



A novel gene selection method using modified MRMR and hybrid bat-inspired algorithm with β -hill climbing

Osama Ahmad Alomari¹ · Ahamad Tajudin Khader¹ · Mohammed Azmi Al-Betar² · Mohammed A. Awadallah³

Published online: 26 June 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

This paper proposed a new gene selection method based on modified Minimum Redundancy Maximum Relevancy (MRMR) as a filtering approach and hybrid bat algorithm with β -hill climbing as an efficient wrapper approach. The gene selection is a process of selecting the discriminative genes that aid in the development of efficient cancer diagnosis and classification. In general, the current filter-based approaches produced gene subset according to its discriminative power. However, one of the deficiencies of single filter approaches is that it has high variability of the classification results. Accordingly, this study aim to improve MRMR through incorporating its with ensemble of filters to increase the robustness and the stability of MRMR. The result of filtering-based approach is a set of discriminative genes. The wrapper-based approach considers the results from the filtering-based approach to formulate the gene selection search space. In wrapper approach, bat algorithm is tailored for gene selection problem and hybridized with a powerful local search method called beta hill climbing to further stress the deep learning side in the search space navigation and thus find a very robust and stable discriminative genes. Bat-inspired algorithm (BA) is a recent swarm-based optimization method while β -hill climbing is an exploratory local search. The proposed method is called Robust MRMR and Hybrid Bat-inspired Algorithm (rMRMR-HBA). To evaluate the proposed method, ten well-known microarray datasets are experimented with. These datasets are varies in terms of number of genes, samples, and classes. For performance evaluation, the proposed filtering-based approach (i.e., rMRMR) is initially tested against the standard MRMR and other well-regard filtering approaches. Thereafter, the wrapper-based approach (i.e., HBA) is evaluated by studying the convergence behavior of BA with and without β -hill climbing. For comparative evaluation, the results of the proposed rMRMR-HBA were compared with state-of-art methods using the same microarray datasets. The comparative results show that our proposed approach achieved outstanding results in two out of ten datasets in terms of clarification accuracy and minimum number of genes.

Keywords Ensemble feature selection · Bat-inspired algorithm · Optimization · Gene selection · MRMR · SVM · Classification

1 Introduction

Detection and classification of cancers are the key issues in microarray technology [1]. The advent of DNA

microarray technology facilitates biologist in providing new insight of biological phenomena and monitoring the activity of thousand of genes (features) in one experiment. Normally, the DNA microarray technology generates gene expression datasets considered as high dimensionality datasets. This dataset consists of huge number of genes and limited number of samples, where the columns represent genes and instances represent clinical patients samples [2, 3]. These data carry precious information, which have great potential for cancer diagnosis and classification. However, the existence of a huge number of genes is a challenging issue for development of an efficient classifier (or called machine learning algorithm) [3]. To address this challenge and to improve the predictive accuracy of diseases, gene selection, also known as feature selection, which is a data preprocessing step in data

✉ Osama Ahmad Alomari
oasa14_com004@student.usm.my

Mohammed Azmi Al-Betar
mohbetar@bau.edu.jo

¹ School of Computer Sciences, Universiti Sains Malaysia (USM), Pulau Pinang, Malaysia

² Department of Information Technology, Al-Huson University College, Al-Balqa Applied University, P.O. Box 50, Al-Huson, Irbid, Jordan

³ Department of Computer Science, Al-Aqsa University, P.O. Box 4051, Gaza, Palestine

mining, can be used to find the subset of most informative genes [4].

Conventionally, methods for gene selection are divided into three categories, namely, wrapper approach, filter approach, and embedded approach [4]. In wrapper approach, the classifier is employed to assess the reliability of genes or gene subsets. In the filter approach, the machine learning algorithm is not used to remove irrelevant and redundant features; instead, the principal characteristics of the training data were applied to rank the significance of the gene subset or genes [5], such as the Minimum Redundancy Maximum Relevancy (MRMR) [6], ReliefF [7], Chi-square [8], KullbackLeibler [9]. The wrapper method provides better results. However, the filter method is computationally less expensive than wrappers, but this method yields less classification accuracy [10]. The embedded method, which is the last category in the gene selection approach, is a hybrid of the filter and wrapper methods. The integration of the features of both approaches leads to detect the informative genes with high classification accuracy [11]. However, the embedded method is still in its infancy. Therefore, further investigations are required for both filter and wrapper methods to develop efficient hybrid/embedded methods for gene selection [3].

In filtering-based approaches, which is one of the core concern of this paper, each filter relies on different metric of various characteristics (distance, probability distribution, information theory, etc.). For a specific dataset, selecting a filter approach is a very crucial for classification-oriented approach where different filtering approach leads to diverse subset of genes [12]. Restate, the filter approach obtained best results for specific data set may not do so for another, which imply that there exist a high variability in the classification outcomes. Therefore, several filtering approaches are combined together in order to find robust subset of genes. Examples of combining filtering-based approaches are i) Seijo-Pardo et al. [13] studied the ensemble of homogeneous and heterogeneous filter-based gene subset selection approaches ii) Ebrahimpour and Eftekhari [14] utilized hesitant fuzzy sets for representing ensemble of ranking algorithms and ensemble of similarity measures for gene subset selection iii) Boln-Canedo et al. [12] presented ensemble of filters and classifiers for microarray data classification.

