A Hybrid Approach for Feature Subset Selection using Ant Colony Optimization and Multi-Classifier Ensemble

Anam Naseer

Department of Computer Science
and Information Technology
The University of Poonch Rawalakot
AJK, Pakistan

Waseem Shahzad

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

Arslan Ellahi

Department of Computer Science
MY University, Islamabad
Pakistan

Abstract—An active area of research in data mining and machine learning is dimensionality reduction. Feature subset selection is an effective technique for dimensionality reduction and an essential step in successful data mining applications. It reduces the number of features, removes irrelevant, redundant, or noisy features, and enhances the predictive capability of the classifier. It provides fast and cost-effective predictors and leading to better model comprehensibility. In this paper, we proposed a hybrid approach for feature subset selection. It is a filter based method in which a classifier ensemble is coupled with Ant colony optimization algorithm to enhance the predictive accuracy of filters. Extensive experimentation has been carried out on eleven data sets over four different classifiers. All of the data sets are available publically. We have compared our proposed method with numerous filter and wrapper based methods. Experimental results indicate that our method has remarkable ability to generate subsets with reduced number of features. Along with it, our proposed method attained higher classification accuracy.

Keywords—Ant colony optimization; predictive; classifier features selection

I. INTRODUCTION

Data mining [1] is the method of removing hidden predictive information from very large texts, databases, and web etc. Mining huge data can extrapolate data and information that can decrease the chances of fraud, improve audit reactions to potential business changes, and ensure that risks are managed in a proactive fashion [1].

Feature subset selection is mostly applied to high-dimensional data which contains a number of features. Such a large number of features make training and testing of classification methods much more difficult. Some of these features may not be important whereas some of the important features may be redundant.So feature selection technique detects most discriminating features which decreases the of data. FSS also increases the predictive accuracy by removing redundant and irrelevant features and decreases the computational time by reducing data dimensionality [2].

A. Feature Subset Selection Process

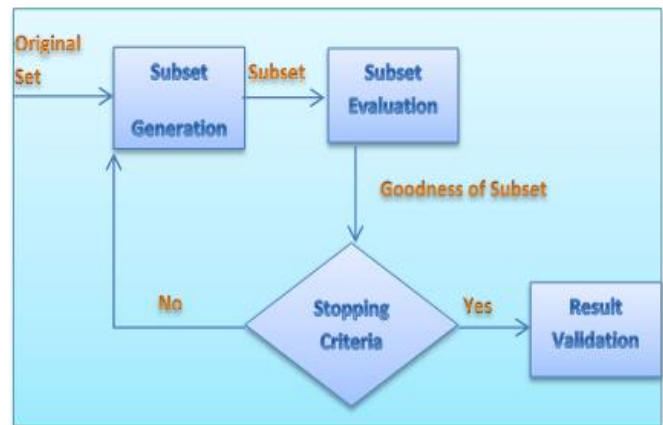Fig. 1 demonstrates the general feature subset selection process.

Fig. 1. Feature subset selection process.

B. Subset Generation

Subset generation is a heuristic search process where search space contains states, each of which specifies a candidate subset for evaluation. Two things must be determined for subset generation, Search starting point and Search strategy [3]. Search starting point can be forward, backward, bi-directional and random. In forwarding selection, thesearch starts with an empty subset and selectively adds those features that are deemed relevant, whereas, in backward elimination, the search starts with full feature subset and selectively discards those features that are useless or irrelevant. In bidirectional selection, search starts with both ends that add and removes features simultaneously and in random search, a feature subset starting point is chosen randomly without any consideration and features are added or deleted as per the requirement. Search may also start with a haphazardly selected feature subset so that it cannot stuck in the local optima [4].

A search strategy must be decided to select the candidate subsets. Different Search strategies have been explored such as exhaustive, heuristic and randomized searching algorithms [5], [6]. The time complexity is exponential for exhaustive search in terms of dimensionality and quadratic for heuristic search. In Random search, complexity can be linear to the number of iterations [5].

C. Subset Evaluation

An evaluation criterion is used to evaluate each newly generated candidate. Based on the dependency on learning algorithms that will be applied to selected feature set an evaluation criterion is categorized into two groups, one is dependent criteria second one is independent criteria [3].

Wrapper model uses dependent criteria and for feature selection it needs a learning algorithm. It applies that learning algorithm on selected subset and uses its performance to determine best feature subset, whereas filter model use independent criteria. Goodness of feature or its subset is measured with the help of significant features of the training data without linking any learning algorithm. Most common independent criteria are information theoretic measures, dependency measures, distance measures, and consistency measures.

