Metaheuristic Computing and Applications, *Vol. 1, No. 2 (2020) 131*-153
DOI: https://doi.org/10.12989/mca.2020.1.2.131

An optimal structure for ensemble feature selection

Amirreza Rouhi[1,2a] and Hossein Nezamabadi-pour[*1]

[1]Intelligent Data Processing Laboratory (IDPL), Department of Electrical Engineering, Shahid Bahonar University of Kerman, Kerman, Iran
[2]Department of Electronics and Information, Politecnico di Milano, Italy

(Received March 3, 2020, Revised August 19, 2020, Accepted December 22, 2020)

Abstract. Today, Gene selection in microarray data is one of the most challenging subjects in the fields of medicine and machine learning. Due to the large number of features and small number of samples in microarray datasets, choosing the desirable genes in these data is a difficult task. Among several methods which have been proposed for gene (feature) selection, ensemble and hybrid methods have attracted more attentions. The purpose of this paper is to find an optimal structure for hybrid-ensemble gene selection method that, by selecting the least number of the genes, yields the desired classification accuracy. For this purpose, the genetic algorithm is used as one of the most popular evolutionary optimization methods to accomplish an optimal hybrid-ensemble feature selection method. The performance of the proposed method is widely tested on 18 microarray datasets, and it is compared to those of the 10 well-known gene selection methods in terms of classification error rates and Gmean. Experimental results demonstrate that the obtained optimal method is considerably superior to the other competing methods over different evaluation methods and datasets.

Keywords: gene selection; high-dimensional data; hybrid methods; metaheuristic; filter methods; ensemble methods

1. Introduction

Feature selection is one of the most fundamental topics in machine learning that has made significant progress in many areas such as medicine, health care and biology [1, 2]. Nowadays, with the increasing importance of data, classical data that has been confined to just a few dozen features has been replaced with high-dimensional and big data which are abundantly found in text processing, combinational chemistry, medicine and informatics [1].

To date, many feature selection methods have been proposed on classical data that have shown good performance, while increasing the data dimension has posed many challenges to feature selection in this data.

One of the most important applications of trait selection is the discovery of the most effective genes in microarray data. These data are one of the most important multidimensional data in medicine and include a large number of genes. Among the thousands of the measured genes in microarray assay, most of them are not correlated with classification validity of the problem, in

[*]Corresponding author, Professor, E-mail: nezam@uk.ac.ir
[a]Assistant Professor, E-mail: amirreza.rouhi@mail.polimi.it

http://www.techno-press.org/?journal=mca&subpage=5 ISSN: 2713-5519 (Print), 2733-8053 (Online)

other words, many of them are irrelevant or redundant, and only a few genes influence the process of cancer diagnosis. As a result, in order to prevent curse of dimensionality the selection of relevant and appropriate genes is a necessary process in dealing with these data [3]. In the meantime, feature selection methods can be used to select effective genes in microarray data. Of this cause, features selection is a critical pre-processing phase in bioinformatics. Treatment of the risk factor for cancer-related deaths by discovering active genes for cancer detection is one of the most significant applications of preference of medicine [1, 4]. Thus, seeking a suitable method for selecting effective genes from thousands of available ones is very useful in such a way that the highest classification precision is achieved. The significance of feature-selection approaches to select insightful genes before the classification stage and diagnosis of cancer detection was illustrated in recent studies [5].

To date, several methods have been proposed for feature selection in high-dimensional data that can be divided into four main categories based on the relation between evaluation function and classifier.: filter methods, wrapper methods, hybrid methods and embedded methods.

Filtering approaches choose sub-sets of genes from the main dataset by utilizing different evaluation criteria focused primarily on statistical and independent approaches. These approaches determine effective features dependent on intrinsic data characteristics without interfering with data mining algorithms. In other words. This method performs the feature selection process faster than other available methods; thus, these methods can be a good option for selecting appropriate features in high dimensional data. On the other hand, the classification accuracy of these methods is not high. Filter methods can be divided into univariate and multivariate classes. Univariate methods, at first evaluates and rates genes independently based on a given criteria, then the final subset is selected as the subset of genes with the highest rankings. Information Gain (IG) and F-Score are of famous univariate filter methods [7]. In the multivariate method, the relation among genes affects the gene selection process [3]. ReliefF [8], FCBF, mRMR, CFS, and INTERACT are common multivariate filter methods.

In 2008, a filter method called MASSIVE [9] is proposed which is suitable for micro-array data feature selection based on an information theory measure called DIRS and the results revealed that on 11 microarray datasets, the suggested method is superior to the 5 state-of-the-art methods on microarray feature selection (based on the accuracy criterion). In [10], a feature selection method for high-dimensional data based on multi-task filter method is proposed. In this paper the result indicated that the multi-task feature selection in conjunction with single-task and multi-task classifiers is very successful. In 2013, a filter method in which weight and redundancy of each feature affect proper feature selection is proposed by [11]. In this method, which is called mRMR, the goal is to obtain the maximum weight and the minimum redundancy. However, the weight of each feature is measures based on filter ranking methods including Laplacian score, Fisher-score and constraint score which shows the importance of a feature. Moreover, redundancy shows the correlation between features which is measured based on Pearson correlation coefficient and mutual information [11]. In 2014, a version of ReliefF method in which there is a trade-off between relevance of a feature and its associated cost is proposed by [14]. The results demonstrated that mRMR leads to significant increases in the accuracy of feature selection and classification.

Wrapper approaches use the output of classifier to determine the importance of the gene subsets to determine the appropriate genes. In these methods, for all possible subsets, a search mechanism is used to find the most efficient subset of features. The greedy search and random (stochastic) search constitute two principal search mechanisms. The classifier evaluates each subset suggested by the utilized search algorithm and then the classifier correction rate is used as the fitness value of the corresponding feature subset. [2, 13, 14].

Greedy search strategies are single-track search approaches that are stuck quickly in local optima. Two examples of important greedy methods are sequential forward selection and sequential backward elimination. Random search methods select the subset of genes randomly. Metaheuristics methods include particle swarm optimization (PSO), ant colony optimization (ACO) and genetic algorithms (GA), as well as gravitational search algorithm (GSA) are the main random search algorithms. These methods have low speed and high complexity and therefore cannot be applied to high dimensional data alone.

In 2006, a wrapper method called BIRS [15] is applied to several micro-array datasets. Comparison of the results of this approach with several other approaches on microarray data has shown that BIRS selects a smaller subset of genes from the main set with similar predictor performance to others. Sharma et.al. proposed a wrapper method called Successive Feature Selection (SFS) [16]. The proposed algorithm is compared to several other feature selection methods and satisfactory results are obtained and also this algorithm was able to select all the genes that were also selected in the original noise free environment. In most feature selection methods, a feature with low rank is not selected while this feature might results in a favorable performance in interaction with a suitable subset of features. In order to resolve this problem, features are first divided into smaller blocks and superior features of each block are specified in the classifier considering their performance, then they are compared to obtain the best subset [16]. In 2015, a wrapper feature selection method based on ant colony optimization called MGSACO is proposed by [3]. The results indicated that a subset of genes with minimum redundancy and maximum relevance was selected by the MGSACO. In addition, the results showed that the classification accuracy of the MGSACO is much higher than that of three well-known and frequently used unsupervised methods for different subsets of genes over three different classifiers: supporting vector machine, naïve Bayes, and decision tree.

Embedded approaches execute the selection of features as an inseparable part of machine algorithms [17]. The learning algorithm and feature selection stage are two impartible elements of these methods. The speed of the feature selection process in these methods is higher than that of wrapper and less than filter methods [18].

Ref. [19] have proposed an embedded method on high-dimensional data which performs feature selection through periodic training of support vector machines with existing features and the elimination of the least significant features. The results showed that, for data overfitting, this approach is much more robust than the baseline method, and the genes chosen by this technique yield improved classifier performance.

Kernel-penalized SVM (KP-SVM) which proposed in [20], selects desired features through penalizing use of the feature in the dual formula of SVM. Experiments have been performed on four real-world benchmark problems by comparing methods with known feature selection methods. Finally, KP-SVM outperformed the alternative approaches and determined consistently fewer relevant features. In 2012, an embedded method is proposed by [21] which employs a metric to determine the least significant features and examine the impact of each feature on classifier performance while disturbed by noise. The comparison of this method with two robust gene selection methods: t-test and SVM-RFE on four microarray datasets, showed acceptable performance of this method.