In wrapper-based approaches, the classification of the gene subset is accomplished through two stages: searching and evaluation. In the search stage, search-based method are utilized to generate a discriminative gene subset based on an efficient classifier [4]. Finding optimal subset of genes has been shown to be NP-hard problem [15]. Therefore, meta-heuristic-based approaches have been devoted as a search method in wrapper-based approaches (i.e., natural inspired evolutionary algorithms) [16]. Several meta-heuristic-based

approaches have been applied to solve gene selection problem such as Harmony search with a Markov blanket (HSA-MB) [17], Binary Particle Swarm Optimization and a Combat Genetic Algorithm (BPSO-CGA) [18], and Genetic Algorithm with Artificial Bee Colony [16]. However, most of these approaches still suffer from the problems of stagnation in local optima [17, 19]. Therefore, a strong search method for identifying near-optimal subset of informative genes for accurate cancer classification is still required. One of the efficient meta-heuristic-based approach that have been recently proposed is bat-inspired algorithm (BA) [20].

BA is a meta-heuristic swarm-based method introduced by Xin-She Yang [20] imitated echolocation behavior of swarming of bats. BA employs the idea of combining population based algorithm and local search to empower the global diverse exploration and local intensive exploitation, which are the key success of meta-heuristic methods. BA has other advantages over meta-heuristic based approaches such that the simplicity, flexibility, adaptability and scalability. Therefore, BA has been successfully adapted for a wide variety of optimization problems, such as scheduling [21, 22], inverse problem and parameter estimation [23, 24], classifications [25], gene selection [26], clustering [27], global optimization problem [28], and image processing [29]. In order to be inline with the complexity of the optimization problems, the theory of BA has been also tweaked where several hybridized versions [30, 31] and modified versions [32, 33] are produced. However, BA is not yet properly investigated for a kind of gene selection problems addressed in present work.

In this paper, a hybrid filter/wrapper gene selection method is proposed based on improved MRMR methods and hybridized BA algorithm [20] with β -hill climbing for cancer classification. In the proposed hybrid method, the filter approach (i.e., improved MRMR) is initially work to find the highly discriminative genes. Thereafter, these genes are used by hybrid BA (HBA) as a strong initial stage to find the most informative subset of genes. To elaborate, the performance of MRMR method is improved via ensemble mechanism in which several filtering-based approaches are combined (i.e., ReliefF, Chi-square, Kullback-Leibler) to obtain a robust gene list. This list is used by MRMR method to improve the filtering stage outcomes. The hybridization version in wrapper stage is suggested in order to improve the performance of BA algorithm. β -hill climbing is hybridized with BA to empower its intensification capabilities to be more suitable for the complexity of gene selection search space. The proposed method is called Robust MRMR and Hybrid Bat-inspired Algorithm (RMRMR-HBA). Note that SVM classifier is employed to evaluate each candidate gene subset in HBA. Specifically, the main contributions of the proposed method are as follows:

- Modified version of MRMR through novel hybridization strategy by combining the filtering mechanism of MRMR with aggregate gene list obtained by ensemble of filter methods to select more robust and stable gene list for high classification accuracy.
- New wrapper approach that embeds of the hybridization BA algorithm with β -hill climbing local search algorithm to guide the search for more informative gene subset.
- To investigate whether the proposed method (rMRMR-HBA) can obtain a gene subset with small number of genes and better classification accuracy when compared with state-of-art gene selection approaches.

Extensive experiments were done on ten well-known microarray datasets to test rMRMR-HBA method for gene selection problem. These datasets characteristics are different in terms of number of genes, samples, and classes. For performance evaluation, the proposed rMRMR is initially tested against the standard MRMR and other well-regard filtering approaches. Thereafter, HBA is evaluated by studying the convergence behavior of BA with and without β -hill climbing. For comparative evaluation, the results of the proposed rMRMR-HBA were compared with state-of-art methods using the same microarray datasets. The comparative results show that the proposed method achieved the best results in two out of ten datasets and competitive results for others.

The remaining parts of this paper is organized as follows: related background is provided in Section 2. The proposed method is discussed in Section 3. The experimental results and comparative evaluation is analyzed in Section 4. Finally, Section 5 concludes the paper and provide a possible future enhancements.

2 Research background

In order to build a self-exploratory paper, the research background of this paper is described to provide the necessary knowledge about the research methodology. Therefore, the Chi-Square, ReliefF, and Kullback-Liebler filtering methods are fully overviews. The fundamental elements of Bat-inspired algorithm together with its Binary version are provided. Finally, β -hill climbing algorithm and Support Vector Machine (SVM) are thoroughly illustrated.

