Feature Subset Selection using Cascaded GA & CFS: A Filter Approach in Supervised Learning

Asha Gowda Karegowda
Siddaganga institute of Technology. Tumkur, India

M.A.Jayaram
Siddaganga institute of Technology. Tumkur, India

A.S .Manjunath
Siddaganga institute of Technology, Tumkur, India

ABSTRACT
Medical data mining has enormous potential for exploring the hidden patterns in the data sets of the medical domain. These patterns can be utilized by the physicians to improve clinical diagnosis. Feature subset selection is one of data preprocessing step, which is of immense importance in the field of data mining. As a part of feature subset selection step of data preprocessing, a filter approach with genetic algorithm (GA) and Correlation based feature selection has been used in a cascaded fashion. GA rendered global search of attributes with fitness evaluation effected by CFS. Experimental results signify that the feature subset recognized by the proposed filter GA+CFS, when given as input to five classifiers, namely decision tree, Naïve Bayes, Bayesian, Radial basis function and k-nearest neighbor classifiers showed enhanced classification accuracy. Experiments have been carried out on four medical data sets publicly available at UCI.

Keywords
Feature selection, filters, Genetic Algorithm, Correlation based feature selection, Decision tree, Naïve Bayes, Bayesian Classifier, Radial Basis Function, K-Nearest Neighbor.

1. INTRODUCTION
Data Mining is the non-trivial extraction of implicit, previously unknown, and potentially useful information about data [1]. In medical and health care areas, due to the availability of computers, a large amount of data is becoming accumulated. Such a large amount of data cannot be processed by the medical experts in a short time, to make diagnosis, prognosis and treatment schedules. Extracting useful knowledge for the diagnosis and treatment of disease from the database increasingly becomes necessary. Medical data mining has enormous potential for exploring the hidden patterns in the data sets of the medical domain. Data preprocessing is a significant step in the knowledge discovery process, since quality decisions must be based on quality data. Data preprocessing includes data cleaning, data integration, data transformation and data reduction [1]. Quality of the data in the medical database enhances the quality of medical diagnosis. The goal of data reduction/ feature subset selection is to find a minimum set of attributes such that the resulting probability distribution of the data classes is as close as possible to the original distribution obtained using all attributes. Mining on the reduced set of attributes has following benefits.

- It reduces the number of attributes appearing in the discovered patterns, helping to make the patterns easier to understand.
- It enhances the classification accuracy.
- It reduces classifier-learning time.

This paper presents use of multivariate filters, which uses GA with CFS as fitness evaluator. The relevant features are provided as input to five classifiers. The results clearly show the enhanced classification by providing the features selected by proposed filter. Section 2 discusses wrapper and filter feature selection methods for both supervised and unsupervised learning algorithms. . Section 3 describes Genetic search algorithm (GA) and Correlation based feature selection (CFS) as subset evaluating mechanism for GA. Performance metrics and dataset used is described in section 4 followed by results and conclusions in section 5 and 6 respectively.

2. FEATURE SELECTION
Feature selection is a process that selects pertinent features as a subset of original features. Feature selection is one of the important and frequently used techniques in data preprocessing for data mining. In real-world situations, relevant features are often unknown a priori. Hence feature selection is a must to identify and remove are irrelevant/redundant features. It can be applied in both unsupervised and supervised learning.

2.1 Feature selection in unsupervised learning
The objective of feature selection for unsupervised learning is to find the smallest feature subset that best uncovers clusters form data according to the preferred criterion [2]. Feature selection in unsupervised learning is much harder problem, due to the absence of class labels. Feature selection for clustering is the task of selecting significant features for the underlying clusters [3]. Feature selection for unsupervised learning can be subdivided into filter methods and wrapper methods. Filter methods in unsupervised learning are defined as using some intrinsic property of the data to select feature without utilizing the clustering algorithm [2]. Entropy measure has been used as filter method for feature selection for clustering [4]. Wrapper approaches in unsupervised learning apply unsupervised learning algorithm to each candidate feature subset and then evaluate the feature subset by criterion functions that utilize the

clustering result [2]. A wrapper method has been proposed where Gaussian mixture model combines a clustering method with a Bayesian inference mechanism for automatically selecting pertinent features [5].

2.2 Feature selection in supervised learning

In supervised learning, feature selection aims to maximize classification accuracy [6]. It is easier to select features for classification/supervised learning than for clustering, since the classification uses class label information. Though domain experts can eliminate few of the irreverent attributes, selecting the best subset of features usually requires a systematic approach. Feature selection method generally consists of four steps described below [7].

(a) Generate candidate subset: The original feature set contains n number of features, the total number of competing candidate subsets to be generated is 2n , which is a huge number even for medium-sized n. Subset generation is a search procedure that produces candidate feature subsets for evaluation based on a certain search strategy. The search strategy is broadly classified as complete/exhaustive (eg. Breadth first search, Branch & bound, beam search, best first), heuristic (forward selection, backward selection, forward and backward selection), and random search (Las Vegas algorithm (LVW), genetic algorithm (GA), Random generation plus sequential selection (RGSS), simulated annealing (SA)).