Hybrid methods can be considered as a combination of filter and wrapper methods. The main feature sets are decreased in the first stage by filter methods, then, by applying the wrapper methods, the final features are selected. These methods have attracted the attention of many researchers because of their acceptable speed and accuracy compared to other feature selection methods [22].

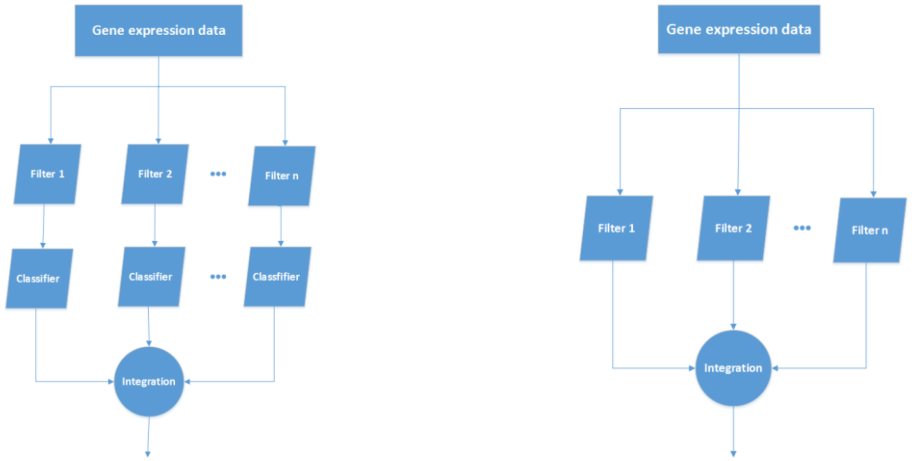
A hybrid approach which is a combining of SVM-RFE and mRMR is proposed in [23]. The method selected less number of genes compared to MRMR or SVM-RFE on most datasets and geneontology analyses have shown that this approach selected genes that have similar functional properties that are important for differentiating cancerous samples. Moreover, in [24], a hybrid method combining information gain (IG) and genetic algorithm is proposed. The experimental results in [24] indicate that this method is capable of achieving high categorization efficiency as calculated by precision, recall, and F-measure. Chuang et al. [25] proposed a combination of correlation-based feature selection (CFS) and the Taguchi-genetic algorithm. Experimental results showed that this method reduced redundant features effectively and achieved superior classification accuracy. In [26], a hybrid method, called R-m-GA, is proposed using a combination of ReliefF, Mrmr, and GA. The comparative study of R-m-GA versus GA and ReliefF-GA has shown that this method is capable of identifying the smallest gene subset with the best classification accuracy. A hybrid method by combining ReliefF, IG, and F-score as the filter stage and improved binary gravitational search algorithm as a wrapper stage is proposed by [27]. The performance of this algorithm is compared with the performance of the system without feature selection, with feature selection by IG, ReliefF, F-score, ensemble of these three filters and also with feature selection by IBGSA on 10 mircoarray datasets. The results show that the higher classification accuracy is obtained with less number of genes.

High-dimensional data not only may be large in terms of the number of features, but they can also face issues with redundancy, noise, and nonlinearity. Therefore, it cannot be said that a method provides good results in all data since many of the methods cannot solve such problems alone. Therefore, researchers have been focusing on using ensemble techniques to take advantage of several methods and combine the results which decrease the probability of choosing a wrong solution and provides better predictions for learning algorithms that may be stuck in local optima [28].

Ensemble techniques have also shown favorable results in feature selection which instead of considering a single feature selection/classification method, outcomes of applying several feature selection/classification methods are combined independently.

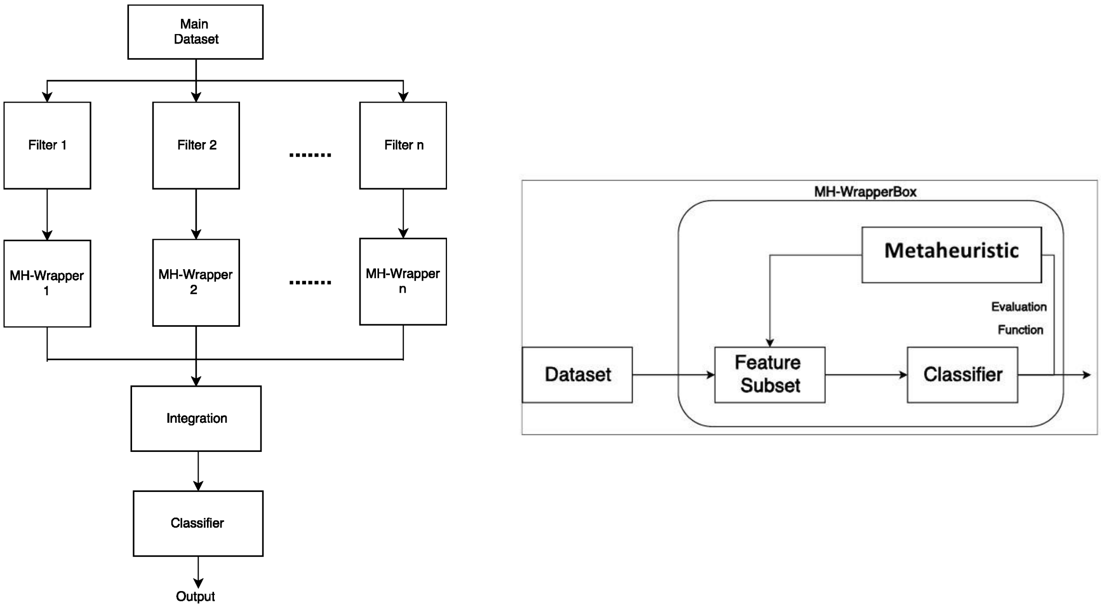
Fig. 1 shows two examples of the use of ensemble methods in feature selection [29]. In Fig. 1(a), which is called “ensemble classification”, multiple filter methods are applied to data separately. Each filter approach selects its own feature subset, then a classifier gets results of each and a classifier classifies the data for each feature selection method. Finally, the classification results are integrated using an integration method [29]. Fig. 1(b) shows “ensemble Feature selection”, which the results of various filters methods on the gene expression data are integrated, thereby producing the final chosen subset.

Abeel et.al. [30] have proposed an ensemble method on micro-array data for cancer diagnosis and then the results are investigated on 4 micro-array datasets and the results showed a significant improvement over the baseline method in terms of classification performance. Yang et.al. [31] proposed an ensemble feature selection method called multi-criteria fusion-based recursive feature elimination (MCF-RFE) and the experimental results on five data sets for gene-expression show that the MCF-RFE algorithm outperforms the widely used benchmark feature selection algorithm SVM-RFE. Bolon et.al [32] proposed an ensemble feature selection using CFS, INTERACT and IG and the results of applying this ensemble method on 10 microarray datasets have shown the



(a) Ensemble classification (b) Ensemble feature selection

Fig. 1 Two Instances of Ensemble Feature Selection [29]



(a) Ensemble classification (b) Ensemble feature selection

Fig. 2 The general diagram of the proposed hybrid-ensemble method. (a) The Hybrid-Ensemble method proposed in [34], (b) Block diagram of the MH-Wrapper box

superiority of this method over the baseline methods used. An ensemble of three filter method IG, CFS, and Relief on high-dimensional data is proposed by [28]. An ensemble of three filter method IG, CFS, and Relief on high-dimensional data is proposed by [28] and the experiments on four high dimensional data sets show that the proposed approach outperforms the single feature selection algorithms (IG, ReliefF, and CFS) as well as two well-known aggregation methods (WMA and CLA) in terms of classification performance. In [33] a gene selection method is proposed which consists of two main stages. At the first stage, stage different filter methods

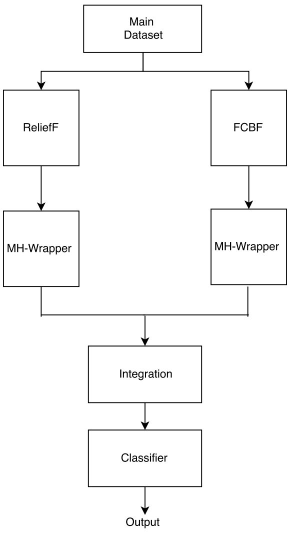


Fig. 3 An instance model of the framework of Fig. 2(a) implemented in [34]

including ReliefF, FCBF, IG are applied, then in the next step the ant algorithm is applied to the results of the previous step. The results of applying this method on five microarray datasets showed the superiority of the method over five filter approaches. In [29], a gene selection method is proposed using ensemble techniques which is a combination of CFS, Cons, IG and Relief filter methods and the results of comparison with baseline methods, and on 7 microarray datasets, showed the optimal performance of this ensemble method.