2.1 Chi-Square filtering-based method

Chi-square is statistical-based measure which is widely used as feature selection filtering-based method. In general, Chi-square individually evaluates each features with respect to its related classes [8]. It should note that the

Chi-Square measurement is work with a feature of discrete nature and therefore features should be discretized form continuous values to discrete intervals. Subsequently, Chi-square measurement can determines a potential merging whether the relative frequencies of the classes in the adjacent intervals are similar enough. Of the N examples, let N_{ij} be the number of samples of the C_i class within the j^{th} interval and M_{ij} is the number of samples in the j^{th} interval. The expected frequency of N_{ij} is $E_{ij} = (M_{ij} * |C_i|) / N$. The Chi-squared statistic of a feature is then calculated as:

$$X^2 = \sum_{i=1}^c \sum_{j=1}^I \frac{(N_{ij} - E_{ij})^2}{E_{ij}} \quad (1)$$

where I is the number of intervals. To put it clearly, this method tries to determine the type of relationship between the gene and class label. If the gene and class are dependent, the frequency of the gene can be used to predict the frequency of the class. The task is to select the gene, of which the frequency is highly dependent on the frequency of the class. If the gene and class are independent, the frequency count is close to the expected frequency count, thus a small chi square X^2 score. So a high value of X^2 score indicates that the hypothesis of independence is incorrect. In other words, the larger the X^2 value is, the more informative of the corresponding feature will be. By means of calculating the X^2 for every feature, the features can be further processed and ranked based on their X^2 score values.

2.2 ReliefF filtering-based method

ReliefF [7] is an extension of the original Relief filtering-based algorithm [34] and its considered as distance-based measure [35]. ReliefF algorithm works as follows: Firstly, perform randomly sampling an instance from the data set. Secondly, from both same and opposite class, nearest neighbor will be located to the sampled instance. Finally, the relevance scores for each feature are updated by comparing the values of the nearest neighbor features to the sampled instance. The rationale behind this is that the meaningful feature should be recognized among instances of different classes which have same value of instances from the same class. ReliefF has advantages over standard Relief: its more robust, better handles incomplete and noisy data, capability of dealing with multiclass problems, can be adapted in all situations, has low bias, facilitates interactions among features, and may be able to capture local dependency which other methods cannot do so.

2.3 Kullback-Liebler filtering-based method

Kullback-Leibler (KL) [9] is probability-based measure attempt to figure out features where distance between the

different class probabilities is big. The KL divergence is a measure of how different two probability distributions (over the same event space) are. The KL divergence or relative entropy of probability distributions P , Q on a finite set X is defined as:

$$MI(X, Y) = KL((P(X, Y)||P(X)P(Y))). \quad (2)$$

$$D(P||Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)} \quad (3)$$

In case the random variable X and Y are independent, then joint probability become equals to the product of marginal probabilities.

$$P(X, Y) = P(X)P(Y) \quad (4)$$

Consequently, the KL divergence between distributions will be zero. This implies that X and Y are independent and Y carries no information about X or X not relevant to Y . Similarly, if X and Y and highly dependent and X carries information about Y then their relevance is higher.

2.4 Bat-inspired algorithm

Bat-inspired algorithm (BA), introduced by Xin-She Yang in 2010 [20], emulates the echolocation behavior of bats. There are many kinds of bats. All of them have similar behavior when navigating and hunting; however, they are different in terms of size and weight. Microbats use echolocation which assists them in seeking for prey and/or avoid obstacles in the complete darkness. In BA, the echolocation features of microbats can be idealized according to the following rules:

1. All bats use echolocation to sense distance and determine the difference between food/prey and background barriers in some magical way;
2. Bats randomly fly with velocity V_i at position X_i with a fixed frequency f_{min} , varying wavelength k and loudness A_0 to seek for prey. They can automatically regulate the wavelength (or frequency) of their emitted pulses and adjust the rate of pulse emission $r \in [0,1]$, depending on the closeness of their target;
3. Although the loudness can vary in many ways, it is assumed that it varies from a large (positive) A_0 to a minimum constant value A_{min} .

• Bat Motion:

The frequency of each bat will be positive integer or float relying on the selected upper bound and lower bound of the frequency. The frequency value is calculated through (5). Determining the upper and lower bound frequencies is based on the domain of interest.

$$F_i = F_{min} + (F_{max} - F_{min}) \times \beta \quad (5)$$

where β is a random number of uniform distribution in $[0, 1]$, F_{max} is upper bound of the frequency, and F_{min} is lower bound of the frequency. The velocity of each bat will be a positive integer number. Each bat will update its velocity according to the following equation.

$$V_i(t + 1) = V_i(t) + (X_i(t) - Gbest) \times F_i \quad (6)$$

Where $Gbest$ is the best solution, F_i represents the frequency of the i th bat and the position X_i of each bat. Each bat's position is updated as shown in (7).

$$X_i(t + 1) = X_i(t) + V_i(t + 1) \quad (7)$$

The BA employed a random walk to improve its capability in exploitation as given below.

$$x_{new} = x_{old} + \epsilon A_t \quad (8)$$

where X_{new} is the new solution, X_{old} is the current solution, and ϵ is random number in $[-1, 1]$.

