

Maximum Relevancy Minimum Redundancy Based Feature Subset Selection using Ant Colony Optimization

Rizwan Mehmood, Waseem Shahzad, Ejaz Ahmed

Computer Science Department
National University of Computer and Emerging Sciences, Islamabad, Pakistan

Received: November 30 2016

Accepted: February 2, 2017

ABSTRACT

In recent years dimensionality reduction of data has gained a lot of interest from the machine learning community, partly due to the huge amount of data available for processing. Classical machine learning algorithms were designed to work with limited amount of data where more emphasis was given to the learning methodology of such algorithms e.g. learning crisp rules from a dataset. Recent explosion of data, proved detrimental to the accuracy of classification algorithms due to lowering storage costs and inexpensive processing power. Therefore, feature selection which is a key technique in dimensionality reduction has become an important frontier in machine learning research. In this paper, we propose a novel filter based feature selection method. Proposed method is based on Ant Colony Optimization (ACO), and maximum Relevance and Minimum Redundancy (mRMR) for efficient subset evaluation. Although wrapper methods frequently use ACO for feature subset generation but ACO is not thoroughly studied in the development of filter methods. Proposed method takes both feature relevancy and feature redundancy in account. Our research ensures selection of features which are highly relevant with the target concept, weakly redundantly with each other and useful predictor for classification algorithms. We have performed an extensive experimentation over eleven publicly available datasets and three popular machine learning classifiers. Experimental results of comparisons show that proposed method achieves higher classification accuracy and employ reduced number of features.

KEYWORDS: Ant Colony Optimization, Feature Selection, Machine Learning, Wrapper Methods, Feature Relevancy and Feature Redundancy

1. INTRODUCTION

Data and information are increasing day by day and research suggests that data normally doubles after every couple of years. Currently, most of the data stored in digital format. We need different tools to manage, store and adjust huge bulk of data. Data mining plays an important role in copying such huge data and manage it in order to gain insights from it. Feature selection is one of the key techniques in dimensionality reduction. Its purpose is to extract the most useful data for the given data mining task e.g. data classification. Moreover, in recent years there has been a lot of thrust from computational intelligence community for developing heuristic based search algorithms which are very helpful to address problems that are inherently very complex, and can not be solved exhaustively in polynomial time. Feature selection is one such problem therefore a number of searching mechanisms have been proposed to generate efficient feature subsets. ACO is bio-inspired natural population and has successful applications in feature selection where it is mostly employed as part of wrapper methods [1, 4]. ACO is used to compute distance related traveling problems [3]. Wrapper methods are a type of feature selection which selects feature subsets on the basis of their classification accuracy on a given classifier [5]. These methods are relatively more accurate than filter methods since a classifier is used as an evaluator. But at the same time wrapper methods are computationally expensive and can not scale very well for very high dimensional datasets. On the other hand, filter methods are inexpensive but relatively less accurate.

We have used three state-of-the-art classifiers or techniques C4.5, KNN and RIPPER to check the worth of subsets selected by our algorithm. The advantage of our proposed method is based on filter based Feature Subset Selection (FSS) and is computationally less expensive than most of the wrapper methods. There are three main advantages of feature selection. It enhances the predictive capability of the classifier; it provides fast and cost-effective predictors; and it improves comprehensibility of the underlying processes. In this paper, we studied the application of ACO for filter methods. An efficient filter method is proposed which used ACO for feature subset generation and a comprehensive mRMR based measure, used to evaluate the generated subsets.

This paper proposed a novel feature selection algorithms based on ACO and mRMR, where mRMR is used as a subset evaluator [2]. Filter methods are usually inferior to wrapper methods in terms of classification accuracy. We propose that if a better searching mechanism is employed which is able to search through all the key feature subsets and evaluate these subsets with a more comprehensive feature evaluation measure then filter methods can perform comparable to the wrapper methods. In today's world we have a lots of features of the data and finding the right

*Corresponding Author: Rizwan Mehmood, Computer Science Department, National University of Computer and Emerging Sciences, Islamabad, Pakistan. manglian@gmail.com

number of features for a problem is a crucial step and it has been applied in many real world applications like agriculture, bioinformatics etc.

Experimentation framework of our algorithm for ACO-mRMR compares selected features on three state-of-the-art classifiers C4.5, KNN and RIPPER. It is also based on comparison with other algorithms with full selected feature sets on same state-of-the-art classifiers. Summary of comparisons of ACO-mRMR with both full features and other algorithms is also included in experimentation.

The rest of the paper is organized as follows; section 2 provides literature review to describe thorough relevancy of research, section 3 provides proposed methodology. Section 4 explains the experimentation and analysis comparisons of proposed technique with other techniques. Finally, section 5 concludes our research work with future work.

2. LITERATURE REVIEW

Feature selection divides into two categories which are, filter based and wrapper based feature selection [5, 14]. Filter based selection accuracy depends on the inter class relevancy and in wrapper based selection algorithm accuracy depends on any type of learning algorithms.

There are several algorithms using information gain to rank features in descending order. This means those features which have more information gain will come at the top. So there will be certain threshold suggested by algorithm or mentioned by user and those features with lower information gain value than the threshold will be removed. It is possible that threshold value will be same for all data sets or it changes from datasets to datasets. If we make a fix threshold then it will not depend on the number of features in any dataset and it may pick those features as well which are either redundant or irrelevant. For example, we create a formula or threshold equals to total feature divided by 2. Now in this formula almost half of the features selected each time despite of their importance.