In order to overcome the problems of selecting suitable and effective genes of microarray data, one solution can be the "Hybrid-Ensemble" approach because wrapper methods result in desirable classification accuracy and filtering methods result in less complexity and high speed. In our previous work [34], a Hybrid-Ensemble method approach for selecting desirable genes in microarray data is presented. Fig. 2 shows the general diagram of the proposed hybrid-ensemble method. In this method, several hybrid methods are applied to data and then the results are combined through different methods. This method might achieve desirable results in high-dimension data because it employs filter, wrapper and ensemble methods.

In previous work to examine the ability of the proposed framework, a two lines instance model through trial and error was implemented which is presented in Fig. 3. In two lines instance model of [34], FCBF and ReliefF filter methods are applied to gene data independently; then improved binary gravitational search algorithm (IBGSA) as the metaheuristic search method incorporation with k-nearest neighbors (k-NN) classifier is applied to the selected genes of each filter method [34]. This instance model outperforms APCES [48], MGSACO [3] and the method proposed by Bolon et al. [50] which is one of the up-to-date methods in gene selection.

However, in the framework of Fig. 2(a), the number of lines, suitable filter and MH-wrapper box (metaheuristic and classifier) methods and the type of aggregator which can guide us to the optimal solution are of existing challenges. In this paper we continue our previous work in [34] to obtain an optimal structure based on Hybrid-Ensemble framework. Thus, finding an optimal structure as an optimization problem and trying to optimize number of lines, type of filter/wrapper methods at each line and type of aggregator for integrating the results of different lines is performed using genetic algorithm. In other words, the purpose of the current work is to optimize hybrid-ensemble framework such that desirable classification accuracy is obtained.

The rest of the paper is organized as follows. In Section 2, basic concepts including filter methods employed in the proposed method, and gravitation search algorithm is described. Section 3 presents the proposed method for optimizing the hybrid-ensemble framework and achieving good solution for feature selection in high-dimensional data. Section 4 investigates the effectiveness of the obtained optimal method. Finally, Section 5 concludes the paper.

2. Background

In this section, filter and wrapper methods employed in feature selection are described.

2.1 Genetic algorithm optimization

In this section, we briefly review the genetic algorithm optimization. Recently genetic algorithms have been very attentive for their potential as a technique of optimization in complex problems. The fundamental concept of the GA is derived from biological Survival and adaptation cycle. Genetic algorithm methodology demands that a coded string of finite length be used to represent the set of decision variables. One codes the decision variable set which describes a trial solution as a binary or dual string or "chromosome" to implement a GA.

Genetic algorithms are probabilistic and not deterministic. These algorithms work by encrypting the solution set, not the solutions themselves. In addition, these algorithms search through a population of solutions, not a single solution, and use the cost function and do not require derivatives.

The main concepts of these algorithms are as follows:

(A) Encoding:

The decision variables of a problem are usually coded into a fixed-length string which could be a binary string or an integer list. For example, 11010101 for binary or 31421234 for integer list.

(B) Selection:

A selection operation is a special phase in a genetic algorithm, since it decides the key areas of the evolutionary search. This aims to improve the survival chances of the fittest individuals. The selection operator operates mostly at chromosome-level. The effectiveness of each individual depends on its fitness. The fitness value may be determined by an objective function or by a problem-specific subjective decision. As the generations go by, the population members should become fitter and more fit. In different conditions different selection mechanisms work well. Appropriate method to increase the optimality of the solution has to be chosen for the specific problem.

(C) Cross over:

The basic idea behind crossover is that if it takes the best traits from each of the parents, the new chromosome could be better than both parents. Crossover operator, which occurs during evolution according to a user definable crossover probability, is a genetic operator which combines (mates) two chromosomes (parents) to generate a new (offspring) chromosome.

(D) Mutation:

Mutation involves modifying the value of each 'gene' of a solution according to a certain probability pm (the probability of mutation). The role of mutations in the genetic algorithm is to restore lost or unknown genetic material in the population to prevent premature convergence of GA to suboptimal solution.

2.2 Filter feature selection methods

Filter methods, as described above, are the most common and applicable feature selection method. Such methods work independently of learning algorithms; thus, their speed is faster than other methods for selecting genes. Some of the filter methods are described in brief.

2.2.1 Relief and ReliefF

Relief [24] is one of the best-known filter methods for nominal and numerical features. Relief looks for features that are statistically correlated with a group; the feature is better based on this algorithm, which allows further distinctions between samples from different groups and the same values for samples with identical groups [24].

2.2.2 FCBH

Fast correlation-based filter method (FCBF) [35] is one of the multivariate filter methods developed based on mutual information for dealing with the high-dimensional data. This approach is focused upon the calculation of symmetrical uncertainty (SU) (Eq. (1)) in order to distinguish appropriate and redundant features and to determine the relationship between feature-feature and feature-class [35].

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

where $H(X)$ and $H(Y)$ are entropies of two features and $IG(X,Y)$ is information gain.

FCBF initially selects a series of features that have a high group association based on SU calculation and retains the features appropriate to a group after eliminating redundant features.

2.2.3 Minimum-Redundancy-Maximum-Relevance (MRMR)

MRMR is also a multi-variate approach for selecting genes based on the correlation between features-feature and feature-class using this measure [14]. This approach selects features that are of greatest class relevance and low redundancy.

2.2.4 Fisher score

The concept of the Fisher score algorithm is to find subsets of features that have the maximum possible distance between data points in different groups and the lowest possible distance between data points in a group [7].

Consider feature Xi as an m-group dataset feature. If sample set of features i is in k-th group X_i^k and $\mid X_i^k \mid = n_k$ where $k = 1,2,\ldots,m$ and $\bar{X}_i^k$ and $\bar{X}_i$ are mean of features in $X_i^k$and X_i, then Fisher score of a feature is [7]:

$$\mathrm{F}(\mathrm{X}_i) = \frac{\sum_{k=1}^{m} n_k(\bar{X}_i^k - \bar{X}_\iota)^2}{\sum_{k=1}^{m}\sum_{x\in X_i^k}(x - \bar{X}_i^k)^2} \tag{2}$$

As can be seen the numerator shows distinction between two groups, while the denominator indicates dispersion in each group. The higher the Fisher score of a feature, the higher the discrimination. After calculating the fisher score of each feature, the features are ranked based on their fisher scores and then the top features are selected based on a predefined threshold.

2.2.5 Information gain

Information Gain (IG) is one of the univariate feature selection methods that evaluates features based on their information gain. In other words, IG of feature xi in Sx is

$$IG(Sx, xi) = H - \sum_{v=values(xi)}^{\frac{|Sxi=0|}{|Sx|}} H(Sxi = v) \tag{3}$$

Which $values(xi)$ is a set of values that can be taken by xi. Entropy $H(Sxi = v)$ can also be defined as follows:

$$\mathrm{H(S)} = -(\mathrm{p}_{+}) \ \log_2(p_{+}) - (p_{-})\log2(p_{-}) \tag{4}$$

In which $p+$ and $p-$ are the ratio of positive samples to total samples and negative samples to total samples, respectively.

Finally using a predefined threshold, the features are ordered by their rank after determining the information gain for each feature.

2.2.6 Correlation based Feature Selection (CFS)

CFS is a multivariate filter feature selection method which presented in [36]. This approach measures features through correlation calculation focused on a heuristic evaluation criterion that is oriented towards subsets that have uncorrelated features with strong correlations to the class.

2.2 Metaheuristic feature selection algorithms

In this section, metaheuristic algorithms employed in the current work are briefly described.

2.2.1 IBGSA

The gravitational search algorithm (GSA) is a metaheuristic algorithm developed in 2009 by [37] inspired by mass and gravity. In 2010, binary version of this algorithm called BGSA was proposed in [38]. IBGSA, an improved version of the BGSA algorithm, was proposed in [39] to avoid being trapped in local Optima to solve feature selection problems. One of the most important differences of IBGSA compared to BGSA is its elitism property. In IBGSA, the agent location changes only when the new location has a fitness function greater than or equal to the previous fitness value.

