

# **A Feature Subset Selection Method based on Conditional Mutual Information and Ant Colony Optimization**

Syed Imran Ali

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

Waseem Shahzad, Ph.D

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

## **ABSTRACT**

Feature subset selection is one of the important problems in a number of fields namely data mining, machine learning, pattern recognition. It refers to the problem of opting for useful features that are neither irrelevant nor redundant. Since most of the data acquired through different sources are not in a proper shape to mine useful patterns from it therefore feature selection is applied over this data to filter out useless features. But since feature selection is a combinatorial optimization problem therefore exhaustively generating and evaluating all possible subsets is intractable in terms of computational cost, memory usage and processing time. Hence such a mechanism is required that intelligently searches for useful set of features in a polynomial time. In this study a feature subset selection algorithm based on conditional mutual information and ant colony optimization is proposed. The proposed method is a pure filter based feature subset selection technique that incurs less computational cost and proficient in terms of classification accuracy. Moreover, along with high accuracy it opts for less number of features. Extensive experimentation is performed based on thirteen benchmark datasets over a number of well known classification algorithms. Empirical results endorse efficiency and effectiveness of the proposed method.

## **Keywords**

Feature Subset Selection; Conditional Mutual Information, Symmetric Uncertainty; Ant Colony Optimization, Classification.

## **1. INTRODUCTION**

In past data was transformed into knowledge manually through data analysis and interpretation. This manual data analysis was highly subjective, slow and costly. But as data generation and recording escalated considerably, manual data analysis became tedious and impractical in many domains. This motivated the need for an efficient and automated knowledge discovery process. It is estimated that information doubles every 20 months in the world. This explosion of data is due to the digital acquisition, generation, storage and retrieval of data. Since data are being generated at a faster pace therefore huge amount of data are not being analyzed due to the shortage of efficient data analysis mechanisms. Moreover, it is very difficult to analyze data in its entirety. Raw data need to be processed in such a way that it helps in analysis and transformation into a more meaningful form i.e. knowledge. In order to analyze data into automatic or semi-automatic manner “Knowledge Discovery in Databases” (KDD) is formulated. Data Reduction is one of the key elements of KDD process. Since data are not gathered with some specific purpose in mind. Hence, these datasets may contain redundant and irrelevant attributes. Inclusion of these attributes can be deterrent to the knowledge discovery, and mislead the process. Moreover, processing time required to analyze these features can increase the overall processing cost.

Feature Subset Selection (FSS) is one of the key types of data reduction. The main objective of this step is to find useful features that represent the data and remove those features that are either irrelevant or redundant.

A useful feature is neither irrelevant nor redundant. Where an irrelevant feature doesn't provide any useful information to predict the target concept and redundant feature doesn't add extra information that might be useful to predict the target concept [1]. FSS helps in a number of ways e.g. it reduces useless features to save computing time and data storage, relevant features improves predictive performance and precludes over-fitting, provides more appropriate description of the target concept. Feature Selection is a combinatorial optimization problem where a feature set containing N number of features can be too large, for exhaustive searching, where N is any integer value.

There are two main categories of selection based algorithms, i.e. filter based methods and wrapper based methods. [1, 2, 3, 4].

Filter based methods are those that perform FSS independently of any learning algorithm using some surrogate classifier measures e.g. statistical measures. Filter based methods are light weight and their running computational cost is quite low as compared to other methods. Filter based methods rely heavily on the statistical measures that exploit the inter-feature relationship among different features.

Wrapper based methods employ learning algorithms to evaluate the usefulness of a selected feature subset moreover they use search methods e.g. population-based searching, to explore feasible feature subset space. Since wrapper methods are computationally expensive and each new subset of features needs to construct a hypothesis. Therefore wrapper methods are considered superior to filter methods in terms of predictive accuracy [5].

## **2. RELATED WORK**

A number of approaches to feature subset selection have been proposed in the literature, a few of them that are based on computational intelligence are referred here.

