Applied Soft Computing 56 (2017) 94–106

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Binary black hole algorithm for feature selection and classification on biological data

Elnaz Pashaei, Nizamettin Aydin*

Yildiz Technical University, Computer Engineering Department, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 14 September 2016
Received in revised form 4 February 2017
Accepted 2 March 2017
Available online 6 March 2017

Keywords:
Feature selection
Black hole optimization algorithm
Decision tree algorithms
Particle swarm optimization
Biomedical data

ABSTRACT

Biological data often consist of redundant and irrelevant features. These features can lead to misleading in modeling the algorithms and overfitting problem. Without a feature selection method, it is difficult for the existing models to accurately capture the patterns on data. The aim of feature selection is to choose a small number of relevant or significant features to enhance the performance of the classification. Existing feature selection methods suffer from the problems such as becoming stuck in local optima and being computationally expensive. To solve these problems, an efficient global search technique is needed.

Black Hole Algorithm (BHA) is an efficient and new global search technique, inspired by the behavior of black hole, which is being applied to solve several optimization problems. However, the potential of BHA for feature selection has not been investigated yet. This paper proposes a Binary version of Black Hole Algorithm called BBHA for solving feature selection problem in biological data. The BBHA is an extension of existing BHA through appropriate binarization. Moreover, the performances of six well-known decision tree classifiers (Random Forest (RF), Bagging, C5.0, C4.5, Boosted C5.0, and CART) are compared in this study to employ the best one as an evaluator of proposed algorithm.

The performance of the proposed algorithm is tested upon eight publicly available biological datasets and is compared with Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Simulated Annealing (SA), and Correlation based Feature Selection (CFS) in terms of accuracy, sensitivity, specificity, Matthews' Correlation Coefficient (MCC), and Area Under the receiver operating characteristic (ROC) Curve (AUC). In order to verify the applicability and generality of the BBHA, it was integrated with Naive Bayes (NB) classifier and applied on further datasets on the text and image domains.

The experimental results confirm that the performance of RF is better than the other decision tree algorithms and the proposed BBHA wrapper based feature selection method is superior to BPSO, GA, SA, and CFS in terms of all criteria. BBHA gives significantly better performance than the BPSO and GA in terms of CPU Time, the number of parameters for configuring the model, and the number of chosen optimized features. Also, BBHA has competitive or better performance than the other methods in the literature.

1. Introduction

Biological data such as microarrays can contain many irrelevant and redundant features (genes). These features may cause misleading in the modeling of algorithms for cancer classification and overfitting with long training times. In order to obtain optimal performance with short training times and reduce memory requirement, the Feature Selection (FS) process should be considered to use as a pre-process step in machine learning before applying classifiers to a dataset [1].

There are many studies based on FS methods. The FS algorithms are broadly categorized into three groups: filter, wrapper, and embedded approaches. This categorization is based on whether or not they are combined with a specific learning algorithm (classifier).

Filter based FS approaches consider the features independently and remove irrelevant features according to the statistical characteristics of the data. The *t*-test, chi-squared test, information gain, and Correlation based FS (CFS) are some well-known filter approaches [2,3].

Wrapper based FS methods apply a specific machine learning algorithm to evaluate the score of selected feature subsets. These methods utilize Cross-Validation (CV) schema to train learning algorithm [4]. Comparing the wrapper methods to the filter

* Corresponding author.
E-mail address: naydin@yildiz.edu.tr (N. Aydin).

http://dx.doi.org/10.1016/j.asoc.2017.03.002
1568-4946/

approaches, wrapper methods are more accurate than the filter approaches because of considering the interactions among the features. However, they are computationally more expensive and the performances of them strongly depend on the given learning algorithm. Embedded based FS methods are special cases of wrapper methods that are characterized by a deeper interaction between the construction of the learning algorithm and the FS. In these methods the FS algorithm is always regarded as a component in the learning model. The Decision Tree (DT) algorithms, such as C4.5 [5] and Classification and Regression tree (CART) [6] are known as the most typical embedded besed FS approaches.

FS is known as an NP-hard and combinatorial problem. Hence, meta-heuristic methods are more appropriate to untie this laborious problem because of their population-based characteristics. Various stochastic global search algorithms have been used to solve FS problem on medical datasets, such as, Genetic Algorithm (GA) [7], Binary Particle Swarm Optimization (BPSO) [8,9], hybrid of GA and PSO algorithms [10], and Simulated Annealing (SA) [11].

GA is a stochastic global search algorithm which has been naturally used for FS. This algorithm works based on two genetic operators; crossover and mutation. GA has the ability to solve complex and non-linear problems. An important disadvantage of GA is its unguided mutation which is the only reason of a very slow convergence of GA. It also has a lot of parameters for tuning [7].

BPSO is a population-based stochastic optimization algorithm which tries to solve the FS problem by simulating the social behavior of fish schooling or bird flocking. BPSO is more computationally efficient than GA and mostly provides better solution. The main disadvantage of PSO is that it is more likely to fall into local optimum. The hybrid of GA and PSO has been developed to solve this problem [10].

Simulated Annealing (SA) is a probabilistic algorithm based on the Metropolis algorithm which was proposed by Kirkpatrick in 1983 and has been specially designed for feature subset selection. GA and SA are more or less equivalent with respect to the quality of the solutions. The SA is not efficient in exploring large solution spaces because of randomly seeding and needs large number of parameters for tuning.

The congenital drawbacks of the mentioned optimization algorithms still puzzle themselves. Therefore, to better address FS problems, a simple and efficient global search technique is needed. The Black Hole Algorithm (BHA) is one of the newest meta-heuristic methods based on the swarm intelligence [12]. This algorithm was discovered by simulating the behavior of black hole in outer space. In the real world black hole is an object of extreme density with intense gravitational attraction. The black hole's gravitational attraction swallows all objects if they come near enough. Because of BHA's characteristics including powerful optimal performance, single parameter, and fast convergence, the BHA has been used for solving a number of problems such as clustering [12], multi-objection reactive power dispatch problem [13], optimization problem [14], spam detection [15], and optimal coordination of digital overcurrent relays problem [16]. A comprehensive study of black hole approach and its applications in different research fields is provided in [17]. However, to the best of our knowledge there is no reported research related to FS using the BHA in literature.

Another version of BHA has been investigated in Ref. [18]. Based on the proposed method in Ref. [18] the binary version of BHA has been introduced in Ref. [19] and used for solving benchmark functions. It considers three parameters for each star; position, mass and electrical charge. It tries to calculate gravitational force (Newton's law) for the global search based on the mass parameter and electricity force (Coulomb's law) for the local search based on the electrical charge parameter. It uses Hawking radiation as mutation step to avoid algorithm from being stuck in local optima. Also, it utilizes a sigmoid function to restrict star's positions. The mechanism of this algorithm is similar to PSO, as it searches for the optimal solutions based on the information from the local best (electricity force) and global best (gravitational force). However, the algorithm has some drawbacks such as high number of variables for computing, the existence of unnecessary constant such as G and K without mentioning the true value of them, and low computational speed. This algorithm lacks one of the main characteristics of Black Hole (event horizon). Therefore we do not analyze in details this algorithm here, as it is not suggested for solving the FS problem.

The main contributions of this paper are as follows: (1) we propose a binary version of the BHA called BBHA based on hyperbolic tangent function for solving discrete problems. Our proposed BBHA is based on the BHA introduced in Ref. [12]; (2) the proposed BBHA apply to wrapper based FS method. The principal idea is to allocate a binary structure for each star in the population which shows a feature whether belonging to the final set of features or not. The accuracy of the supervised classifier is used as a fitness function of this optimization algorithm. Each star is evaluated by training a classifier with the chosen features that is encoded as the position of stars; (3) we test the effectiveness of BBHA wrapper based FS method with two classifiers, Random Forest (RF) and Naive Bayes (NB), on twelve benchmark datasets from different domains (biological, text, and image). Repeated 10-fold-CV method is used to justify the performances of classifiers. RF is chosen as an evaluator of feature subsets in the proposed wrapper approach because it follows a linear procedure; easy to learn; do not need any tuning; highly interpretable; insensitive to noise, outliers, and overfitting. It also achieves better performance than the five popular DT algorithms (Bagging, C5.0, Boosted C5.0, C4.5, and CART). To find the robust and best DT classifier, the performance of these popular DT algorithms were also compared in this study; (4) we compare the performance of the proposed BBHA wrapper based approach with the GA, BPSO, SA, and CFS in terms of Matthews' Correlation Coefficient (MCC), accuracy, specificity, sensitivity, area under ROC curve (AUC), CPU Time, the number of parameters for configuring the model, and the number of chosen optimized features.

The rest of the paper is organized as follows: In Section 2, related work on the gene selection by DT algorithms and hybrid of them with optimization algorithms is presented. In Section 3 the method used is introduced. It includes continuous BHA, proposed binary version of BHA, proposed wrapper approach for FS based on BBHA, and brief overviews of GA, PSO, SA, and CFS. The characteristics of datasets, experimental design and setting are described in Section 4. The experimental results on twelve datasets from different domains and summery of discussion are presented in Section 5. Finally, Section 6 concludes with some directions for future research.

2. Related works

Further to briefly reviewing the FS technique based on DT algorithms, this section presents a summary of the main approaches related to FS methods based on integration of optimization algorithms with DT algorithms documented in the literature.

In these methods, DT algorithms have been served as fitness function of nature inspired metaheuristic algorithms to evaluate the quality of candidate feature subsets.

A greedy algorithm is a basic method for DT algorithms, which are based on top-down recursive divide-and-conquer manner. Greedy strategies are preferred to utilize as they are easy and efficient to implement. At each node of the tree in DT algorithms, all possible splits are evaluated. Each split has own information gain. If an information gain of the split is highest among the others, it should be chosen as a divider of data into binary parts. The algorithm runs until the stop condition is met. Iterative Dichotomiser 3

(ID3) is the first series of algorithms created by Ross Quinlan based on greedy strategies to generate DTs. Finding the optimal size of the final tree in a DT algorithm is known as the horizon effect problem. A common strategy for solving this problem is to grow the tree until each node contains a small number of samples then use pruning to replace irrelevant branches with leaf nodes. Pruning reduce the size of a DT without reducing predictive accuracy. It removes nodes that do not provide additional information.

