature Set S'

Algorithm:

Steps: Rank Aggregation

2: **for** $j = 0$ to $|Q'| - 1$ **do**

$S'' = S$ $\triangleright S''$ is a temporary variable

4: Rank S'' using FET_j where $FET_j \in Q'$

add S'' to list L

6: **end for**

$S' =$ aggregated feature set from L using equation 1

Algorithm 2

K step feature subset selection

Input:

Feature Set S' , Dataset D , Classifiers set M

Output:

feature subset of size k

Algorithm:

set $S^* \subset S'$, $m_i \in M$ and $\sum_i |m_i| = |M|$ and $K_j \in S'$ with $K_1 < K_2 \dots K_{j-1} < K_j$ where K is the feature subset, $1 > j > k$

2: **for** $1 < i < |M|$ **do**

for $1 < j < K$ **do**

4: add feature set K_j to S^*

 learn S^* using M_i

6: Calculate accuracy of M_i and store it in list $temp_j$

end for

8: search $temp_j$ for $1 < j < K$ with the highest predictive power for M_i and store in K^*_i

end for

10: select the $MAX(K^*_i)$ where $1 < i < |M|$

Algorithm 3

Robustness Index calculation

Input:

- 1) Classification models M_i , where $1 \leq i \leq m$, m is the number of classifiers used,
- 2) Feature set f_k generated from p feature selection techniques, where k is number of top features

Output:

Robustness Index r_p , for each p feature selection technique

Algorithm:

```

for  $i = 0$  to  $m - 1$  do                                ▷ For each  $M_i$ 
2:   for  $j = 0$  to  $p - 1$  do
        $C_p$  = classification error with  $f_k$ 
4:   end for
       Rank each  $p$  based on  $C_p$  score and save
6:   end for
       for  $i = 0$  to  $p - 1$  do
8:     Aggregate the ranks across  $M_i$  for  $1 < i < m$  using equation 1
       Assign  $r_p$  = aggregated ranks
10:  end for

```