2.2.2 ACO, BACO & ABACO$_H$

Ant colony optimization (ACO) is a metaheuristic algorithm inspired by the behavior of ants in search of food [13]. The ants can find the shortest route from nest to food by tracking the remaining pheromones by communicating with each other and transmitting path information.

By defining the feature selection problems as a graph, in which features are used as nodes, we can use ACO to find the best features. In recent years, several methods have been presented to select best feature based on ACO algorithm. Binary ACO (BACO) proposed based on the ACO algorithm, including optimal classification rate and higher speed than ACO.

In recent years, a version of BACO called ABACO$_H$ [40] was proposed, which incorporates both BACO and discrete ACO. ABACO$_H$ allows the ant to search among all existing features. This approach allows an ant to pick or reject the observed features, unlike previous ACO approaches, where the observed feature was chosen. Other features of this method are the ability to see all the features not seen before.

2.2.3 BPSO

Particle swarm optimization (PSO) [41] is a metaheuristic population-based search algorithm inspired by the social behavior of birds. In this approach, a population of candidate solutions moves around in the search space in order to find the optimal solution. One of the most prominent birds' behaviors is their group's ability to locate a desirable position in the given area [41]. Suitable computational complexity, few parameters, and global search ability are among the advantages of this algorithm which make PSO one of the most successful existing algorithms. Because the PSO has the powerful search strength, many researchers have used its binary version (BPSO) as a feature subset generator and achieved a good result. The BPSO is first proposed in 1997. In this algorithm, each solution is a particle in the swarm which has a location in the search space and the algorithm is stopped when the desired solution is obtained, or the number of iterations reaches the predetermined value (maximum iteration).

3. The proposed method

As mentioned, increasing dimension in high-dimensional data like micro-array data, which have a small number of samples might cause the curse of dimensionality and confusing the classifier. Thus, feature selection is a necessary procedure for working with such data. Employing advantages of filter along with wrapper methods in hybrid-ensemble methods can improve feature selection performance.

There are several challenges in the hybrid-ensemble framework, such as determining number of ensemble lines, selecting effective filter and wrapper methods and selecting suitable integration method. In this section, by defining the problem as an optimization task, it will be tried to find an (sub) optimal solution by the genetic algorithm to optimize number of lines, type of filter/wrapper technique at each line and integration type of lines' outcomes.

3.1 Optimization using genetic algorithm

In this paper, the genetic algorithm is applied to obtain an optimal hybrid-ensemble feature selection method. To do so, first we should encode the problem. Therefore, we need to determine the final framework which should be optimized. Therefore, the framework shown in Fig. 2-a has been used as the basic framework.

3.2 Chromosome representation

As mentioned, to implement the genetic optimization algorithm, two chromosome representations including binary or integer can be used. In this work, the integer chromosome representation is used. Fig. 4 is used to describe the basic representation of the solution, which

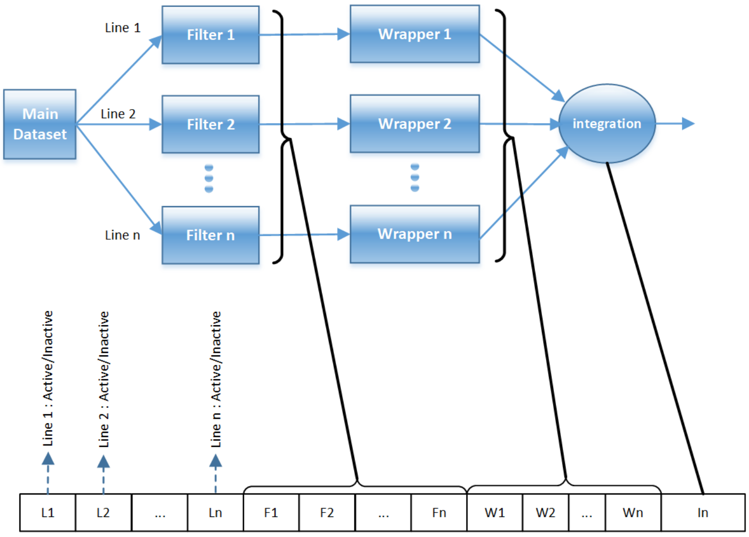


Fig. 4 The block diagram of the proposed method

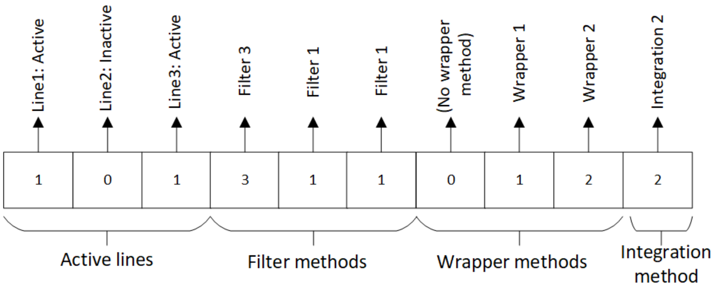


Fig. 5 An example of chromosome string

shows the hybrid-ensemble method used for optimization in this paper. As can be seen, dimensions of the main dataset are reduced using several filter methods in n separate lines, each of which is applied separately to the main dataset, and then MH-Wrapper methods are applied to each line and then results are combined. If the number of lines in the ensemble framework is n, the maximum number of filter and MH-Wrapper methods which can be employed in the hybrid-ensemble framework is n since, for each filter algorithm, one or no wrapper algorithm can be considered.

For hybrid-ensemble method optimization, a string of chromosomes of length $3n$+1 is considered where the first n genes indicate active/inactive lines, second n genes show type of filter methods for lines i.e., each gene corresponds to each line, third n genes show corresponding MH-Wrapper methods for lines and the last gene is used for combination (integration) method. An overview of considered chromosome string is shown in Fig. 4.

In Fig. 4 the variable Li indicates that line i is active or inactive. If $Li = 0$, line i and the corresponding filter/wrapper methods are ignored. Fi and Wi represent the filter and wrapper methods used in line i, respectively. Moreover, $Wi = 0$ indicates that no wrapper method is used in the considered block (line) and filter method is applied to data alone.

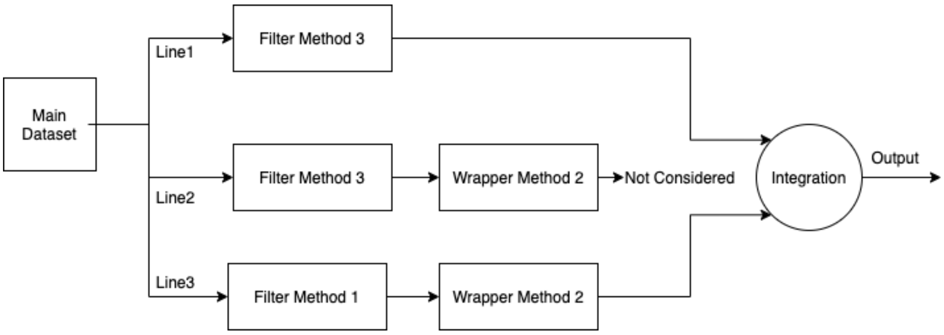


Fig. 6 Implement the example of chromosome string mentioned in Fig. 5

Table 1 Microarray datasets used for benchmarking

NO	Dataset	#Features	#Samples	#Classes
1	11_Tumors	12533	174	11
2	9_Tumors	5726	60	9
3	Brain_Tumors1	5920	90	5
4	Brain_Tumors2	10367	50	4
5	Breast	24481	60	2
6	CNS	7129	72	2
7	Colon	2000	62	2
8	Leukemia	7129	72	2
9	Leukemia1	5327	72	3
10	Leukemia2	11225	72	3
11	Lung_Cancer	12600	253	2
12	Lung_Cancer2	12533	181	2
13	SRBCT	2308	83	4
14	Prostate_Tumor	10509	102	2
15	Prostate_Cancer2-26-12600	12600	136	2
16	DLBCL	5469	83	4
17	DLBCL2-4026	4026	77	2
18	Ovarian	15154	253	2

For example, if $n = 3$, which means that the desired hybrid ensemble framework consists of 3 parallel feature selection layers (lines), and we have the chromosome string as Fig. 5, the block diagram of the desired feature selection framework would be as shown in Fig. 6.