• Variations of loudness and pulse rates:

Once a bat finds its prey, the loudness usually decreases and the rate of pulse emission increases. In this case, the loudness can be chosen as any value of convenience. Loudness A and pulse emission rate r are updated according to (9) and (10).

$$A_i(t + 1) = \alpha A_i(t) \quad (9)$$

$$r_i(t + 1) = r_i(0)(1 - e^{(-\gamma \times t)}) \quad (10)$$

where α and γ are constant parameters that lies between 0 and 1 and used to update loudness rate A_i and pulse rate (r_i). The pseudo code of the algorithm is presented in the following pseudo-code. Note that $f(X_i)$ is that fitness function value of X_i .

Algorithm 1 bat-inspired algorithm

```

1: Initialize the bat population  $X_i(i = 1, 2, \dots, n)$  and  $V_i$ .
2: Define pulse frequency  $F_i$ .
3: Initialize pulse rate  $r_i$  and the loudness  $A_i$ .
4: while  $t < Maxnumberofiterations$  do
5:   Generate new solutions by adjusting frequency.
6:   Updating velocities and positions [(7) to (6)].
7:   if  $rand > r_i$  then
8:     Select a solution  $X$  among the best solutions randomly.
9:     Generate a local solution around the selected best solution
10:  end if
11:  if  $rand < A_i$  and  $f(X_i) < f(x^*)$  then
12:    Accept the new solutions
13:    Increase  $r_i$  and reduce  $A_i$ 
14:  end if
15:  Rank the bats and find the current  $Gbest$ 
16: end while

```

2.5 Binary bat-inspired algorithm

In the continuous version of BA, the artificial bat can be moved around the search space utilizing positions

and velocity vectors (or updating position vectors) within continuous domain. However, in dealing with binary search space, the position or solution is a series of 0's and 1's binary bits. The updating position process cannot be performed using (6) because the binary search space deals with only two numbers (0 and 1). Therefore, a transfer strategy should be determined to reflect the velocity vector value in changing the elements of position vector from 0 to 1 or vice versa. The solution in gene selection problem is modified to be as a binary vector and the binary bat algorithm is effective in dealing with it.

Nakamura et al. [36] have introduced a binary version of the Bat Algorithm restricting the new bat's position to only binary values using a sigmoid function as follows:

$$S(V_i^j) = \frac{1}{1 + e^{-v_i^j}} \tag{11}$$

Therefore, (11) can be replaced by:

$$x_i^j \leftarrow \begin{cases} 1, & U(0, 1) > \sigma \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

in which $\sigma = U(0, 1)$. Therefore, (12) can provide only binary values for each bat's coordinates in the Boolean lattice, which represent the presence of absence of the features.

2.6 β -hill climbing algorithm

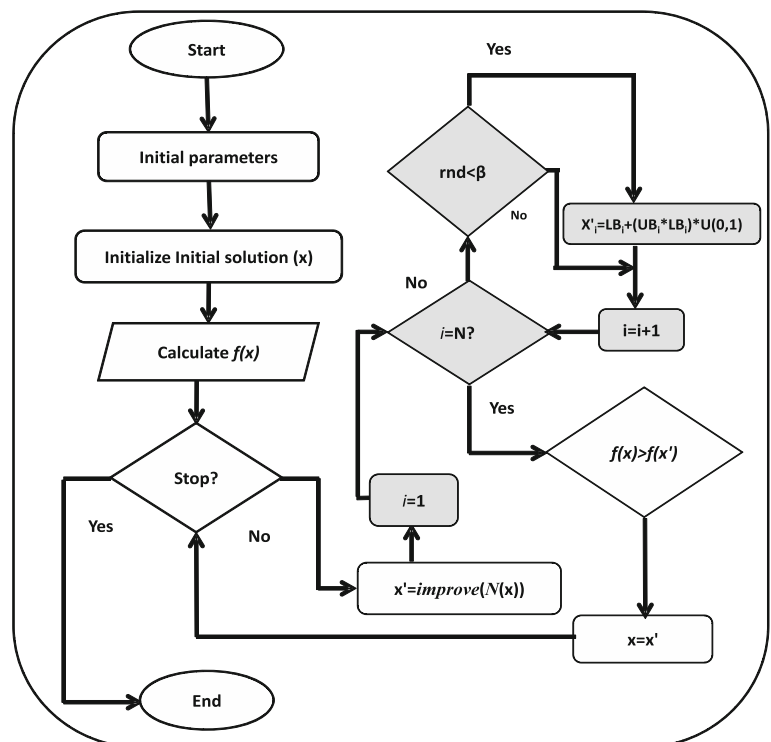
Hill climbing is the base algorithm of the local search-based techniques [37]. It doesn't involve an explorative

strategy during its search process. Thus, it is easily trapped in local optima. As a consequence, several extensions of hill climbing have been developed to handle this problem. Some extended algorithms tried to compensate the weakness of explorative strategies by embedding neighborhood functions to define systematically different search space regions. Other methods like simulating annealing (SA) [38] and greedy deluge (GD) [39] utilized a relaxed acceptance method to deal with uphill moves. Unlike other methods, β -hill climbing [40, 41] is a recent extension method to the basic hill climbing proposed by Al-Betar in 2016, wherein a new explorative operator called (β) has been utilized based on an idea inspired by the uniform mutation operator of GA [42]. At each iteration, the current solution is moved around in the search space based on function called neighborhood search \mathcal{N} and unbounded search space is defined based on the β operator. Figure 1 shows the flow chart of the β -hill climbing algorithm.

