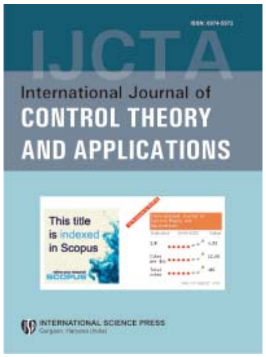


International Journal of Control Theory and Applications

ISSN : 0974–5572

© International Science Press

Volume 10 • Number 23 • 2017

A Clustering and Genetic Algorithm based Feature Selection (CLUST-GA-FS) for High Dimensional Data

S. DeepaLakshmi[a] and T. Velmurugan[b]

[a]Research Scholar, Bharathiar University, Coimbatore, India.

E-mail: deepa.dgvc@gmail.com,

[b]Associate Professor, PG and Research Department of Computer Science, D.G.Vaishnav College, Chennai, India.

E-mail: velmurugan_dgvc@yahoo.co.in

Abstract : Feature selection involves finding the relevant and useful features that provide enhanced classification results as the original set of features. Feature selection in a high dimensional data is even more significant as it finds the optimal set of features from a vast number of features. A clustering and genetic algorithm based feature selection (CLUST-GA-FS) algorithm is proposed and evaluated experimentally in this research work. The proposed algorithm has three stages namely irrelevant feature removal, redundant feature removal, and optimal feature generation. Mutual information filter method is used to remove the irrelevant features in the first stage. The features are grouped into clusters by using graph based clustering method in the second stage. Also, the redundant features are removed and a representative feature is selected from each cluster in the second stage. In the final stage, the genetic algorithm is applied to the representative set of features and an optimal set of features is generated. The effectiveness of the proposed method is ensured using the well known classifiers, namely Naïve Bayes, Multilayer Perceptron and Adaboost. The CLUST-GA-FS method is implemented on four publicly available high dimensional medical dataset and text data. The result demonstrates that CLUST-GA-FS improves the performance of the classifiers.

Keywords: *Feature Selection, Mutual Information, Feature Clustering, Genetic Algorithm.*

1. INTRODUCTION

Data Mining is the task of identifying interesting patterns from large amounts of data. Mining high dimensional data has some challenges including the curse of dimensionality and the meaningfulness of the similarity measure in the high dimensional space. High dimensional data contains redundant and irrelevant features that slow down the mining process, reduce the accuracy of data mining algorithms, lead to problems in retrieval and storage and it is hard to interpret [1]. Feature selection or attribute selection is the method of selecting features that are relevant from a large number of features. Some of the benefits of feature selection are facilitating data visualization and data understanding, reducing the measurement and storage requirements, reducing training and utilization times of the final model, defying the curse of dimensionality, to improve prediction and performance.

Feature selection is divided into filters, wrappers and embedded methods. Filter feature selection methods rank features based on the score obtained by applying a statistical measure to each feature [2]. Common filter methods used are correlation coefficient scores, chi squared test, mutual information, information gain etc. The correlation coefficient is the statistical measure of linear relationship between two variables or data values. Pearson correlation coefficient, spearman rank correlation, kendall tau rank correlation etc. are the types of the correlation coefficient. Mutual information measures the mutual dependence between two variables. In cluster analysis, graph based methods are used in many applications. In graph based clustering methods, similar data are represented in a graph. The highly connected sub graph forms the clusters [3]. The elements in a cluster are highly similar to each other. Wrapper feature selection methods select a subset of features using the learning algorithm as part of the evaluation function. Some of the wrapper methods are the genetic algorithm, simulated annealing, ants colony etc. Genetic Algorithm is a search technique used to find true or approximate solutions to optimization and search problem. Clustering and genetic algorithm based feature selection (CLUST-GA-FS) works in three steps. In the first step, irrelevant features are removed using filter method. The relevant features are grouped into clusters using graph based clustering methods in the second step. Also, from each cluster, one strong representative feature is selected. In the third step, the set of features is given as input to a genetic algorithm to generate a set of optimal features. The rest of the paper is organized as follows: in section 2, the literature survey is elaborated, and in section 3, the proposed feature selection algorithm is presented. In section 4, experimental results that support the proposed algorithm is reported, and section 5 summarizes the research contribution and concludes.

2. LITERATURE SURVEY

The curse of high dimensionality has made classification a difficult task. A number of approaches to feature subset selection have been proposed in the literature of which only a few references are discussed here. Mutual information based multi-label feature selection was proposed using interaction method which is able to measure dependencies among multiple variables [4]. A feature selection method called MINT based on max-relevance and min-redundancy was proposed and applied to genetic trait prediction problem [5]. The mutual information based feature selection method is transformed into a global optimization problem which is based on single-objective and multi-objective optimization and was applied to both synthetic and real datasets [6]. mRMR filter was used in the SVM-RFE method to improve the classification performance [7]. A Hierarchical feature clustering algorithm was proposed in which several consecutive features are grouped into one cluster, and mutual information is used to predict the most relevant ones from the clusters [8].

A new information-theoretic divisive algorithm for clustering words was proposed, and it was applied to text classification [9]. Graph based clustering method to cluster proteins was proposed by Kawaji et al.[10]. A hybrid feature selection algorithm which uses mutual information filter method to find the relevant and remove redundant features and a wrapper method to find the best feature subset was proposed by Gert van Dijck et al.[11]. A standard genetic algorithm with rank based selection in which the fitness function combines the accuracy of classification function and the cost of performing the classification was proposed by Jihoon Yang et al.[12]. Lu et al. proposed a dynamic genetic algorithm based feature selection in which the statistical features were extracted using wavelet transform [13]. Fisher ratio was used to evaluate the features and the genetic algorithm was used to find the optimal features.

Zexuan Zhu et.al proposed a hybrid wrapper-filter method which incorporates filter method in the genetic algorithm [14]. A wrapper method based on binary differential evolution was combined with rank based filter method [15]. Sebban et al. proposed a hybrid method of feature selection which constructs a minimum spanning tree which is replaced by 1 nearest-neighbour graph [16]. A hybrid filter-wrapper method which combines spectral feature selection using laplacian score and a modified calinski-harabasz index was proposed by Solorio et al. [17]. Peng Wen and Lie Wen-xia proposed a method based on niche genetic algorithm and minimum spanning tree for substation planning [18]. A multi-operator genetic algorithm of the genotype-phenotype class was proposed by Contreras-boltan et al.[19]. The genotype represents selected vertex in clusters and phenotype is a minimum spanning tree. A revised genetic algorithm based on spanning tree was proposed by Wang et al.[20]. A fast clustering based feature selection algorithm was proposed by Quinbao song et al. Symmetric Uncertainty is used to remove irrelevant features and graph theoretic clustering is used to cluster the features [21]. The framework of the proposed algorithm is discussed in the next section.