To obtain the final framework, the real genetic optimization algorithm is applied to suggest an optimal feature selection method. Filter methods considered to be employed in the hybrid-ensemble framework include: ReliefF, FCBF, Mrmr, Fisher score and Information Gain (IG). Investigated wrapper methods include ACO, BACO, ABACO, GA, BPSO and IBGSA which perform well in high-dimensional data. In addition, two logic operators AND and OR are used for integration.

3.3 Fitness function

Desired metrics of an optimal method for feature selection include high classification accuracy and small number of features. Moreover, reducing the number of ensemble classes might be an effective factor in feature selection since it reduces computational complexity. Thus, a fitness function by combining these metrics can be used in GA to find the optimal structure of the Fig. 2(a).

4. Experiments and results

In order to solve feature selection problem using the metaheuristic algorithms (i.e., in the MH-Wrapper box), each solution of the problem is defined as a string of 0 and 1, whose length is equal to number of the features; 1 indicates that the corresponding feature is selected and 0 indicates that the feature is not selected. In order to determine the fitness of each solution, features of each solution are given to a classifier. Feature subset which results in higher classification rate, achieve a higher score. All experiments are performed on a 2.30 GHz Intel Core-i5 CPU with 4GB of RAM.

4.1 datasets

In order to perform the experiments, 15 micro-array data whose general characteristics are given in Table 1, are employed. This table indicates the variety of samples, number of features and number of groups. Supplementary information and datasets used in references can be seen beside each one. The datasets are accessible in [43] and [44].

As mentioned, micro-array data have many features and a few number of samples, which can be seen in Table 1. Among data sets of Table 1, the minimum number of features is 2000 which corresponds to Colon data set with 47 samples, and maximum number of features is 15154 which corresponds to the Ovarian dataset with 253 samples. It should be mentioned that in this paper, multi-class data sets like Lung-cancer and SRBCT are investigated too.

4.2 Adjusting parameters

In execution of GA for optimization, the population size, crossover probability and mutation probability are set to 50, 0.9 and 0.05, respectively. Moreover, the parameters of six employed metaheuristic algorithms (used in MH-Wrapper box) are as follows: in all methods, the initial population was set to 50. The parameters of BGA algorithm and different ACO algorithms including ACO, BACO, ABACO are set as reported in [40] which yield remarkable results. In IBGSA and BPSO the parameters, are set as reported in [39]. It should be mentioned that in all of the above metaheuristic methods, k-NN classifier ($k = 1$) is used as classifier for evaluating features set and in all of the above metaheuristic algorithms, the number of features are determined while executing the algorithm.

4.3 Performance evaluation metrics

In order to evaluate the proposed algorithm and compare it with other algorithms, classification accuracy rate is used. Classification accuracy rate is obtained as follows:

$$acc = \frac{number\ \ of\ \ correct\ \ classified\ \ sample}{Total\ \ test\ \ samples} \tag{5}$$

A higher classification accuracy rate indicates that the selected features are more suitable. Another important metric for comparing the performance of different methods is $Gmean$. This metric is a standard metric for comparing different methods because it considers the effect of dimension reduction rate Fr and classification accuracy acc. This metric is calculated as follows

$$Gmean = \sqrt{acc \times Fr} \tag{6}$$

Since the classification accuracy rate alone cannot be a criterion for the superiority or weakness of a method, criteria such as the feature reduction should be used alongside it. In this paper, the geometric mean of these two criteria is used to consider the effect of two criteria of the classification accuracy rate and the parameter of feature reduction simultaneously. In Eq. (7), Fr is dimension reduction parameter which is calculated using the following equation

$$Fr = \frac{p - q}{p} = 1 - \frac{q}{p} \tag{7}$$

In this equation, p is the total number of features and q is the number of selected features. The closer is Fr to 1, the number of features is reduced more. Considering equations 6 and 7, the higher is $Gmean$ of a method, that method is more desirable. In this paper, in addition to classification accuracy rate, $Gmean$ is also used as a performance evaluation metric.

4.4. Fitness function

As mentioned before, designing fitness function has an effective role in optimizing the structure of Fig. 2(a). In this paper, the fitness function employed in GA is as follows

$$fit = \frac{1}{Nst}(c_1 \times acc + c_2 \times Fr) \tag{8}$$

where c_1 and $c2$ are coefficients that determine the effect of classification accuracy and dimension reduction and Nst is the number of selected ensemble lines in the optimization process. The purpose of applying optimization is to maximize the fitness values of Eq. (8).

4.5. Optimized structure

To find the optimized structure, the genetic optimization algorithm is applied to 4 datasets including Brain-Tumor1, Brain-Tumor2, Colon and Lung-Cancer. In order to evaluate the efficacy of methods, the classification accuracy through 30 independent executions are reported and 5-CV evaluation method is employed. It should be mentioned that in the above experiments, the threshold of filter methods like ReliefF, IG, Fisher-Score and mRMR is set to 0.005.

In Tables 2 to 4, five superior results obtained from applying GA on 4 mentioned datasets and their accuracy rate are given. Considering results, in these four data sets 4, 3, 3 and 5 solutions obtained in independent runs of GA for Colon, Lung-Cancer, Brain-Tumor1 and Brain-Tumor2 are respectively belong to the structures with one line. This is due to fitness function of Eq. (8) which causes the number of lines decreases and we design it to reduce the computational

Table 2 The results of the implementation of the genetic optimization on COLON dataset

	Method	Integration	ACC
1	FCBF+ ABACO	-	0.9901
2	FCBF+IBGSA	-	0.9801
3	ReliefF+BACO	-	0.9821
4	FCBF+GA	-	0.9679
5	ReliefF+GA Mrmr+GA	OR	0.9538

Table 3 The results of the implementation of the genetic optimization on LUNG_CANCER dataset

	Method	Integration	ACC
1	FCBF + ABACO	-	0.9804
2	FCBF+ABACO F-SCORE+ABACO	OR	0.9802
3	FCBF+IBGSA	-	0.9801
4	FCBF+BACO ReliefF+BACO IG+BACO	OR	0.9755
5	F-SCORE+BACO	-	0.9753

Table 4 The results of the implementation of the genetic optimization on BRAIN_TUMOR 1 dataset

	Method	Integration	ACC
1	FCBF+ABACO	-	0.9667
2	F-SCORE+ ABACO, FCBF + ABACO	OR	0.9660
3	FCBF+IBGSA	-	0.9654
4	F-SCORE+ ABACO, FCBF + ABACO	OR	0.9654
5	F-SCORE+ BPSO, FCBF + BPSO	OR	0.9654

Table 5 The results of the implementation of the genetic optimization on BRAIN_TUMOR 2 dataset

	Method	Integration	ACC
1	ReliefF+BPSO	-	1
2	ReliefF+ABACO	-	1
3	ReliefF+GA	-	1
4	FCBF+ABACO	-	0.9833
5	FCBF+IBGSA	-	0.9833

complexity. Before discussing the results given in these tables, it is necessary to describe the results. For example, the solution 5 in Table 2 shows the optimum solution is a two lines ensemble structure where in the first line ReliefF is used as filter method and GA is used as metaheuristic in

Table 6 Comparing FABACO with three up-to-date methods with 1/3 test and 2/3 training data

Dataset	DRF0-CFS [45]	E-Min [46]	E-geoMean [47]	FABACO
Breast	73.68	-	-	86.88
CNS	70	60	75	88.5
Colon	90	85	80	93.81
Leukemia	91.18	91.18	91.18	95.2
Lung_Cancer	98.66	--	-	98.66
Lung_Cancer2	-	99.33	96.64	97.22
Prostate_Cancer2	85.29	73.53	73.53	96
DLBCL2	93.33	86.67	86.67	93.5
Ovarian	100	100	100	100