The β -hill climbing starts with random initial solution $x = (x_1, x_2, \dots, x_N)$. At each iteration, new solution $\hat{x} = (x'_1, x'_2, \dots, x'_N)$ is generated based on two operators: neighborhood navigation (i.e., \mathcal{N} -operator) and β -operator. In \mathcal{N} -operator stage, a random neighboring solution of the solution x is adopted by means of function *improve*(\mathcal{N}) along with 'random walk' acceptance rule which only move one step without check the change in the objective function, as follow:

$$x'_i = x_i \pm U(0, 1) \times bw \quad \exists! i \in [0, 1]$$

Fig. 1 Flowchart of β -hill climbing algorithm



In the β -operator stage, with a probability range of β , each single variable of the new solution is assigned values base either existing values of the current solution or randomly from available range, as follows:

$$\mathbf{x}'_i \leftarrow \begin{cases} x_r & U(0, 1) > \text{rnd} \leq \beta \\ x_i & \text{otherwise.} \end{cases} \quad (13)$$

where $\beta \in [0, 1]$, $x_r \in X_i$ is the possible range for decision variable x'_i , and rnd generate a uniform random number between 0 and 1. By the aforementioned operators, the convergence of β -hill climbing is boosted toward the optimal solution. Systematically, \mathcal{N} -operator navigates the neighboring solutions of the current one and return the one of the solutions with better objective function value. In β -operator, the convergence can be achieved by constructing a controlled portion of the current solution, and therefore, the convergence rate could be boosted. The β -operator can be considered as source of exploration while the \mathcal{N} -operator can be source of exploitation. In terms of search space, β operator is very useful that empower β -hill climbing algorithm in jumping from search space region to another in the same level or less. By means of this ability, β -hill climbing algorithm can escape the trap of local minima by trying stochastic values for some decision variables.

Algorithm 2 β -hill climbing pseudo code

```

1:  $x_i = LB_i + (UB_i - LB_i) \times U(0, 1), \forall i = 1, 2, \dots, N$  The
   initial solution  $x$ 
2: Calculate  $f(x)$ 
3:  $itr=0$ 
4: while  $itr \leq \text{Max of } itr$  do
5:    $x' = \text{improve}(N(x))$ 
6:   for  $i = 1, \dots, N$  do
7:     if  $\text{rnd} \leq \beta$  then
8:        $x'_i = LB_i + (UB_i - LB_i) \times U(0, 1)$ 
9:     end if  $\text{rnd} \in [0, 1]$ 
10:  end for
11:  if  $f(x') \leq f(x)$  then
12:     $x = x'$ 
13:  end if
14:   $itr = itr + 1$ 
15: end while

```

2.7 Support vector machine

SVM is used as an evaluator for the candidate gene subset solution, which will be used as an objective function in the proposed method as will be seen in Section 3. SVM is a supervised learning algorithm used for classification and regression [43]. It is a powerful classification algorithm due to its efficient performance in pattern recognition domain. For example, SVM classifier was successfully applied to classify high-dimension data, such as microarray gene expression data [2].

SVM constructs hyperplanes or a set of hyperplanes in a high-dimensional space, which can be utilized for

classification, regression, or other tasks. It has the capability to deal with linear and nonlinear datasets. In a linear data, SVM tries to find an optimal separating hyperplane that maximizes the margin between the training examples and the class boundary. In a nonlinear data, such as the microarray gene expression data used in this paper, the definition of a feature mapping function is required where $X \mapsto \phi(x)$. The mechanism that defines a feature mapping process is called kernel function.

1) Polynomial kernel function

$$k = (x_i, x_j) = (x_i \cdot x_j + a)^b \quad (14)$$

2) Radial basis kernel function (RBF)

$$k(x_i, x_j) = e^{-(\|x_i - x_j\|^2 / 2\sigma^2)} \quad (15)$$

3) Sigmoidal kernel function

$$k(x_i, x_j) = \tanh(ax_i \cdot x_j - b) \quad (16)$$

where x_i and x_j are vectors, and a and b are parameters that define the kernel's behavior. Polynomial, RBF, and Sigmoidal are the nonlinear kernels. The generation of a predictive model for cancer classification using microarray dataset is considered a nonlinear classification task [16]. In this paper, RBF kernel will be adopted for cancer classification.

3 Proposed method for gene selection

The proposed method for gene selection has two different stages, namely, the filter approach stage and the wrapper approach stage, as shown in Fig. 2. Each stage will be thoroughly discussed in this paper.