Ratanamahatana and Gunopulos [10] have a feature selection algorithm based on decision trees. The name of the method is Selective Bayesian Classifier (SBC). It is a two-step process; firstly 10% of the training datasets gives to c 4.5 classifier. Only those features which are present on the first three level of the decision tree are selected. After the selection of these features they now send to Naive Bayes to get the final classification accuracy. The main problem of this type of process is its computational cost and over fitting [9]. Also it selects on first three level of the feature which is same like fixed threshold for all data sets which we have mentioned earlier.

Bai-Ning Jiang *et al.* have a hybrid feature selection algorithm. This algorithm also had two steps. Firstly Symmetric Uncertainty (SU) is used to get the features value. After SU those feature which has less value than a certain threshold will be removed. Secondly Genetic Algorithm as a searching mechanism used on those features which are selected on the basis of SU value. Naïve Base classifier used to predict the goodness of selected feature subsets. After that Naïve Base used with SU for Optimization of subset [6].

J. Zhou *et al.* have an algorithm with Mutual Information (MI) and ACO based feature selection. Mean squared error and Regression estimation are used to evaluate feature subsets. Classification accuracy and MI are used to optimize subsets [11].

There is another algorithm based on ACO and Mutual Information (MI) based FSS algorithm proposed by Chun-Kai Zhang *et al.* MI used in as a heuristic function and artificial neural network based classifier and it is used as an evaluator of feature subset. Both ANN and MI are used for subset optimization as well. [12].

R. Jenses *et al.* have an algorithm based on fuzzy rough set using ACO and C 4.5 proposed a fuzzy-rough data reduction using any colony optimization and C4.5. Where ACO is used to find fuzzy-rough set. Where a fuzzy-rough based measure is used as a stopping criteria. Once the maximum dependency reaches and it will stop creating solution [13].

Xiangyang Wang *et al.* have an algorithm using Particle Swarm Optimization (PSO) with rough set based feature selection. As basic PSO is used as a standard PSO converted into binary PSO. Here PSO is used to find rough set and LEM2 is used for subset optimization and rule induction with the help of rough sets [14].

M. H.Aghdam *et al.* proposed ACO based algorithm for text feature selection. They used accuracy of classifier as heuristics and also as evaluator of subset feature. Only KNN used as a classifier[8].

M. F. Zaiyadi *et al.* [15] also create ACO based text feature selector. In this algorithm they use IG as subset evaluator but they did not provide any comprehensive results to prove their theory.

R K Sivagaminathan *et al.* have a hybrid algorithm. This algorithm based on ACO and ANN [16].

Shahla Nemati *et al.* algorithm based on ACO and GA based featured sub set selection to predict protein prediction. KNN used to evaluate subset [17].

Hai-Hua Gao *et al.* presents an ACO based feature selection algorithm to detect network intrusion Support Vector Machine (SVM) used as classifier [18].

Mohammad Ehsan Basiri *et al.* presents feature selection algorithm based on ACO and it will predict the post-synaptic activity in proteins. KNN used to evaluate subsets and classification accuracy as heuristic. [19].

Yannis Marinakis *et al.* used an algorithm, applied feature selection on financial classification problem this algorithm based on ACO and PSO [21].

ant constructs a potential solution. Once all the solutions have been generated they stored in the local memories of the ants. After that each subset is evaluated using some evaluation function. When all solutions are evaluated then stopping criteria will be checked. In this paper the stopping criteria is the m-number of ant converges to a path or maximum number of epochs completed. If criteria do not meet pheromone then it will be updated and ants will generate for next generation. Figure 2 depicts main steps of the proposed algorithm.

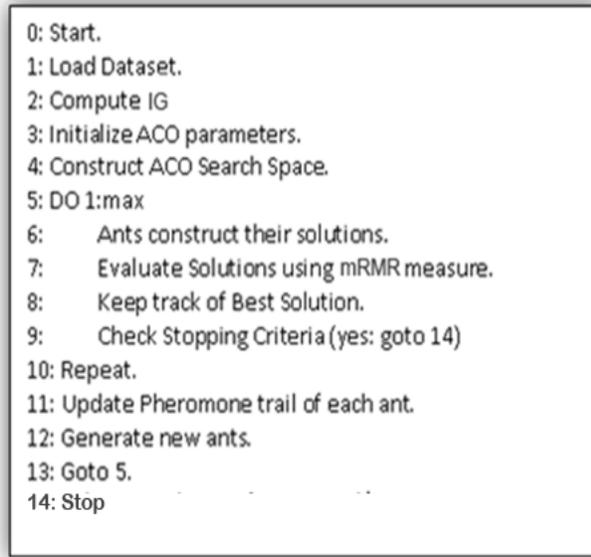


Figure 2: Algorithm of the proposed ACO-mRMR.

3.1 Feature Importance Measures

There are different subset evaluation techniques. Here, we discuss only those techniques which are relevant to our proposed technique. Wrapper based techniques are based on learning algorithm while filter based approaches are the techniques from the statistic domain. Mark A Hall used correlation based feature selection mechanism [25, 30]. Mark said that for classification tasks one can use correlation base measures. There are two types of correlation base measures which are Information theoretic measures and linear correlation measures. Some famous and important feature subset evaluation measures are: mutual information, symmetric uncertainty and information gain are the popular and efficient feature subset evaluation methods [6, 15, 17]. Whereas, linear correlation based measures are also used to evaluate and check redundancy among the features [7, 9].

Ajay Kumar Tanwani et al. provided empirical results and proposed that information gain of a dataset is directly proportional to the classification accuracy. Moreover, it is also proposed that classification accuracy of an algorithm is related with the nature of the dataset rather than the algorithm itself [26]. It is also proposed that the nature of the dataset used has more strong influence on the classification accuracy rather than the selection of a learning algorithm [27].