D. Stopping Criteria

In feature subset selection a stopping criterion governs when feature selection process will stop. Some of the commonly used stopping criteria are as follows:

- Exhaustive search completes.

- A bondedshe could be used as stopping criteria where a bound can be a specified number. It can be a maximum number of iterations or minimum number of features.

- If a successive addition or removal of any feature does not affect results feature selection process could be stopped.

- If a satisfactorily good subset is selected.

E. Result Validation

At the end, results are validated by using classification error rate of classifiers as a performance indicator. Experiments are conducted to equate the classification error rate on the full set of the classifier learned on features and that trained on the selected feature subset [7], [8].

Feature subset selection techniques are of two types. Selection based reduction and transformation based reduction. Selection based reduction reduce data using original set of features whereas form new set of features by transforming an original set of features. Proposed approach is based on selection based reduction. Selection based algorithms have two categories, Filters and wrappers. Filter model feature subset evaluation methods are those that perform feature selection using some independent selection criterion, independently of any learning algorithm [9]. The computationalcost offilter-based feature subset evaluation methods are less as equated to other methods. Filter based

methods depend on the independent measures that shows the relationships among different features.

Wrapper based feature subset evaluation methods induce learning algorithms during evaluation step to measure the goodness of a selected feature subset based on the algorithm's accuracy so are computationally expensive as compared to filters. In terms of predictive or classification, accuracy wrapper methods are considered superior to filter [10].

The methodology proposed is a hybrid filter based selection method algorithm, where ACO is coupled with the Gain ratio for the first time to cope with biases of other information theoretic measures towards multi-valued attributes. A multi-classifier ensemble is used iteratively for selecting the best subset of different convergence threshold value and also for final subset selection in a novel way. Our proposed approach has used gain ratio as the subset evaluator and an ensemble of classifiers for selecting final best. If the independent measure fails to capture important features, an ensemble of classifiers captures those features. In proposed approach ensemble of classifiers are used iteratively for only selecting a final subset and not used for subset optimization as in wrapper based methods. So proposed algorithm is computationally less expensive as compared to wrapper approaches and yields higher accuracy with many reduced subsets.

The paper is organized as follows. Section II describes some of existing techniques of related to feature subset selection. A detailed description of proposed approach is described in Section III. Section IV presents the results of our experimental studies, methodology, and a comparison with other existing feature selection techniques. Section V provides Conclusion and future work directions.

II. LITERATURE REVIEW

Literature studied shows that much of the work has already been done on feature subset selection different techniques. The present techniques are grouped into two categories: filters and wrappers on the basis of search strategy and subset evaluation method [9], [10]. Some existing filter and wrapper based approaches are described here:

Feng Tan et al. presented a framework for feature subset selection based on genetic algorithm [11]. The proposed algorithm rank features using entropy and T-statistics as a ranking criterion and select features on the basis of their rank. Top-ranked features are provided as input to GA and then evaluation is done on the basis of fitness function.

Bai Jiang et al. gave hybrid algorithm for feature subset selection [12]. It is composed of two step process. Symmetric Uncertainty of the each individual features is calculated in the first step and features with SU more than the threshold value is selected and features with SU less than the threshold value are discarded. In the second step GA based searching is carried out for the left over features. To assess the quality of the feature subsets Naive Bayes classifier is used by 10 fold cross validation. For subset optimization, Naïve Bayes is also used along with symmetric uncertainty [24].

Li-Yeh Chuang et al. proposed hybrid filter-wrapper approach [13] in which an improved binary particle swarm optimization is used as a wrapper feature selection for which information gain is used as filtered model; for the performance evaluation of classification selected gene subsets were used.

Shailendra Kumar Shrivastava proposed a new ensemble technique [14]. The motivation of this approach is to enhance the performance of multiple K-nearest neighbor classifiers. This approach combines multiple K-nearestneighbor classifiers. Each classifier uses a different subset. These subsets are selected through ACO based search procedure. A subset which gets high classification accuracy on a majority of K-Nearest Neighbour classifiers is selected as a final subset.

Md.Monirul Kabir et al. proposed a hybrid technique using ant colony optimization that takes the advantage of both wrapper and filter approaches [15]. Information gain is used as filter approach and neural network as wrapper approach. This research has focused on generating reduced sized subsets. The proposed approach has used a subset size determination scheme that emphasizes not only the selection of a subset of relevant features but also on selecting features of reduced number.