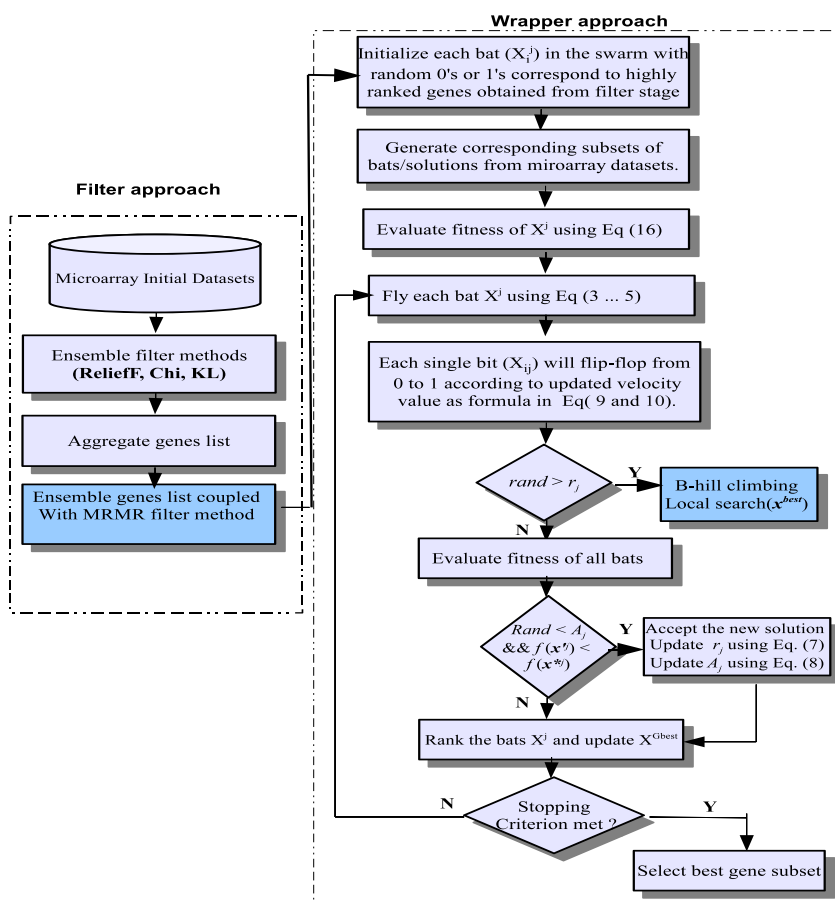
3.1 Stage I: Filter approach

In this stage, the Minimum Redundancy Maximum Relevancy (MRMR) as filter approach [44] is modified by hybridizing its filtering process with ensemble of different filter methods. Initially, the original MRMR is defined in Section 3.1.1. Thereafter, the modified version of MRMR is proposed in Section 3.1.2.

3.1.1 Minimum redundancy maximum relevancy (MRMR)

The minimum redundancy maximum relevancy (MRMR) as a filter approach has been developed by Hanchuan Peng [6]. The MRMR tries to find the most relevant features based on its correlation with class label and to minimize the redundancy of the features themselves [44]. This filtering process reveals the features that have maximum relevancy and minimum redundancy. To quantify both relevancy and redundancy, mutual information is used to estimate the

Fig. 2 Flowchart of the proposed approach for gene selection



mutual dependency of two variables. Mutual information is defined in (17)).

$$I(x_i, c) = \int \int P(x_i, c) \log \frac{P(x_i, c)}{P(x_i)P(c)} \quad (17)$$

where $I(x_i, c)$ is the mutual information between the feature x_i and class label c . $P(x_i)$ and $P(c)$ are marginal probability functions, and $P(x_i, c)$ is the joint probability distribution. The mutual information value for two completely independent random variables is 0 [45].

Given x_i , which represents the feature i , and the class label c , the Maximum-Relevance method selects the top m features in the descent order of $I(x_i; c)$, i.e. the best m individual features that are relevant to the class labels.

$$\max_S \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c) \quad (18)$$

To eliminate the redundancy among features, a Minimum-Redundancy criterion is defined as:

$$\min_S \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i; x_j) \quad (19)$$

A sequential, incremental search algorithm can be used to solve the simultaneous optimizations of optimization criteria of (1) and (19). Suppose we already have S_{m-1} ,

the feature set with $m-1$ features; the task is to select the m -th feature from the set $G - S_{m-1}$. This feature is chosen by maximizing the single-variable relevance minus the redundancy function, as follow:

$$\max_{x_i \in G - S_{m-1}} \left(I(x_i; c) - \frac{1}{m-1} \sum_{x_j \in S_{m-1}} I(x_i; x_j) \right) \quad (20)$$

The method utilizes a series of intuitive measures of relevance and redundancy to pick out promising features of both continuous and discrete datasets, as shown in (17), (18) and (19).