3. MATERIALS AND METHODS

Feature selection algorithms must identify the irrelevant and redundant features as it affects the accuracy of the learning machines. Good feature subsets contain features highly correlated with class, yet uncorrelated with each other [22]. This research work has a close look at developing a novel algorithm that removes irrelevant and redundant features and selects the optimal set of features. A framework of the algorithm is given and explained by appropriate methods.

3.1. Framework of the proposed algorithm

The feature selection algorithm has three components: irrelevant feature removal, redundant feature removal, and optimal feature generation. The first component removes the irrelevant features by using mutual information, a filter method. The second component removes the redundant features by choosing the representatives from each cluster. The genetic algorithm is used as the third component to find the optimal set of features. The irrelevant feature removal obtains features relevant to the class by eliminating the features which are irrelevant to the target class. Feature relevance is measured in terms of feature correlation. Relevant features have a strong correlation with the target class. Mutual Information measures the mutual dependence between two variables and can handle both linear and non-linear relationship. If two features are independent, the mutual information between them is zero, and if the two features are highly dependent, the mutual information is large [23]. So, mutual information is chosen as the measure of correlation between the feature and the class variable. The mutual information is defined as follows:

$$\mathrm{MI}(\mathrm{X}\,;\mathrm{Y}) \;=\; \sum_{y\in \mathrm{Y}}\sum_{x\in \mathrm{X}} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{1}$$

Where X and Y are two features, $p(x, y)$ is the joint probability distribution function of X and Y, $p(x)$ and $p(y)$ are the probability distribution functions of X and Y. Mutual Information is always greater than or equal to zero. The features whose mutual information values are greater than a particular threshold value comprise the relevant feature subset.

Redundant feature removal removes features that are redundant in 3 steps: Constructing a minimum spanning tree from the relevant features, grouping the features in the forest into clusters and selecting the representative feature from each cluster. Mutual information between each pair of features f_i' and f_j' is calculated as MI(f_i', f_j'). A complete Graph G = (V, E) is constructed where V is the set of features from relevant feature subset and E is the mutual information MI(f_i', f_j') ($i \neq j$) which is the weight of the edge between the vertices. For a high dimensional data, the graph G is heavily dense, and thus a minimum spanning tree is built. The mutual information between each pair of features MI(f_i', f_j') is compared with the mutual information between each feature and the class variable. If MI(f_i', f_j') is less than the mutual information MI(f_i',C) and MI(f_j',C) where C is the target class, then the edge (f_i', f_j') is removed.

Each deletion results in a disconnected tree and a forest is obtained. A biograph is constructed for visualization of the forest. Each collection of disconnected trees in the forest represents a cluster. The forest is then traversed, and the mutual information of the features in each cluster with the class variable is determined. The feature that has the maximum MI(f_i',C) is selected as the representative feature from each cluster. The set of the representative feature from each cluster forms the subset of features which is strongly relevant to the class variable. Due to high dimensionality, the subset of features obtained is still large, and thus a genetic algorithm is used to obtain the optimal subset of features from the set of features.

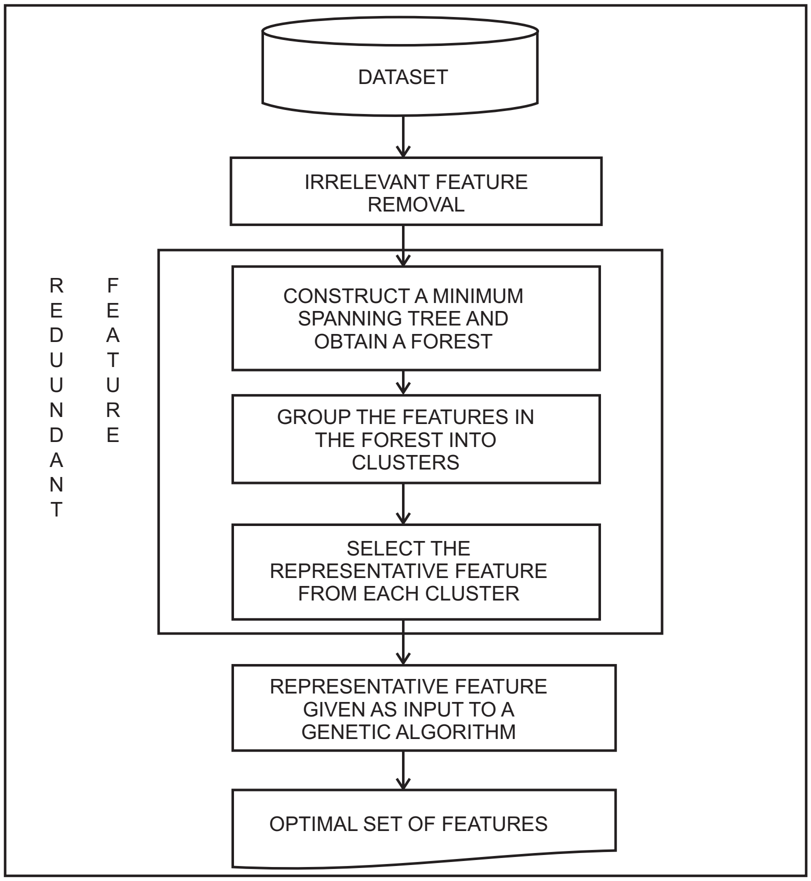

DATASET
IRRELEVANT FEATURE REMOVAL
REDUNDANT
FEATURE
CONSTRUCT A MINIMUM SPANNING TREE AND OBTAIN A FOREST
GROUP THE FEATURES IN THE FOREST INTO CLUSTERS
SELECT THE REPRESENTATIVE FEATURE FROM EACH CLUSTER
REPRESENTATIVE FEATURE GIVEN AS INPUT TO A GENETIC ALGORITHM
OPTIMAL SET OF FEATURES