An ensemble of unpruned DTs using bagging and bootstrap techniques is known as RF, which was introduced by Leo Breiman in 2001 [20]. RF constructs multiple DTs with randomly selected features and samples. The final classification of RF can be obtained by combining the classification results from the individual DTs. No needing for pruning trees, automatically generation of accuracy and variable importance, being robust to overfitting and outliers, high speed even in prediction, are some of the properties of RF [21]. As demonstrated by several bioinformatics studies, RF is well suited for high-dimensional data and have been increasingly applied for gene selection and classification [22]. For instance, in Ref. [23] RF was utilized as fitness function of GA-tuned PSO and was applied for prediction of o-glycosylation sites in proteins. The GA-tuned PSO has achieved higher classification accuracy in terms of AUC with comparison to PSO-RF and several tools for predicting the O-glycosylation sites in proteins.

RF is used in Ref. [24] against nine multi-class microarray data sets as a gene selection method and it yealds very small sets of genes while preserving predictive accuracy. Also, it shows high performance compared to Direct Linear Discriminant Analysis (DLDA), K-Nearest Neighbors (KNN), and SVM classifiers.

In Ref. [25] a new algorithm based on RF namely Balanced Iterative Random Forest (BIRF) is proposed to select informative genes from four imbalanced microarray data sets. BIRF has ability in handling the class-imbalanced data and has outperformed the predictive performance of SVM-Recursive Feature Elimination (SVM-RFE), Multi-class SVM-RFE, RF and Naive Bayes (NB) classifiers.

Bagging was proposed by Breiman in 1996 [26]. Bagging averages the predictions of classification trees over a group of bootstrap samples. It helps to avoid overfitting. Bagging improves stability and accuracy of DT methods by reducing variance. In Ref. [27] a new method based on hybrid of gene selection and bagging classifier namely select-bagging is proposed for classification of high-dimensional and balanced datasets in bioinformatics.

A C4.5 DT is an improved form of the ID3 algorithm and was introduced in [5]. The performance of C4.5 is high. In main memory algorithms, training data are completely loaded into main memory, and are thus severely limited in the number of examples they can learn from. Classical tree-based models such as ID3, Classification and Regression tree (CART), LDA, and quadratic discriminant analysis (QDA) are some example of main memory algorithms.

C4.5 Comparing to the main memory algorithms is quicker [28]. In Ref. [29] the C4.5 was used as fitness function of PSO namely PSODT on 5 small medical datasets from UCI Machine Learning Repository. The C4.5 classifier is also adopted as a fitness function of PSO for gene selection in [8,30] against eleven benchmark gene expression microarrays.

By improving the algorithm of C4.5, Ross Quinlan introduced C5.0 [31]. Missing value and numeric attributes can be handled by C5.0. Lower error rate, high speed, less memory, and support for boosting can be mentioned as characteristics of C5.0 [32]. In BoostedC5.0, Ada-boost algorithm can be used to improve the accuracy of C5.0 [33]. In Refs. [34,35] Boosted C5.0 was used as fitness function of PSO and was applied on small medical datasets and some benchmark microarrays to improve the performance of PSODT.

Leo Breiman in 1984 introduced CART classifier [6]. It uses Gini index for node impurity, allows only binary outcomes, and prunes tree based on the complex model [36]. In Ref. [37] the researchers proposed a new approach based on CART algorithm namely Sequential CART (S-CART) for gene selection on binary microarrays. They proved that the performance of S-CART is better than Stochastic Search Variable Selection (SSVS) and RF classifiers in terms of speed and accuracy.

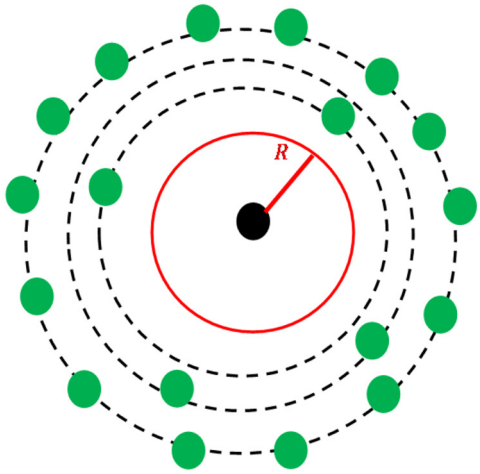


Fig. 1. Black Hole schema.

Previous research all indicates that DTs which are listed in the top 10 most influential data mining algorithms [38] in combination with optimization algorithms and on their own are promising to solve the feature (gene) selection and classification problem.

This paper puts forward a gene selection method for classification of several biological data based on combining of proposed binary version of BHA and DT algorithm (RF) and investigates its performance.

3. System and methodology

In this section, we give a detailed description of the continuous Black Hole Algorithm (BHA), proposed Binary Black Hole Algorithm (BBHA) and BBHA wrapper based approach, which is introduced for solving discrete problems including FS. We also present some brief overviews of GA, PSO, SA, and CFS.

3.1. Continuous black hole optimization algorithm (BHA)

The black hole optimization algorithm is a robust stochastic optimization technique based on simulation of the behavior of black hole in outer space.

The below steps explain manner of simulating BHA from black hole phenomenon:

Step 1: Outer space is full of known and unknown stars. In real space black hole is formed by collapsing individual stars so BHA begins with the population of stars that located arbitrarily in the explore space. In BHA each star has a fitness value, which is evaluated by a fitness function to be optimized. The best star that has the best fitness value is selected as the black hole. It is called “black” because it absorbs all the light and reflects nothing. Fig. 1 shows BHA schema. The black circle is the black hole and green circles are stars. They placed randomly in search space.

Step 2: In the real space, a black hole is an object of extreme density with an intense gravitational attraction. This leads to a great amount of gravitational force pulling stars around it. BHA has followed the same behavior. By Eq. (1) all the stars began moving toward the black hole.

Step 3: The sphere shaped bound of a black hole in outer space is known as the event horizon. The event horizon radius is called as the Schwarzschild radius. The red circle in Fig. 1 shows the event horizon of black hole. In the real space the Schwarzschild radius is computed by Eq. (2) and in BHA is computed by Eq. (3).

Step 4: Because of extreme density and strong gravitational attraction of black hole when a star crosses the event horizon, it will be swallowed by the black hole and disappear. In the region of event horizon the escapee speed is tantamount to the speed of the light, so nothing can get away from within the event horizon. In BHA, the Euclidean distance between black hole and star is computed. If this distance is less than Schwarzschild radius, substitute it with a fresh star in the random location in the search space.

Step 5: In BHA if a star reaches a location with lower cost than the black hole, in that case theirs locations should be replaced [12,39,40].

$$X_i(t+1) = X_i(t) + rand \times (X_{BH} - X_i(t)), \quad for\ i = 1, 2, \ldots, N \tag{1}$$

$$R = 2GM/C^2 \tag{2}$$

$$R = \frac{f_{BH}}{\sum_{i=1}^{N} f_i} \tag{3}$$

where $X_i(t)$ and $X_i(t+1)$ signify the locations of the i^{th} star at *iterations* t and $t+1$, respectively. *Rand* indicates uniform distribution with a range from 0 to 1. N denotes the number of stars. X_{BH} points the location of the black hole in the exploration space. M, G, and C signify the mass of the black hole, the gravitational constant, and the speed of light respectively. f_i denotes the fitness value of the i th star and f_{BH} indicates the fitness value of the black hole. Based on the above explanation the framework of the BHA method is presented in Algorithm 1.

Algorithm 1. The continuous black hole algorithm

```
Input
 number of stars(N), number of iteration
Output
Black hole
The fitness value of black hole
Begin
Initialize a population of stars
For j = 1 to numbers of stars
     calculate the objective function of the star(j) and save in fitness array(f)
Next j
The star with the most remarkable fitness value is chosen as the black hole
While (max iteration or convergence criteria is not met) do
      For a = 1 to numbers of stars
           X_a^new = X_a^old + rand × (X_BH − x_a^old)
            Evaluate fitness value of the star(X_a)
                 If fitness of X_a > fitness of X_BH Then
                     X_BH = X_a
                 End if
             Replace the new fitness value of the star ( X_a ) with the previous value
             Update fitness array (f) and Calculate : R = f_BH / Σ_{i=1}^{N} f_i
             If √((X_BH − X_a)^2) <R Then
                 replace X_a with a new star in an optional location in the search scope
             End if
       next a
end while
End
```

3.2. The proposed binary black hole algorithm (BBHA)

The BHA was originally developed for continuous valued spaces. But there exist a number of discrete combinatorial optimization problems, such as FS, in which the values are not continuous numbers but rather discrete binary integers. For this reason, we have introduced binary version of BHA and called it BBHA. Binarization techniques can be categorized into two groups: two steps binarization and continuous-binary operator transformation. Our proposed binarization technique belongs to the first group. In the first group without any modifications in the operators, only two steps is added after the continuous iteration.

In solving FS problem the search space must be modeled as a d-dimensional Boolean lattice, where the i^{th} star moves around the d-dimensional space.

Since the problem is to select or not select of a given feature, the position of a star only takes the values 1 or 0. Therefore, a transfer function is needed to forces stars to move in a binary space. Transfer functions define the probability of changing position's elements from 0 to 1 and vice versa. In the proposed approach, Hyperbolic Tangent function is utilized to modify the position of stars as in Eqs. (4) and (5).

$$S(X_{id}(t+1)) = abs(\tanh(X_{id}(t+1))) \tag{4}$$

$$X_{id}(t+1) = \begin{cases} 1 & \text{if } S\ (X_{id}(t+1)) > rand \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where *rand* is a uniform random number between 0 and 1. In Eq. (5), instead of *rand* threshold 0.6 can also be considered. Hyperbolic Tangent function belongs to the group of v-shaped transfer functions. It has been used here because it shows good performance compared to the other transference functions such as sigmoid function [41]. In addition, in the proposed algorithm we may face with situation that one star with small number of features has the same fitness value with black hole. In this situation we should change their positions.