3.2 Minimal-Redundancy-Maximal-Relevance (mRMR)

In feature selection there are two important factors. Firstly, is to remove redundant data and secondly consider only relevant data. So this technique considers these two factors and it make feature selection on the basis of these two points. Maximum relevance implemented using mean value of mutual information of all feature with class. The formula of maximum relevance is given below.

$$\max D(S, c), \quad D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c). \quad \dots \text{Eq. 1}$$

If we just implement relevance there are very bright chances that the dependency between features could be increased as shown in equation 1. That is why minimum redundancy should be implemented in such a way its relevance may not be disturbed. So following formula of minimum redundancy should be used to get minimum redundancy between features.

$$\min R(S), R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j).$$

...Eq. 2

Combing both equations 1 and 2 will optimize. Following formula of equation 3 will depict the working of mRMR.

$$\max \Phi(D, R), \Phi = D - R.$$

.... Eq. 3

A search method can be used here with this technique e.g. incremental search to get best features. We are using ACO to get best features.

3.3 Ant Colony Optimization and maximum Relevancy and Minimum Redundancy (ACO-mRMR)

ACO-mRMR employs MI to measure redundancy between the features and relevancy with the class[3]. Feature subsets are generated through ACO. Following formula evaluates the worth of the selected subset evaluation.

$$\text{Merit}(S') = (N - |S'| * (\sum_{i=1}^N |C_{ir}| - \alpha \sum_{i=1} \sum_{j=1} |C_{ij}|)) / N \quad \dots \text{Eq. 4}$$

Where |C_{ir}| is the feature-class correlation between feature i and class c, and |C_{ij}| is the inter-feature correlation between feature i and feature j. α is the scaling factor as given in equation 4. Figure 3 depicts an algorithm for the ACO-mRMR.

```

0: Start.
1: Load Dataset.
2: Compute IG
3: Initialize ACO parameters.
4: Construct ACO Search Space.
5: DO 1:max
6:     Ants construct their solutions.
7:     Evaluate Solutions using mRMR measure.
8:     Keep track of Best Solution.
9:     Check Stopping Criteria (yes: goto 14)
10: Repeat.
11: Update Pheromone trail of each ant.
12: Generate new ants.
13: Goto 5.
14: Select 10 subsets of converged/max generation.
15: Run Ensemble classifiers.
16: Opt highest accuracy subset through majority voting.
17: Stop.
    
```

Figure3: Algorithm of ACO-mRMR.

As it is illustrated in the above figure 3, in first step dataset is loaded into the program. Once dataset is loaded, its Information gain is calculated for each attribute. All the parameters of ACO are including number of ants, α and β values of node transition probability, convergence threshold value, maximum number of generations and evaporation rate are initialized. Along with search space construction that takes place in the subsequent steps. Search space consists of nodes corresponding to the number of features in the dataset. Entire search space is represented using mesh topology. Each generation has a fixed number of ants representing candidate solutions in a generation. After each generation, candidate solution evaluation takes place. Following formula equation 1 is used for evaluating a selected subset. Subset evaluator is based on a formula that computes IG between selected features and the class, and uses following formula of Fitness calculation of subset S' by ACO-mRMR to compute the merit of the selected subset.

$$\text{Merit}(S') = \frac{(N-|S'|) * (\sum_{i=1}^n |IG_i|)}{N} \quad \dots \text{Eq. 5}$$

Where in equation 5, S' is the current selected subset, N is the total number of features present in the dataset, $|IG_i|$ is the information gain of feature in the subset S' . For example, there are 9 features in a dataset three subsets are selected having 3, 4, and 2 features respectively. Suppose all the three subsets have the same IG e.g. 0.90. Then according to the above formula their merit will be computed as follows.

$$\begin{aligned} \text{Merit (3)} &= ((9-3)*(0.90))/9 = 0.60. \\ \text{Merit (4)} &= ((9-4) * (0.90))/9 = 0.50 \\ \text{Merit (2)} &= ((9-2)*(0.90))/9 = 0.70 \end{aligned}$$

As demonstrated, the above mentioned formula favors those subsets that have small number of features. Hence, the subset that has two features is preferable. Once subset evaluation of candidate solutions takes place, best solution of a generation is preserved. And stopping criteria is checked. Stopping criteria consists of two conditions i.e. maximum number of epochs reached or convergence threshold is met. Once stopping criteria is met algorithm stops. Otherwise, pheromone trail of each ant is updated according to the fitness value of its solution. Fitness value of an ant is computed using above mentioned formula. New set of ants are generated and entire process takes place iteratively until maximum number of epochs is reached or ACO convergence to a solution. Once algorithm has converged or reached to its maximum iterations, 10 best ants are gathered in a set, and the best subset is selected using majority voting.

ACO-mRMR is developed to resolve the problems faced by traditional feature weighting and ranking algorithms that require a threshold value to select final subset. ACO-mRMR adaptively selects an optimal feature subset and no threshold value is required. Since a threshold value can be different for different datasets and in advance an optimal threshold value is not known therefore ACO-mRMR iteratively selects an optimal feature subset and hence can be applied on different datasets.

Since filter based method employs independent statistical measures therefore classification accuracy of the selected subset is indirectly targeted. Our proposed method ACO-mRMR does not require any classifier and hence incur less computational cost. Moreover, feature subsets produced does not decrement accuracy of the classifier; in most of the cases it enhances classifier's accuracy while substantially reducing the dataset.

4. Experimentation and Analysis

For the experimentation, we have used eleven datasets. Following table 1, elaborates characteristics about the datasets used in the experimentation. All the datasets used are publicly available and taken from UCI repository [28].