Figure 1: Framework of the proposed feature

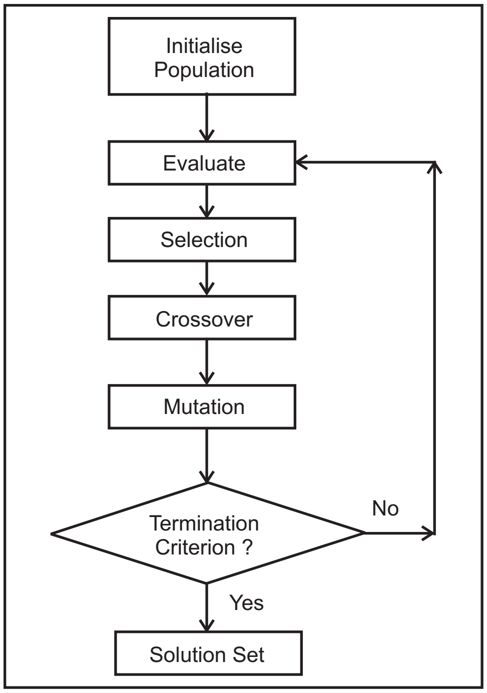


Figure 2: GA based selection algorithm

The set of representative feature from each cluster, the class variable and the number of features desired is provided as input to a genetic algorithm. Genetic algorithm (GA) is an adaptive heuristic search that uses optimization techniques to find true or approximate solutions based on the evolutionary ideas of natural selection and genetics. It is inspired by evolutionary biology such as mutation, selection, and crossover. GA begins with a set of chromosomes called the population. New populations are evolved by mutating solutions according to their fitness value. The solutions that are better are selected for the mating and mutation operations and they carry the genetic code from one generation to another. Genetic Algorithms can solve every optimization problem, problems with multiple solutions, multi-dimensional, non-differential, non-continuous and non-parametrical problems and are easily transferred to existing simulations and models.

The fitness function used in this proposed method is based on the principle of min-redundancy and max-relevance. The relevance of a feature is the average of mutual information values of all the features with the class variable C. The redundancy is the average of the mutual information values between feature f_i and f_j. The feature set of size k is randomly chosen to compose the chromosome. The chromosomes compose a population. Each chromosome is evaluated by the joint conditional entropy which is the fitness function used. During each generation, the most appropriate chromosome is selected. If the fitness function of the succeeding chromosome does not change over the generations, the process will be terminated. The relevance of a feature set S is given as

$$\mathrm{D}(\mathrm{S}, \mathrm{C}) = \frac{1}{|\mathrm{S}|}\sum_{f_{i \in s}} \mathrm{I}(f_i, \mathrm{C}) \tag{2}$$

The redundancy of all features in S is given as

$$\mathrm{R}(\mathrm{S}) = \frac{1}{|\mathrm{S}|^2}\sum_{f_i\, f_{j \in s}} \mathrm{I}(f_i, f_j) \tag{3}$$

The minimum redundancy-maximum-relevance (mRMR) is given as

$$\text{Fitness Function} : m\mathrm{RMR} = \max_{\mathrm{S}}\left[\frac{1}{|\mathrm{S}|}\sum_{f_{i \in s}} \mathrm{I}(f_i, \mathrm{C}) - \frac{1}{|\mathrm{S}|^2}\sum_{f_i f_{j \in s}} \mathrm{I}(f_i, f_j)\right] \tag{4}$$

The set of representative features from the clusters and the desired number of features is provided as input to the genetic algorithm. The new population is generated by mutation, and each chromosome is evaluated using the fitness function. The process is terminated when the fitness function of the new chromosomes selected is identical to the old chromosomes and the optimum set of features is generated.

3.2. CLUST-GA-FS Algorithm

This new innovative proposed algorithm consists of three steps. 1) removing irrelevant features 2) removing redundant features and 3) generating the optimum set of features using a genetic algorithm. For a data set S with m features $\{f_1, f_2, \ldots, f_m\}$ and class C, the mutual information MI(f_i, C) value for each feature is determined. The features whose mutual information values are less than a predefined threshold value θ are the irrelevant features, and hence these features are removed. The relevant feature subset is $\mathrm{F}' = \{ f'_1, f'_2, \ldots, f'_l\} (l \le m)$ where l features are relevant features selected from m feature set.

The mutual information value MI(f'_i, f'_j) for each pair of features in F′ is calculated. A graph G = (V, E) is constructed with features as the vertices V and mutual information values as the edges E. A minimum spanning tree is constructed for the graph G. The edges whose weights MI(f'_i, f'_j) are smaller than MI(f'_i, C) and MI (f'_j, C) are removed. The deletion of edges results in disconnected trees forming a forest. Each of the disconnected trees in the forest is a cluster. From each cluster, the feature with the maximum MI(f'_i,C) value is selected as the representative feature. The representative feature set is $\mathrm{F}'' = \{ f'_1, f'_2, \ldots, f'_k\} (k \le l)$ where k features are selected from l features.

The representative feature set which is strongly relevant to the class C is given as input to a genetic algorithm. The genetic algorithm takes as input parameters the representative feature set, the class C and the desired number of features. The population is composed of the feature set of size *k*. mRMR is the fitness function used and the chromosomes are generated at each generation. The process is terminated when the chromosomes are identical over the generations, and the optimal set of features is generated.