Gang Wang gave a hybrid ensemble method for credit risk assessment problem [16]. In ensemble method, multiple classifiers are used to solve the same problem and also to boost many weak learners. The approach proposed in this paper works through integrating two popular ensemble strategies i.e. bagging and random subspace.

Shunmugapriya Palanisamy et al. gave a hybrid algorithm ABCE [17]. It is the combination of Artificial Bee Colony algorithm with ensemble classifier. This multi ensemble classifier is composed of support vector machine classifier, decision tree classifier, and naïve Bayes classifier. The author used ABC for generating and selecting feature. For evaluating subsets an ensemble made up of Decision Tree (DT), Naïve Bayes (NB) and Support Vector Machine (SVM) is used.

In 2012 Syed Imran Ali et al. gave a feature subset selection mechanism based on ant colony optimization algorithm and symmetric uncertainty [18]. It is a pure filter based approach which investigated the role of ACO in filter approaches. In this technique, ACO is introduced to generate optimal feature subsets. And symmetric uncertainty is used as an independent statistical measure for subset evaluation. Proposed algorithm selects fewer features and produces comparably higher accuracy.

III. PROBLEM STATEMENT

Different filter and wrapper techniques and a number of classifier ensemble methodologies for feature selection have been proposed and implemented so far in order to improve the classification accuracy. Filter approaches applied so far mostly used statistical measures to evaluate feature and to measure the goodness of feature subset. Most of the existing techniques have used information gain as a goodness measure. The main limitation of this measure is its biases towards attributes with large number of distinct values. So this drawback should be normalized.

Secondly, most of the existing techniques used learning algorithms or wrapper approach to improve classification accuracy of filter approaches. Some of these approaches have used classifier ensemble to evaluate the fitness of feature subset. These approaches increase classification accuracy but also increase computational complexity. So we will deal with two problems in this thesis: 1) We will compensate drawbacks of information gain. 2) We will try to improve classification accuracy of filter approach using a classifier ensemble without increasing computational complexity.

IV. PROPOSED SOLUTION

The proposed approach ACO-CE employs Ant colony optimization algorithm as a population based feature subset selection mechanism. Proposed approach is a hybrid filter-based feature selection, where ACO is coupled with the gain ratio for the first time as a filter solution. The gain ratio is used to normalize biases of some of the already used statistical measures towards multi-valued attributes such as information gain and mutual information etc. high split information is penalized using gain ratio an ensemble of the classifier is used iteratively over different convergence threshold values for final subset selection, not for subset optimization.

On each convergence thresh hold value, some best subsets are selected using gain ratio based fitness function and these subsets are provided to classifier ensemble and on the basis of average mean accuracy one best subset is selected and saved. Then this process is repeated by changing convergence threshold ten times and each time one best subset is selected and saved at the end from ten saved subsets on ten different convergence threshold values one subset with highest average accuracy of classifier ensemble is selected as final subset.

A. Gain Ratio

The gain ratio is normalized or compensates the biases of information gain towards attributes with a large number of values. It is basically a refinement of information gain. It takes into account split information of every attribute. Large numbers of small partitions in every split are penalized. The gain ratio is defined as:

$$\text{Gain ratio } (X) = \frac{Information gain\ (X)}{Split information (X)} \tag{1}$$

$$\text{Information\_Gain} = H(X) - H(X|Y) \tag{2}$$

Here $H(X)$ is the entropy of a random variable X and $H(X|Y)$ is the conditional entropy of X given Y. following are the equations for entropy and conditional entropy of a variable. Where split information is defined as:

$$\text{Split\_information} - \sum_{i=0}^{n} \frac{n_i}{n} log \frac{n_i}{n} \tag{3}$$

A feature that will get a high value of information gain and low value of split information will be preferred. Its goal is to maximize information gain and minimize the number of its values.

B. Ant Colony Optimization

Ant colony optimization (ACO) a population-based probabilistic Meta-Heuristics ACO is based on ants foraging behavior [19]. Foraging behavior of ant is an interesting

phenomenon by which ant colonies find the shortest path between food source and nest through indirect communication called stigmergy. Ants, like many other social insects, communicate with each other by dropping a chemical substance on their path. This chemical substance is called pheromone. It provides a positive feedback mechanism to attract other ants. Those paths which have a higher value of pheromone have a high probability of being selected. Whereas the paths that are not selected their pheromone is decreased by an evaporation process.