Filter based methods employ feature ranking based on some similarity metric. There are a number of similarity metrics proposed in literature. These metrics are categorized into four sub-types i.e. distance measures, information theory measures, dependency measures, and consistency based measures [6]. Wrapper based feature selection employs classification accuracy of some learning algorithm [7]. Bai-Ning Jiang et al. [8] proposed a hybrid feature selection algorithm. It is a two step process. In the first step Symmetric Uncertainty (SU) of the individual features is calculated and those features that have less SU than the threshold value are discarded. In the second

step Genetic Algorithm based searching is employed on the left over features. Naive Bayes classifier and SU are used to evaluate goodness of the feature subsets by 10 fold-cross validation. J. Zhou et al. [9] proposed an Ant Colony Optimization and Mutual Information based feature selection for equipment fault diagnosis. Regression estimation model and mean squared error are used for the feature subset evaluation. Both mutual information and classification accuracy are used for subset optimization.

Chun-Kai Zhang et al. [10] proposed Ant Colony Optimization and Mutual Information based hybrid feature subset selection algorithm for weather forecasting. R. Jenses et al. [11] proposed a fuzzy-rough data reduction using any colony optimization and C4.5. Where ACO is used to find fuzzy-rough set reducts. Where a fuzzy-rough based dependency measure is defined as the stopping criterion.

Xiangyang Wang et al. [12] proposed a rough set based feature selection using Particle Swarm Optimization. Since standard PSO is used for continuous optimization problems therefore it is formulated into Binary PSO. LEM2 is used for rule induction and subset optimization along with size of the reduct. A-R Hedar et al. [13] proposed a Tabu search for attribute reduction in rough set theory. H. Liu et al. [14] proposed a consistency based feature selection mechanism. It evaluates the worth of a subset of attributes by the level of consistency in the class values when the training instances are projected onto the subset of attributes. Consistency of any subset can never be lower than that of the full set of attributes; hence the usual practice is to use this subset evaluator in conjunction with a Random or Exhaustive search which looks for the smallest subset with consistency equal to that of the full set of attributes. Genetic Algorithm is used for feature subset generation [15] along with aforementioned consistency based subset evaluation.

### 3. PROPOSED METHOD

In this section the proposed method, Ant Colony Optimization with Conditional Mutual Information (ACO-CMI), is elaborated. Since feature subset selection is a combinatorial optimization problem therefore ACO is well suited for this task. The proposed method employs ACO as a population-based feature subset selection mechanism where selected subsets are evaluated on the basis of an information theoretic measure. Population based feature subset selection is predominately used in the wrapper based approach where a learning algorithm is used to provide goodness measure of the selected features whereas filter based approaches relied heavily on the feature ranking and selective sampling approaches for selecting a final feature subset. This study investigates the role of meta-heuristics i.e. ACO, in filter based feature selection for the adaptive selection of a feature subset.

#### 3.1 Symmetric Uncertainty and Conditional Mutual Information

As mentioned above for a filter based feature selection mechanism there needs to have some surrogate classifier measure that weighs the worth a selected feature subset. The proposed method has used information theoretic measures called symmetric uncertainty (SU) and conditional mutual information (CMI) in order to evaluate the worth of constructed solutions. There are a number of benefits for using SU i.e. it is symmetric in nature therefore  $SU(i,j)$  is same as that of  $SU(j,i)$  hence it reduces the number of comparisons required, it is not influenced by multi-valued attributes as that is in the case of information gain, and its values are normalized. Following is the equation for symmetric uncertainty.

$$SU(X, Y) = 2 \left[ \frac{IG(X|Y)}{H(X)+H(Y)} \right] \quad (1)$$

Where  $IG(X|Y)$  is the information gain of feature X, that is an independent attribute and Y is the class attribute.  $H(X)$  is the entropy of feature X and  $H(Y)$  is the entropy of feature Y. Information gain has a desired property, i.e. it is symmetric. The amount of Information given by a feature Y about another feature X is effectively the same as that of the information given of feature X and the feature Y.