In BBHA we only need to set number of stars. The proposed algorithm does not suffer from some of other optimization algorithms difficulties such as the slow convergence rate and adjusting several parameters. Compared with other optimization algorithms, BBHA is easier to implement, depend on a single parameter for configuring the model, requires much less memory, and converges more rapidly.

3.3. The proposed wrapper approach based on BBHA for FS

In this section, we present details on the process used to enable FS with BBHA. At the beginning of BBHA, the primary population of the star's position is initialized randomly. Each star encodes a candidate feature subset based on a bit string. The length of the string is equivalent to the total number of features in the dataset of interest. In the binary encoding, a bit of one implies the feature is chosen and a bit of zero means that the feature is not chosen. Similar to other optimization algorithms, the fitness value of each star is calculated by using an evaluator. Here, two classifiers; RF and NB serve as the evaluators of our proposed algorithm.For biological data accuracy of RF classifier and for text and image datasets accuracy of NB classifier are used. The proposed wrapper approach based on integration of BBHA with RF is called BBHA-RF and based on combination with NB classifier is called BBHA-NB.

In the part of evaluating fitness value of stars, when two founded stars have identical fitness value, the one with smaller number of features is chosen as the best star (black hole).

The procedure stops once stopping criteria (maximum number of iterations) is met. The parameters for BBHA specify 25 iterations of population consisting of 10 stars. At the end of the BBHA wrapper based FS algorithm, the star with the best performance is selected. The position of this star gives the selected features. By using the subset of data that contains these selected features, again the performance of RF or NB model is assessed using 10-fold-CV for five different metrics. In order to avoid producing random results and provide an assurance for impartial comparison of the classification performances, assessing the efficiency of RF or NB model for selected features by optimization algorithms is executed 100 times. Average across 100 times of run is considered as a last result of one execute of whole procedure. The whole procedure runs 5 times for biological data and 3 times for text and image data. In each time, different subsets of features are selected by optimization algorithms. Average of these 5 times or 3 times of runs for whole procedure is reported in this study. Algorithm 2 illustrates the procedure of applying BBHA for FS.

Algorithm 2. Pseudo code of Binary BHA for FS

Input
Rand, number of stars, number of iteration
Output
Black hole
The fitness value of black hole
Begin
Initialize a population of stars
For $j = 1$ to numbers of stars
    Evaluate fitness value of the star(j) by 10-fold-CV RF or NB and save in fitness array(f)
Next j
The star with the most remarkable fitness value is chosen as the black hole
While (max iteration or convergence criteria is not met) do
    For $a = 1$ to numbers of stars
        Evaluate fitness value of the star(X_a) by 10-fold-CV RF or NB
        If (fitness of X_a > fitness of X_{BH}) **Then**
            $X_{BH} = X_a$
        Else if ((fitness of X_a == fitness of X_{BH}) and ($|X_a| < |X_{BH}|$)) **Then**
            $X_{BH} = X_a$
        End if
        Replace the new fitness value of the star (X_a) with the previous value
        Update fitness array (f) and Calculate : $R = \frac{f_{BH}}{\sum_{i=1}^{N} f_i}$
        If $\sqrt{(X_{BH} - X_a)^2} < R$ **Then**
            replace X_a with a new star in an optional location in the search scope
        End if
    next a
    For $i = 1$ to numbers of stars
        For $d = 1$ to number of features
            $X_{id}^{new} = X_{id}^{old} + rand \times \left(X_{BH\,d} - x_{id}^{old}\right)$
            If $abs(tanh(X_{id}^{new})) >$ rand **Then**
                $X_{id}^{new} = 1$
            Else
                $X_{id}^{new} = 0$
            end if
        next d
    next i
end while
End

By following this algorithm, we attempt to find optimal feature subset, which could improve the classification accuracy of medical data.

3.4. Binary particle swarm optimization algorithm

We integrate Binary PSO algorithm with the RF classifier to address the FS problem. The important features are proposed using BPSO algorithm. The efficiency of these features is investigated by RF with 10-fold-CV scheme that is employed as a fitness function of the BPSO algorithm. Algorithm 3 shows the outline of the BPSO for FS.

Algorithm 3. Pseudo code of BPSO for FS

```
Input
    c1, c2, r1, r2, w, v_min, v_max, number of particles, number of iteration
Output
Global Best (gb)
The fitness value of gb
Begin
Initialize a population of particles
For j = 1 to numbers of particles
    Evaluate fitness value of the particle(j) by 10-fold-CV RF or NB
Next j
The particle with the most remarkable fitness value is chosen as the Global Best
While(max iteration or convergence criteria is not met) do
    For i = 1 to numbers of particles
        r1= Rand ; r2= Rand ; r3= Rand
        For d = 1 to number of features
            v_id^new = w × v_id^old + c1 r1(pb_id^old − x_id^old) + c2 r2 (gb_d^old − x_id^old)
            If v_id^new > v_max Then
                v_id^new = v_max
            End if
            If v_id^new < v_min Then
                v_id^new = v_min
            End if
            sigmoid(v_id^new) = 1/(1+e^(−v_id^new))
            If sigmoid(v_id^new) > r3 Then
                x_id^new = 1
            Else
                x_id^new = 0
            End if
        next d
    next i
    For a = 1 to numbers of particles
        Evaluate fitness value of the particle(X_a) by 10-fold-CV RF or NB
        If fitness X_a > fitness of pb_a Then
            pb_a = X_a
        End if
        If fitness of X_a > fitness of gb Then
            gb = X_a
        End if
    Next a
end while
End
```

Where v_{id}^{new} and v_{id}^{old} are the particle velocities,pb and gb are local and global fitness values, factors r1 and r2 are random numbers between 0 and 1, c1 áis cognitive learning factor, c2is social learning factor, vmax is upper bound of velocity, vminis the lower bound of velocity and w is inertia weight.

3.5. Genetic algorithm

FS with GA needs to consider the process of FS as an optimization problem and then mapping it to the genetic structure of stochastic variation and natural selection. In the first step of GA algorithm, a primary population of chromosomes is generated randomly. The chromosomes are modeled as the binary vectors. Then, fitness value of each chromosome is evaluated by using a classifier. Two chromosomes with the best fitness value are selected. Then, for these chromosomes, a split point is chosen randomly. In the next step the front of one chromosome is mapped to the back of the other (and vice versa) in order to generate two offspring chromosomes with combined genes. In the last step these two offspring chromosomes are mutated randomly according to predetermined probability. The process would end if maximum iteration was met.

3.6. Simulated annealing

The trope of SA comes from the annealing specifications in metal processing. The annealing process contains the control of heat and its cooling rate. Compared to other optimization algorithms, SA has an advantage of being able to avoid the algorithm from being stuck at the local minimum. SA uses random chain in terms of Markov chain. In the FS based on SA, a primary solution is chosen randomly and it is supposed to be the optimal solution. Afterward, the value of the primary solution is calculated using the fitness function. Whereas heat T does not meet the end condition, a neighboring solution of the current optimal solution is chosen and its fitness value is calculated. If the fitness value of the freshly chosen neighboring solution is greater than or equal to the current optimal solution, the current optimal solution is substituted with a freshly chosen neighbor solution. If the fitness value of the neighboring solution is less than the current optimal solution, a random number is generated in the range of (0, 1). In this situation, the substitution of the optimal solution is allowed only if a generated random number is less than Eq. (6). Then the heat is reduced by Eq. (7). The process would finish if maximum iteration was met.

$$e^{-\frac{cost\ (\text{neighbor solution})-cost\ (\text{optimal solution})}{T}} \tag{6}$$

$$T \longleftarrow r \times T \tag{7}$$

3.7. Correlation based feature selection (CFS)

By using correlation and entropy measures, CFS algorithm quickly identifies irrelevant, redundant, and noisy features, and finds informative feature subsets. This algorithm for searching the feature subset space uses best first search strategy.

4. Experiments

4.1. Experimental design

We have tested the performances of different DT algorithms to find out which of them has higher performance than the others on medical data sets to choose it as fitness function of optimization algorithms. Then, we have examined the relative performance of the combined BBHA and RF method (best DT algorithm as fitness function) denoted as BBHA-RF for true classification of medical data, with a series of repeated 10-fold-CV experiments (repeat for 5 times to avoid bias). We have compared the performance of the proposed BBHA-RF method with GA-RF, BPSO-RF, SA-RF, and CFS. The results are reported in terms of accuracy, sensitivity, specificity, MCC, and AUC. Larger values of these criteria represent good classification performance. These measures are defined as follows:

4.1.1. Accuracy

Accuracy represents the percentage of correct predictions. Let TN, TP, FN, FP denotes true negative, true positive, false negative and false positive, respectively. The Accuracy is:

$$Accuracy = (TP + TN) / (TP + TN + FN + FP) \tag{8}$$

4.1.2. Sensitivity & specificity

To see how the accuracy is distributed over the classes, sensitivity and specific values are presented. Sensitivity is the ability of the classifier to find all the positive samples. Specificity is the ability of the classifier to find all the negative samples.

$$Sensitivity = TP/(TP + FN) \tag{9}$$

$$Specificity = TN/(TN + FP) \tag{10}$$

4.1.3. Matthews correlation coefficient

In machine learning, the quality of unbalance binary (two-class) classifications can be obtained by Matthew's correlation coefficient. If the classes are unbalanced (not equal), computing the MCC of classification system can be so much more appropriate than computing accuracy. The Matthews correlation coefficient (MCC) is:

$$MCC = \frac{TN \times TP - FN \times FP}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{11}$$

The aim of Matthew's correlation coefficient is to measure the quantity of correlation between predictions and real target values. The answer is bounded between the range +1 and −1.

The answer in the range of (0, +1] shows that predictions are positively related to the target values. A zero shows that the prediction is completely random. The answer in the range of [−1, 0) shows that predictions are negatively related to the target values.