(b) Subset evaluation function to evaluate the subset generated in the previous step (generate candidate subset) by using filter/ wrapper approach. Filter and Wrapper approach differ only in the way in which they evaluate a subset of features. The filter approach (information gain, gain ratio, CFS, Principal Component Analysis (PCA), Chi-square Feature Evaluation) is independent of the learning induction algorithm. Wrapper strategies for feature selection use an induction algorithm to estimate the merit of feature subsets. Wrappers often achieve better results than filters due to the fact that they are tuned to the specific interaction between an induction algorithm and its training data. Filters are described in section 2.2.1 and wrappers are described in section 2.2.2.

(c) Stopping Condition: Since the number of subsets can be enormous, some sort of stopping criterion is necessary. Stopping criteria may be based on a generation procedure/ evaluation function.

Stopping criteria based on generation procedure include:

- Whether a predefined number of features are selected
- Whether a predefined number of iterations reached.

Stopping criteria based on an evaluation function can be:

- Whether addition (or deletion) of any feature does not produce a better subset
- Whether an optimal subset according to some evaluation function is obtained.

(d) Validation procedure to check whether the feature subset selected is valid. Usually the result of original feature set is compared with the feature selected by filters/wrappers as input to some induction algorithm using artificial/real-world datasets. Another approach for validation is to use different feature selection algorithm to obtain relevant features and then compare the results by using classifiers on each relevant attribute subset. The above four steps are shown in the figure 1.

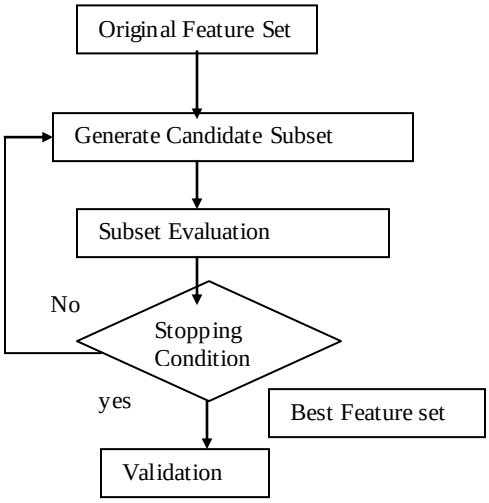

Figure 1. Steps for feature selection

2.2.1 The Filter Approach for Feature Selection

The filter approach actually precedes the actual classification process. The filter approach [figure 2], is independent of the learning induction algorithm, computationally simple fast and scalable. The filter method uses the intrinsic prosperities of data and the target class to be learned for feature selection. Using filter method, feature selection is done once and then can be provided as input to different classifiers. Various feature ranking and feature selection techniques have been proposed such as Correlation-based Feature Selection (CFS), Principal Component Analysis (PCA), Gain Ratio (GR) attribute evaluation, Chi-square Feature Evaluation, Fast Correlation-based Feature selection (FCBF), Information gain, Euclidean distance, i-test, Markov blanket filter.

Some of these filter methods do not perform feature selection but only provide ranking to features, hence they are combined with search method when one needs to find out the appropriate number of attributes. Such filters are often used with forward selection, backward elimination, bi-directional search, best-first search, genetic search and other methods [8-10].

FOCUS algorithm used forward selection strategy carries out exhaustive search until it finds a minimal combination of features .It is limited to binary, noise-free data [11]. A continuous extension of FOCUS is C-FOCUS is developed to deal with discrete and continuous features [12]. Kira and Rendell [13] described a statistical feature selection algorithm called RELIEF that uses instance based learning to assign a relevance weight to each Feature, which is to denote the relevance of the feature to the target. High order information gain has been used for feature selection [14]. The PRESET algorithm [3] is heuristic feature selector that uses the theory of Rough sets to heuristically rank the features in noise-free binary domain. The Selection Construction and Ranking using Attribute Pattern (SCRAP) [15] is an instance based filter approach; uses sequential search to identify the features that change at the decision boundaries and include them in the feature sub set. The decision tree has been used as filter

approach to provide the relevant features as input to neural network classifier [16]. The split attributes (non leaf nodes) in the decision tree are identified as the relevant attributes. Further Correlation based feature selection has been used in a cascaded fashion with GA as filter to provide relevant inputs to neural networks classifier [17].

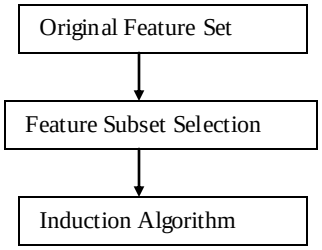

Figure 2. Filter approach for feature selection

2.2.2 The Wrapper Approach for Feature Selection

Wrapper model approach uses the method of classification itself to measure the importance of features set; hence the feature selected depends on the classifier model used. Wrapper methods generally result in better performance than filter methods because the feature selection process is optimized for the classification algorithm to be used.