CMI is a paired-wise correlation measure based on the same principles as that of information gain. It measures correlation between two independent features, when the value of a third feature is known. CMI is used to evaluate inter-feature correlation within a selected subset. This measure is minimized in order to reduce redundancy among the selected features. Calculating CMI between X and Y when Z is given is as follows:

$$CMI(X; Y|Z) = H(XYZ) - H(XZ) - H(YZ) + H(Z) \quad (2)$$

#### 3.2 Ant Colony Optimization

Ant Colony Optimization was proposed by Marco Dorigon in his seminal work [16]. Ants are simple agents that construct a candidate solution iteratively. An ant represents a complete solution. In each generation there are a number of ants. Each ant constructs a solution probabilistically based on the pheromone value on the path and the quality of the solution. After a generation, all the solutions are evaluated, and the traversed paths are updated. Gradually in subsequent generations solutions get more optimized.

Following are the main considerations for applying ACO to an optimization problem.

##### 3.2.1 Graph Representation

Feature subset space is represented graphically. Each node is required to represent a feature. Since there is no ordering significance of the attributes collected in a feature set therefore configuration of the graph is based on mesh topology i.e. each node is fully connected with every other node present in the graph. Terminal represents end of the traversal of the graph. Since each ant represents an individual solution therefore it can terminate before traversing the complete tour of all the nodes. Each link is associated with two values i.e. pheromone concentration and heuristic desirability. Following is the figure for search space designed for feature subset selection problem.

##### 3.2.2 Feedback process and heuristic desirability

All the experience of the preceding generation of ants is approximately stored on the graph in the form of pheromone concentration. This process of transferring experience to the subsequent generations of ants is called positive feedback process. Another important consideration for the ants in order to decided their traversing rout is the desirability of the nodes. This information is also stored on the associated links between nodes. Symmetric uncertainty is used as heuristic.

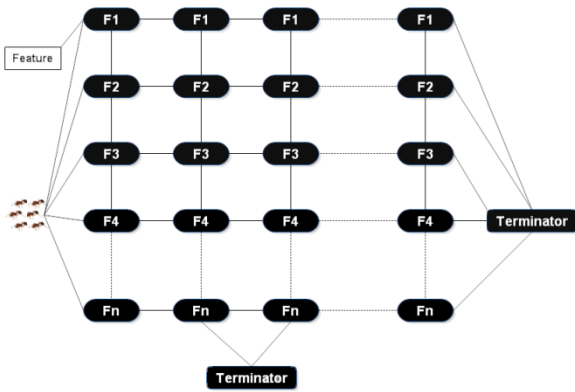


Figure 1: Search space for feature subset selection problem.

Following equation is used to calculate the traversing probability of an ant from node  $i$  to node  $j$ .

$$P_{ij} = \frac{[\tau(i,j)]^\alpha [\eta(i,j)]^\beta}{\sum_{k \in S} [\tau(i,k)]^\alpha [\eta(i,k)]^\beta} \quad (3)$$

Where  $\alpha$  and  $\beta$  are scaling factors,  $[\tau(i,j)]$  represents the concentration of pheromone and  $[\eta(i,j)]$  represents heuristic desirability of node  $j$  and  $k$  is the subset of those nodes that are directly reachable from the node  $i$ .

Once all the ants construct their solution, their respective solutions are tested against a fitness function. All the ants are required to reinforce their path for the subsequent generations according to the quality of their respective solutions. Following equation is used to evaporate pheromone concentration value. Where  $\rho = 0.15$ .

$$\tau_{ij}(t) = (1 - \rho) \cdot \tau_{ij}(t) \quad (4)$$

And following formula is used to update the pheromone value according to the new fitness values received from each ant's solution.

$$\tau_{ij}(t+1) = \tau_{ij}(t) + \left[1 - \left(\frac{1}{1+Q}\right)\right] \cdot \tau_{ij}(t) \quad (5)$$

Where "Q" is the quality of the solution and quality is based on the fitness value of the solution.

### 3.2.3 Solution construction and constraint satisfaction

In each generation a number of solutions are constructed. These solutions are parallel to the number of ants in a generation. Normally this number is fixed for each generation. Each ant traverses through the graph and records all the traversed nodes into its local memory. Once an ant completes its tour, its collected features are evaluated. This process of evaluating the feasibility of constructed solutions is called constraint satisfaction.