4.1.4. Area under ROC curve (AUC)

The impact of a threshold on the false positives and false negatives (FP/FN) tradeoff can be visualized by Receiver Operating Characteristic (ROC) curve. The functions of the threshold can be described by the coordinates of ROC curve points:

$$threshold = \theta \in \mathbb{R}, \quad \text{here } \theta \in [0, 1] \tag{12}$$

$$ROC_X(\theta) = FPR(\theta) = \frac{FP(\theta)}{FP(\theta) + TN(\theta)} = \frac{FP(\theta)}{\#N} \tag{13}$$

$$ROC_Y(\theta) = TPR(\theta) = \frac{TP(\theta)}{FN(\theta) + TP(\theta)}$$

$$= \frac{TP(\theta)}{\#P} = 1 - \frac{FN(\theta)}{\#P} = 1 - FNR(\theta) \tag{14}$$

No false positives and all true positives (FPR, TPR) = (0, 1) is the optimal point on the ROC curve. AUC can be achieved by computing the area of the convex shape under the ROC curve. When AUC reaches to 1 this means that ROC is reached to the optimal point of perfect prediction.

4.2. Dataset

To evaluate the performance of proposed method twelve datasets, which belong to completely different domains, are employed. These domains are biological (life and microarray), text, and image.

Text datasets are Chess and Email word subject which are two classes and obtained from UCI Machine Learning Repository at the website: http://www.ics.uci.edu/-mlearn/MLRepository.html. Image datasets are warpAR10P and warpPIE10P which are multiclass and taken from the http://featureselection.asu.edu/datasets.php. The summary of these data sets are given in Table 1.

For the biological datasets, three small medical datasets with a variety of complexity and five widely-used binary microarrays were used. Small medical datasets are Wisconsin diagnostic breast cancer, Parkinson's, and Heart Statlog, which are obtained from UCI Machine Learning Repository. Colon Tumor, Central Nervous System, Leukemia, Breast Cancer and Ovarian Cancer are microarray datasets that are available for download at the website: http://csse.szu.edu.cn/staff/zhuzx/Datasets.html. The characteristics of these medical data sets are shown in Table 2. The datasets are diverse in terms of the number of samples and features. The number of classes is two for all biological datasets.

4.3. Experimental setup

Our experiment results consist of two parts. Firstly, we compared the performance of BBHA with BPSO, GA and CFS on the text and image datasets. We used NB classifier with 10-fold-CV as a fitness function and feature subset evaluator on these datasets. Then the performances of six well-known DT classifiers (Random Forest, Bagging, C5.0, Boosted C5.0, C4.5, and CART) are compared with each other to identify the best one of them. Finally, we considered the best DT (RF) classifier with 10-fold-CV as fitness function and gene subset evaluator of BBHA. Then, the proposed BBHA is conducted on a set of eight well-known medical datasets and compared with BPSO, GA, SA, and CFS.

In the process of 10-fold-CV, the samples of data are divided into 10 equally subsets. Each time, 9 subsets are located next to each other to create the training set and the remained one subset is utilized as the test set. Then the average accuracy across all 10 trials is calculated. Since one try of the 10-fold CV is generally biased and in order to have statistically meaningful conclusion we repeated 10-fold-CV for 100 times and reported average and standard deviation of them. The experiments are carried out on a laptop with Windows 7, 2.40 GHz CPU and 4 GB of RAM, using R version 3.2.1. For RF classifier we used 'randomForest' package. For Bagging and C4.5 classifiers, 'RWekajars', 'rJava', and 'RWeka' packages have been used. 'C50′ package has been utilized for C5.0 classifier and Boosted C5.0 (trials=10). For CART classifier 'rpart' package, for CFS filter approach 'FSelector' package, and for Naive Bayes (NB) classifier 'e1071′ package have been employed. Also, for GA and SA optimization algorithms 'caret' package has been used.

For all datasets, the number of particles for BPSO, the number of stars for BBHA, and the number of chromosomes for GA are set to 10. Most of the genes in the high dimensional data like microarrays are irrelevant and not useful for classification problems. Selection of top ranked genes as a preparation step for microarrays by removing a large number of irrelevant, redundant and noisy genes can provide a better classification accuracy [42]. We have selected 50

Table 1
Main characteristics of the image and text datasets.

Dataset	Number of Features	Number of data objects	Number of class	Domain
Chess	36	3196(1669,1527)	2	Text
Email word subject	242	64(35,29)	2	Text
WraoAR10P	2400	130	10	Image, face
WrapPIE10P	2420	210	10	Image, face

Table 2
Main characteristics of the biological datasets.

Dataset	Number of Features	Number of data objects	Number of class	Domain
Wisconsin diagnostic breast cancer	31	569(357,212)	2	Life
Parkinson's	22	195(48,147)	2	Life
Heart-Statlog	13	270(150,120)	2	Life
Colon Tumor	2000	62(40,22)	2	Microarray
Central Nervous System	7129	60(39,21)	2	Microarray
ALL-AML (Leukemia)	7129	72(47,25)	2	Microarray
Breast Cancer	24481	97(51,46)	2	Microarray
Ovarian Cancer	15154	253(91,162)	2	Microarray

top ranked genes by Chi- Squared statistic with leave-one-out cross validation method. The parameters of Binary PSO are adjusted as follow [8];(v_{min}), (v_{max}), (c_1), (c_2) and (w) are set at −4, 4, 2, 2 and 0.4 respectively. The crossover and mutation probability of GA are set to 0.8 and 0.1 respectively. The process would stop if maximum iteration was met. Here, a maximum number of iteration is set at 25 because by increasing the number of iterations the improvement in results was insignificant.

5. Results and analyses

5.1. Experimental results on text and image datasets using NB classifier

The summary of four text and image datasets considered here are given in Table 1. In this study, NB classifier with 10-fold-CV was considered as a fitness function of BBHA, BPSO, and GA optimization algorithms and also for feature subset evaluation of CFS. The average classification accuracy of BBHA-NB, BPSO-NB, GA-NB and CFS on the text and image datasets are reported in Table 3. In particular, the number of selected features and CPU times are tabulated. Due to the stochastic nature of BBHA, GA, and BPSO the average FS results of them for three independent runs are reported. Here, the maximum iteration is set to 30 and as a pre-process step 250 top ranked features are chosen by Chi- Squared statistic for image datasets. In Table 3, among the four algorithms on each dataset the best average classification accuracies are highlighted in bold typeface. Table 3 shows that BBHA-NB outperforms the other three algorithms in terms of accuracy on all datasets except the Email word subject dataset. GA-NB gives better accuracy than BBHA-NB for this dataset. The accuracies of BBHA-NB and GA-NB are not significantly different from each other on 3 out of four datasets. BBHA-NB performs significantly better than GA-NB only on the WraoAR10P dataset. With regard to the number of selected features, CFS chooses the least number of features but it does so at the cost of low classification accuracy. After CFS, BBHA-NB chooses fewer number of features but with a classification accuracy that is superior to the BPSO-NB, GA-NB, and CFS. CFS filter method uses less time than the other 3 wrapper approaches. BBHA-NB is the second approach which costs less time than BPSO-NB and GA-NB.

5.2. Experimental results on biological datasets using RF classifier

In order to determine which DT algorithm is more robust and has higher performance than the others to be used as fitness function of optimization algorithms, we have compared six well-known DT classifiers on eight medical datasets. The computational results of this experiment are shown in Table 4. The classification accuracy, sensitivity, specificity, MCC, and AUC along with standard deviations of them are presented in this Table. According to Table 4, RF classifier has higher performance in almost all datasets. Beside the high performance, the robustness is an important factor in evaluating a classifier. The standard deviation of all criteria for RF in all datasets is small. This shows that Random Forest is a robust classifier.

The Fig. 2 displays average AUC, classification accuracy, and MCC of six DT classifiers on eight medical datasets, respectively. As can be seen from this figure, it is clearly found that the performance of RF is better than other classifiers. Therefore, we choose this classifier as a fitness function of four optimization algorithms (i.e. BBHA, BPSO, GA, and SA). After RF, Boosted C5.0 has higher performance than the others. Bagging, C5.0, and C4.5 are placed in the next positions, respectively. CART is the classifier which has worse performance than the others.

To evaluate the effectiveness of our proposed method, we compare the results of BBHA-RF with BPSO-RF, GA-RF, SA-RF, and CFS. The average AUC, classification accuracy, sensitivity, specificity and MCC along with standard deviations of them for 100 independent runs of the selected features in 5 executions of the whole procedure and CPU time are presented in Table 5. In order to illustrate the good performance of the proposed FS method, Table 6 reports an average number of the selected features from the entire data set by BBHA-RF, BPSO-RF, GA-RF, SA-RF, and CFS.

As can be seen from Table 5, the proposed BBHA-RF outperformed BPSO-RF, GA-RF, SA-RF, and CFS in terms of all criteria. Table 6 demonstrates that the proposed approach is significantly better than all wrapper approaches and CFS filter approach in term of the number of selected optimized features. BBHA-RF selects features approximately 5 times fewer than BPSO-RF and GA-RF. Compared with SA-RF and CFS the proposed method selects 3 times less features.

The computational efficiency of BBHA-RF is comparable to SA-RF and is better than BPSO-RF and GA-RF. BBHA-RF converges approximately 3 times faster than BPSO-RF and approximately 6 times faster than GA-RF. For all biological datasets BBHA-RF can achieve high performance with least number of features in short time.

Fig. 3 shows the average solution quality of BBHA-RF, BPSO-RF, GA-RF, SA-RF, and CFS on eight biological datasets. We can observe that the proposed wrapper approach (BBHA-RF) compared with other FS algorithms maximizes solution quality while using fewer features. All medical datasets except heart-statlog and breast cancer microarray are unbalanced. So consideration of MCC is more appropriate than accuracy. Average MCC of BBHA-RF is signifi-

Table 3
Average accuracy, number of features, and computational efficiency of each wrapper method for 3 independent runs.