Proposed Algorithm: CLUST-GA-FS

```
Inputs:  Data set S{ $f_1, f_2, \ldots, f_m$ C}, C – class, θ – threshold value
Output:  Representative feature subset
         //===== Part 1 : Irrelevant feature removal=======
Step 1:  for each feature $f_i$ in the data set
         compute MI( $f_i$, C)
         if MI( $f_i$, C) < θ
         remove the feature $f_i$
         end
         end
         relevant feature set F′ = {$f'_1, f'_2, \ldots, f'_1$}(1 ≤ $m$).
         //===== Part 2 : Removal of Redundant feature=====
Step 2:  for each feature $f_i$ in F′
         construct graph G = (V, E) with feature $f_i$ as vertices
         and MI($f_i$ , $f_j$) as edges
         end
         generate the minimum spanning tree of G
         forest = minimum spanning tree of G
         for each edge in forest
         if MI( $f_i, f_j$) < MI( $f_i$, C ) and MI($f_i, f_j$ ) < MI($f_j$, C)
         remove edge from the forest
         end
         end
         for each tree in the forest
         find the maximum MI($f_i$, C)
         select $f_i$ as the representative feature of the cluster
         end
         representative feature set is F″ = { $f'_1, f'_2, \ldots, f'_k$}($k$ ≤ 1).
         //======= Part 3 : Generating Optimum set of features using Genetic Algorithm======
Step 3:  Input: Representative feature set F″ ’={ $f'_1$, $f'_2, \ldots, f'_k$},
         Class C, desired no of features
         Output: Optimal set of features
         Max_gen = no.of generations desired
         find the entropy of F″ – $H_f$ and C – $H_C$ and mutual information between the features – $MI_{ff}$
```

```
and mutual information between the features and class C – $\mathrm{MI}_{fC}$
generate the population consisting of the feature set
while generation is less than max_gen
find the fitness function = $\max_S \left[ \frac{1}{|F''|} \sum_{f_i \in F''} \mathrm{MI}(f_i, C) - \frac{1}{|S|^2} \sum_{f_i f_j \in F''} \mathrm{MI}(f_i, f_j) \right]$
rearrange the population according to their fitness values
create a new generation
if the chromosomes generated are identical
sel = population rearranged
end
end
sel = optimal set of features
```

Steps1 and 2 generate the relevant set of features and remove the redundant feature. The representative set of features is given as input to a genetic algorithm that determines the optimal set of features.

4. RESULTS AND DISCUSSION

The CLUST-GA-FS algorithm is executed using MATLAB software, and the results are verified by classification algorithms. This verification is done by using WEKA software. Three classification algorithms are employed to classify datasets before and after the feature selection. The classifiers used are Naïve Bayes, Multilayer Perceptron and Adaboost. Naïve Bayes Classifier is a probabilistic classifier which is based on Bayes' theorem that assumes a strong independence between the features. The data sets used, the number of features selected by the algorithm and results at various phases of the algorithm are discussed.

4.1. Data Source

For the implementation of the CLUST-GA-FS algorithm, four publicly available data sets were used. The datasets used for the evaluation of algorithms contains microarray and text data as shown in table 1.

Table 1
Data Set Description

S.No.	*Data set*	*No. of features*	*No. of instances*	*No. of Classes*	*Domain*
1.	Colon	2000	62	2	Microarray
2.	Leukemia	7129	38	2	Microarray
3.	Arcene	10001	200	2	Microarray
4.	SMS spam	1833	5574	2	Text

4.2. Proportion of Selected Features

The CLUST-GA-FS algorithm selects optimal features from a large data set. The main purpose of this work is to classify the data set with a minimum number of features. The desired number of features is specified in the genetic algorithm and the number of features which give high classification accuracy is given in table 2. The proportion of the selected features is specified. The proposed algorithm selects 30 optimal features from 7129 features for the leukemia dataset which is approximately 0.42%. Similarly, optimal features selected for the arcene dataset is 32(0.32%), for the colon dataset is 50(2.5%) and for the smsspam dataset is 27(1.47%). Figure 3 compares the number of the features selected for leukemia, Arcene, colon and SMS spam dataset.

Table 2
Proportion of selected features using the proposed algorithm

Data Set	*Proportion of selected features (%)*
Leukemia	0.42
Arcene	0.32
Colon	2.50
SMS spam	1.47

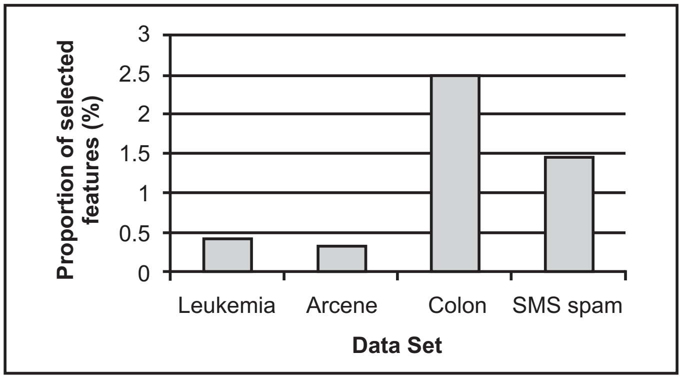


Figure 3: Proportion of selected features

Figure 3 gives a graphical representation of the proportion of the selected features. The number of features is very high for leukemia and arcene dataset.