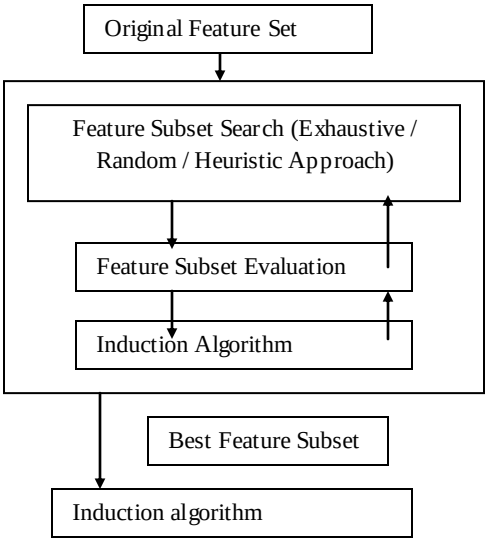

Figure 3. Wrapper approach for feature selection

However, wrapper methods are too expensive for large dimensional database in terms of computational complexity and time since each feature set considered must be evaluated with the classifier algorithm used [7][9][10]. The working of wrapper approach is shown in figure 3.

3. PROPOSED METHOD

3.1 Genetic Algorithms

GA is a stochastic general search method, capable of effectively exploring large search spaces, which is usually required in case of attribute selection. Further, unlike many search algorithms, which perform a local, greedy search, GAs performs a global search. The Gas simulates the processes in natural systems for evolutions based on the principle of "survival of the fittest" given by Charles Darwin [18].

A genetic algorithm mainly composed of three operators: reproduction, crossover, and mutation. Reproduction selects good string (subset of input attributes); crossover combines good strings to try to generate better offspring's; mutation alters a string locally to attempt to create a better string. The string consists of binary bits: 1 to represent selection of attribute else 0 to drop that attribute. 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 three GA operators and then re-evaluated. This process is repeated for specified number of generation.

3.2 The Fitness Function

In this paper WEKA GA is used as search method with CFS as subset evaluating mechanism (fitness function). The features selected by filter GA-CFS have been experimented with five classifiers. The proposed method is shown in figure 4.

The downside of univariate filters for eg information gain is, it does not account for interactions between features, which is overcome by multivariate filters for eg CFS. CFS evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Correlation coefficient is used to estimate correlation between subset of attributes and the target class label, as well as inter-correlations between the features. Relevance of a group of features grows with the correlation between features and classes, and decreases with growing inter-correlation [8]. CFS is used to determine the best feature subset and can be combined with search strategies such as forward selection, backward elimination, bi-directional search, best-first search and genetic search. Equation for CFS is given is equation 1. Authors have GA as search method with CFS as fitness function.

$$r_{zc} = \frac{k\overline{r_{zi}}}{\sqrt{k + k(k-1)\overline{r_{ii}}}} \qquad (1)$$

Where rzc is the correlation between the summed feature subsets and the class variable, k is the number of subset features, rzi is the average of the correlations between the subset features an the class variable, and rii is the average inter-correlation between subset features [8].