Dataset	Criteria	BPSO- NB	GA-NB	CFS-NB	BBHA-NB
Chess	*Accuracy*	93.93 ± 0.04	94. 33 ± 0.03	90.42	**94.66** ± 0.017
	# of features	14 ± 1.41	14 ± 2.82	**3**	**6** ± 1
	CPU time	1090.22	3233.11	2.31	510.27
Email word subject	*Accuracy*	91.23 ± 4.57	**93.37** ± 0.01	92.18	92.28± 4.05
	# of features	118.5 ± 7.77	111.66 ± 3.21	**2**	**25.66** ±7.09
	CPU time	591.11	2102.99	6.17	127.33
WraoAR10P	*Accuracy*	74.63 ± 1.69	76.40 ± 0.44	74.93	**80.46** ± 1.41
	# of features	124 ± 1.41	93.66 ± 18.82	19	**14.33** ± 2.51
	CPU time	1091.71	3473.11	130.87	396.33
WrapPIE10P	*Accuracy*	92.32 ± 0.95	93.96 ± 0.72	91.29	**94.91** ± 0.89
	# of features	126.5 ± 6.36	120 ± 8.18	**32**	37 ± 2.64
	CPU time	1282.07	3561.11	463.21	422.34

Table 4
Solution quality of each decision tree classifier on medical data sets.

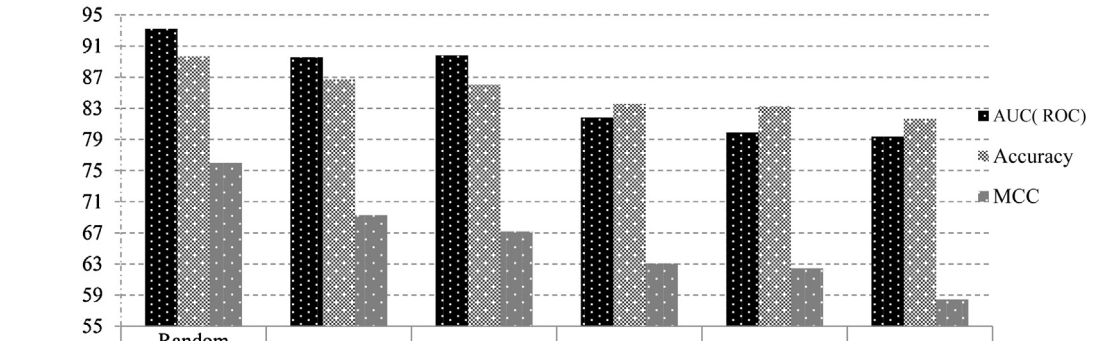
Dataset	Criteria	RF	Bagging	C5.0	Boosted C5.0	C4.5	CART
Wisconsin diagnostic breast cancer	*Accuracy*	96.08 ± 0.31	94.66 ± 0.57	93.71 ± 0.79	95.93 ± 0.53	93.39 ± 0.74	92.41 ± 0.71
	Sensitivity	93.75 ± 0.67	91.90 ± 0.98	90.59 ± 1.42	93.38 ± 0.94	90.54 ± 1.42	89.46 ± 1.55
	Specificity	**97.48 ± 0.34**	96.37 ± 0.66	95.38 ± 0.80	97.46 ± 0.58	95.36 ± 0.92	94.13 ± 0.82
	MCC	**91.60± 0.84**	88.73 ± 1.27	86.49 ± 1.63	91.26 ± 1.13	86.18 ± 1.45	83.82 ± 1.54
	AUC(ROC)	**99.07 ± 0.13**	98.51 ± 0.28	96.24 ± 0.64	98.97 ± 0.25	92.69 ± 1.27	93.80 ± 0.87
Parkinson's disease	*Accuracy*	**90.70 ± 0.88**	87.73 ± 1.45	84.16 ± 2.05	89.61 ± 1.49	85.07 ± 1.95	86.13 ± 2.08
	Sensitivity	**72.70 ± 3.95**	65.52 ± 5.93	69.64 ± 6.55	71.82 ± 4.50	69.62 ± 6.02	65.44 ± 5.89
	Specificity	**96.90** ± 0.75	94.88 ± 1.35	89.14 ± 2.14	95.73 ± 1.26	89.34 ± 2.36	93.63 ± 1.65
	MCC	**74.16** ± 3.87	63.25 ± 5.01	56.55 ± 5.76	70.25 ± 4.95	59.39 ± 4.95	60.77 ± 6.29
	AUC(ROC)	**96.09** ± 1.92	92.74 ± 2.56	82.02 ± 3.68	93.98 ± 2.86	80.19 ± 3.74	84.28 ± 3.42
Heart-Statlog	*Accuracy*	**82.95 ± 1.02**	81 ± 1.29	77.96 ± 1.59	79.94 ± 1.48	78.15 ± 1.57	80.27 ± 1.57
	Sensitivity	**77.56 ± 1.34**	74.80 ± 2.38	73.14 ± 2.68	76.05 ± 2.26	72.77 ± 2.74	74.64 ± 2.39
	Specificity	**87.53 ± 1.42**	85.60 ± 1.84	82.41 ± 2.26	83.88 ± 2.01	82.01 ± 2.45	85.21 ± 2.03
	MCC	**65.66 ± 1.86**	61.45 ± 2.86	55.65 ± 4.17	60.25 ± 3.31	56.93 ± 3.52	60.01 ± 3.38
	AUC(ROC)	**90.34 ± 0.74**	88.31 ± 1.29	82.08 ± 1.59	87.78 ± 1.11	78.28 ± 2.04	82.41 ± 1.83
Colon Tumor	*Accuracy*	**85.58 ± 1.46**	81.35 ± 3.96	81.91 ± 3.41	82.08 ± 2.63	83.03 ± 3.19	76.09 ± 3.96
	Sensitivity	87.72 ± **2.38**	**90.47** ± 4.62	88.93 ± 4.13	88.17 ± 3.01	88.65 ± 3.55	84.81 ± 4.84
	Specificity	**82.35 ± 5.68**	66.03 ± 7.76	70.74 ± 8.65	69.73 ± 7.47	74.88 ± 6.97	61.97 ± 8.76
	MCC	**65** ± 7.71	52.71 ± **7.54**	56.73 ± 10.5	56.12 ± 8.15	57.87 ± 9.27	44.68 ± 8.85
	AUC (ROC)	**86.61** ± 7.29	81.70 ± 7.36	75.96 ± 7.97	83.56 ± 7.66	76.35 ± 7.84	68.48 ± **6.38**
Central Nervous System	*Accuracy*	**84.86 ± 1.93**	76.96 ± 4.20	78.93 ± 4.06	79.61 ± 3.67	74.13 ± 4.31	71.46 ± 4.35
	Sensitivity	**92.49 ± 2.07**	90.09 ± 4.18	87.30 ± 6.01	90.61 ± 3.69	82.85 ± 5.83	83.44 ± 6.24
	Specificity	**71.99 ± 5.93**	53.34 ± 10.8	60.78 ± 10	59.71 ± 9.49	57.51 ± 9.88	50.63 ± 10.3
	MCC	**62.03 ± 7.68**	44.14 ± 10.1	48.66 ± 11.6	48.95 ± 10.1	40.39 ± 10.8	30.97 ± 11.1
	AUC(ROC)	**87.61 ± 6.95**	78.90 ± 8.43	69.40 ± 7.98	79.08 ± 7.92	64.13 ± 7.63	62.70 ± 7.40
ALL-AML (Leukemia	*Accuracy*	**97.98 ± 0.86**	94.7 ± 1.61	86.07 ± 2.84	92.01 ± 2.56	85.08 ± 2.68	83.71 ± 2.60
	Sensitivity	**99.41 ± 0.92**	96.68 ± 2.32	86.40 ± 4.01	95.12 ± 2.93	86.61 ± 3.68	85.71 ± 3.92
	Specificity	**95.53 ± 2.31**	90.31 ± 5.02	84.98 ± 6.74	87.50 ± 5.54	82.88 ± 6.36	80.58 ± 6.16
	MCC	**91.28 ± 6.58**	82.84 ± 7.97	67.68 ± 8.22	79.16 ± 6.33	65.09 ± 7.13	62.28 ± 7.97
	AUC(ROC)	**96.02** ± 5.24	95.71 ± **4.86**	82.47 ± 6.18	90.28 ± 6.44	80.95 ± 6.35	78.23 ± 6.43
Breast Cancer	*Accuracy*	**80.04 ± 2.02**	74.52 ± 3.36	67.67 ± 4.17	75.80 ± 3.07	69.05 ± 3.92	65.97 ± 4.21
	Sensitivity	**81.57 ± 3.58**	76.48 ± 4.76	71.57 ± 5.91	76.89 ± 5.08	71.30 ± 6.58	68.15 ± 6.11
	Specificity	**79.40 ± 3.51**	74.54 ± 4.95	66.79 ± 6.65	75.34 ± 5.29	67.89 ± 7.14	63.50 ± 7.24
	MCC	**60.19 ± 4.72**	49.86 ±7.03	36.97 ± 8.33	51.15 ± 6.83	37.86 ± 9.45	31.42 ± 8.80
	AUC(ROC)	**90.07 ± 3.20**	83.37 ± 3.68	67.92 ± 5.17	83.55 ± 4.10	68.47 ± 5.53	68.13 ± 5.11
Ovarian Cancer	*Accuracy*	**99.12** ± 0.40	97.45 ± 0.44	98.19 ± 0.43	98.64 ± 0.59	98.06 ± 0.65	97.14 ± 0.39
	Sensitivity	**99.54** ± 0.45	98.56 ± 0.40	98.15 ± 0.74	98.90 ± 0.60	97.53 ± 0.67	98.03 ± 0.53
	Specificity	98.20 ± **0.95**	95.23 ± 1.16	98.01 ± 1.12	98.38 ± 1.17	**98.86** ± 1.26	95.67 ± **0.68**
	MCC	**98.05 ± 0.90**	94.51 ± 1.25	95.93 ± 1.19	97.16 ± 1.19	95.96 ± 1.39	93.68 ± 1
	AUC(ROC)	**99.94 ± 0.08**	99.18 ± 0.55	98.46 ± 0.43	99.40 ± 0.43	98.14 ± 0.77	96.86 ± 0.42

cantly better than all mentioned FS algorithms. Fig. 4 displays the computational time in seconds for each of the filter and wrapper approaches. The speed of CFS is much higher than mentioned wrapper approaches. The rapidity of BBHA-RF is approximately similar to SA-RF. The proposed approach converges much faster than BPSO-RF and GA-RF.

In the following, the performance of BBHA-RF on microarray datasets is compared with eight state-of-the-art methods from literature. Table 7 reports the results of BBHA and different feature (gene) selection methods for five microarrays. The classification accuracy (first value in every table cell) and the number of selected features (the value in parenthesis) are used as criteria for comparing the performances of methods.