NO.	Dataset	Total Features	Instances	Classes
1	Iris	4	150	3
2	Liver Disorder	6	345	2
3	Diabetes	8	768	2
4	Breast Cancer- W	9	699	2
5	Vote	16	435	2
6	Labor	16	57	2
7	Hepatitis	19	155	2
8	Colic-Horse	22	368	2
9	Ionosphere	34	351	2
10	Lymph	18	148	4
11	Lung Cancer	56	32	3

Table 1: Sample Datasets

The proposed algorithm works with categorical attributes therefore continuous attributes need to be discretized in a preprocessing step. We used unsupervised discretization filter of Weka-3.6 machine learning tool for discretizing continuous attributes. Filter first computes the intervals of continuous attributes from the training dataset and then uses these intervals for discretization.

4.1 Experimentation Framework for ACO-mRMR

The number of ants is equal to the number of features in the dataset. All the other parameters of ACO are that of a standard ACO i.e. α and β are 1. Maximum number of epochs is 1000. Path convergence threshold is 200. In the experimentation, worth of a FSS algorithm is evaluated on two key measures, i.e. predictive accuracy of the selected subset and number of features selected.

We have implemented ACO-mRMR in Matlab 2010. We have used standard implementation of [7, 10, 13, 14, 23, 51]. Implementation of these algorithms is provided by data mining software Weka [59]. All the algorithms are used with their default values and no tweaking is done with the methods. Since most of these algorithms are implemented

by their respective authors therefore it is assumed that parameter setting is already incorporated. Following table shows number of features selected by the feature selection algorithms.

4.1.1 ACO-mRMR and Selected Features

Information about the total number of features in full feature set, features selected by other feature selection algorithms and our proposed method ACO-mRMR are shown in table 2.

Dataset	Total	ACO-FRS	PSO	Tabu	GA	ACO-mRMR
Iris	4	4	4	3	2	2
Liver Disorder	6	5	5	5	5	3
Diabetes	8	8	8	6	8	3
Breast Cancer- W	9	4	5	9	7	4
Vote	16	14	12	12	11	6
Labor	16	6	6	7	7	6
Hepatitis	19	14	13	13	12	8
Colic-Horse	22	13	16	11	12	3
Ionosphere	34	26	19	16	21	10
Lymph	18	8	7	7	9	6
Lung Cancer	56	5	5	5	13	16

Table 2: Reduced Datasets, ACO-mRMR

In figure 3, our proposed method consistently selected smaller number of features as compared with other algorithms except in lung cancer dataset where a number of features are highly correlated with each other. Lung cancer dataset is a high-throughput dataset where number of instances is less than number of features. It can be observed that proposed method consistently selects small number of features when adequate number of instances is present in the dataset.

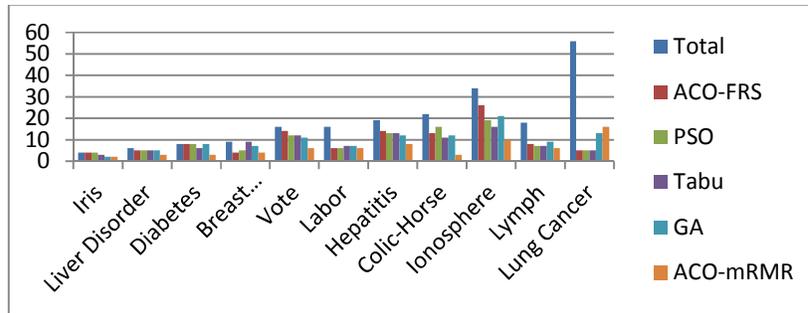


Figure3: Reduced dataset, ACO-mRMR

In the following table 3, selected features of each algorithm are presented.

No	ACO-FRS	PSO	Tabu	GA	ACO-mRMR
1	1,2,4,3	1,2,3,4	2,3,4	3,4	4,3
2	2,3,4,5,6	2,3,4,5,6	2,3,4,5,6	2,3,4,5,6	1,4,6
3	All	All	2,4,5,6,7,8	All	2,6,7
4	All	All	All	1,2,3,6,7,8,9	7,6,12
5	1,2,3,4,5,7,9,10,11,12,13,14,15,16	1,2,3,4,5,9,10,11,12,13,15,16	1,2,3,4,5,7,10,11,1,2,13,15,16	1,2,3,4,5,9,10,11,12,13,15	12,3,14,10,5,4,11
6	2,3,6,11,14,16	3,6,7,8,12,16	1,2,7,10,12,14,16	1,2,7,10,12,14,16	15,5,9,11,3,4,12
7	1,2,3,4,5,7,8,9,11,13,16,17,18,19	1,2,3,4,5,6,7,9,11,1,2,15,18,19	1,2,3,4,5,6,7,8,9,1,14,17,18,19	1,2,3,5,7,9,11,12,14,17,18,19	14,18,17,6,19,11,12
8	1,2,3,5,6,7,8,11,13,15,17,18,19,22	1,2,3,5,6,7,8,9,10,11,12,13,15,16,17,22	1,3,5,8,10,11,12,1,3,16,18,22	1,2,6,8,10,12,13,16,18,20,21,22	12,20,1
9	1,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,19,20,21,22,24,26,28,30,33,34	1,4,6,7,8,9,10,12,14,17,20,21,22,24,26,28,30,32,33	1,4,6,7,10,12,14,1,6,17,20,21,22,24,2,6,30,33	6,7,8,9,12,13,14,15,1,6,17,20,21,22,23,24,25,26,29,30,31,34	15,29,1,23,7,33,31,6,13,5
10	1,2,3,10,11,13,14,15	3,8,11,12,13,14,15	1,2,5,11,13,14,15	1,2,5,6,8,12,13,14,17	11,13,8,15,7,9
11	6,8,12,29,35,39	10,26,29,34,37	3,6,12,34,40	3,9,15,18,20,21,26,30,35,42,43,49,55	32, 44, 6, 17, 3, 56, 48, 19, 23, 14, 35, 13, 27, 7, 37, 53

Table 3: Selected paths, ACO-mRMR

4.1.2 Comparison of ACO-mRMR with Full Feature Set over C4.5

In the following table 4, predictive accuracy of proposed algorithm and full feature set is presented. C4.5 is used for classification with 10-fold cross validation. Bold value represents highest accuracy achieved by a best performing algorithm.