In ACO each ant constructs a complete solution using two things (1) node transition probability function which is based on the quantity of pheromone spread by ants and heuristic information about the importance and quality of each individual solutions and (2) already traversed solutions memory. As generations get completed, solutions constructed by each ant are evaluated using some evaluation criteria. After that pheromone evaporation and update,mechanism is also used which evaporates intensity of pheromone from the paths with low fitness value and hence discarded gradually. The ACO algorithm requires specifying the following aspects for implementation:

1) Representation of the problem domain in such a way that it lends itself to incrementally building a solution for the problem, usually in the form of a graph.

2) Node transition probability rule based on the amount of pheromone value and of the heuristic function we have employed gain ratio as a heuristic function. Following is the equation for calculating the probability of each node:

$$P_j^i = \frac{[\tau(i,j)]^\alpha [\eta(i,j)]^\beta}{\sum_{k \epsilon S}[\tau(i,j)]^\alpha [\eta(i,j)]^\beta} \qquad (4)$$

Where P_j^i is the probability of the *ith*ant to move from node i to node *j at time t.*P_i^j (t) =0 means that ants are not allowed to move to any node In the neighbor.

$[\tau(i,j)]^\alpha$is the amount of pheromone on the edge connecting i and j, where α is a constant which is used to control relative importance of pheromone information. After each iteration, this pheromone information is updated by all the ants and in some versions of ACO only best ant is allowed to update pheromone.

$[\eta(i,j)]^\beta$is the heuristic function that denotes the heuristic value of edge connecting i and j. usually, theheuristic value does not change during execution of the algorithm. In this paper we have used gain ratio to denote heuristic value. $\beta$is a constant which is used to control relative importance of heuristic value.

3) A heuristic evaluation function called fitness function dependent on the problem, which provides a goodness measurement for the different solution components. We have used fitness function is based on gain ratio to normalize the biasness of information gain and mutual information towards multi-valued attributes. Following formula being used to compute the value of the selected subset.

$$Fitness(S) = \frac{(F-S)*(\sum_{i=1}^{n}(GR))}{F} \qquad (5)$$

Where S is reduced subset selected by ACO, GR is the gain ratio of feature i in the subset S and F is the total number of features present in the dataset . It will select feature subset with high gain ratio value and with less number of features.

4) Pheromone evaporation and updating rule which takes into account the evaporation and reinforcement of the paths. Once subsets are evaluated using fitness function, pheromone trails are updated. Firstly using an evaporation rate ρ the pheromone trails on the edges are evaporated or decreased to minimize the effect of a sub-optimal feature to which the ants have previously converged. Secondly amountof pheromone on the edges is updated with amounts proportional to the fitness of the solution. Some approaches for pheromone updating allowed all the ants to update their paths according to the fitness of their solution and in some approaches only best ant is allowed to update pheromone value on its path. In this thesis former approach is used in which all the ants update their path according to the fitness of their solution.

For the pheromone evaporation and updating following equations are used.

$$\tau = (1 - \rho) * \tau \text{ where } \rho \text{ is } 0.15 \qquad (6)$$

$$\tau = \tau + (\tau * Q) \qquad (7)$$

$$\text{and } Q = \left[1 - \left(\frac{1}{1+Fitness}\right)\right] \qquad (8)$$

Equation (7)-(8): Pheromone evaporation and updating.

5) Where "Fitness" is the value of the selected subset through an independent statistical measure.

6) Stoppping/convergence criterion that decides when the algorithm terminates usually depends on maximum number of iterations.

C. Proposed ACO-CE

This is our proposed approach. In this approach ACO is used for selecting most optimal feature subsets along with Gain Ratio where Gain ratio is used as heuristic function for selecting most relevant features. Fitness function or subset evaluation is also based on gain ratio. It's a pure filter approach, along with it we have also used classifier ensemble to improve predictive performance of filter approaches comparable to the wrapper approaches.

In proposed approach first of all dataset is loaded. Once dataset is loaded, gain ratio of each feature/attribute in data set is computed. Then all the parameters of ant colony optimization algorithm are initialized. Such as number of ants, α and β values of node transition probability function , path convergence threshold value, pheromone evaporation rate ρ. and maximum number of generations. A search space is constructed that consists of nodes proportional to the number of features in the dataset. Fixed numbers of ants are generated in each iteration where each ant generates a candidate solution. After each generation, generated solutions are evaluated using a subset evaluator. Subset evaluator is based

on Gain Ratio between selected features and the class. After subset evaluation best solution is gained on the basis of maximum fitness value and is preserved. Then termination criteria of the algorithmare checked which is based on two conditions i.e. on a maximum number of generations and convergence threshold. If termination criteria are not met each ant updates its pheromone value to the quality of solution generated by each ant. Otherwise, if any termination/stopping criterion is met algorithm outputs ten best subsets.