Following is the complete algorithm of the proposed method.

- ```

0: Start.
1: Load Dataset.
2: Compute SU of each feature.
3: Initialize ACO parameters.
4: Construct ACO Search Space.
5: Do 1:max
  5.1: Each Ant constructs its solution.
  5.2: Evaluate all the solutions.
  5.3: Keep track of Best Solution.
  5.4: Check Stopping Criteria (yes: Goto 10)
6: Repeat.
7: Update Pheromone Trail of each ant.
8: Generate new ants.
9: Goto 5.
10: Select subset the highest fitness value.
11: Output the selected subset.
12: Stop.
  
```

As it is mentioned, once a dataset is loaded, first operation is to measure the symmetric uncertainty value of each attribute. Subsequently, an ACO compliant search space is constructed along with initialization of parameters. Each ant in a generation constructs a solution. And its worth is evaluated through a fitness function. Since it is possible to converge at a pre-mature solution, all the ants are used to update pheromone trail of their paths according to their respective fitness values. If best solution of a generation is not changed after a number of consecutive generations then loop terminates and best solution so far obtained is produced. Following fitness function is used to evaluate the worth of a selected feature subset.

$$Merit(S) = \frac{[N-s] \sum_{i=1}^N |C_{ic}| - \alpha \sum_{i=1}^N \sum_{j=1}^N |C_{ij}|}{N} \quad (6)$$

Where "S" is a constructed solution comprised of a local memory of an ant, "N" is total number of features in a dataset, "s" is the number of features in solution "S",  $|C_{ic}|$  is the correlation between a feature "i" and the class "c", it is computed using  $SU(i;c)$  and one by one symmetric uncertainty of all the features in the selected subset is computed and aggregated.  $|C_{ij}|$  is the intra- feature correlation; it is computed using conditional mutual information. Moreover, " $\alpha$ " is the scaling factor that determines the importance of intra-feature correlation. After each generation pheromone concentration on the selected paths is updated according to the quality of the solution. Once maximum epoch is reached or best solution is found then search terminates. Now the selected subset has highest fitness value as compared with other subsets of its generations. And this selected subset is treated as a final subset. In most of the filter based feature selection methods statistical measures are used to rank the features in some order. Then a threshold value is required to pick the top performing features. Since neither all the datasets are of equal size nor each feature can be distinctly measures as useful or useless therefore a fixed value of threshold e.g.  $\theta = 0.5$  where all the features that measured greater than or equal to 0.5 are to be opted, is not a very good criterion for the selection of features in a selected feature subset. Since value of  $\theta$  may differ from dataset to dataset. And if 10%, 20%, or 25%, etc, of the top performing features are selected, still it will not provide an optimal feature subset since there is a chance that top features may only comprise of 5% or 30% of the total features. In former case, a number of selected features may not be required and in latter case some of the vital features may be ignored that hinder the predictive classification accuracy of the classifier. The

proposed technique searches for an optimal subset amongst the generated subsets. Hence, nature of the dataset dictates how many features are to be accounted for in a final subset. Moreover, the proposed method prefers subsets with small number of features.