Dataset	All (C4.5)	ACO-mRMR
Iris	97.33	97.33
Liver Disorder	57.39	68.09
Diabetes	65.75	68.68
Breast Cancer- W	92.70	95.56
Vote	96.32	97.09
Labor	82.45	75.13
Hepatitis	76.77	84.58
Colic-Horse	85.05	85.86
Ionosphere	90.88	87.46
Lymph	75.65	79.72
Lung Cancer	50	65.52

Table 4: Comparison of ACO-mRMR with full feature set over C4.5

C4.5 has its internal method of feature selection therefore such a method can be termed as an embedded method. It uses information gain ratio for node splitting and the feature with highest information gain ratio is selected as a root node. Our proposed method performs comparable or better on ten datasets over C4.5.

4.1.3 Comparison of ACO-mRMR with full Feature set over K-Nearest Neighbor

In table 5, predictive accuracy of proposed algorithm and full feature set is presented. K-Nearest Neighbor is used for classification with 10-fold cross validation. For all Experiments using KNN in this paper the value $k=3$ is chosen. Bold value represents highest accuracy achieved by a best performing algorithm.

Dataset	All (KNN)	ACO-mRMR
Iris	95.33	97.33
Liver Disorder	56.23	58.55
Diabetes	65.75	67.83
Breast Cancer- W	95.42	96.13
Vote	93.10	96.13
Labor	85.96	78.94
Hepatitis	82.58	84.51
Colic-Horse	77.98	85.32
Ionosphere	84.90	88.03
Lymph	82.43	80.40
Lung Cancer	40.62	46.87

Table 5: Comparison of ACO-mRMR with full feature set over KNN

In the above mentioned table our method performs comparable or better on nine datasets.

4.1.4 Comparison of ACO-mRMR with full feature set over RIPPER

Predictive accuracy of proposed algorithm and full feature set is presented as shown in table 6. RIPPER is used for classification with 10-fold cross validation. Bold value represents highest accuracy achieved by a best performing algorithm.

Dataset	All (RIPPER)	ACO-mRMR
Iris	96.66	97.33
Liver Disorder	56.52	58.55
Diabetes	67.83	67.83
Breast Cancer- W	94.70	95.70
Vote	96.09	96.86
Labor	91.22	64.51
Hepatitis	81.93	82.52
Colic-Horse	83.42	85.59
Ionosphere	90.88	85.18
Lymph	75	81.08
Lung Cancer	46.87	53.12

Table 6: Comparison of ACO-mRMR with full feature set over RIPPER

RIPPER is a rule induction algorithm. Their mechanism is based on two important loops, outer loop adds one rule at a time in the rule base whereas inner loop adds conditions to it iteratively, this process continues until rules perform negative coverage of the instances. Hence, RIPPER has its own mechanisms for internal feature selection based on

information gain and rule coverage, low quality features do not contribute in rule construction. In the above mentioned table our method performs comparable or better on nine datasets.

4.1.5 Summarized Comparison of ACO-mRMR and Full Feature Set

In the following table7, predictive accuracy of proposed algorithm and full feature set is presented in a summarized manner.

CLASSIFIER	ALL	mRMR
C4.5	2	9
KNN	2	9
RIPPER	2	9
TOTAL	6	27

Table7: Summarized comparison of ACO-mRMR and full feature set over three classifiers

Following figure 4 is the graphical representation of the comparison of ACO-mRMR and Full feature set using eleven datasets over three classifiers.

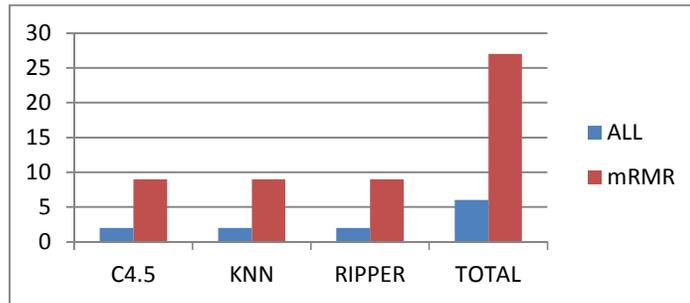


Figure 4: Comparison of ACO-mRMR with full feature set over three classifiers

As it can be observed in the above figures that our proposed method consistently performs better when compared with the full feature set. Some of the classification algorithms employ embedded techniques of feature selection, e.g. state-of-the-art classifiers. Hence our method is also compared with embedded techniques of classification algorithms state-of-the-art classifiers.

4.1.6 Comparison of ACO-mRMR with other algorithms over C4.5

In the following table 8, predictive accuracy of proposed algorithm and other feature selection algorithms are presented. C4.5 is used for classification with 10-fold cross validation. Bold value represents highest accuracy achieved by a best performing algorithm.