From the results of Table 7, one observed that BBHA-RF except breast cancer microarray gives highly competitive results compared with these reference methods. The most remarkable result for BBHA-RF concerns the ovarian cancer microarray. We obtain 99.82% accuracy with average 2.8 genes while the previous methods reach a prediction rate no greater than 99.44% with at least 4 genes.

For Wisconsin diagnostic breast cancer dataset 97.38% accuracy with average 6.4 features is obtained by BBHA-RF. The proposed method outperformed the results of the literature in [43–45]. For this dataset 100% classification accuracy with only 3 features is obtained by a method based on modified correlation rough set FS and MLP classifier with 80-20 train-test scheme [46].



	Random Forest	Boosted C5.0	Bagging	C5.0	C4.5	CART
AUC(ROC)	93.21	89.57	89.8	81.81	79.9	79.36
Accuracy	89.66	86.7	86.04	83.57	83.24	81.64
MCC	75.99	69.28	67.18	63.08	62.45	58.45

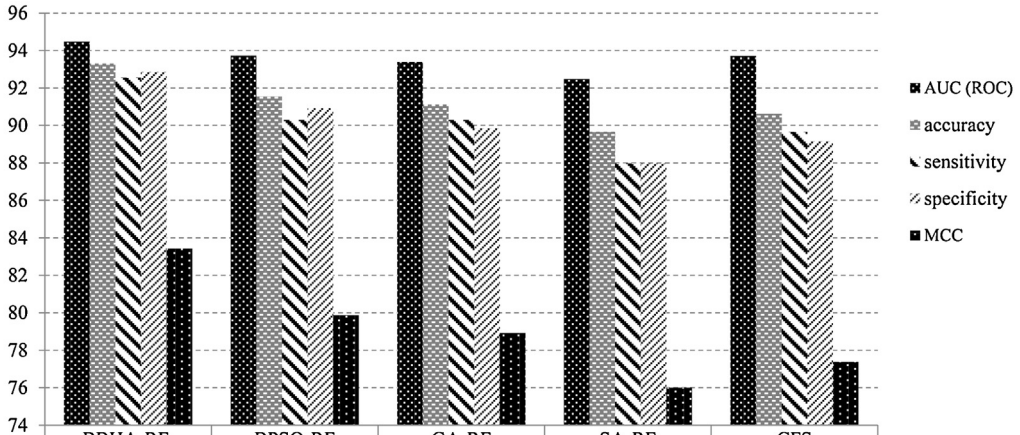
Fig. 2. Average classification AUC (ROC), Accuracy, and MCC of 6 well-known decision tree classifiers on 8 biological datasets.

Table 5
Best, average solution quality and computational efficiency of each wrapper method for 5 independent runs.

Dataset	Criteria	BPSO- RF	GA-RF	SA-RF	CFS-RF	BBHA-RF	Best BBHA (# features)
Wisconsin diagnostic breast cancer	*Accuracy*	96.92 ± 0.39	96.21 ± 0.33	95.90 ± 0.30	95.86	97.38 ± 0.28	97.85 (5)
	Sensitivity	94.66 ± 0.66	93.65 ± **0.65**	93.27 ± 0.68	93.50	**95.79** ± 0.87	96.71
	Specificity	98.41 ± 0.33	97.62 ± 0.36	97.43 ± **0.28**	97.27	**98.57** ± 0.30	99.421
	MCC	93.11 ± 0.75	91.75 ± 0.65	91.25 ± 0.64	91.11	**93.85** ± **0.71**	95.53
	AUC(ROC)	99.30 ± 0.11	99.08 ± 0.13	98.84 ± 0.15	98.93	**99.47** ± **0.09**	99.71
	CPU Time	4212.114	6226.58	800.28	**2.16**	2079.33	2070
Parkinson's	*Accuracy*	92.30 ± **0.77**	93.37 ± 0.86	92.98 ± 0.80	91.20	**93.91** ± 0.77	95.78 (3)
	Sensitivity	78.37 ± 2.80	79.44 ± 4.35	80.54 ± 3.89	76.28	**84.79** ± **2.76**	88
	Specificity	97.33 ± **0.67**	**98.10** ± 0.64	97.19 ± 0.53	96.23	97.95 ± 1.13	98.61
	MCC	80.27 ± 16.27	81.28 ± 3.64	79.79 ± **3.36**	75.07	**82.61** ± 3.52	89.89
	AUC(ROC)	96.15 ± 2.46	97.08 ± 2.07	95.91 ± 2.15	95.19	**97.22** ± **1.80**	99.02
	CPU Time	649.53	2401.26	115.66	1.03	373.35	320
Heart-Statlog	*Accuracy*	84.44 ± 1.03	83.54 ± 0.99	83.57 ± 0.95	81.17	**85.75** ± **0.44**	86.29 (3)
	Sensitivity	**80.88** ± 1.66	80.19 ± 1.29	78.88 ± 1.62	75.13	80.74 ± **0.65**	82.06
	Specificity	87.67 ± 1.44	86.03 ± 1.10	87.46 ± 1.26	86.20	**89.71** ± 0.89	91.40
	MCC	67.03 ± 2.05	66.55 ± 2.04	66.46 ± 2.11	61.92	**71.09** ± 1.07	73.55
	AUC(ROC)	**90.75** ± 0.69	90.37 ± 0.69	89.25 ± 0.81	88.82	88.55 ± **0.67**	90.23
	CPU Time	628.69	517.4	132.46	0.36	314.34	300
							5, 11
Colon Tumor	*Accuracy*	86.40 ± 1.63	86.56 ± 0.92	85.70 ± 1.33	88.08	**91.41** ± **1.3**	93.33 (3)
	Sensitivity	87.85± 2.11	88.69 ± 1.28	89.43± 0.99	90.05	**95.57** ± **1.91**	100
	Specificity	83.65 ± 3.69	83.86 ± 0.77	80.18± 3.59	84.51	**85.90** ± 4.83	96.29
	MCC	65.77 ± 4.22	66.48 ± 2	64.36 ± 3.19	68.39	**76.68** ± **3.97**	92.5
	AUC(ROC)	86.71 ± 1.43	86.08± 1.78	85.34 ± 1.48	**89.32**	87.68 ± 1.94	100
	CPU Time	361.42	776.88	76.27	2.52	114.48	102
Central Nervous System	*Accuracy*	90.27 ± 1.97	87.96 ± 1.76	84.47 ± 4	87.93	**91.85** ± 1.96	93.33 (5)
	Sensitivity	96.26 ± 0.46	94.93 ± 0.52	91.88 ± 3.03	96.83	**96.90** ± 1.26	98.33
	Specificity	80.26 ± 4.49	76.77 ± 4.86	71.72 ± 6.53	73.10	**83.85** ± 5.27	93.51
	MCC	72.45 ± 8.11	66.83 ± 5.09	60.49 ± 7.63	66.29	**75.41** ± **7.50**	89.14
	AUC(ROC)	89.44 ± 2.75	88.19 ± 1.47	86.24 ± 3.39	89.27	**93.06** ± **3.2**	100
	CPU Time	333.05	1009.03	139.93	7.22	107.47	116.68
							8,14
ALL-AML (Leukemia)	*Accuracy*	98.55 ± 1.20	98.11 ± 0.73	96.59 ± 0.73	98	**98.61** ± 1.23	100 (2)
	Sensitivity	**99.93** ± 0.37	99.63 ± 0.68	99.01 ± 0.72	99.10	98.88 ± 1.43	100
	Specificity	96.56 ± 2.83	95.29 ± 1.18	92.55 ± 1.95	95.47	**98.77** ± 1.99	100
	MCC	92.47 ± 1.73	91.57 ± 1.23	88.74 ± 1.11	91.68	**92.69** ± 1.34	100
	AUC(ROC)	96.30 ± 1.44	95.72 ± 0.52	96.15 ± 0.49	95.79	**96.38** ± 1.38	100
	CPU Time	268.39	1013.63	87.47	1.95	101.84	91.71
Breast Cancer	*Accuracy*	83.94 ± 1.85	83.72 ± 0.98	79.56 ± 2.25	84.22	**87.77** ± 1.78	91.11 (6)
	Sensitivity	84.84 ± 3.53	86.24 ± 2.01	80.84 ± 1.47	87.57	**87.74** ± 3.18	94.66
	Specificity	83.72 ± 3.67	81.96 ± 1.77	79.60 ± 3.55	82.17	**88.49** ± 3.17	94.66
	MCC	68.61 ± 2.04	67.93 ± 2	59.99 ± 3.90	67.46	**75.45** ± 3.83	83.85
	AUC(ROC)	91.23 ± 2.62	90.66 ± 1.04	88.25 ± 1.50	92.50	**93.47** ± 2.83	97.38
	CPU Time	407.37	909.25	105.89	6.02	129.26	102
							25,47
Ovarian Cancer	*Accuracy*	99.64 ± 0.35	99.52 ± 0.11	98.58 ± 0.96	98.63	**99.82** ± 0.34	100 (3)
	Sensitivity	99.62 ± 0.43	99.66 ± 0.20	99.17 ± 0.80	98.77	**100** ± 0.0	100
	Specificity	**99.77** ± 0.51	99.31 ± 0.68	97.57 ± 1.11	98.30	99.58± 0.85	100
	MCC	99.29 ± 0.82	99 ± 0.24	96.85 ± 2.01	97.16	**99.69** ± 0.61	100
	AUC(ROC)	99.99 ± 0.009	99.97 ± 0.03	99.91 ± 0.11	99.95	**100** ± 0.0	100
	CPU Time	743.96	1086.28	170.17	1.99	266.60	100

Table 6
Average number of selected feature.