#### 4. RESULTS AND DISCUSSIONS

In this study thirteen datasets are obtained from the UCI repository [17] for the experiments. The main characteristics of these datasets are shown in Table I. All the datasets are discretized and missing values are imputed. Datasets comprised of binary class to multi-class problems with varying number of features. All the experiments are performed using ten-fold cross validation method i.e. predictive classification values are averaged over ten independent runs of a classifier using ten disjoint training and testing samples out of the main dataset. The numbers of ants in a generation are 20,  $\alpha$  and  $\beta$  values are 1 and pheromone evaporation rate  $\rho$  is 0.15. Proposed method is compared with some feature selection methods based on computational intelligence. The proposed method is compared with the standard implementation of Ant Colony Optimization using Fuzzy Rough Sets (ACO-FRS), Particle Swarm Optimization using Fuzzy Rough Sets (PSO-FRS), Tabu search using Fuzzy Rough Sets (Tabu-FRS), Genetic Algorithm using Consistency measure (GA-Cons). These aforementioned methods are implemented in a machine learning open source software package called WEKA [18]. All the standard parameters of the above mentioned algorithms are used. Explorer option of WEKA is used for classifiers' results compilation. These feature selection methods are evaluated over a well known classifier i.e. K-Nearest Neighbor with  $K = 5$  in Table II. Bold values represent highest classification value for that respective dataset. As it can be seen that the proposed method out-performed all the other methods. The proposed method achieved better or comparable results on eight of the datasets. KNN is an instance based classifier method. Hence, any irrelevance in the feature due to noise or mediocre nature of the feature may affect the predictive classification of KNN classifier. With the reduction in the number of features so the reduction in the complexity of the feature space since every feature serves as a dimension over the feature space in KNN.

Table III presents comparison results of the proposed method with other feature selection methods over Support Vector Machine. And the proposed method performed better or comparable over seven datasets. Although GA-Cons. also performed better in a number of datasets but it opted for more features at an average than the proposed method.

In Table IV the proposed method along with other feature selection methods is compared over a well known classifier called Ripper. The proposed method outperformed other feature selection techniques over seven datasets. Ripper utilizes a rule induction algorithm for the classification of a dataset. Rules are built along with their respective conditions as long as no negative coverage takes place. Hence, only those rules are built that perform well over the dataset. The proposed method not only selected the relevant features it also helped reduce the clutter and uncertainty in the dataset, that in turn may have helped Ripper to build more efficient rules.

Lung Cancer dataset is very challenging one since it has more number of features than the instances. The proposed method successfully dealt with this issue, and retained higher accuracy with small number of features. Table V mentions number of features selected by the proposed method along with other feature selection methods over thirteen datasets. As it can be

observed that at an average the proposed method achieved very encouraging results.

**Table 1. Characteristic of the Datasets.**

| Dataset          | Features | Instances | Classes |
|------------------|----------|-----------|---------|
| Iris             | 4        | 150       | 3       |
| Liver Disorder   | 6        | 345       | 2       |
| Diabetes         | 8        | 768       | 2       |
| Breast Cancer- W | 9        | 699       | 2       |
| Vote             | 16       | 435       | 2       |
| Labor            | 16       | 57        | 2       |
| Hepatitis        | 19       | 155       | 2       |
| Colic-Horse      | 22       | 368       | 2       |
| Ionosphere       | 34       | 351       | 2       |
| Lymph            | 18       | 148       | 4       |
| Dermatology      | 34       | 366       | 6       |
| Lung Cancer      | 56       | 32        | 3       |
| Audiology        | 69       | 226       | 24      |

**Table 2. Classification results on K Nearest Neighbor classifier using 10 Fold Cross Validation.**

| Dataset          | All          | ACO - CMI    | ACO -FRS     | PSO - FRS    | Tabu -FRS    | GA- Cons     |
|------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Iris             | 95.33        | <b>97.33</b> | 95.33        | 95.33        | 78.66        | <b>97.33</b> |
| Liver Disorder   | 56.23        | <b>58.55</b> | 57.39        | 57.39        | 57.39        | 57.39        |
| Diabetes         | 65.75        | <b>68.09</b> | 65.75        | 65.75        | 67.57        | 65.75        |
| Breast Cancer- W | 95.42        | 94.99        | <b>95.56</b> | 93.84        | 95.42        | 95.13        |
| Vote             | 93.10        | <b>94.94</b> | 93.10        | 92.87        | 93.33        | 94.48        |
| Labor            | 85.96        | 85.96        | <b>89.47</b> | 82.45        | 78.94        | 78.94        |
| Hepatitis        | 82.58        | <b>83.87</b> | <b>83.87</b> | <b>83.87</b> | 81.93        | 82.58        |
| Colic-Horse      | 77.98        | 81.82        | <b>83.69</b> | 81.52        | 81.79        | 81.52        |
| Ionosphere       | 84.90        | <b>86.03</b> | 85.47        | 84.33        | 82.62        | 82.62        |
| Lymph            | <b>82.43</b> | <b>82.43</b> | 77.02        | 79.02        | 79.02        | 77.70        |
| Audiology        | 60.61        | 68.14        | 67.69        | 64.60        | <b>69.02</b> | 68.58        |
| Dermatology      | <b>95.62</b> | 88           | 90.71        | 76.77        | 89.89        | 86.61        |
| Lung Cancer      | 40.62        | <b>53.12</b> | <b>53.12</b> | 34.37        | 50           | 34.37        |