5. RESULT ANALYSIS

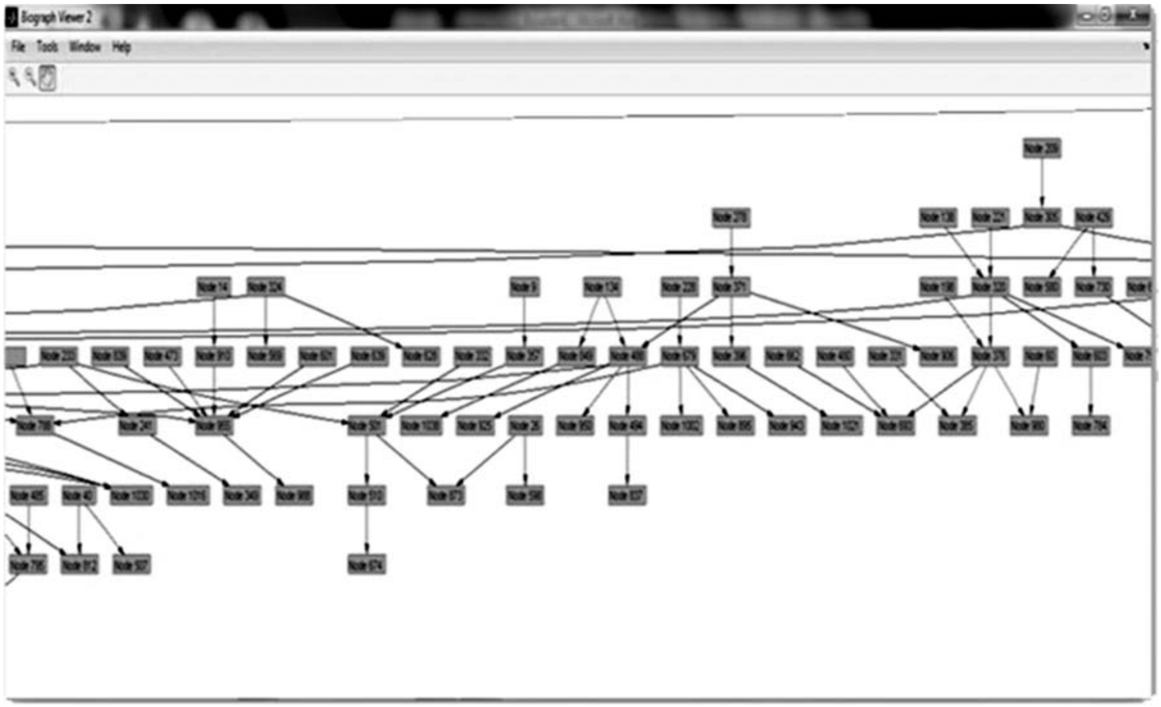


Figure 4: Biograph image showing the minimum spanning tree of the Arcene dataset

The CLUST-GA-FS algorithm is executed in three stages. The first stage removes the irrelevant features from the dataset. The second stage finds the redundant data and removes it. A graph is constructed with the features as the vertices and the mutual information between the features as the edges. Minimum spanning tree of the graph G is constructed. Biograph is a tool used in MATLAB to display the minimum spanning tree as shown in figure 4. The arcene dataset has 10001 features and 200 instances. The irrelevant features are removed, and a minimum spanning tree with relevant features is constructed. Figure 4 shows the minimum spanning tree constructed for the arcene dataset with 1067 nodes and 1045 edges. The nodes indicate the features and the edges indicate the mutual information between the features. A part of the minimum spanning tree is shown in Figure 4, in which the node numbers specify the feature number.

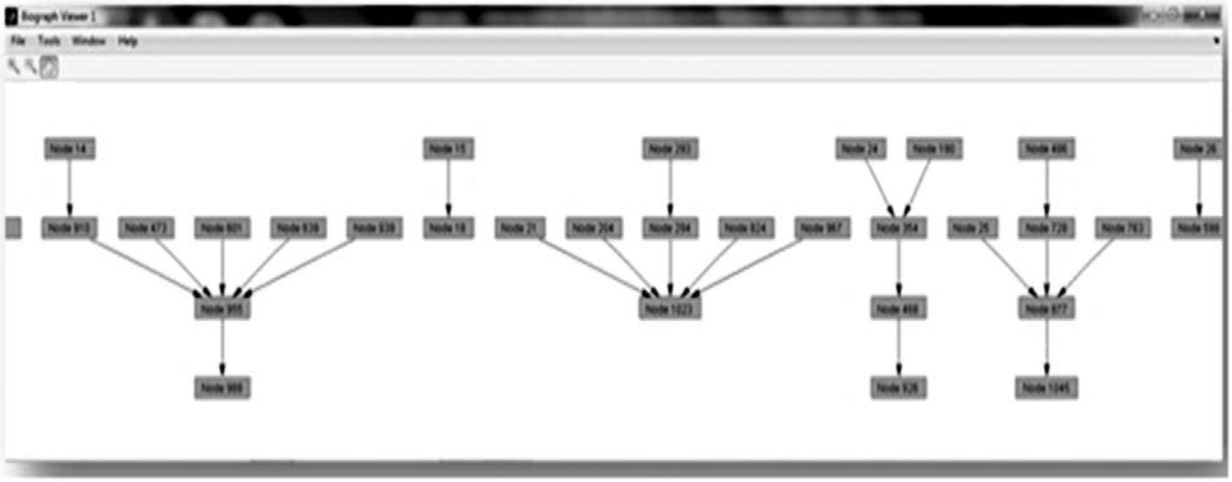


Figure 5: Biograph image showing the forest of the Arcene dataset

A representative feature is selected from each cluster in the forest and the set of representative features are given as input to a genetic algorithm that generates the optimum feature set. The fitness function used is the minimum redundancy-maximum-relevance (mRMR). The population is generated till the chromosomes are identical. The Leukemia and Arcene dataset had 81 generations, Colon dataset had taken 28 generations, and SMS Spam dataset had taken only 3 generations for the fitness function to have a constant value. The plot of the fitness function and the number of generations for leukemia and arcene dataset are given in figure 6 and for the dataset colon and SMS spam are given in figure 7.

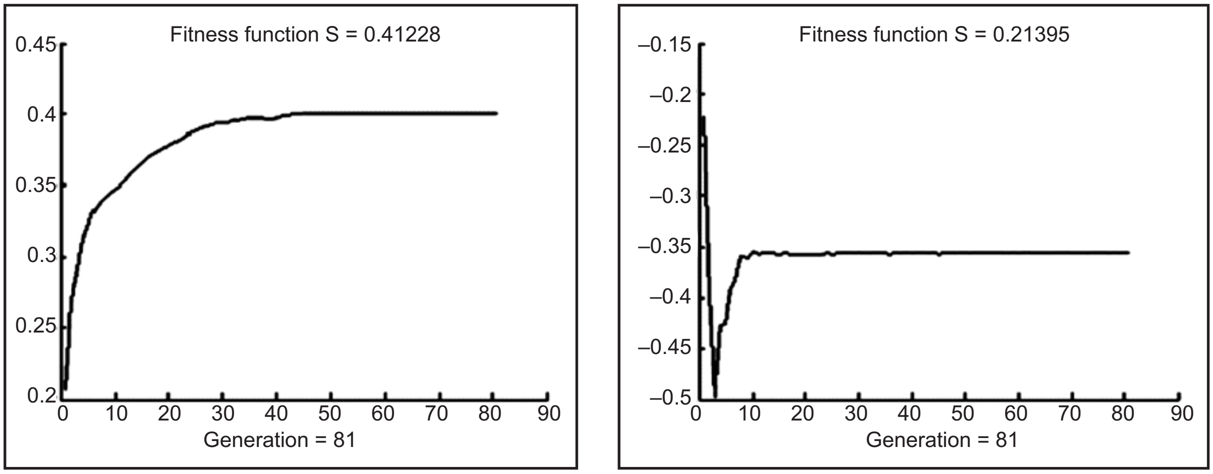


Figure 6: Plot of the fitness function and number of generations for Leukemia and Arcene dataset