Dataset (C4.5)	ACO-FRS	PSO	Tabu	GA	ACO-mRMR
Iris	97.33	97.33	78.66	97.33	97.33
Liver Disorder	57.39	57.39	57.39	57.39	68.09
Diabetes	65.75	65.75	68.09	65.75	68.68
Breast Cancer- W	92.13	92.70	92.70	95.27	95.56
Vote	96.32	96.32	96.32	96.32	97.09
Labor	80.70	77.19	80.70	77.19	75.13
Hepatitis	85.80	78.70	80.64	76.77	84.58
Colic-Horse	84.51	84.78	85.05	85.32	85.86
Ionosphere	90.88	88.31	86.32	87.17	87.46
Lymph	77.02	78.75	77.70	79.72	79.72
Lung Cancer	40.62	46.87	75	40.62	65.52

Table 8: Comparison of ACO-mRMR with other algorithms over C4.5

Following figure 5, is the graphical representation of the comparison of algorithms using C4.5 classifier.

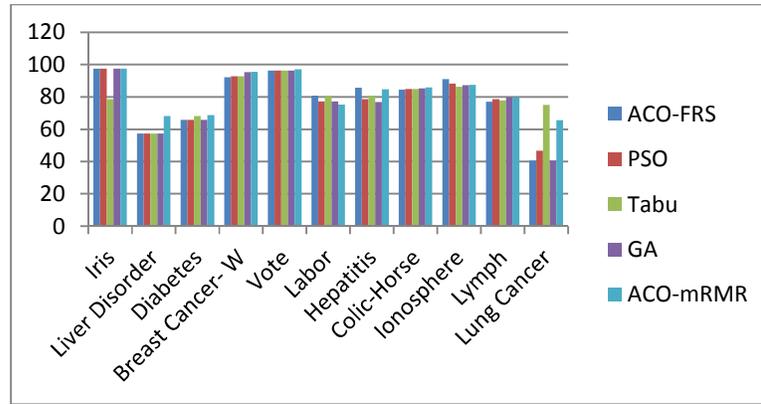


Figure 5: Comparison of ACO-mRMR and other algorithms over C4.5

In iris dataset, although accuracy of ACO-FRS, PSO and GA is same as that of ACO-mRMR but number of features selected by our method is less than these algorithms except GA where number of features selected are equal to features of our method. In diabetes, our method performs better due to smaller number of features selected as compared to tabu search. Collectively on eight of the datasets our proposed method performs better.

4.1.7 Comparison of ACO-mRMR with other Algorithms over K-Nearest Neighbor

In the following table 9, predictive accuracy of proposed algorithm and other feature selection algorithms is presented. K-Nearest Neighbor is used for classification with 10-fold cross validation. Bold value represents highest accuracy achieved by a best performing algorithm.

Dataset (KNN)	ACO-FRS	PSO	Tabu	GA	ACO-mRMR
Iris	95.33	95.33	78.66	97.33	97.33
Liver Disorder	57.39	57.39	57.39	57.39	58.55
Diabetes	65.75	65.75	67.57	65.75	67.83
Breast Cancer- W	95.56	93.84	95.42	95.13	96.13
Vote	93.10	92.87	93.33	94.48	96.13
Labor	89.47	82.45	78.94	78.94	78.94
Hepatitis	83.87	83.87	81.93	82.58	84.51
Colic-Horse	83.69	81.52	81.79	81.52	85.32
Ionosphere	85.47	84.33	82.62	82.62	88.03
Lymph	77.02	79.02	79.02	77.70	80.40
Lung Cancer	53.12	34.37	50	34.37	46.87

Table 9: Comparison of ACO-mRMR with other algorithms over KNN

Following figure 6, is the graphical representation of the comparison of algorithms using K-Nearest Neighbor as a classifier.

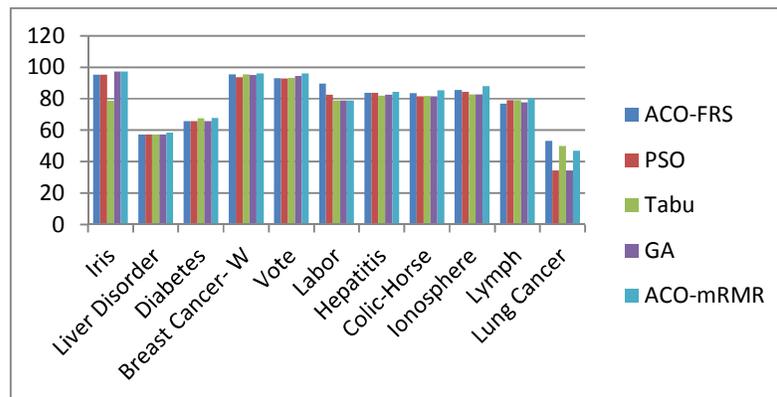


Figure 6: Comparison of ACO-mRMR and other algorithms over KNN

In Ionosphere our algorithm performs better due to smaller number of features selected. Collectively our algorithm performs comparable and/or better in nine datasets.

4.1.8 Comparison of ACO-mRMR with other Algorithms over RIPPER

In the following table 10, predictive accuracy of proposed algorithm and other feature selection algorithms is presented. RIPPER is used for classification with 10-fold cross validation. Bold value represents highest accuracy achieved by a best performing algorithm.