Then these subsets are provided to classifier ensemble and these subsets are provided to classifier ensemble consisting of C4.5 decision tree classifier, Naïve Bayes, and K-Nearest Neighbor classifier.

Then one subset is selected on the basis of the highest average weighted accuracy of classifier ensemble and saved. Then again convergence threshold is checked. If it is less than 500, the whole process is repeated again. Otherwise, the algorithm stops and one best subset with the highest accuracy is selected from all saved subsets and is considered as final subset Then new ants are produced and this complete process goes iteratively till highest number of epochs is reached or the algorithm convergence to a solution.

Our proposed approach as shown in Fig. 2 is a filter approach which selects features on the basis independent gain ratio measure. So some features that might be less important in terms of independent relevance to class but for a classifier, such features could be important. Therefore, ACO-CE uses classifier ensemble on different convergence threshold values and classification accuracy of subsets is used to provide final feature subset. So our approach improves classification accuracy of filter approaches.

TABLE I.   PROPOSED ALGORITHM

1.   Start
2.   Load the dataset
3.   Compute gain ratio (heuristic value)of all the attributes/features in dataset
4.   Initialize ACO parameters
5.   Set convergence threshold(50:50:500)
6.   Do 1: maximum generations
7.   Each ant generate solutions
8.   Evaluate each solution
9.   Keep track of best solutions
10.   Check stopping criteria(yes: go to 14)
11. Update pheromone for all ants.
12. Generate new ants
13. Go to 6
14. Select 10 best subsets of converged /maximum generations
15. Run multi classifier ensemble
16. Select subset with the highest average accuracy
17. Save it
18. Check if convergence threshold >500(yes: go to 20)
19. Go to 5
20. Select one best subset from all the saved subsets that has got highest average classification accuracy of classifier ensemble
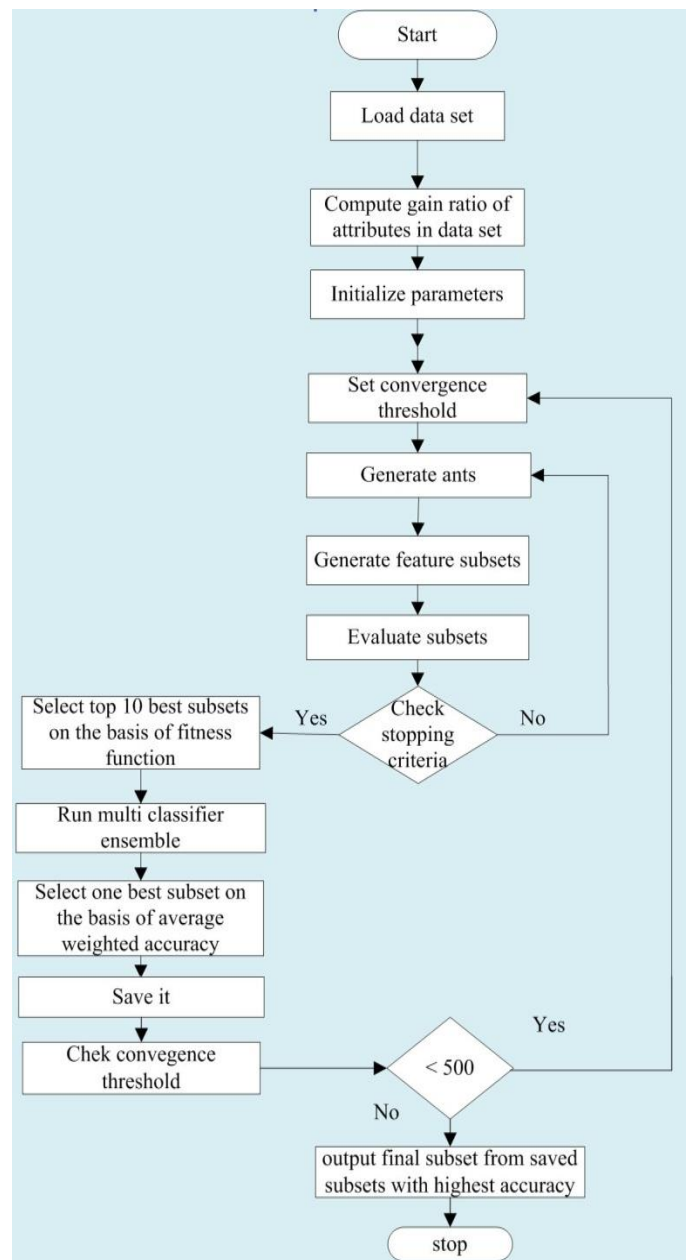21. Output best subset as final subset
22. Stop.