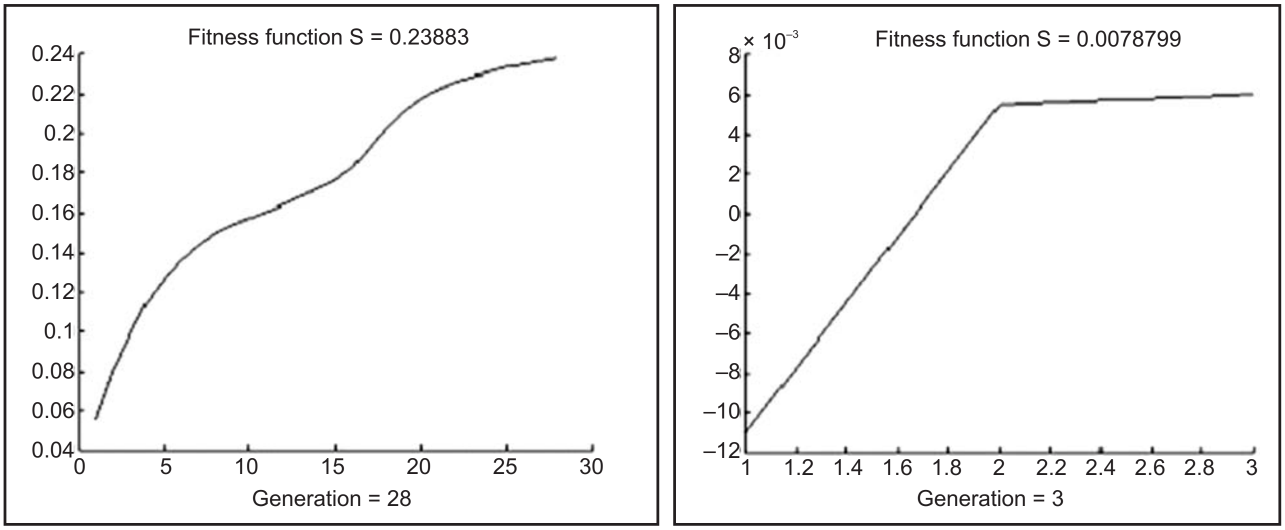


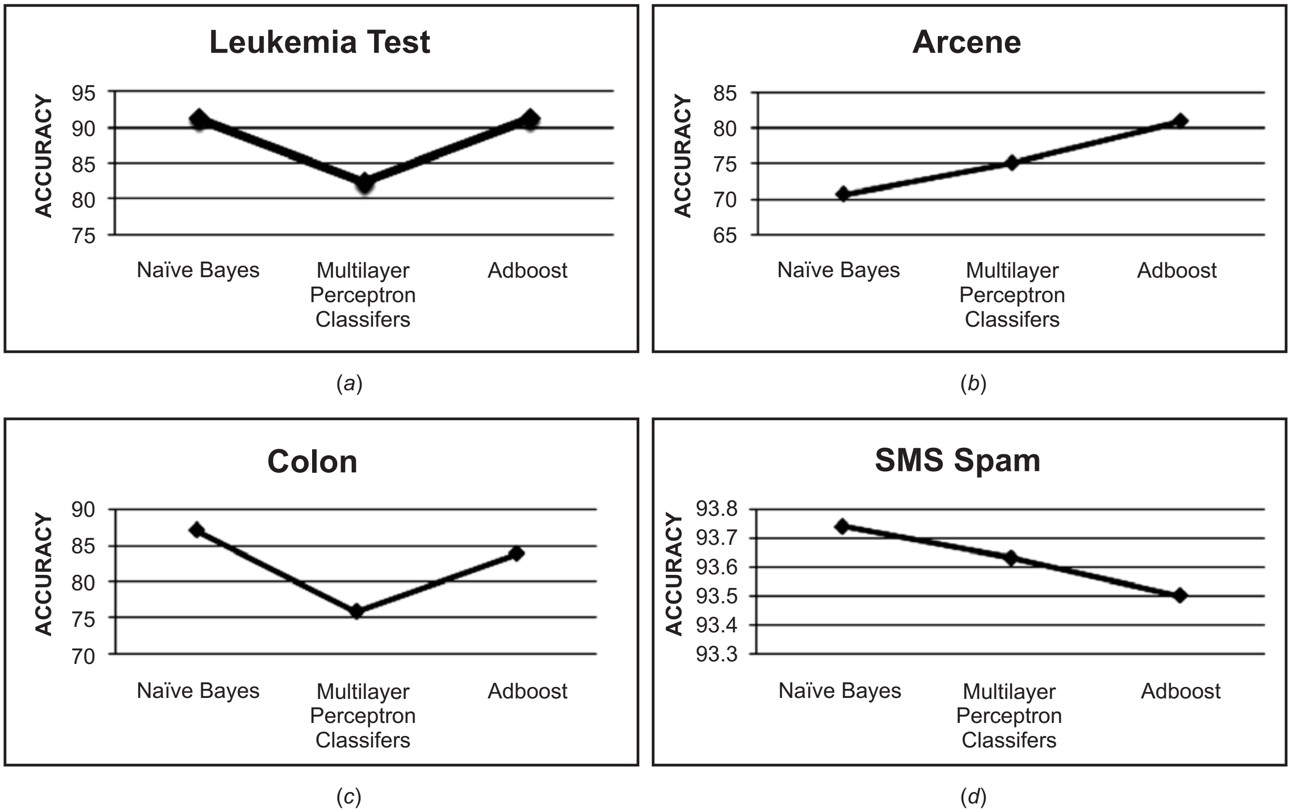
Figure 7: Plot of the fitness function and number of generations for Colon and SMS Spam dataset

To analyze the performance of the proposed algorithm, classifiers bayesnet, adaboost and multilayer perceptron are used to classify the dataset with the selected proportion of features given in table 2. A 10-fold cross validation was performed on the classifiers on the specified four data sets. The criterion used to assess the performance of the classifiers is the scalar characteristics like the accuracy, error rate, precision, recall, and *f* – measure. The accuracy of the classifiers with the optimal feature set selected by the proposed method is given in table 3. Figure 8 shows a graphical representation of the accuracy of the selected data sets. Comparing the accuracy obtained by the classifiers for all the datasets from table 3, it can be observed that Naïve Bayes has performed well for Leukemia, Colon and SMS spam dataset. Adaboost ranks second with good classification accuracy for Leukemia, Colon and arcene data set. Multilayer perceptron has an average performance on all the datasets.