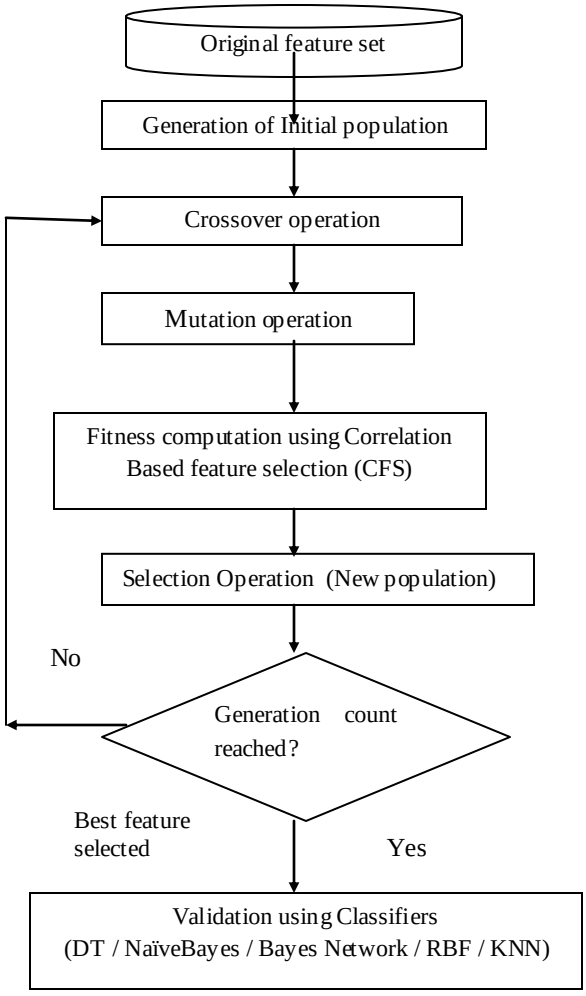

Figure 4. Proposed filter

4. DISCUSSION

4.1 Data used for the Model

PIDD includes the following attributes 8 input attributes and target variable takes two values: tested negative and tested positive. A total of 768 cases are available in PIDD. 5 patients had a glucose of 0, 11 patients had a body mass index of 0, 28 others had a diastolic blood pressure of 0, 192 others had skin fold thickness readings of 0, and 140 others had serum insulin levels of 0. After deleting these cases there were 392 cases with no missing values (130 tested positive cases and 262 tested negative) [19]. The Heart Statlog dataset consist of 270 instances, 13 input attributes and the output class variable to be predicted is presence or absence of heart disease. The Breast Cancer Dataset consist of 286 instances, 9 input attributes and the output class variable to be predicted is no-recurrence-events or recurrence-events. The multi class Dermatology dataset consists of 366 instances, 34 inputs and the output class variable to be predicted has 6 class labels. The differential diagnosis of erythemato-squamous diseases is a real problem in dermatology. They all share the clinical features of erythema and scaling, with very little differences. The diseases in this group are psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, cronic dermatitis, and pityriasis rubra pilaris. These 6 class labels are used as c1,c2,… c6 in table 9. The mentioned medical data sets are available at http://www1.ics.uci.edu/~mlearn/MLSummary.html.

4.2 Performance metrics

The evaluation is based on a set of performance metrics. For the sake of completeness few of the performance metrics have been discussed. True positive (TP) corresponds to the number of positive examples correctly predicted by the classifier. False negative (FN) corresponds to the number of positive examples wrongly predicted as negative by the classifier. False positive (FP) corresponds to the number of negative examples wrongly predicted as positive by the classifier. True negative (TN) corresponds to the number of negative examples correctly predicted by the classifier.

The true positive rate (TP rate) or sensitivity is the fraction of positive examples predicted correctly by the model. TP Rate = TP/(TP + FN). The false positive rate (FP rate) is the fraction of negative examples predicted as a positive class. FP Rate = FP/(TN + FP) Precision is the fraction of records that actually turns out to be positive in the group the classifier has declared as a positive class. Precision = TP / (TP + FP). Recall is the fraction of positive examples correctly predicted by the classifier. Recall = TP / (TP + FN).F-measure is used to examine the tradeoff between recall and precision. F-measure=2*TP/(2*TP+FP + FN). The above measures are usually used for binary classification.

For the multiclass problem two common approach which extends the binary classifiers to handle multiclass problems are one-against-rest (1-r) approach and one-against-one(1-1) approach. Consider a multiclass problem with m classes. Y = { y1,y2, ….. ym}. With m-class problem, in one-against-rest approach, the multiclass problem is decomposed no m binary problems. For each class yj, all instances belonging to yj are considered as positive example, while remaining instances , belonging to other classes , are considered negative examples. In one-against-one approach, m(m-1)/2 binary classifiers are constructed between a pair of classes ,(yk yj). While constructing a binary classifier with class yk and yj , the instances which do not belong to yk or yj are neglected[20].

A receiver operating characteristics (ROC) curve is a graphical approach for displaying the tradeoff between TP rate and FP rate of a classifier. The area under the ROC curve is the measure of accuracy of the model. The model with the perfect accuracy will have an area of 1. The model, which performs random guessing or has less accuracy, has area closer to 0.5. Further the Root mean squared error, Relative absolute error, Root relative squared error and Mean absolute error has been computed.

5. RESULTS

As a part of feature selection step the multivariate filter: Genetic algorithm with Correlation based feature selection as subset evaluating mechanism has been used with four medical datasets from the UCI Machine Learning Repository: Pima Indians Diabetes , Heart Statlog , Breast Cancer and Dermatology dataset. For GA, population size is 20, number of generation is

20, crossover rate is 0.6 and mutation rate is 0.033. The number of relevant features selected by proposed filter GA with CFS for the four medical dataset is shown in table 1. The five Weka classifiers Decision tree C3.4, Naïve bayes, K-NN, RBF and Bayesian classifier has been tested on four medical datasets from the UCI Machine Learning Repository using the relevant feature as identified by the proposed filter. Weka version j4.8 of C4.5 decision tree has been used. K-NN was experimented with different values of K neighbors. K value corresponding to best accuracy of KNN is shown in tables. For RBF ,K-means clustering algorithm has been used to obtain k basis functions for each class.