Dataset (RIPPER)	ACO-FRS	PSO	Tabu	GA	ACO-mRMR
Iris	96.66	96.66	96.66	97.33	97.33
Liver Disorder	56.23	56.23	56.23	56.23	58.55
Diabetes	67.83	67.83	67.83	67.83	67.83
Breast Cancer- W	92.99	94.13	94.70	94.56	95.70
Vote	95.17	95.40	95.40	95.86	96.86
Labor	78.94	91.22	91.22	91.22	64.51
Hepatitis	83.22	82.58	83.22	81.93	82.52
Colic-Horse	85.05	84.51	85.32	84.51	85.59
Ionosphere	90.88	86.89	87.46	87.46	85.18
Lymph	77.70	76	75	75.67	81.08
Lung Cancer	40.62	46.87	59.37	46.87	53.12

Table 10: Comparison of ACO-mRMR with other algorithms over RIPPER

Following figure 7, is graphical representation of the comparison of algorithms using RIPPER as a classifier.

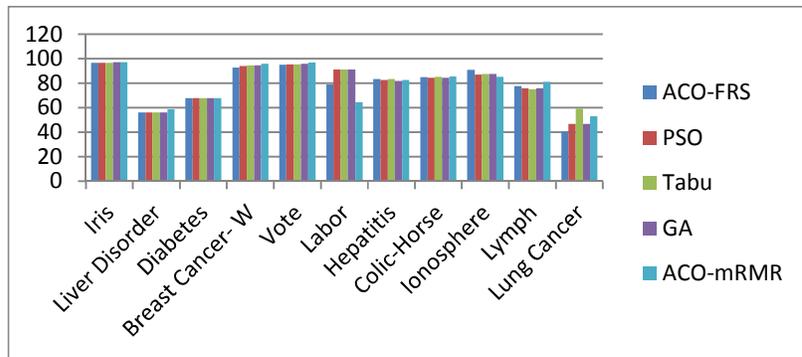


Figure 7: Comparison of ACO-mRMR and other algorithms over RIPPER

In lymph our algorithm performs better due to smaller number of features selected. Collectively our algorithm performs comparable and/or better in seven datasets.

4.1.9 Summarized Comparison of ACO-mRMR and other Algorithms

In the following table 11, predictive accuracy of proposed algorithm and other feature selection algorithms is presented in a summarized manner.

Classifiers	ACO-FRS	PSO	Tabu	GA	ACO-mRMR
C4.5	3	2	1	3	8
KNN	1	0	0	1	9
RIPPER	3	2	4	4	7
TOTAL	7	4	5	8	24

Table 11: Summarized comparison of ACO-mRMR with other algorithms

Following figure 8 is the graphical representation of the summarized comparison of proposed algorithm and other feature selection algorithms.

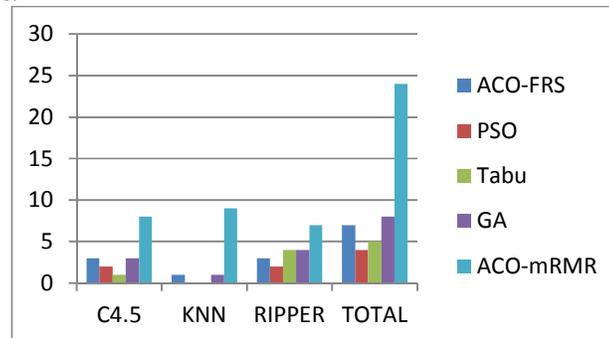


Figure8: Summarized comparison of ACO-mRMR over three classifiers

Our proposed algorithm consistently opts for smaller number of features. In figure 8 it can be observed that our method achieves better predictive classification accuracy over three classifiers. Our method employs mRMR as a subset evaluation method. It can be observed that this measure can optimize subsets optimally.

5. Conclusion and Future Works

In this paper, we propose a new feature selection method based on ACO and mRMR, our proposed method ACO-mRMR is compared with a full feature set over state-of-the-art three classifiers followed by comparison of ACO-mRMR with other feature selection methods. Experimentation results showed that our proposed method performed better in terms of predictive classification accuracy and number of feature selected. Our research work achieves better predictive classification accuracy over three classifiers.

Our research work can be extended in a number of directions. First we are looking forward for hybrid searching mechanisms where one technique may compensate for the demerits of the other. For example, ACO can be combined with other population based searching mechanisms, differential evolution [29]. Moreover, parameter tuning can yield better optimized search. Moreover, feature selection can be extended to include more challenging task e.g. data stream mining. We can extend our algorithm in the domain of data stream classification. Population based feature selection is given very less attention in this domain.