Table 3
Accuracy of the Classifiers

Data Set	*Naïve Bayes*	*Multilayer Perceptron*	*Adaboost*
Leukemia Train	94.74	100	100
Leukemia Test	91.18	82.36	91.18
Arcene	70.59	75	80.89
Colon	87.10	75.80	83.87
SMS Spam	93.74	93.63	87.85

For microarray and text data, the classification accuracy of Naïve Bayes has been improved by the proposed algorithm. The scalar characteristics for performance evaluation do not provide enough information. The true positive rate, false positive rate and the receiver operating characteristic (ROC) are useful for evaluation of dichotomic classifiers performance. ROC is a plot of false positive and true positive rate. The performance of the classifiers is also compared with ROC as shown in table 4.



(*a*) (*b*)

(*c*) (*d*)

Figure 8: Accuracy of Leukemia, Arcene, SMS Spam and Colon dataset

Table 4
ROC measure of the Classifiers

Data Set	*Naïve Bayes*	*Multilayer Perceptron*	*Adaboost*
Leukemia Train	0.946	1	1
Leukemia Test	0.871	0.921	0.918
Arcene	0.806	0.846	0.817
Colon	0.908	0.847	0.913
SMS Spam	0.799	0.808	0.663

The ROC values show that Multilayer Perceptron has performed well than Naïve Bayes and Adaboost for all the datasets. ROC curve is an effective method of evaluating the performance of the classifiers. AUC is the area under the ROC curve, and it can have a value between 0 and 1. The closer the value of AUC is to 1, the better the performance of the classifier.

Table 5
Mean accuracy of the classifiers

	Mean Accuracy	
Classifier	*MicroArray Data*	*Text Data*
Naïve Bayes	85.9	93.74
Multilayer Perceptron	83.29	93.63
Adaboost	88.98	87.85

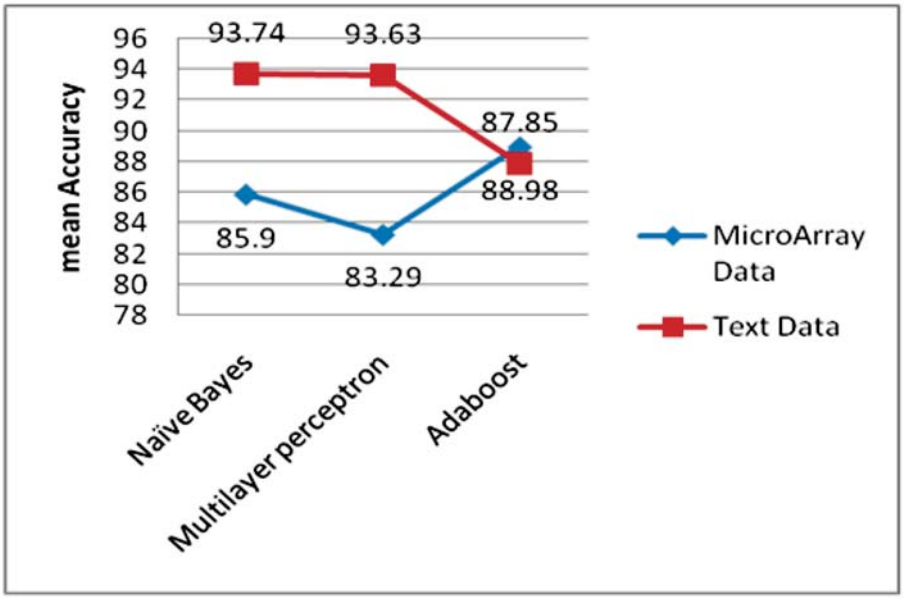


Figure 9 : Mean Accuracy of the Classifiers

Table 6
Mean ROC of the classifiers

	Mean ROC	
Classifier	*MicroArray Data*	*Text Data*
Naïve Bayes	0.8827	0.799
Multilayer Perceptron	0.9035	0.808
Adaboost	0.912	0.663

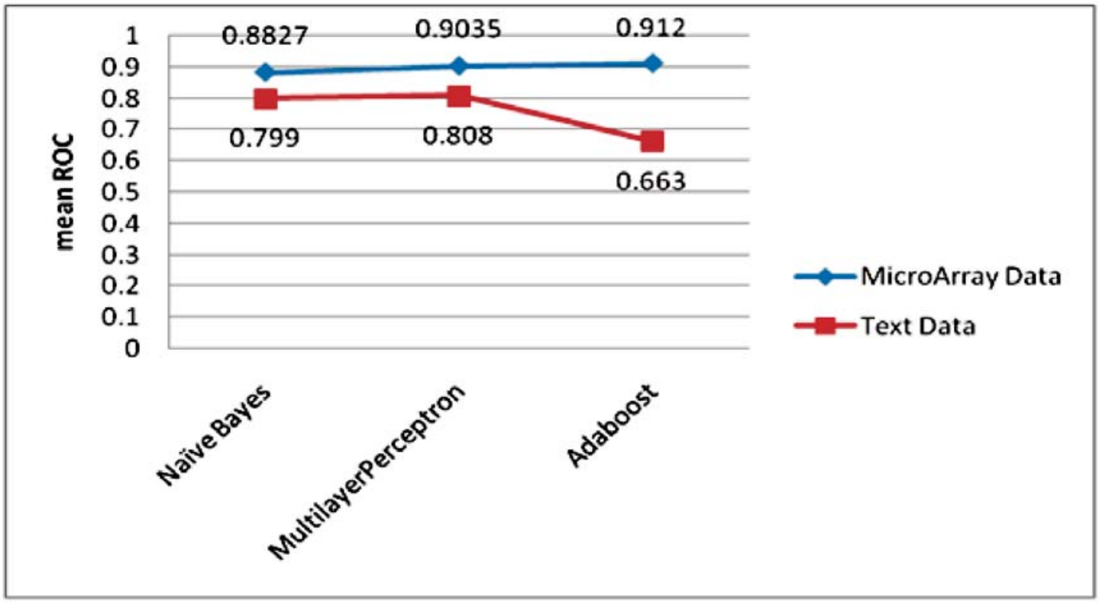


Figure 10 : Mean ROC of the Classifiers

The mean accuracy and mean ROC of the classifiers for microarray data and text data are given in the table 5 and table 6. It is evident from the figure 9 that the accuracy of naïve bayes and multilayer perceptron for text data is higher than for microarray data. The accuracy of adaboost for microarray data is more than that of text data. Figure 10 clearly depicts that the mean ROC value of the classifiers naïve bayes, multilayer perceptron and adaboost is higher for microarray data than for text data. Though accuracy is a good measure of performance of the classifiers, ROC is an effective measure of evaluating the performance of the classifiers. Based on accuracy and ROC measure, it can be concluded that the classifiers naïve bayes, multilayer perceptron and adaboost have good performance for microarray data.