3.1.2 Modified MRMR

The main work in this stage is to modify MRMR to overcome the weakness induced by single filter-based approaches (as mentioned in Section 1) and thus improve the robustness and stability of MRMR. The MRMR is modified via novel hybridization strategy by combining the filtering mechanism of MRMR with aggregate gene list obtained by the ensemble of filter methods. The benefits of ensemble filters is to introduce diversity and reduce this variability associated with single-based filter methods [46]. In this manner, the hybridization of filtering mechanism of

MRMR with an ensemble of filters may lead to improve its robustness and stability. The modified version of MRMR is called rMRMR. Moreover, the pseudo-code for the proposed rMRMR is given in Algorithm 3.

Among the wide range of filters available in the literature, three filters were ensembled according to previous studies [14, 35]. Each of which uses different metrics. Thus, the ensemble method will be a combination of the following filters: ReliefF, Chi-Square and Kullback-Liebler. Thereafter, each individual filter method will be executed and the ranks/scores of genes are combined/aggregated by using ‘Mean’ of the scores for each of the ranking list obtained from each individual filter method. The ranking list and the scores of genes ($R(G_i)$) will be incorporated with filtering process of MRMR as follows: Firstly, the ranking list is used to formulate the dimensionality of search space used by MRMR. The main merit of incorporating process is to empower the outcome of MRMR by mean of initiating its process with high informative genes produced by ensembled methods. Secondly, while MRMR computes its gene scores, the ensemble scores of genes will be used with relevancy computation process, as shown in (21). This is to emphasize the calculation of each individual gene score subject to different metrics (which is avoided the bias result of single metric). Doing so, the robustness and stability of MRMR will be enhanced. (21).

$$\text{Relevancy} = I(G_i, c) * R(G_i) \quad (21)$$

Lastly, on this stage, SVM classifier is used to calculate the classification accuracy of gene list resulting from rMRMR. To further explore reduced gene subset and identify a subset of informative genes, hybrid BA with β Hill climbing and SVM classifier are combined to seek for the better genes subset using the wrapper method.

3.2 Stage II: Wrapper approach

In this stage, we proposed a hybridized version of Bat-inspired as a global optimizer with β -hill climbing as a local optimizer. As a β -hill climbing is a local-area algorithm which can deal with a single solution at a time, they can easily stagnate in some local optima. On the other hand, BA as a population based algorithm can deal with multi-solutions at the same time, however, it performs a wider search without concentration on single area in the search space. Therefore, complementing their capability in a single hybrid optimization framework can be produced an efficient search method. The Hybrid Bat-inspired Algorithm (HBA) is proposed to select the most informative subset of genes from the top ranked genes obtained by rMRMR. As an evaluator to candidate subset of genes, SVM classifier is iteratively used as a fitness function. The goal of our proposed method is to maximize classification accuracy

and minimize the size of genes subset. The pseudo-code of the proposed method is shown in Algorithm 4. The components and process of the proposed method in Stage II are illustrated in the following sections:

Algorithm 3 Hybrid MRMR with ensemble of filter methods

```

1: Input:
2:  $\overline{D}(G_1, G_2, \dots, G_m)$  Dataset with m genes.
3: class c, no of genes to select n.
4: k: number of selected filters.
5: k selected filters  $F \in [\text{Chi-Square, ReliefF, Kullback-Liebler}]$ 
Ensemble of filters
6: for  $i \in \{1, \dots, k\}$  do
7:   for  $j \in \{1, \dots, m\}$  do
8:     Employ  $F_i$  to compute score of gene  $G_i$ 
9:   end for
10: Rank genes according to the score of  $G_i$ , and get new ranking  $F_i(G_j), j = 1, \dots, m$ 
11: end for
12: Create gene ranking list  $\mathbf{R}$  by combining k different filters  $F_i(G_j), i = 1, \dots, k \forall j$  with arithmetic mean. MRMR
13:  $S_{ALL} \leftarrow 1, 2, \dots, G$ 
14:  $S \leftarrow \phi$ 
15:  $S_a \leftarrow \phi$ 
16: for each  $G \in \{1, \dots, m\}$  do
17:    $I(G_i, c) = R(G_i)$ 
18: end for
19:  $MAX = \text{Min}(\text{Length}(S_{ALL}), 1000) // \text{length}(S_{ALL}) \text{return the number of elements in } S_{ALL}$ 
20: for  $i = 1$  to  $MAX$  do
21:    $S_a \leftarrow S_a \cup \underset{i \in S_{ALL} \setminus S_a}{\text{argmax}} I(G_i, c)$ 
22: end for
23:  $S \leftarrow \underset{i \in S_{ALL}}{\text{argmax}} I(G_i, c)$ 
24: while  $\text{Length}(S) < n$  do
25:   for each  $x \in S_a \setminus S$  do
26:      $\text{Relv}(G_x) = I(G_x, c) * R(G_i) //$ 
Enhance Relevancy computing
27:      $\text{Red}(G_x) = \frac{1}{\text{Length}(S)} \sum_{y \in S} I(G_x, G_y)$ 
28:   end for
29:    $S \leftarrow S \cup \underset{x}{\text{argmax}} (\text{Relv}(G_x) - \text{Red}(G_x))$ 
30: end while
Output
31: Build SVM classifier based on top high ranked genes.
```

3.2.1 Solution representation

Gene subset selection problems can be expressed as a combinatorial optimization problem, in which the search space involves a set of all possible subsets [47, 48]. This problem is known to be NP-hard problem [49] which has a highly combinatorial search space in nature. The number of solutions in the search space exponentially increases when the number of genes increases; and there are $[2^N]$ possible subsets of genes, where N represents the number of genes.