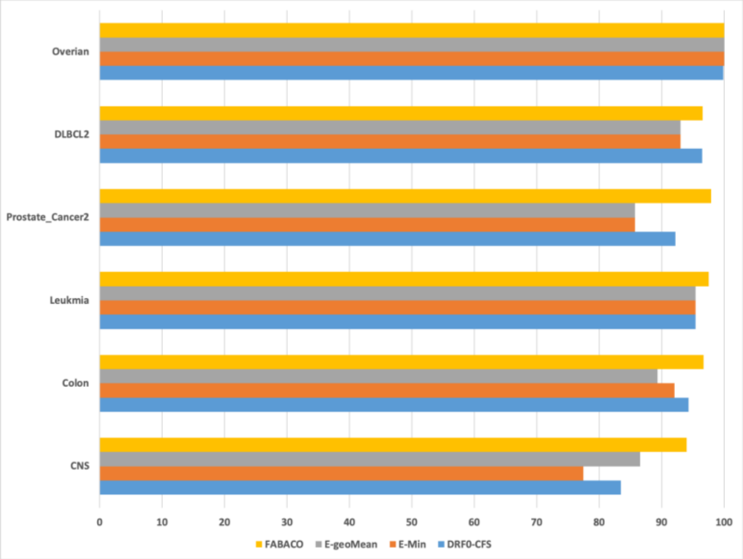


Fig. 7 Gmean obtained by applying DRF0-CFS, E-Min, E-geoMean and FABACO on 6 micro-array datasets using LOOCV evaluation method

MH-Wrapper box and in the second line Mrmr is used as filter and GA is used as metaheuristic in MH-Wrapper box. Furthermore, in this structure the “OR” is used for results integration.

The results presented in these tables show that FCBF+ABACO (one line structure including FCBF filter method and ABACO as metaheuristic in MH-Wrapper box) has given best results in all 4 datasets. As can be seen, this hybrid method is introduced as the optimal solution in COLON, Lung-Cancer and Brain-Tumor1 data sets and, in Brain-Tumor2 data set, it achieves ranked third. Moreover, FCBF+IBGSA is also seen in superior results of all 4 datasets. As a final result, we select one line hybrid method of FCBF+ABACO as final structure of the proposed method.

4.6 Comparison with State-of-the-Art algorithms and discussion

Considering the results obtained in the previous section, FCBF+ABACO is obtained as the

Table 7 Comparing Classification accuracy in FABACO and IG-ISSO and MOBBBO by LOOCV evaluation method

Dataset	MOBBBO [48]	IG-ISSO [47]	FABACO
11_Tumors	92.41	95.92	98.85
9_Tumors	80.50	91.67	98.33
Brain_Tumors1	96.67	98	98.89
Brain_Tumors2	99.80	99.8	98
Leukemia1	100	100	100
Leukemia2	100	100	100
Lung_Cancer	98.47	99.41	98
SRBCT	100	100	100
Prostate_Tumor	98.33	98.82	99.02
DLBCL	100	100	100

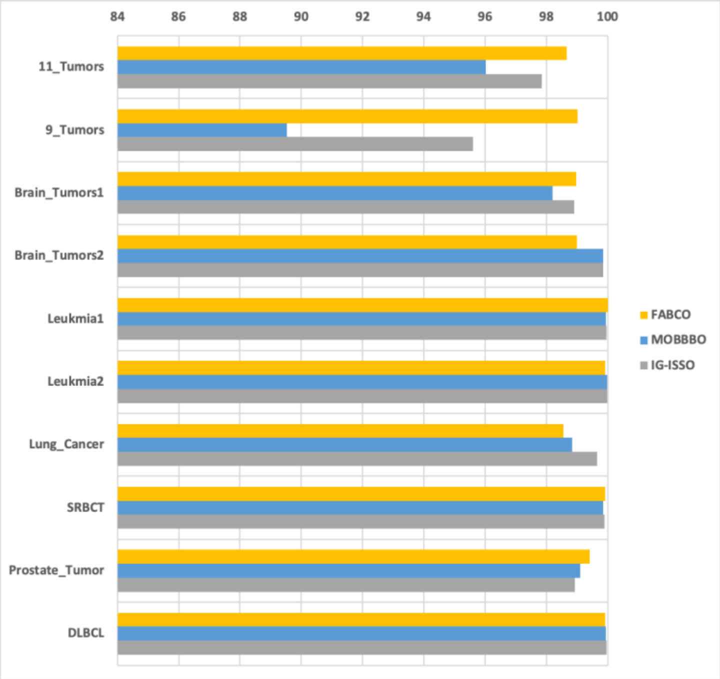


Fig. 8 Gmean obtained by applying MOBBI, IG-ISSO and FABACO to 10 micro-array data sets using LOOCV evaluation method

optimal method based on the fitness function of Eq. (8) after applying GA. This method is called FABACO in brief. In this section, the performance of this method is investigated through comparing its performance with several up-to-date methods. It should be mentioned that all results of the proposed method are obtained by k-NN classifier.

Table 6 shows results of applying FABACO to nine micro-array data sets which are divided randomly into 1/3 and 2/3 for test and training data. Obtained results of the proposed FABACO are

Table 8 Comparing classification accuracy of FABACO with several up-to-date methods using 5-CV methods

Dataset	APCES [49]	T-SS [50]	FABACO
Breast	73.38	89.3	94.64
CNS	-	65.5	94.4
Colon	84.4	87.12	96.69
Leukemia	97.5	97	96.38
Lung_Cancer2	94.2	-	98.01
Prostate_Cancer2	-	-	95.24
DLBCL2	-	-	96.7
Ovarian	99.0	-	100

Table 9 Comparing classification accuracy of FABACO with several up-to-date methods using 10-CV methods

Dataset	E1-CP [51]	Meta-ensemble [53]	BDE-C45$_{\text{Rank}}$ [52]	FABACO
Breast	-	79.87	-	98
CNS	70	90.19	-	95.67
Colon	85	99.21	93.8	98.88
Leukemia	85.29	94.12	94.1	98.75
Lung_Cancer2	98.66	97.99	96.1	98.8
Prostate_Cancer2	97.06	52.94	73.5	95.78
DLBCL2	93.33	96.13	100	97.42
Ovarian	100	99.95	97.5	100

compared with those of DRF0-CFS [45], E-Min [46] and E-geoMean [47].

It is clear that in Table 6, FABACO has a better performance rather than other three methods. It should be mentioned that Breast and Prostate-Cancer2 data sets with 24481 and 12600 features are of big challenges in machine learning i.e., the selection of desired features with high classification accuracy is difficult for these two data sets. However, FABACO has obtained a desirable classification accuracy in these datasets.

In Fig. 7, the 4 mentioned methods are compared in terms of G-mean in 6 common datasets. As mentioned, G-mean is one of the important metrics in evaluating the performance of the method which considers dimension reduction and accuracy.

As can be seen in Fig. 7, FABACO has obtained better results compared to other three methods in terms of $Gmean$. This superiority can be seen for CNS, Prostate-Cancer2 and DLBCL, obviously.

Table 7 shows results of applying FABACO to 10 datasets, which are compared using LOO-CV evaluation method with IG-ISSO [47] and MOBBBO [48].

Considering Table 7, FABACO using LOOCV evaluation method, which is one of the most accurate evaluation methods, has obtained better results in eight data sets compared to MOBBBO [48] and IG-ISSO [47]. Fig. 8 shows $Gmean$ for employed data sets and methods in this case.

Considering Fig. 8, the proposed FABACO has obtained desirable results in comparison with

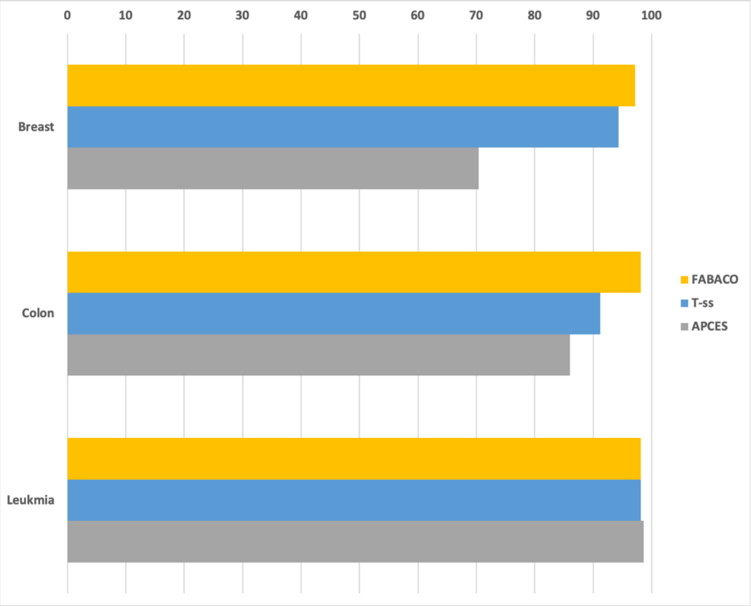


Fig. 9 Gmean obtained from applying APCES-Tss and FABACO on six micro-array datasets using 5-CV evaluation method