6. CONCLUSION

The curse of high dimensionality in any field poses a challenge for machine learning techniques. The problem of feature selection must be efficiently addressed to derive essential knowledge from the overwhelming data. In this research work, a clustering and genetic algorithm based feature selection algorithm for high dimensional data is proposed. The algorithm involves removing the irrelevant features, removing redundant features by constructing a minimum spanning tree, splitting the minimum spanning tree into clusters. Also, finding the representative feature from each cluster and finally finding the optimal set of features using a genetic algorithm. The proposed CLUST-GA-FS algorithm has selected the optimum set of features from large dimensional datasets, and the classifiers have obtained a good accuracy with less number of features. The datasets used are high dimensional microarray and text data. The accuracy of the classifiers for text data is higher than the accuracy obtained for microarray data. But, the proportion of the number of features obtained for text data is higher than that for the microarray data. Hence, the proposed algorithm has performed well for microarray data in terms of the proportion of features and the ROC value obtained. Naïve Bayes obtains the best classification accuracy for both microarray and text data. The datasets used in this work are two-class classification problems, and in future, this work can be extended to handle data sets with multiple classes.

REFERENCES

[1] Yu, L., & Liu, H. “Efficiently handling feature redundancy in high-dimensional data.”, Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2003,pp.685.

[2] Guyon, I., & Elisseeff, A,” An Introduction to Variable and Feature Selection. Journal of Machine Learning Research (JMLR)”,vol.3 issue 3, pp.1157–1182, 2003 https://doi.org/10.1016/j.aca.2011.07.027

[3] Hartuv, Erez, and Ron Shamir, “A clustering algorithm based on graph connectivity.” Information processing letters, vol. 76, pp.175-181,2000.

[4] Lee, J., & Kim, D. W,” Mutual Information-based multi-label feature selection using interaction information. Expert Systems with Applications, vol.42, issue 4, pp. 2013–2025, 2014

[5] He, D., Rish, I., Haws, D., & Parida, L,” MINT: Mutual Information Based Transductive Feature Selection for Genetic Trait Prediction”, IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol.13, issue3, pp. 578–583, 2016.

[6] Han, M., & Ren, W, ”Global mutual information-based feature selection approach using single-objective and multi-objective optimization”, Neurocomputing, vol.168, pp.47–54, 2015.

[7] Mundra, P. A., & Rajapakse, J. C, ”SVM-RFE with MRMR filter for gene selection”, IEEE Transactions on Nanobioscience, vol.9, issue 1, pp. 31–37, 2010.

[8] C. Krier, D. François, F.Rossi, M. V, ”Feature clustering and mutual information for the selection of variables in spectral data”, European Symposium on Artificial Neural Networks, pp. 157–162, 2010

[9] Dhillon, Inderjit S., Subramanyam Mallela, and Rahul Kumar. “A divisive information-theoretic feature clustering algorithm for text classification.”Journal of machine learning research 3, pp. 1265-1287,2003.

[10] Kawaji, H., Yamaguchi, Y., Matsuda, H. and Hashimoto, A., " A graph-based clustering method for a large set of sequences using a graph partitioning algorithm", Genome Informatics, vol. 12, pp.93-102, 2001.

[11] Van Dijck, Gert, and Marc M. Van Hulle. "Speeding up the wrapper feature subset selection in regression by mutual information relevance and redundancy analysis." In International Conference on Artificial Neural Networks, pp. 31-40, 2006.

[12] Yang, Jihoon, and Vasant Honavar. "Feature subset selection using a genetic algorithm." In Feature extraction, construction and selection, Springer US , pp. 117-136, 1998.

[13] Lu, L., Yan, J. and Meng, Y., "Dynamic Genetic Algorithm-based Feature Selection Scheme for Machine Health Prognostics", Procedia CIRP,vol.56, pp.316-320, 2016.

[14] Zhu, Zexuan, and Yew-Soon Ong. "Memetic algorithms for feature selection on microarray data." In International Symposium on Neural Networks, Springer Berlin Heidelberg , pp. 1327-1335., 2007.

[15] Apolloni, Javier, Guillermo Leguizamón, and Enrique Alba. "Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments." Applied Soft Computing, vol. 38, pp.922-932, 2016.

[16] Sebban, Marc, and Richard Nock. "A hybrid filter/wrapper approach of feature selection using information theory." Pattern Recognition, vol. 35, issue 4, pp. 835-846, 2002.

[17] Solorio-Fernendez, S., Carrasco-Ochoa, J. A., & Martinez-Trinidad, J. F, "A new hybrid filter-wrapper feature selection method for clustering based on ranking", Neurocomputing, vol. 214, pp. 866–880, 2016.

[18] Wen, Peng, and Liu Wenxia. "Niche Genetic Algorithm and Minimum Spanning Tree for Substation Planning." In Intelligent Systems, 2009. GCIS'09. WRI Global Congress on, IEEE, vol. 3, pp. 61-65, 2009.

[19] Contreras-Bolton, C., Gatica, G., Barra, C.R. and Parada, V., "A multi-operator genetic algorithm for the generalized minimum spanning tree problem", Expert Systems with Applications, vol. 50, pp.1-8, 2016.

[20] Wang, Hsiao-Fan, and Hsin-Wei Hsu. "A closed-loop logistic model with a spanning-tree based genetic algorithm." Computers & operations research, vol. 37, issue 2, pp. 376-389, 2010.

[21] Song, Q., Ni, J., & Wang, G, "A fast clustering-based feature subset selection algorithm for high-dimensional data", IEEE transactions on knowledge and data engineering, vol. 25, issue 1, pp. 1-14, 2013.

[22] Hall, Mark A. "Correlation-based feature selection for machine learning." PhD diss., The University of Waikato, 1999.