Experiment results show that by employing feature subset selection enhances the classification accuracy of all the five classifier for diabetic dataset. Table 1 illustrates the improvement in classification accuracy of the five classifiers on four medical dataset as result of feature selection. For heart statlog dataset, classification accuracy of DT , Bayesian classifier remained same with all inputs as well as with relevant features as identified by proposed filter, which illustrates the fact that elimination of 6 irrelevant features did not worsen the classification accuracy. Further for Breast cancer dataset the classification accuracy of Naïve Bayes, RBF and K-NN was substantially improved. The removal of 4 irreverent attributes did not worsen the classification accuracy of Bayesian classifier. Exceptional case found was the classification accuracy of DT or breast cancer dataset with reduced features as input declined by 2%. ROC area clearly illustrates the substantial improvement in classification accuracy. The objective of the paper is not to find best classifier, instead to illustrate the significance of feature selection. The experiment results clearly show an appreciable improvement in accuracy for KNN, followed by for RBF and Naïve Bayes classifier. There was not major improvement in classification accuracy of Bayesian classifier, but it was noted that the reduction in irrelevant attribute did not decrease the accuracy of classifier. The behavior of DT was different for the all the three dataset. The TP rate, FP rate, Precision, Recall and F-measure for the five classifier with all inputs and reduced inputs by proposed filter for the diabetic, heart and breast cancer medical dataset is shown in table 3,5 and 7 respectively. Further the Root means squared error, relative absolute error, root relative squared error, mean absolute and ROC area for the five classifier using all inputs and inputs selected by proposed filter for diabetic, heart and breast cancer medical dataset is shown in table 2,4 and 6 respectively.

For the multiclass dataset, Dermatology, one-against-rest (1-r) approach has been used for estimate the performance metrics. The predictor error measure for Dermatology dataset is shown in Table 8. Table 9 shows how TP Rate, FP Rate, precision, recall, F-measure and Roc area is computed considering 6 binary classifiers. For the dermatology data, the relevant attributes identified by GA_CFS have indeed improved classification accuracy of RBF, DT, NB and K-NN. Further the performance of Bayesian does not decrease with the relevant attributes as input by proposed filter. The high value of F-measure for all the dataset except for breast cancer dataset proves both the precision and recall are reasonably high.

6. CONCLUSIONS

The proposed filter, GA with CFS as subset-evaluating mechanism has been experimented with four medical datasets. While GA ensures global search, CFS results in reduced feature subset. In addition CFS is highly correlated with the class have low intercorrelation. The experimental results clearly illustrate that the proposed filter GA_CFS improves classification accuracy of Naïve bayes, K-NN and RBF classifier for all the four medical dataset. The Bayesian classifier performance did not improve appreciably, neither did not decline with less number of relevant inputs provided by GA_CFS. The performance of DT improved for diabetic and dermatology dataset, remained same for heart statlog dataset, but marginally decreased for breast cancer dataset.

7. REFERENCES

[1] J. Han And M. Kamber," Data Mining: Concepts and Techniques", San Francisco, Morgan Kauffmann Publishers, 2001.

[2] Jennifer G. , "Feature Selection for Unsupervised Learning", Journal of Machine Learning, Vol. 5,pp845-889, Dec 2004.

[3] Maciej Modrzejewski,"Feature selection using Rough Sets Theory", In Proceedings of the European Conference on Machine learing,pp 213-226,1993.

[4] Manoranjan Dash, Kiseiok Choi, Petr Scheuermann, Huan Liu. ,"Feature Selection for Clustering – a Filter Solution", In Proceedings of the Second International Conference on Data Mining, 2002.

[5] Volfer Rotz, and Tilman Lange," Feature Selection in Clustering Problems", In Advances in Neural Information Processing Systems 16,2003.

[6] Ron Kohavi, George H. John ,"Wrappers for feature subset Selection", Artificial Intelligence, Vol. 97, No. 1-2. pp. 273-324, 1997.

[7] M. Dash, H. Liu," Feature Selection for Classification", Intelligent Data Analysis, pp 131–156, March1997.

[8] Mark A. Hall,"Correlation-based Feature Selection for Machine Learning", Dept of Computer science, University of Waikato. http://www.cs.waikato.ac.nz/ mhall/thesis.pdf, 1998.

[9] Shyamala Doraisamy ,Shahram Golzari ,Noris Mohd. Norowi, Md. Nasir B Sulaiman , Nur Izura Udzir,"A Study on Feature Selection and Classification Techniques for Automatic Genre Classification of Traditional Malay Music", ismir2008.ismir.net/papers/ISMIR2008 256.pdf, 2008.

[10] Y.Saeys, I.Inza, and P. LarrANNaga," A review of feature selection techniques in bioinformatics", Bioinformatics, 23(19) pp.2507-2517,2007.

[11] Hussein Almuallim and Thomas G. Dietterich," Learning Boolean concepts in the presence of many irrelevant features", Artificial Intelligence, 69(1-2): 279–305, 1994.

[12] Antonio Arauzo, Jose Manuel Benitez, and Juan Luis Castro," C-focus: A continuous extension of focus", In proceedings of the 7th online World Conference on Soft Computing in Industrial Applications, 2002.

[13] Kira, Kenji, and Larry A. RENDELL, " A practical approach to feature selection", In: Derek H. SLEEMAN and Peter EDWARDS, eds. ML92, Proceedings of the Ninth International, Conference on Machine Learning. San Francisco, CA, USA: Morgan Kaufman Publishers Inc., pp. 249–256, 1992.

[14] H. Liu and W.X. Wen,"Concept Learning Through Feature Selection", In Proceedings of First Australian and New Zealand Conf. Intelligent Information Systems, 1993.