Fig. 2.   Flow Chart of ACO-CE.

V.   EXPERIMENTS AND RESULTS

Extensive experiments have been carried out on ACO-CE in order to find out the effectiveness of ACO-CE for feature selection. Feature selection using ACO-CE have been implemented in Matlab 2009. We have used standard parameters of ACO i.e. α = β a= 1.A number of ants in proposed ACO-CE are equal to the number of attributes in the dataset. Maximum epochs are 500 and path convergence threshold starts from 50 and stops on 500 with incrementing threshold value 10 times with 50 and Classifier ensemble consists of C4.5 decision trees, K-Nearest Neighbor and Naïve Bayes.

TABLE II. DATASET

Dataset	Total Features	Instances	Class
Iris	4	150	3
Diabetes	8	768	2
Liver disorder	6	345	2
Hepatitis	19	155	2
Colic horse	22	368	2
Ionosphere	34	351	2
Dermatology	34	366	6
Breast cancer	10	699	2
Lymphography	18	148	4
Vote	16	435	2
Labor	16	57	2

We have tested the performance of our technique with a standard implementation of three existing feature selection techniques: Genetic algorithm with consistency measure for subset evaluation [22], PSO using fuzzy rough sets as subset evaluation method [20] and ACO using fuzzy rough sets as subset evaluation method [21]. These algorithms have already been implemented in Weka [24], data mining software. Most of these algorithms are implemented by their respective authors so we have used these with their default values without doing any modification.

A. Data Sets

We have used eleven datasets as shown in Table II which are publically available in UCI machine learning repository [23]. Table I shows details about data sets used for experimentation. All of these datasets are discretizedusing weka 3.7.11 [24].

B. Results and Discussion

Table III shows the total features that were selected by our proposed methodology in comparison with the features that are selected by other eleven datasets. It is observed that ACO-CE selected a small number of features for all other datasets having more features.

TABLE III. NUMBER OF FEATURES SELECTED

Dataset	Total	PSO	GA	ACO	ACO-CE
Iris	4	4	4	4	2
Diabetes	8	8	8	8	2
Liver disorder	6	5	5	5	2
Hepatitis	19	15	10	16	7
Colic horse	22	18	9	20	5
Ionosphere	34	17	19	26	7
Dermatology	34	9	14	21	12
Breast Cancer	10	10	7	10	4
Lymphography	18	8	9	8	7
Vote	16	11	10	14	5
Labor	16	10	6	11	6

TABLE IV. CLASSIFICATION ACCURACY ON C4.5

Dataset	PSO	GA	ACO	ACO-CE
Iris	**97.33**	**97.33**	**97.33**	**97.33**
Diabetes	65.75	65.75	65.75	**68.09**
Liver disorder	57.39	57.39	57.39	**57.68**
Hepatitis	83.22	83.22	81.93	**83.87**
Colic horse	85.32	**85.86**	85.32	85.59
Ionosphere	86.32	85.75	90.88	**91.16**
Dermatology	89.89	88.79	**93.98**	**93.98**
Breast cancer	94.42	94.42	94.42	**94.70**
Lymphography	**81.75**	79.72	77.02	81.08
Vote	**96.32**	96.32	96.32	95.63
Labor	70.17	75.43	70.17	**78.94**

Table IV presents the comparison of the classification accuracy of ACO-CE with all algorithms over C4.5 classifier. Classification accuracy is checked in weka by using 10 folds cross-validation process. The Bold value in every column represents the highest value of accuracy. Proposed approach is better in 8 data sets. In iris, all the algorithms have same predictive accuracy but our approach has gained same accuracy with smaller feature set as compared to other approaches.

Table V presents the comparison of the classification accuracy of ACO-CE with all algorithms over K Nearest Neighbor classifier. Classification accuracy is checked by using 10 folds cross-validation process. Proposed approach is better in 9 d.