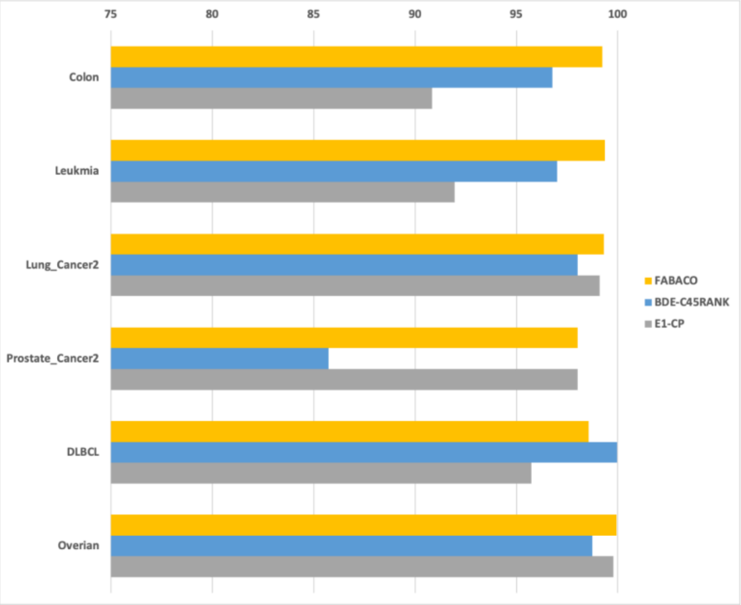


Fig. 10 Gmean obtained from applying E1-CP, BDE-C45Rank and FABACO on six micro-array datasets using 10-CV evaluation method

two competing methods except in Brain-Tumors 2 and Lung-Cancer data sets. As shown by this figure, FABACO has a uniform path for $Gmean$ criterion, which shows more stability of this method compared to competing methods.

Tables 8 and 9 compare the performance of the proposed method with APCES [49], T-SS [50], E1-CP [51], BDE-C45RANK [52] and Meta-ensemble [53] using 5 cross-fold validation (5-CV) and 10 cross-fold validation (10-CV).

Tables 8 and Table 9 indicate FABACO has better performance and obtained better classification accuracy compared to 5 other methods. Furthermore, the ability of the proposed method is shown in Breast data set which is a high-dimensional problem. In the case of 5-CV evaluation method (Table 8), among competing methods, FABACO has only less classification accuracy in Lukemia data set. In addition, in the case of 10-CV evaluation method (Table 9), among competing methods, the proposed method has obtained significant superiority in Breast, CNS, Leukemia, Lung-Cancer2 and Ovarian data sets. It should be mentioned that above comparisons, method type (filter, wrapper, hybrid or ensemble) is ignored and it should be considered that filter methods have higher speed compared to hybrid and wrapper methods and hybrid methods have better speed than wrapper methods. Figs. 9 and 10 compare of the competing methods in this case based on 5-CV and 10-CV methods.

Figs. 9 and 10 show that FABACO has obtained desirable results in terms of $Gmean$. FABACO has obtained lower $Gmean$ value in DLBCL2 compared to BDE-C45Rank, while it has obtained higher values in other data sets. According to Figure 10, BDE-C45Rank has faced the significant reduction of $Gmean$ in Prostate-Cancer 2, while FABACO has shown suitable performance without any changes in $Gmean$.

The experimental results reveal that FABACO not only performs well in classification accuracy, which is a common metric, but also it performs well in terms of $Gmean$. This method has shown desirable performance on data sets like Prostate-Cancer2 and Breast.

5. Conclusions

Today, with the increase of data dimensions, variety of contexts including machine learning, informatics, and medicine are encountered with big challenges. However, reducing data dimension is the basic method in handling high-dimensional data, because by reducing dimensions, applying many of the operations on data is facilitated. With the advent of medical data such as micro-array data, it has become important to reduce the data dimension and to select effective features. Micro array data are a kind of large-scale data with few samples that has been one of the challenges in medicine and bioinformatics. Therefore, the selection of effective genes in these data is one of the most complicated and significant processes for many researchers.

The purpose of this paper is to search and find an optimal feature selection method among filter, wrapper and hybrid methods through ensemble approaches. Thus, genetic optimization algorithm is first applied to several data sets to search for the optimal method. Then, the optimal method is extracted through comparing superior results of sample data sets. In this paper, FABACO is suggested as the optimal method through performing the above experiments. To evaluate the ability of the FABACO, it is applied to 18 micro-array data sets with different evaluation metrics and results are compared with 10 up-to-date methods. After comparing different methods using classification accuracy and Gmean metrics, the performance of FABACO is investigated. One of the important points in these comparisons is the acceptable performance of this method in execution on challenging datasets like Prostate-Cancer2 and Breast which has shown desirable performance, unlike other methods. Results obtained from applying this method to other micro-array data shows the capability of this method in handling high-dimensional data.

References

Abeel, T., Helleputte, T., Van de Peer, Y., Dupont, P. and Saeys, Y. (2010), "Robust biomarker identification for cancer diagnosis with ensemble feature selection methods", *Bioinformatics*, **26**(3), 392-398. https://doi.org/10.1093/bioinformatics/btp630.

Apolloni, J., Leguizamón, G. and Alba, E. (2016), "Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments", *Appl. Soft Comput.*, **38**, 922-932. https://doi.org/10.1016/j.asoc.2015.10.037.

Bolón-Canedo, V., Remeseiro, B., Sánchez-Maroño, N. and Alonso-Betanzos, A. (2014), "mC-ReliefF", *In Proceedings of the 6th International Conference on Agents and Artificial Intelligence.*

Bolón-Canedo, V., Sánchez-Marono, N. and Alonso-Betanzos, A. (2014), "Data classification using an ensemble of filters", *Neurocomputing*, **135**, 13-20. https://doi.org/10.1016/j.neucom.2013.03.067.

Bolón-Canedo, V., Sánchez-Maroño, N. and Alonso-Betanzos, A. (2015), "Distributed feature selection: An application to microarray data classification", *Appl. Soft Comput.*, **30**, 136-150. https://doi.org/10.1016/j.asoc.2015.01.035.

Bolón-Canedo, V., Sánchez-Maroño, N. and Alonso-Betanzos, A. (2012), "An ensemble of filters and classifiers for microarray data classification", *Pattern Recognition*, **45**(1), 531-539. https://doi.org/10.1016/j.patcog.2011.06.006.

V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "Data classification using an ensemble of filters," *Neurocomputing,* vol. 135, pp. 13-20, 2014.

Bolón-Canedo, V., Sánchez-Marono, N., Alonso-Betanzos, A., Benítez, J. M. and Herrera, F. (2014), "A review of microarray datasets and applied feature selection methods", *Inform. Sci.*, **282**, 111-135. https://doi.org/10.1016/j.ins.2014.05.042.

Brahim, A.B. and Limam, M. (2013), "Robust ensemble feature selection for high dimensional data sets. *In 2013 International Conference on High Performance Computing & Simulation (HPCS)*.

Canul-Reich, J., Hall, L.O., Goldgof, D.B., Korecki, J.N. and Eschrich, S. (2012), "Iterative feature perturbation as a gene selector for microarray data" *Int. J. Pattern Recognition Artificial Intelligence*, **26**(5), 1260003. https://doi.org/10.1142/S0218001412600038.

Chuang, L.Y., Yang, C.H., Wu, K.C. and Yang, C.H. (2011), "A hybrid feature selection method for DNA microarray data", *Comput. Biology Medicine*, **41**(4), 228-237. https://doi.org/10.1016/j.compbiomed.2011.02.004.

Dataset Repository (2014), Bioinformatics Research Group, http://www.upo.es/eps/bigs/datasets.html.

Gu, Q., Li, Z. and Han, J. (2012), "Generalized fisher score for feature selection. arXiv preprint arXiv:1202.3725.