[15] Baranidharan Raman, Thomas R, Ioerger, "Instance Based Filter for Feature Selection", 2002.– http://citeseerx.ist.psu.edu.

[16] M.A.Jayaram, Asha Gowda Karegowda," Integrating Decision Tree and ANN for Categorization of Diabetics Data", International Conference on Computer Aided Engineering, IIT Madras, Chennai, India, December 13-15, 2007.

[17] Asha Gowda Karegowda and M.A. Jayaram," Cascading GA & CFS for Feature Subset Selection in Medical Data Mining", International Conference on IEEE International Advance Computing Conference (IACC'09), Thapar University, Patiala, Punjab India, March 6-7, 2009

[18] D. Goldberg," Genetic Algorithms in Search, Optimization, and Machine learning", Addison Wesley, 1989.

[19] Joseph L.Breault, "Data Mining Diabetic Databases: Are rough Sets a Useful Addition?" 2001. www.galaxy.gmu.edu/interface/I01/.../JBreault/JBreault-Paper.pdf

[20] Pang-Ning Tan, Michael Steinbach, Vipin Kumar, "Introduction To Data Mining", Pearson Education, Third Impression, 2009.

Table 1. Classification accuracy using proposed filter for different Medical dataset

Approach for Attribute selection Method	Medical dataset	Number of Attributes	Classifiers Accuracy (%)				
			Decision Tree C4.5	Naïve Bayes	Bayesian classifier	RBF	K_NN
GA+CFS	Pima diabetic	4	85.71	83.46	84.21	87.97	85.70(k = 15)
With all inputs		8	82.71	79.70	82.71	81.20	84.21(k = 15)
GA+CFS	Breast Cancer	5	66.00	72.16	70.10	70.10	74.50(k = 20)
With all inputs		9	68.04	71.13	70.10	68.04	70.10(k = 15)
GA+CFS	Heart	7	76.08	84.78	82.16	83.70	85.87(k=30)
With all inputs		13	76.09	83.70	82.61	82.61	82.60(k =20)
GA+CFS	Dermatology	21	97.5806	98.39	99.13	98.39	97.58(k =15)
With all inputs		34	94.35	97.58	99.13	95.96	95.96(k= 15)

Table 2. Predictor error measures for Diabetic dataset

Classifier	Approach for attribute selection	Root mean squared error	Relative absolute error	Root relative squared error	Mean absolute error	ROC Area
Naïve Bayes	GA+CFS	0.3471	53.1042	77.2479	0.2323	0.881
	All Inputs	0.3815	50.7517	84.9014	0.222	0.867
Bayesian	GA+CFS	0.3203	51.1888	71.2738	0.2239	0.898
	All Inputs	0.3418	50.9211	76.0612	0.2227	0.886
RBF	GA+CFS	0.332	59.5762	73.8709	0.2606	0.897
	All Inputs	0.3695	66.1438	82.2216	0.2893	0.841
Decision Tree C3.4	GA+CFS	0.3495	62.6902	77.7732	0.2742	0.79
	All Inputs	0.4026	60.412	89.5957	0.2642	0.685
K-NN	GA+CFS	0.3339	60.4229	74.2929	0.2643	0.901
	All Inputs	0.3459	60.9956	76.9707	0.2668	0.86

Table 3. Classifier Accuracy measures for Diabetic dataset

Classifier	Approach for attribute selection	Sensitivity	FP Rate	Precision	Recall	F-measure	Class
Naïve Bayes	GA+CFS	0.901	0.375	0.883	0.901	0.892	tested_negative
		0.625	0.099	0.667	0.625	0.645	tested_positive
	All Inputs	0.851	0.375	0.878	0.851	0.864	tested_negative
		0.625	0.149	0.571	0.625	0.597	tested_positive
Bayesian	GA+CFS	0.901	0.344	0.892	0.901	0.897	tested_negative
		0.656	0.099	0.677	0.656	0.667	tested_positive
	All Inputs	0.871	0.313	0.898	0.871	0.884	tested_negative
		0.688	0.129	0.629	0.688	0.657	tested_positive
RBF	GA+CFS	0.95	0.344	0.897	0.95	0.923	tested_negative
		0.656	0.05	0.808	0.656	0.724	tested_positive
	All Inputs	0.901	0.469	0.858	0.901	0.879	tested_negative
		0.531	0.099	0.63	0.531	0.576	tested_positive
Decision Tree C3.4	GA+CFS	0.911	0.313	0.902	0.911	0.906	tested_negative
		0.688	0.089	0.71	0.688	0.698	tested_positive
	All Inputs	0.901	0.406	0.875	0.901	0.888	tested_negative
		0.594	0.099	0.655	0.594	0.623	tested_positive
K-NN	GA+CFS	0.931	0.375	0.887	0.931	0.908	tested_negative
		0.625	0.069	0.741	0.625	0.678	tested_positive
	All Inputs	0.911	0.375	0.885	0.911	0.898	tested_negative
		0.625	0.089	0.69	0.625	0.656	tested_positive