REFERENCES

1. M.M. Kabir, M. Shahjahan and K. Murase, A New Hybrid Ant Colony Optimization Algorithm for Feature Selection, *Experts Systems with Applications: An International Journal*, vol. 39, no. 3, pp. 3747–3763, 2012.
2. Hanchuan Peng, Fuhui Long, and Chris Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 8, pp.1226-1238, 2005.
3. Hashim Ali, Muhammad Haris, Fazl Hadi, Ahmadullah, Salman and Yasir Shah, Solving Traveling Salesman Problem through Optimization Techniques Using Genetic Algorithm and Ant Colony Optimization, *Journal of Applied Environmental & Biological Sciences*, ISSN: 2090-4274, Vol. 6(4S), pp. 55-62, March 2016.
4. E A. R. Baig, W. Shahzad, and S. Khan: Correlation as a Heuristic for Accurate and Comprehensible Ant Colony Optimization Based Classifiers. *IEEE Transactions on Evolutionary Computation* 17(5): 686-704 Oct, 2013
5. Richard Jensen, *Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches*, Wiley-IEEE Press 2008.
6. Bai-Ning Jiang Xiang-Qian Ding Lin-Tao Ma "A Hybrid Feature Selection Algorithm: Combination of Symmetrical Uncertainty and Genetic Algorithms" *The Second International Symposium on Optimization and Systems Biology*, Lijiang, China, pp. 152–157, 2008.
7. Lei Yu and Huan Liu. "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution", In *Proceedings of the Twentieth International Conference on Machine Learning*, pp. 856-863, 2003.
8. Isabelle Guyon, "An introduction to variable and feature selection ", *Journal of Machine Learning Research*, Vol. 3, pp. 1157-1182, 2003.
9. M.H. Aghdam, N. Ghasem-Aghaee, and M.E. Basiri, "Text feature selection using ant colony optimization", *Expert Systems with Applications*, Vol. 36, No. 3, pp.6843-6853, 2009.
10. Francois Fleuret, and Isabelle Guyon, "Fast Binary Feature Selection with Conditional Mutual Information", in *Journal of Machine Learning Research*, Vol. 5, pp. 1531-1555, 2004.
11. Chotirat Ann Ratanamahatana and Dimitrios Gunopulos, "Selective Bayesian Classifier: Feature Selection for the Naive Bayesian Classifier Using Decision Trees", in *Proceedings of the 3rd International Conference on Data Mining Methods and Databases for Engineering*, pp. 613-623, September 2002.
12. J. Zhou, R. Ng, and X. Li, "Ant colony optimization and mutual information hybrid algorithms for feature subset selection in equipment fault diagnosis", in *Proceeding of 10th International Conference on Control, Automation, Robotics and Vision*, pp.898-903, 2008.
13. Chun-Kai Zhang, and Hong Hu, "Feature selection using the hybrid of ant colony optimization and mutual information for the forecaster", in *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics*, Vol.3, pp. 1728–1732, 2005.

14. R. Jensen and Q. Shen, "Fuzzy-rough data reduction with ant colony optimization", presented at *Fuzzy Sets and Systems*, Vol. 149, pp.5-20, 2005.
15. X. Wang, J. Yang, X. Teng, W. Xia, and R. Jensen, "Feature selection based on rough sets and particle swarm optimization", presented at *Pattern Recognition Letters*, pp.459-471, 2007.
16. M. F. Zaiyadi and B. Baharudin, "A Proposed Hybrid Approach for Feature Selection in Text Document Categorization," in *World Academy of Science, Engineering and Technology*, Vol. 72, pp. 137-141, 2010.
17. R. K. Sivagaminathan and Sreeram Ramakrishnan, "A hybrid approach for feature subset selection using neural networks and ant colony optimization", *International Journal of Expert Systems with Applications*, Vol. 33, No. 1, July, 2007.
18. S. Nemat, M.E. Basiri, N. Ghasem-Aghae, and M.H. Aghdam, "A novel ACO-GA hybrid algorithm for feature selection in protein function prediction", presented at *Expert Syst. Appl.*, Vol. 36, No. 10, pp.12086-12094, 2009.
19. Hai-Hua Gao, Hui-Hua Yang, and Xing-Yu Wang, "Ant colony optimization based network intrusion feature selection and detection", in Proceedings of *International Conference on Machine Learning and Cybernetics*, Vol. 6, pp. 3871–3875, 2005.
20. Mohammad Ehsan Basiri, Nasser Ghasem-aghae, and Mehdi Hosseinzadeh Aghdam, "Using Ant Colony Optimization-Based Selected Features for Predicting Post-synaptic Activity in Proteins", *EvoBIO*, pp 12-23, 2008.
21. K. R. Robbins, W. Zhang, J. K. Bertrand and R. Rekaya, "The ant colony algorithm for feature selection in high-dimension gene expression data for disease classification", in *Math Med Biol*, Vol. 24, No. 4, pp 413-426, 2007.
22. Y. Marinakis, M. Marinaki, M. Doumpos, and C. Zopounidis, "Ant colony and particle swarm optimization for financial classification problems", presented at *Expert Syst. Appl.*, pp.10604-10611, 2009.
23. Hedar, J. Wang, and M. Fukushima, "Tabu search for attribute reduction in rough set theory", *Soft Computing*, pp.909-918, 2008.
24. H. Liu and R. Setiono, "A probabilistic approach to feature selection - A filter solution", the *13th International Conference on Machine Learning*, pp. 319-327, 1996.
25. David E. Goldberg, *Genetic algorithms in search, optimization and machine learning*, Addison-Wesley, 1989.
26. Hall, M. A., *Correlation-based Feature Subset Selection for Machine Learning*, PhD dissertation, 1999, Department of Computer Science, University of Waikato.
27. Ajay Kumar Tanwani, Jamal Afridi, M. Zubair Shafiq, and Muddassar Farooq, "Guidelines to Select Machine Learning Scheme for Classification of Biomedical Datasets", in Proceedings of *EvoBIO, LNCS*, pp. 128–139, 2009.
28. S. Hettich, and S.D. Bay, "*The UCI KDD Archive*". Irvine, CA: Department of Information Computer Science, University, California, 1996 [Online]. Available: <http://kdd.ics.uci.edu>.
29. Waseem Shahzad, Ahsan Yawar, Ejaz Ahmed, "Drug Design and Discovery using Differential Evolution", *Journal of Applied Environmental and Biological Sciences (JAEBS)*, ISSN: 2090-4274, Vol. 6(12), pp. 16-26, ISI Indexed JIF 1.72, December 2016
30. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten, "The WEKA Data Mining Software: An Update", *SIGKDD Explorations*, Vol. 11, No. 1, pp. 10-18, 2009.
31. Qinbao Song, Jingjie Ni, and Guangtao Wang, "A Fast Clustering-Based Feature Subset Selection Algorithm for High-Dimensional Data", *IEEE Transactions on Knowledge and Data Engineering* Vol. 25(1), January 2013.