Dataset	BPSO-RF	GA-RF	SA-RF	CFS	BBHA-RF
Wisconsin diagnostic breast cancer	13.5 ± 3.53	25± 0.70	13 ± 1.41	9 ± 0.0	5.4 ± 2.40
Parkinson's	9 ± 1.41	10 ± 1	5.5± 0.7	9 ± 0.0	4 ± 0.70
Heart-Statlog	8.2 ± 0.83	9 ± 2.08	7 ± 2.94	6± 0.0	4.8 ± 1.78
Colon Tumor	20.8 ± 1.30	18.6± 2.96	11.8 ± 1.92	14 ± 0.0	3.4 ± 0.54
Central Nervous System	24.4 ± 5.45	27 ± 5.14	14.02 ± 1.92	28 ± 0.0	8.4 ± 3.20
ALL-AML (Leukemia)	26 ± 2.16	16.4 ± 9.20	16 ± 1.87	7 ± 0.0	5.6 ± 2.70
Breast Cancer	24.2 ± 2.16	22 ± 8.15	14.02 ± 1.30	26 ± 0.0	6.2 ± 1.78
Ovarian Cancer	24.8 ± 1.92	22.5 ± 7.18	14.25 ± 4.71	9± 0.0	2.8 ± 0.44



	BBHA-RF	BPSO-RF	GA-RF	SA-RF	CFS
AUC (ROC)	94.47	93.73	93.39	92.48	93.72
accuracy	93.31	91.55	91.12	89.66	90.63
sensitivity	92.55	90.3	90.3	87.97	89.65
specificity	92.85	90.92	89.84	87.96	89.15
MCC	83.43	79.87	78.92	75.99	77.38

Fig. 3. Average solution quality of one filter and four wrapper approaches on 8 medical datasets.

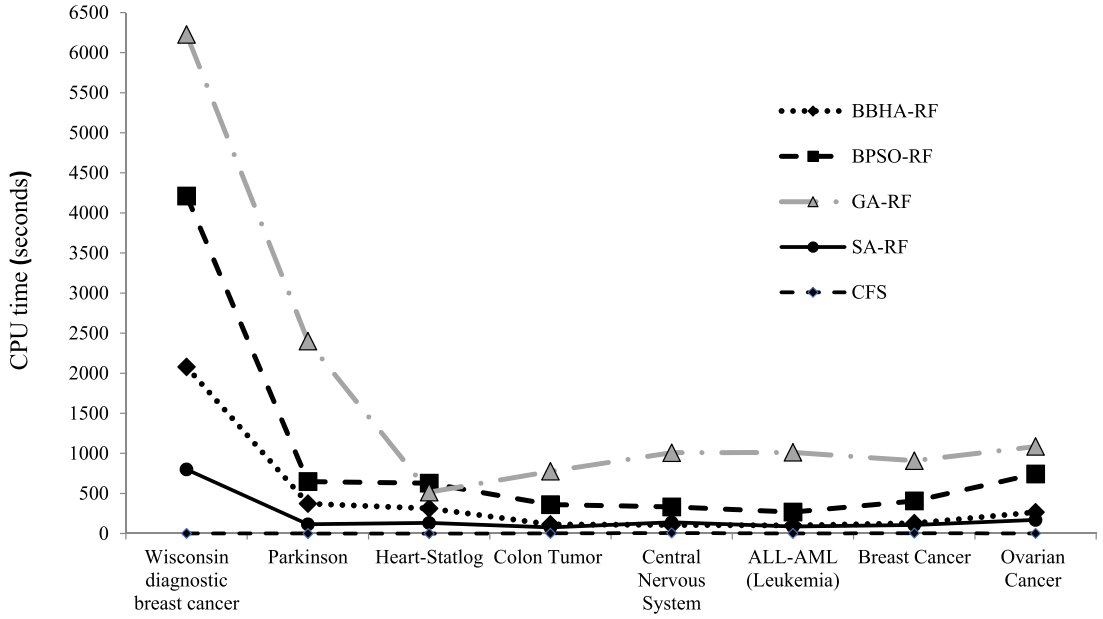


Fig. 4. Computational efficiency of one filter and four wrapper approaches on 8 medical datasets.

For Parkinson's dataset BBHA-RF obtained 94.20% accuracy with average 4 features. The proposed method gave better performance than other approaches reported in [47–49]. For this dataset the highest classification accuracy (98.12%) with 11 features is obtained by a method based on minimum redundancy maximum relevance FS and complex-valued artificial neural network classifier with 10-fold-CV scheme [50].

Table 7
Comparison of relevant works on cancer classification with our proposed method BBHA-RF.

Dataset	BBHA-RF	[59]	[10]	[55]	[60]	[56]	[61]	[57]	[58]
Colon Tumor	91.41 (3.4)	**100(2)**	91.9(18.0)	93.32(8)	93.55(6)	96.67 (20)	99.44(5)	90(2)	93.5(9)
Central Nervous System	**91.85 (8.4)**	–	–	–	–	–	–	90(2)	86.6(7)
ALL-AML(Leukemia)	98.61 (5.6)	97.38(3)	97.2(18.7)	98.61(7)	98.74(4)	**100 (17)**	99.10(10)	–	100(5)
Breast Cancer	87.77 (6.2)	95.86(4)	93.4(26.9)	–	–	96 (12)	**100 (20)**	–	–
Ovarian Cancer	**99.82 (2.8)**	99.44(4)	–	–	–	–	–	–	98.8(19)

Table 8
Best subsets of genes which found by BBHA-RF.

Dataset	Accuracy (# of genes)	Name of genes
Colon Tumor	93.33 (3)	X12671, M16937, M91463
Central Nervous System	93.33 (5)	S71824_at, D83542_at, AF002020_at, HG2417-HT2513_at, U43747_s_at
ALL-AML(Leukemia)	100 (2)	L09209_s_at, M92287_at
Breast Cancer	91.11 (6)	Contig24311_RC, Contig7258_RC, NM_005192, Contig38726_RC, Contig14882_RC, NM_003450
Ovarian Cancer	100 (3)	MZ2.8234234, MZ418.49538, MZ435.46452

For Heart-Statlog dataset, the best accuracy (89.96%) with only 3 features is obtained by a method which uses self-regulated learning PSO as FS and extreme learning machine as a classifier with 70-30 train-test scheme [51]. BBHA-RF obtained 85.40% accuracy with average 4.8 features. Our proposed algorithm performs better than the results of the literature in [52–54].

5.3. Discussion

In summary, from the experimental results on text and image data sets it is worth noting that CFS selects the lowest number of features on 3/4 datasets in the least amount of running time but suffers in terms of classification accuracy. GA-NB and BPSO-NB obtains good classification accuracy but require more running time and select many more features. Compared to mentioned methods, BBHA with NB is able to find significantly least number of features and to provide better classification performance in a sensible CPU time.

Also, the computational results on medical datasets confirmed that RF gives better performance compared to other five DT classifiers therefore is chosen as a fitness function of optimization algorithms and evaluator of feature subsets. From the experimental results on biological data sets, it is inferred that BBHA-RF approach not only improves classification accuracy of RF by selecting the most informative features, but also obtains better performance in terms of all eight evaluation criteria when compared to BPSO-RF, GA-RF, SA-RF, and CFS filter method. The eight evaluation criteria are classification accuracy, MCC, AUC (ROC), sensitivity, specificity, the number of selected features, CPU time, and robustness. Because 6/8 biological datasets are unbalanced, considering MCC criteria is more suitable. Average MCC of BBHA-RF is significantly better than all mentioned algorithms and selects significantly fewer feature subsets on all datasets in a reasonable time. BBHA-RF converges much faster than GA-RF and BPSO-RF. The speed of proposed approach is comparable to SA-RF. Moreover, the standard deviations of the computational results are relatively small, indicating that the repeated 10-fold-CV is reliable and appropriate for classification of medical data. The comparison of BBHA-RF with other approaches in the literature suggests that BBHA-RF has competitive or better performance. Ultimately, a summary of the best subsets of genes found for each microarray by BBHA-RF is listed in Table 8.

6. Conclusion

Feature selection is an important approach that used before applying classifiers to a data set in order to select informative features. A good FS method by selecting significant features helps to successfully and meaningfully modeling with low computational cost and high classification accuracy. During the past years, several metaheuristic algorithms such as GA, firefly, PSO, binary bat algorithm and ACO make an effort to design the FS as a combinatorial optimization problem. However, almost all of the existing methods have a lot of parameters for configuring the model and are computationally expensive. Therefore, proposing a FS approach with a few parameters, high computational speed, and simplicity are necessary for classification.

In this paper, the BHA is, for the first time, being used to solve a FS problem. By applying the hyperbolic tangent function, a new binary version of BHA called BBHA is used to solve FS in text, image, and biomedical data. Two classifiers (RF and NB) serve as the evaluators of our proposed algorithm. In addition, to confirm that RF is the best DT classifier, the performances of six popular DT algorithms were compared in this study.

Experimental results demonstrate that RF is the best DT algorithm and the proposed BBHA wrapper based FS approach outperforms the performances of BPSO, GA, SA, and CFS in terms of AUC, accuracy, MCC, sensitivity, specificity, and the number of selected optimized features. Furthermore, if the computational cost is taken into account, BBHA wrapper approach performs much faster than BPSO and GA. BBHA only needs a single parameter for configuring the model and is simple to understand. In the future, the proposed method could be applied in other areas such as pattern recognition and bioinformatics for biomarker discovery. Also, the improvement in BBHA by applying a new initialization strategy could be treated as a research subject in the future.

Acknowledgement

The first author thanks, Dr. Abdolreza Hatamlou for his helpful consultation.

References

[1] B. Cao, D. Shen, J.T. Sun, Q. Yang, Z. Chen, Feature selection in a kernel space, in: International Conference on Machine Learning (ICML), USA, 2007, pp. 121–128.
[2] N.S. Maroño, A.A. Betanzos, M.T. Sanromán, Filter methods for feature selection –a comparative study, Intelligent Data Engineering and Automated Learning (IDEAL) (2007) 178–187.
[3] M.A. Hall, Correlation-based Feature Selection for Machine Learning, NewZealand, 1999.
[4] R. Kohavi, G.H. John, Wrappers for feature subset selection, Artificial Intelligence – Special Issue on Relevance 97 (1997) 273–324.
[5] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann, San Francisco, 1993.
[6] L. Breiman, H.J. Friedman, R. Olshen, C. Stone, Classification and Regression Trees, CRC Press, 1984.