Table 4. Predictor error measures for Heart Statlog dataset

Classifier	Approach for attribute selection	Root mean squared error	Relative absolute error	Root relative squared error	Mean absolute error	ROC Area
Naïve Bayes	GA+CFS	0.3582	36.7902	69.3799	0.186	0.904
	All Inputs	0.3675	37.4102	71.1913	0.1895	0.908
Bayesian	GA+CFS	0.353	41.4722	68.3958	0.2101	0.915
	All Inputs	0.3636	41.0937	70.4273	0.2081	0.91
RBF	GA+CFS	0.3572	45.0695	69.1858	0.228	0.917
	All Inputs	0.3677	46.8065	71.2232	0.2371	0.906
Decision Tree C3.4	GA+CFS	0.4535	55.6033	87.8373	0.2816	0.742
	All Inputs	0.4553	52.9524	88.201	0.2682	0.735
K-NN	GA+CFS	0.3684	55.9531	71.3562	0.2834	0.908
	All Inputs	0.3714	56.0326	71.9412	0.2838	0.897

Table 5. Classifier Accuracy measures for Heart Statlog dataset

Classifier	Approach for attribute selection	Sensitivity	FP Rate	Precision	Recall	F-measure	Class
Naïve Bayes (heart)	GA+CFS	0.907	0.204	0.796	0.907	0.848	Absent
		0.796	0.093	0.907	0.796	0.848	present
	All Inputs	0.907	0.204	0.796	0.907	0.848	Absent
		0.796	0.093	0.907	0.796	0.848	present
Bayesian	GA+CFS	0.907	0.245	0.765	0.907	0.83	Absent
		0.755	0.093	0.902	0.755	0.822	present
	All Inputs	0.837	0.184	0.8	0.837	0.818	Absent
		0.816	0.163	0.851	0.816	0.833	present
RBF	GA+CFS	0.884	0.204	0.792	0.884	0.835	Absent
		0.796	0.116	0.886	0.796	0.839	present
	All Inputs	0.884	0.224	0.776	0.884	0.826	Absent
		0.776	0.116	0.884	0.776	0.826	present
Decision Tree C3.4	GA+CFS	0.93	0.338	0.678	0.93	0.784	Absent
		0.612	0.07	0.909	0.612	0.732	present
	All Inputs	0.953	0.408	0.672	0.953	0.788	Absent
		0.592	0.047	0.935	0.592	0.725	present
K-NN	GA+CFS	0.953	0.224	0.788	0.953	0.863	Absent
		0.776	0.047	0.95	0.776	0.854	present
	All Inputs	0.907	0.245	0.765	0.907	0.83	Absent
		0.755	0.093	0.902	0.755	0.822	present

Table 6 Predictor error measures for Breast cancer dataset

Classifier	Approach for attribute selection	Root mean squared error	Relative absolute error	Root relative squared error	Mean absolute error	ROC Area
Naïve Bayes	GA+CFS		80.2565	99.6049	0.3442	0.691
	All Inputs	0.4825	79.9872	100.9522	0.3431	0.676
Bayesian	GA+CFS	0.4827	81.105	100.9965	0.3478	0.681
	All Inputs	0.4902	81.0012	102.5618	0.3474	0.659
RBF	GA+CFS	0.4517	86.9976	94.5167	0.3731	0.66
	All Inputs	0.4729	87.36	98.9447	0.3747	0.66
Decision Tree C3.4	GA+CFS	0.4782	99.8245	100.0659	0.4281	0.5
	All Inputs	0.4879	92.4804	102.0849	0.3966	0.603
K-NN	GA+CFS	0.4474	86.1367	93.6179	0.3694	0.688
	All Inputs	0.4512	88.3017	94.4007	0.3787	0.684

Table 7. Classifier Accuracy measures for Breast Cancer dataset

Classifier	Approach for attribute selection	Sensitivity	FP Rate	Precision	Recall	F-measure	Class
Naïve Bayes	GA+CFS	0.828	0.485	0.768	0.828	0.797	no-recurrence-events
		0.515	0.172	0.607	0.515	0.557	recurrence-events
	All Inputs	0.828	0.515	0.757	0.828	0.791	no-recurrence-events
		0.485	0.172	0.593	0.485	0.533	recurrence-events
Bayesian	GA+CFS	0.813	0.515	0.754	0.813	0.782	no-recurrence-events
		0.485	0.188	0.571	0.485	0.525	recurrence-events
	All Inputs	0.813	0.515	0.754	0.813	0.782	no-recurrence-events
		0.485	0.188	0.571	0.485	0.525	recurrence-events
RBF	GA+CFS	0.844	0.576	0.844	0.74	0.788	no-recurrence-events
		0.424	0.156	0.583	0.424	0.491	recurrence-events
	All Inputs	0.859	0.667	0.714	0.859	0.78	no-recurrence-events
		0.333	0.141	0.55	0.333	0.415	recurrence-events
Decision Tree C3.4	GA+CFS	1	1	0.66	1	0.795	no-recurrence-events
		0	0	0	0	0	recurrence-events
	All Inputs	0.875	0.697	0.709	0.875	0.783	no-recurrence-events
		0.303	0.125	0.556	0.303	0.392	recurrence-events
K-NN	GA+CFS	0.984	0.727	0.724	0.984	0.834	no-recurrence-events
		0.273	0.016	0.9	0.273	0.419	recurrence-events
	All Inputs	1	0.879	0.688	1	0.815	no-recurrence-events
		0.121	0	1	0.121	0.216	recurrence-events