TABLE V. CLASSIFICATION ACCURACY ON KNN

Dataset	PSO	GA	ACO	ACO-CE
Iris	96.66	**97.33**	**97.33**	**97.33**
Diabetes	65.88	65.88	65.88	**68.09**
Liver disorder	56.23	56.23	56.23	**58.55**
Hepatitis	85.80	**87.09**	83.87	86.45
Colic horse	79.61	84.51	83.47	**85.86**
Ionosphere	82.33	83.47	85.18	**88.31**
Dermatology	87.70	87.97	**95.08**	94.53
Breast cancer	95.27	94.84	95.27	**95.99**
Lymphography	77.70	72.97	77.70	**82.43**
Vote	93.56	93.33	92.41	**96.09**
Labor	77.19	78.94	73.68	**82.45**

TABLE VI. CLASSIFICATION ACCURACY ON NAIVE BAYES

Dataset	PSO	GA	ACO	ACO-CE
Iris	**97.33**	**97.33**	**97.33**	**97.33**
Diabetes	**69.01**	**69.01**	**69.01**	68.09
Liver disorder	**58.55**	**58.55**	**58.55**	**58.55**
Hepatitis	82.58	84.51	83.87	**86.45**
Colic horse	79.61	81.35	80.16	**86.68**
Ionosphere	76.92	79.48	88.03	**91.16**
Dermatology	89.07	91.25	**97.26**	96.44
Breast cancer	**97.28**	96.85	**97.28**	**97.28**
Lymphography	80.40	77.70	82.43	**84.45**
Vote	92.87	92.41	91.03	**94.94**
Labor	84.21	82.45	84.21	**89.47**

TABLE VII.   CLASSIFICATION ACCURACY ON RIPPER

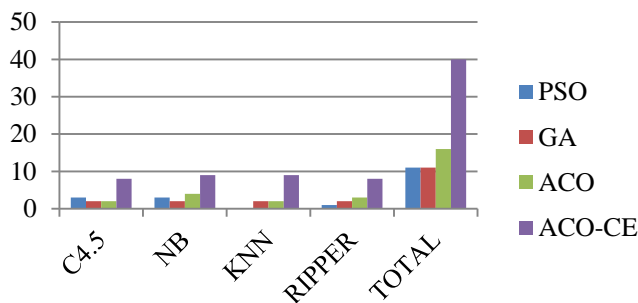| Dataset	PSO	GA	ACO	ACO-CE
Iris	**97.33**	**97.33**	**97.33**	**97.33**
Diabetes	67.83	67.83	67.83	**68.09**
Liver disorder	56.23	56.23	56.23	**58.55**
Hepatitis	83.22	83.22	80.00	**83.87**
Colic horse	85.05	**85.59**	**85.59**	85.05
Ionosphere	86.03	87.74	90.88	**91.16**
Dermatology	80.05	85.51	**93.98**	92.07
Breast cancer	94.56	94.56	94.56	**94.84**
Lymphography	76.35	**78.37**	77.70	75.67
Vote	95.63	95.17	95.63	**95.86**
Labor	84.21	80.70	82.45	**85.96**

Fig. 3.   Summarized comparison.

Classification accuracy is checked by using 10 folds cross-validation process. Proposed approach is better in 8 data sets in both Tables VI and VII. It has been observed that proposed approach has outperformed as compared to other approaches over all classifiers. Fig. 3 is the graphical representation to present the performance rate of ACO-CE and all algorithms. It has shown that our proposed approach has performed much better as compared to all algorithms over all four classifier.

VI.   CONCLUSIONS

A new hybrid method of ACO-CE has been proposed and implemented in this paper. ACO-CE is proposed by combining the ACO with a Classifier Ensemble (CE) and has been used to optimize the feature subset selection process.

Results showed that proposed approach has outperformed in terms of dimensionality reduction and classification accuracy as compared to other approaches. The Gain Ratio is used as a heuristic measure in ACO-CE which has normalized the biases of another heuristic measure towards multi-valued attributes and selected features that are highly relevant to the class. Secondly, the classifier ensemble has been used in a novel way with ACO. It checks the classification accuracy of

subsets achieved on different convergence threshold. Classifier ensemble helps to opt important features that are not selected by independent measure. We have not used classifier ensemble for optimizing results rather we have used it only for selecting a subset with the highest accuracy so our approach is not computationally costly.

Results showed that our approach has performed superior as compared to other feature selection techniques.