**Table 3. Classification results on Support Vector Machine classifier using 10 Fold Cross Validation.**

| Dataset          | All          | ACO - CMI    | ACO -FRS     | PSO - FRS    | Tabu -FRS    | GA - Cons    |
|------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Iris             | 94           | <b>97.33</b> | 94           | 94           | 95.33        | <b>97.33</b> |
| Liver Disorder   | 57.68        | <b>58.84</b> | 57.39        | 57.39        | 57.39        | <b>58.84</b> |
| Diabetes         | 66.66        | <b>68.09</b> | 66.66        | 66.66        | 67.96        | 66.66        |
| Breast Cancer- W | 95.56        | <b>95.70</b> | 93.27        | 93.84        | 95.56        | 94.70        |
| Vote             | <b>95.63</b> | <b>95.63</b> | <b>95.63</b> | <b>95.63</b> | <b>95.63</b> | <b>95.63</b> |
| Labor            | 85.96        | 85.96        | 73.68        | <b>91.22</b> | 87.71        | <b>91.22</b> |
| Hepatitis        | 82.58        | <b>86.16</b> | 82.58        | 82.58        | 83.22        | 85.80        |

|             |              |              |              |              |              |       |
|-------------|--------------|--------------|--------------|--------------|--------------|-------|
| Colic-Horse | 83.69        | 80.97        | 83.96        | <b>84.51</b> | 83.96        | 83.15 |
| Ionosphere  | 88.31        | <b>90.60</b> | <b>90.60</b> | 88.60        | 84.90        | 85.75 |
| Lymph       | <b>86.48</b> | 83.78        | <b>86.48</b> | 77.70        | 85.81        | 82.43 |
| Audiology   | <b>81.85</b> | 69.02        | 58.40        | 58.84        | 62.38        | 56.35 |
| Dermatology | 95.35        | 92.07        | <b>97.26</b> | 84.15        | 95.90        | 94.53 |
| Lung Cancer | 40.62        | 56.25        | 43.75        | 43.75        | <b>65.62</b> | 18.75 |

**Table 4. Classification results on Ripper classifier using 10 Fold Cross Validation.**

| Dataset          | All          | ACO - CMI    | ACO - FRS    | PSO - FRS    | Tabu-FRS     | GA - Cons    |
|------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Iris             | 96.66        | <b>97.33</b> | 96.66        | 96.66        | 96.66        | <b>97.33</b> |
| Liver Disorder   | 56.52        | <b>58.55</b> | 56.23        | 56.23        | 56.23        | 56.23        |
| Diabetes         | 67.83        | <b>68.09</b> | 67.83        | 67.83        | 67.83        | 67.83        |
| Breast Cancer- W | <b>94.70</b> | <b>94.70</b> | 92.99        | 94.13        | <b>94.70</b> | 94.56        |
| Vote             | 96.09        | <b>96.63</b> | 95.17        | 95.40        | 95.40        | 95.86        |
| Labor            | <b>91.22</b> | 85.96        | 78.94        | <b>91.22</b> | <b>91.22</b> | <b>91.22</b> |
| Hepatitis        | 81.93        | 80.64        | <b>83.22</b> | 82.58        | <b>83.22</b> | 81.93        |
| Colic-Horse      | 83.42        | 80.70        | 85.05        | 84.51        | <b>85.32</b> | 84.51        |
| Ionosphere       | 90.88        | <b>91.45</b> | 90.88        | 86.89        | 87.46        | 87.46        |
| Lymph            | 75           | <b>78.37</b> | 77.70        | 76           | 75           | 75.67        |
| Audiology        | 70.79        | 68.58        | <b>71.68</b> | 64.15        | 65.48        | 67.69        |
| Dermatology      | 86.88        | 86.88        | <b>92.62</b> | 75.13        | 88.25        | 86.33        |
| Lung Cancer      | 46.87        | 46.87        | 40.62        | 46.87        | <b>59.37</b> | 46.87        |