Table 8 Predictor error measures for Dermatology dataset

Classifier	Approach for attribute selection	Root mean squared error	Relative absolute error	Root relative squared error	Mean absolute error
Naïve Bayes	GA+CFS	0.0529	2.4445	14.5189	0.0065
	All Inputs	0.0717	2.6373	19.6991	0.007
Bayesian	GA+CFS	0.0519	3.0665	14.2493	0.0082
	All Inputs	0.0511	2.9978	14.0442	0.008
RBF	GA+CFS	0.0694	4.3451	19.053	0.0116
	All Inputs	0.1105	5.3225	30.3408	0.0142
Decision Tree C3.4	GA+CFS	0.097	6.3507	26.6473	0.0169
	All Inputs	0.1363	9.6864	37.42	0.0258
K-NN	GA+CFS	0.0893	9.011	24.5245	0.024
	All Inputs	0.1019	11.3531	27.9861	0.0303

Table 9. Classifier Accuracy measures for Dermatology dataset

Classifier	Approach for attribute selection	TP Rate	FP Rate	Precision	Recall	F-measure	ROC Area	Class
Naïve Bayes	GA+CFS	0.905	0	1	0.905	0.95	1	C1
		1	0	1	1	1	1	C2
		1	0	1	1	1	1	C3
		1	0	1	1	1	1	C4
		1	0.019	0.905	1	0.95	0.999	C5
		1	0	1	1	1	1	C6
	All Inputs	0.857	0	1	0.857	0.923	0.989	C1
		1	0	1	1	1	1	C2
		1	0	1	1	1	1	C3
		1	0	1	1	1	1	C4
		1	0.019	0.905	1	0.95	1	C5
		1	0	1	1	1	1	C6
Bayesian	GA+CFS	1	0.01	0.955	1	0.977	0.999	C1
		1	0	1	1	1	1	C2
		1	0	1	1	1	1	C3
		1	0	1	1	1	1	C4
		0.947	0	1	0.947	0.973	0.999	C5
		1	0	1	1	1	1	C6
	All Inputs	1	0.01	0.955	1	0.977	0.999	C1
		1	0	1	1	1	1	C2
		1	0	1	1	1	1	C3
		1	0	1	1	1	1	C4
		0.947	0	1	0.947	0.973	0.999	C5
		1	0	1	1	1	1	C6
RBF	GA+CFS	0.905	0	1	0.905	0.95	0.995	C1
		1	0	1	1	1	1	C2
		1	0	1	1	1	1	C3
		1	0	1	1	1	1	C4
		1	0.019	0.905	1	0.95	0.994	C5
		1	0	1	1	1	1	C6
	All Inputs	0.857	0.01	0.947	0.857	0.9	0.951	C1
		1	0.012	0.976	1	0.988	0.994	C2
		1	0	1	1	1	1	C3
		0.917	0.009	0.917	0.917	0.917	0.977	C4
		0.947	0.01	0.947	0.947	0.947	0.957	C5
		1	0.009	0.875	1	0.933	0.999	C6
Decision Tree C3.4	GA+CFS	1	0.019	0.913	1	0.955	0.994	C1
		1	0	1	1	1	1	C2
		1	0.01	0.96	1	0.98	0.995	C3
		1	0	1	1	1	1	C4
		0.842	0	1	0.842	0.914	0.934	C5
		1	0	1	1	1	1	C6
	All Inputs	0.857	0.01	0.947	0.857	0.9	0.94	C1
		0.976	0.06	0.889	0.976	0.93	0.958	C2
		1	0.01	0.96	1	0.98	0.995	C3
		1	0	1	1	1	1	C4
		0.842	0	1	0.842	0.914	0.934	C5
		1	0	1	1	1	1	C6
K-NN	GA+CFS	0.905	0.01	0.95	0.905	0.927	0.996	C1
		1	0	1	1	1	1	C2
		1	0	1	1	1	1	C3
		1	0	1	1	1	1	C4
		0.947	0.019	0.9	0.947	0.923	0.997	C5
		1	0	1	1	1	1	C6
	All Inputs	0.905	0.029	0.864	0.905	0.884	0.995	C1
		0.976	0	1	0.976	0.988	1	C2
		1	0	1	1	1	1	C3
		1	0	1	1	1	1	C4
		0.947	0.019	0.9	0.947	0.923	0.997	C5
		0.857	0	1	0.857	0.923	1	C6