Hall, M.A. (1999), *Correlation-based feature selection for machine learning*, Ph.D. Dissertation, The University of Waikato, Hamilton, NewZealand.

I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning,* vol. 46, pp. 389-422, 2002. https://doi.org/10.1016/j.knosys.2011.04.014.

Kamkar, I., Gupta, S.K., Phung, D. and Venkatesh, S. (2015), "Stable feature selection for clinical prediction: Exploiting ICD tree structure using Tree-Lasso", *J. Biomedical Informatics*, **53**, 277-290. https://doi.org/10.1016/j.jbi.2014.11.013.

Kashef, S. and Nezamabadi-pour, H. (2015), "An advanced ACO algorithm for feature subset selection", *Neurocomputing*, **147**, 271-279. https://doi.org/10.1016/j.neucom.2014.06.067.

Kira, K. and Rendell, L.A. (1992), "The feature selection problem: Traditional methods and a new algorithm", *In Aaai*, **2**, 129-134.

Kononenko, I. (1994), "Estimating attributes: analysis and extensions of RELIEF. *In European conference on machine learning*, Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-57868-4_57.

Lai, C.M., Yeh, W.C. and Chang, C.Y. (2016), "Gene selection using information gain and improved simplified swarm optimization", *Neurocomputing*, **218**, 331-338.

https://doi.org/10.1016/j.neucom.2016.08.089.

Li, X. and Yin, M. (2013), "Multiobjective binary biogeography based optimization for feature selection using gene expression data", *IEEE Transactions NanoBiosci*., **12**(4), 343-353. https://doi.org/10.1109/TNB.2013.2294716.

Mahendran, N., Vincent, P.D.R., Srinivasan, K. and Chang, C.Y. (2020), "Machine learning based computational gene selection models: A survey, performance evaluation, open issues, and future research directions.", *Front. Genetics*, 11. https://dx.doi.org/10.3389%2Ffgene.2020.603808.

Maldonado, S., Weber, R. and Basak, J. (2011), "Simultaneous feature selection and classification using kernel-penalized support vector machines", *Inform. Sci*., **181**(1), 115-128. https://doi.org/10.1016/j.ins.2010.08.047.

Meng, J., Hao, H. and Luan, Y. (2016), "Classifier ensemble selection based on affinity propagation clustering", *J. Biomedical Informatics*, **60**, 234-242. https://doi.org/10.1016/j.jbi.2016.02.010.

Meyer, P.E., Schretter, C. and Bontempi, G. (2008), "Information-theoretic feature selection in microarray data using variable complementarity", *IEEE J. Selected Topics Signal Processing*, **2**(3), 261-274. https://doi.org/10.1109/JSTSP.2008.923858.

Mundra, P.A. and Rajapakse, J.C. (2009), "SVM-RFE with MRMR filter for gene selection", *IEEE Transactions Nanobiosci*., **9**(1), 31-37. https://doi.org/10.1109/TNB.2009.2035284.

Mundra, P.A. and Rajapakse, J.C. (2016), "Gene and sample selection using T-score with sample selection", *J. Biomedical Informatics*, **59**, 31-41. https://doi.org/10.1016/j.jbi.2015.11.003.

Nagi, S. and hattacharyya, D.K. (2013), "Classification of microarray cancer data using ensemble approach", *Network Model. Analysis Heal. Inform. Bioinform*., **2**(3), 159-173. https://doi.org/10.1007/s13721-013-0034-x.

Nezamabadi-pour, H., Rostami-Shahrbabaki, M. and Maghfoori-Farsangi, M. (2008), "Binary particle swarm optimization: challenges and new solutions", *CSI J. Comput. Sci. Eng*., **6**(1), 21-32.

Peng, H., Long, F. and Ding, C. (2005), "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy", *IEEE Transactions on pattern analysis and machine intelligence*, **27**(8), 1226-1238. https://doi.org/10.1109/TPAMI.2005.159.

Prasartvit, T., Banharnsakun, A., Kaewkamnerdpong, B. and Achalakul, T. (2013), "Reducing bioinformatics data dimension with ABC-kNN", *Neurocomputing*, **116**, 367-381. https://doi.org/10.1016/j.neucom.2012.01.045.

Rashedi, E., Nezamabadi-Pour, H. and Saryazdi, S. (2009), "GSA: a gravitational search algorithm", Information sciences, **179**(13), 2232-2248. https://doi.org/10.1016/j.ins.2009.03.004.

Rashedi, E., Nezamabadi-Pour, H. and Saryazdi, S. (2010), "BGSA: binary gravitational search algorithm", *Nat. Comput*., **9**(3), 727-745. https://doi.org/10.1007/s11047-009-9175-3.

Rouhi, A. and Nezamabadi-pour, H. (2016), "A hybrid method for dimensionality reduction in microarray data based on advanced binary ant colony algorithm", *In 2016 1st Conference on Swarm Intelligence and Evolutionary Computation (CSIEC)*.

Rouhi, A. and Nezamabadi-pour, H. (2017), "A hybrid-ensemble based framework for microarray data gene selection", *Int. J. Data Min. Bioinform*., **19**(3), 221-242. https://doi.org/10.1504/IJDMB.2017.090987.

Rouhi, A. and Nezamabadi-pour, H. (2017), "A hybrid feature selection approach based on ensemble method for high-dimensional data", *In 2017 2nd Conference on Swarm Intelligence and Evolutionary Computation (CSIEC)*,16-20. https://doi.org/10.1109/CSIEC.2017.7940163.

Rouhi, A., Spitale, M., Catania, F., Cosentino, G., Gelsomini, M. and Garzotto, F. (2019). "Emotify: emotional game for children with autism spectrum disorder based-on machine learning", *In Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion*. 31-32.

Ruiz, R., Riquelme, J.C. and Aguilar-Ruiz, J.S. (2006), "Incremental wrapper-based gene selection from microarray data for cancer classification", *Pattern Recognition*, **39**(12), 2383-2392. https://doi.org/10.1016/j.patcog.2005.11.001.

Saeys, Y., Inza, I. and Larrañaga, P. (2007), "A review of feature selection techniques in bioinformatics", *Bioinformatics*, **23**(19), 2507-2517. https://doi.org/10.1093/bioinformatics/btm344.

Seijo Pardo, B., Bolón-Canedo, V., and Alonso-Betanzos, A. (2016), "Using a feature selection ensemble on

DNA microarray datasets", In *ESANN*.

Sharma, A., Imoto, S. and Miyano, S. (2011), "A top-r feature selection algorithm for microarray gene expression data", *IEEE/ACM Transactions Comput. Biology Bioinformatics*, **9**(3), 754-764. https://doi.org/10.1109/TCBB.2011.151.

Shreem, S.S., Abdullah, S., Nazri, M.Z.A. and Alzaqebah, M.A.L.E.K. (2012). "Hybridizing ReliefF, MRMR filters and GA wrapper approaches for gene selection", *J. Theor. Appl. Inf. Technol*, **46**(2), 1034-1039.

Statnikov, A., Aliferis, C.F., Tsamardinos, I. (2005), Gems: Gene Expression Model Selector, http://www.gems-system.org.

Tabakhi, S., Najafi, A., Ranjbar, R. and Moradi, P. (2015), "Gene selection for microarray data classification using a novel ant colony optimization", *Neurocomputing*, **168**, 1024-1036. https://doi.org/10.1016/j.neucom.2015.05.022.

Uğuz, H. (2011), "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm", *Knowledge-Based Systems*, **24**(7), 1024-1032.

Wang, J., Wu, L., Kong, J., Li, Y. and Zhang, B. (2013), "Maximum weight and minimum redundancy: a novel framework for feature subset selection.", *Pattern Recognition*, **46**(6), 1616-1627. https://doi.org/10.1016/j.patcog.2012.11.025.

Yang, F. and Mao, K.Z. (2010), "Robust feature selection for microarray data based on multicriterion fusion", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **8**(4), 1080-1092.

You, W., Yang, Z. and Ji, G. (2014), "PLS-based recursive feature elimination for high-dimensional small sample", *Knowl.Based Syst*., **55**, 15-28. https://doi.org/10.1016/j.knosys.2013.10.004.

Yu, L. and Liu, H. (2003), "Feature selection for high-dimensional data: A fast correlation-based filter solution. *In Proceedings of the 20th international conference on machine learning (ICML-03)*, 856-863.

PA