**Table 5. Number of selected features in each dataset by different feature selection methods.**

| Dataset          | ACO - CMI | ACO - FRS | PSO - FRS | Tabu - FRS | GA - Cons |
|------------------|-----------|-----------|-----------|------------|-----------|
| Iris             | <b>2</b>  | 4         | 4         | 3          | <b>2</b>  |
| Liver Disorder   | <b>2</b>  | 5         | 5         | 5          | 5         |
| Diabetes         | <b>2</b>  | 8         | 8         | 6          | 8         |
| Breast Cancer- W | <b>4</b>  | <b>4</b>  | 5         | 9          | 7         |
| Vote             | <b>5</b>  | 14        | 12        | 12         | 11        |
| Labor            | <b>5</b>  | <b>6</b>  | <b>6</b>  | 7          | 7         |
| Hepatitis        | <b>7</b>  | 14        | 13        | 13         | 12        |
| Colic-Horse      | <b>4</b>  | 13        | 16        | 11         | 12        |
| Ionosphere       | <b>9</b>  | 26        | 19        | 16         | 21        |
| Lymph            | 8         | 8         | <b>7</b>  | <b>7</b>   | 9         |
| Audiology        | 14        | 38        | 22        | <b>13</b>  | 35        |
| Dermatology      | 18        | 20        | <b>10</b> | 11         | 13        |
| Lung Cancer      | <b>5</b>  | <b>5</b>  | <b>5</b>  | <b>5</b>   | 13        |

## 5. CONCLUSION

When large amount of data are accumulated it becomes a daunting task to mine meaningful patterns. Noise in the data also undermines the capabilities of the leading state of the art classification methods. In this regard feature subset selection plays a vital role. This study has introduced a very efficient method of feature selection. The proposed method addresses the issue of threshold value selection in filter method by introducing subset generation capability in the filter methods. Experiments regarding the proposed method yielded very encouraging results. A number of feature selection methods are compared with the proposed method. Simplicity and ease of implementation along with predicative accuracy of the selected features are the strong motivations for the proposed method.

## 6. REFERENCES

- [1] M.Dash, H. Liu, "Feature Selection for Classification", *Intelligent Data Analysis (IDA)*, Vol. 1, No. 3, pp. 131-156, 1997.
- [2] Richard Jensen, *Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches*, Wiley-IEEE Press 2008.
- [3] Y. Saeyns, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics", *Bioinformatics*, Vol. 23, No. 19, pp. 2507-2517, 2007.
- [4] Isabelle Guyon, "An introduction to variable and feature selection", *Journal of Machine Learning Research*, Vol. 3, pp. 1157-1182, 2003.
- [5] 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.
- [6] M. Dash and H. Liu, "Consistency-based search in feature selection", presented at *Artificial Intelligence*, Vol. 151, No. 1-2, pp.155-176, 2003.
- [7] Ron Kohavi and Dan Sommerfield, " Feature subset selection using the wrapper method: Over-fitting and Dynamic Search Space Technology", in Proceedings of *2nd Int. Conf. on Knowledge Discovery and Data Mining*, 1995.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] A. Hedar, J. Wang, and M. Fukushima, "Tabu search for attribute reduction in rough set theory", *Soft Computing*, pp.909-918, 2008.
- [14] 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.
- [15] David E. Goldberg, *Genetic algorithms in search, optimization and machine learning*, Addison-Wesley, 1989.
- [16] M. Dorigo, *Optimization, Learning and Natural Algorithms*, PhD thesis, Politecnico di Milano, Italy, 1992.
- [17] 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>.
- [18] 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