[7] J.J. Liu, G. Cutler, W. Li, Z. Pan, S. Peng, T. Hoey, L. Chen, X.B. Ling, Multiclass cancer classification and biomarker discovery using GA-based algorithms, Bioinformatics 21 (2005) 2691–2697.
[8] K.H. Chen, K.J. Wang, M.L. Tsai, K.M. Wang, A.M. Adrian, W.C. Cheng, et al., Gene selection for cancer identification: a decision tree model empowered by particle swarm optimization algorithm, BMC Bioinf. (2014).
[9] H.M. Harb, A.S. Desuky, Feature selection on classification of medical datasets based on particle swarm optimization, Int. J. Comput. Appl. (2014) 14–17.
[10] S. Li, X. Wu, M. Tan, Gene selection using hybrid particle swarm optimization and genetic algorithm, Soft Comput. (Springer) 12 (2008) 1039–1048.
[11] F. Gonzalez, L.A. Belanche, Feature selection for microarray gene expression data using simulated annealing guided by the multivariate joint entropy, Comput. Syst. 18 (2014) 275–293.
[12] A. Hatamlou, Black hole: a new heuristic optimization approach for data clustering, Inf. Sci. 222 (2013) 175–184.
[13] K. Lenin, B.R. Reddy, M.S. Kalavathi, Black hole algorithm for solving optimal reactive power dispatch problem, Int. J. Res. Manage. Sci. Technol. 2 (2014) 2321–3264.
[14] M. Farahmandian, A. Hatamlou, Solving optimization problem using black hole algorithm, J. Adv. Comput. Sci. Technol. 4 (2015) 68–74.
[15] J. Zhang, K. Liu, Y. Tan, X. He, Random black hole particle swarm optimization and its application, in: IEEE International Conference on Neural Networks and Signal Processing, China, 2008, pp. 359–365.
[16] N. Ghaffarzadeh, S. Heydari, Optimal coordination of digital overcurrent relays using black hole algorithm, TI J. World Appl. Program. 5 (2015) 50–55.
[17] S. Kumar, D. Datta, S.K. Singh, Black hole algorithm and its applications, Comput. Intell. Appl. Model. Control 575 (2014) 147–170.
[18] M. Nemati, R. Salimi, N. Bazrka, Black hole algorithm: a swarm algorithm inspired of black holes for optimization problem, IAES Int. J. Artif. Intell. 2 (2013).
[19] M. Nemati, H. Momeni, N. Bazrkar, Binary black hole algorithm, Int. J. Comput. Appl. 79 (2013) 36–42.
[20] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[21] G. e. Biau, Analysis of a random forests model, J. Mach. Learn. Res. 13 (2012) 1063–1095.
[22] J.R. Carr, Statistical self-affinity, fractal dimension, and geologic interpretation, Eng. Geol. 48 (1997) 269–282.
[23] H. Hassan, A. Badr, M. Abdelhalim, Prediction of O-glycosylation sites using random forest and GA-tuned PSO technique, Bioinf. Biol. Insights 9 (2015) 103–109.
[24] R. Díaz-Uriarte, S. Alvarez de Andrés, Gene selection and classification of microarray data using random forest, BMC Bioinf. (2006).
[25] A. Anaissi, P.J. Kennedy, M. Goyal, D.R. Catchpoole, A balanced iterative random forest for gene selection from microarray data, BMC Bioinf. (2013).
[26] L. Breiman, Bagging predictors, Mach. Learn. 24 (1996) 202–207.
[27] D.J. Dittman, T.M. Khoshgoftaar, A. Napolitano, A. Fazelpour, Select-Bagging: effectively combining gene selection and bagging for balanced bioinformatics data, in: IEEE International Conference on Bioinformatics and Bioengineering (BIBE), USA, 2014, pp. 413–419.
[28] T.S. Lim, W.Y. Loh, Y.S. Shih, A comparison of prediction accuracy, complexity and training time of thirty-tree old and new classification algorithms, Mach. Learn. 40 (2000) 203–228.
[29] M.C. Tsai, K.H. Chen, C.T. Su, H.C. Lin, An application of PSO algorithm and decision tree for medical problem, in: 2nd International Conference on Intelligent Computational Systems (ICS'2012), Indonesia, 2012, pp. 124–126.
[30] K.H. Chena, K.J. Wanga, K.M. Wang, M.A. Angelia, Applying particle swarm optimization-based decision tree classifier for cancer classification on gene expression data, Appl. Soft Comput. 24 (2014) 773–780.
[31] M. Kuhn, K. Johnson, Applied Predictive Modeling, Springer, 2013.
[32] RuleQuestResearch, Is See5/C5.0 Better Than C4.5? 2012 (Available:) http://rulequest.com/see5-comparison.html.
[33] R.E. Schapire, Y. Freund, Boosting: Foundations and Algorithms, 2013.
[34] E. Pashaei, M. Ozen, N. Aydin, Improving medical diagnosis reliability using boosted C5.0 decision tree empowered by particle swarm optimization, in: 37th Annual International Conference of the Engineering in Medicine and Biology Society, Milano, 2015.
[35] E. Pashaei, M. Ozen, N. Aydin, A novel gene selection algorithm for cancer identification based on random forest and particle swarm optimization, in: Presented at the Computational Intelligence in Bioinformatics and Computational Biology, Niagara Falls, Canada, 2015.
[36] W.Y. Loh, Classification and regression trees (overview) Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. 1 (2011) 14–23.
[37] C.D. Bastian, G.A. Rempala, Gene selection with sequential classification and regression tree algorithm, Biostat. Bioinf. Biomath. 2 (2011) 157–186.
[38] X. Wu, V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, et al., Top 10 algorithms in data mining, Knowl. Inf. Syst. 14 (2008) 1–37.
[39] L. Kaper, E. Heuvel, P. Woudt, R. Giacconi, Black hole research past and future, in Black Holes in Binaries and Galactic Nuclei: Diagnostics, Demography and Formation, Springer, Berlin/Heidelberg, 1999, pp. 3–15.
[40] C.A. Pickover, Black Holes: A Traveler's Guide, John Wiley & Sons, United States of America, 1996.
[41] S. Mirjalili, A. lewis, S-shaped versus V-shaped transfer functions for binary particle swarm optimization, Swarm Evol. Comput. 9 (2013) 1–14.
[42] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (1999) 531–537.
[43] D. Lavanya, K.U. Rani, Analysis of feature selection with classification: breast cancer datasets, Indian J. Comput. Sci. Eng. (IJCSE) (2011).
[44] D. Lavanya, Ensemble decision tree classifier for breast cancer data, Int. J. Inf. Technol. Convergence Serv. 2 (1) (2012) 17–24.
[45] A. Darzi, M. AsgharLiaei, Feature selection for breast cancer diagnosis: a case-based wrapper approach, Eng. Technol. Int. J. Med. Health Biomed. Bioeng. Pharm. Eng. 5 (5) (2011).
[46] T. Sridevi, A. Murugan, A novel feature selection method for effective breast cancer diagnosis and prognosis, Int. J. Comput. Appl. 88 (11) (2014) 28–33.
[47] M.F. Caglar, B. Cetisli, I.B. Toprak, Automatic recognition of parkinson's disease from sustained phonation tests Using ANN and adaptive neuro-fuzzy classifier, J. Eng. Sci. Des. 1 (2) (2010) 59–64.
[48] A. Ozcift, A. Gulten, Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms, Comput. Methods Prog. Biomed.(elsevier)e 104 (2011) 443–451.
[49] I. Psorakis, T. Damoulas, M.A. Girolami, Multiclass relevance vector machines: sparsity and accuracy, IEEE Trans. Neural Netw. 21 (10) (2010) 1588–1598.
[50] M. Peker, B. Şen, D. Delen, Computer-Aided diagnosis of Parkinson's disease using complex-valued neural networks and mRMR feature selection algorithm, J. Healthcare Eng. 6 (3) (2015) 281–302.
[51] C.V. Subbulakshmi, S.N. Deepa, Medical dataset classification: a machine learning paradigm integrating Particle Swarm Optimization with extreme learning machine classifier, Sci. World J. (2015) (Hindawi Publishing).
[52] F.J. Martíınez-Estudillo, C. Hervías-Martíınez, P.A. Gutiíerrez, Evolutionary product-unit neural networks classifiers, Neurocomputing 72 (1–3) (2008) 548–561.
[53] F. C.Hervías-Martíınez, J.Martíınez-Estudillo, andM.Carbonero-Ruz, Multilogistic regression by means of evolutionary product-unit neural networks, Neural Netw. 21 (7) (2008) 951–961.
[54] P. Jaganathan, R. Kuppuchamy, A threshold fuzzy entropy based feature selection for medical database classification, Comput. Biol. Med. 43 (12) (2013) 2222–2229.
[55] A. El Akadi, A. Amine, A. El Ouardighi, D. Aboutajdine, A two-stage gene selection scheme utilizing MRMR filter and GA wrapper, Knowl. Inf. Syst. 26 (3) (2011) 487–500.
[56] N.Y. Moteghaed, K. Maghooli, S. Pirhadi, M. Garshasbi, Biomarker discovery based on hybrid optimization algorithm and artificial neural networks on microarray data for cancer classification, J. Med. Signals Sens. (2015) 88–96.
[57] X. Wang, O. Gotoh, Microarray-based cancer prediction using soft computing approach, Cancer Informatics7 (2009) 123–139.
[58] E. Huerta, B. Duval, J. Hao, Gene Selection for Microarray Data by a LDA-Based Genetic Algorithm, Springer, 2008, pp. 252–263.
[59] E. Alba, J. Garcıa-Nieto, L. Jourdan, E. Talbi, Gene selection in cancer classification using PSO/SVM and GA/SVM hybrid algorithms, IEEE Transactions on Evolutionary Computation (2007) 284–290.
[60] M. Abdi, S. Hosseini, M. Rezghi, A novel weighted support vector machine based on Particle Swarm Optimization for gene selection and tumor classification, Corporation Computational and Mathematical Methods in Medicine(Hindawi Publishing) (2012).
[61] B. Sahua, D. Mishra, A novel feature selection algorithm using Particle Swarm Optimization for cancer microarray data, International Conference on Modeling Optimization and Computing (ICMOC-2012) 38 (2012) 27–31.