REFERENCES

[1]   Y.Ramamohan, K.Vasantharao, C.Kalyana, A.S.K Ratnam, "A Study of Data Mining Tools in Knowledege Discovery Process" , International Journal of Soft Computing and Engineering vol. 2, Issue 3, pp. 74-79, July 2012.

[2]   ChulminYun, Byonghwa Oh, JihoonYang andJonghio Nang, " Feature Subset Selection Basedon Bio-Inspired Algorithms", Journal of Information Science and Engineering 27, 1667-1686 , 2011.

[3]   Haun Liu, Lei Yu, "Toward Integrating Feature Selection Algorithms for Classification and Clustering", IEEE Transactions on Knowledge and Data Engineering pp. 491-502, Vol. 17, No. 4 April 2005.

[4]   J. Doak, "An Evaluation of Feature Selection Methods and Their Application to Computer Security," technical report, Univ. of California at Davis, Dept. Computer Science, 1992.

[5]   Roberto Ruiz, Jos´e C. Riquelme, and Jes´us S. Aguilar-Ruiz; "Heuristic Search over a Ranking for Feature Selection"; IWANN, LNCS 3512, pp. 742–749; 2005.

[6]   Yao-Hong Chan; "Empirical comparison of forward and backward search strategies in L-GEM based feature selection with RBFNN"; International Conference on Machine Learning and Cybernetics (ICMLC), Vol. 3, pp. 1524 - 1527 ; 2010.

[7]   I.H. Witten and E. Frank, Data Mining-Pracitcal Machine Learning Tools and Techniqueswith JAVA Implementations.Morgan Kaufmann, 2000.

[8]   I.A. Gheyas and L.S. Smith. Feature subset selection in large dimensionality domains, Pattern Recognition, 2010.

[9]   P.A. Est_evez, M. Tesmer, C.A. Perez, and J.M. Zurada. Normalized mutual information feature selection. Neural Networks, IEEE Transactions on, 20(2):pp189-201, 2009.

[10]   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 Leaning*, pp. 856-863, 2003.

[11]   Feng Tan, XuezhengFu ,Yanqing Zhang, "A genetic algorithm-based method for feature subset selection", Volume 12, Issue 2, pp. 111–120 Springer-Verlag 2007.

[12]   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.

[13]   Li-Yeh Chuang, Chao-HsuanKe, and Cheng-Hong Yang, "A Hybrid Both Filter and Wrapper Feature Selection Method for Microarray Classification", Proceedings of the International MultiConference of Engineers and Computer Scientists pp. 146-150, Vol I, 2008.

[14]   Shailendra Kumar Shrivastava, "ACO Based Feature Subset Selection for Multiple k-Nearest Neighbor Classifiers", International Journal on Computer Science and Engineering, pp. 1831-1838, Vol. 3 No. 5 May 2011.

[15]   Md.MonirulKabir, Ms. Shahjahan Kazuyuki Murase, "A New Hybrid Ant Colony Optimization Algorithm for Feature SelectionVolume 39, Issue 3, 15 February 2012, Pages 3747–3763.

[16]   Gang Wang, JianMa,"A Hybrid Ensemble Approach For Enterprise Credit Risk Assessment Based on Support Vector Machine", Volume 39, Issue 5, April 2012, pp. 5325–5331, Elsevier journal

[17]   Shunmugapriya Palanisamy1 and Kanmani S2, "Classifier Ensemble Design using Artificial Bee Colony based Feature Selection based Feature Selection" IJCSI International Journal of Computer Science Issues, pp522-529, Vol. 9, Issue 3, No 2, May 2012.

[18] Syed Imran Ali, and WaseemShahzad "Feature Subset Selection Method Based on Symmetric Uncertainty and Ant Colony Optimization," in IEEE International Conference on Emerging Technologies (ICET), pp. 1-6, October 2012.

[19] Hazem Ahmed," Swarm Intelligence Concepts Models and Applications", Technical Report, School of Computing Queen's University Kingston, Ontario Canada, Feburary 2012.

[20] 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, 2006.

[21] 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.

[22] David E. Goldberg, *Genetic algorithms in search, optimization and machine learning*, Addison-Wesley, 1989.

[23] S. Hettich, and S.D. Bay, *"The UCI KDD Archive"*. Irvine, CA: Dept. Inf. Comput. Sci., Univ. California, 1996 [Online]. Available: http:// kdd.ics.uci.edu.

[24] 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.