The complete set of genes (i.e., solution \mathbf{x}) is represented by a binary string of length N , $\mathbf{x} = (x_1, x_2, \dots, x_N)$, where a bit x_i in the solution is set to 1 if the corresponding gene is preserved and set to 0 if it is to be discarded. N is the original number of all genes.

3.2.2 Fitness function

HBA initially starts with a set of generated bats or solutions. Each bat is a series of 0's and 1's bit, where bit value of 1 indicates that this gene is selected while 0 is discarded. Each bat is evaluated during the search using (22) where the evaluation function combines classification accuracy and gene subset length [50].

$$\alpha \times \gamma R(D) + \beta \times \frac{|C| - |R|}{|C|} \quad (22)$$

where $\alpha R(D)$ is the average of classification accuracy rate obtained by performing ten multiple cross-validation with SVM classifiers, on the training dataset with gene set R chosen from each bat in the population to decision D . $|R|$ is the length of the selected gene subset. $|C|$ is the total number of genes. α and γ are two parameters related to the significance of classification quality and subset length. $\alpha \in [0, 1]$ and $\gamma = (1-\alpha)$. The classification accuracy is more important than the subset length. The quality of each bat is gaged by this objective function. In this paper, α is set to 0.8 [3, 50, 51].

3.2.3 Hybrid bat-inspired and β -hill climbing algorithm (HBA)

In this stage, BA algorithm is hybridized with β -hill climbing(β HHC) to empower its exploitation capabilities. Consequently, we named our hybridization method Hybrid Bat-inspired algorithm(HBA). Therefore, HBA algorithm was employed as a gene subset selection method to figure out the near optimal gene subset from the reduced set generated by filter approach stage. Initially, since the gene selection problem is a binary in nature, Nakamura et al. [36] introduced a binary version of the BA as illustrated in Section 2.5. The pseudo-code of the proposed HBA algorithm is presented in Algorithm 4.

After the bat population is initialized and each candidate solution is evaluated, BA iterates toward the optimal solution by means of generating new solutions according to equations formulated by the BA operators, as shown in (5), (6) and (7) as discussed in Sections 2.4 and 2.5. With a probability range of pulse rate r_i (Line 7 in Algorithm 4), instead of using the local search strategy in Basic BA algorithm, we propose to use β HHC to utilize an efficient exploitation capabilities around the current best solution (Lines 10 to 21 in Algorithm 4). In other words, β HHC iteratively generates a new solution around the current best solution based on two operators: Neighborhood navigation (i.e. \mathcal{N} -operator) and β operator. Note that in both operators, the flip-flop process is used in \mathcal{N} -operator and a random digit either 0 or 1 is generated to replace the original one in β operator with a probability of β parameter where $\beta \in [0, 1]$.

During the search process, the function $improve(N(X))$ (i.e., Line 11 in Algorithm 4) explores the neighborhood solutions X' of the current solution X using flip-flop process. In flip-flop process, a random variable is selected from X and its value is flipped to be 1 if it 0 and vice-versa. In β -operator (Line 13 in Algorithm 4), new solution (i.e., X'') can be generated from the current solution (X') as follow: With a probability range of β parameter, any variable in the new solution (i.e., X'') can be assigned by a value of 0 or 1. This is shown in Algorithm 4, line 13. In the algorithm, the function rnd a random distribution number of the range $rnd \in [0, 1]$. The new solution X'' is accepted which has an equal or higher fitness value, i.e., $f(X'') \geq f(X)$. The search process is repeated until the maximum number of generations are achieved.

Algorithm 4, Line 23 shows the loudness process. In case the loudness parameter a value is higher than a random number generated by rnd , where $rnd \in [0, 1]$. The second condition will be checked. The second condition is satisfied when the fitness function of new solution is better the best solution (i.e., X^*), then the new solution (i.e., X'') will be included in the bat population and the current solution (i.e., X) will be excluded from bat population. In addition, the loudness rate is decreased and the rate of pulse emission is increased. In the meanwhile the best bat location will be iteratively updated as pseudocoded in Algorithm 4, Line 27. Finally, in case the termination condition is satisfied (Lines 10 to 28 in Algorithm 4), therefore, best predictive genes subset which represents highest fitness function is returned.

Algorithm 4 Hybrid bat-inspired algorithm

```

1: Initialize the bat population  $X_i (i = 1, 2, \dots, n)$  and  $V_i$ .
2: Define pulse frequency  $F_i$ .
3: Initialize pulse rate  $r_i$  and the loudness  $A_i$ .
4: while  $t < Maxnumberofiterations$  do
5:   Generate new solutions by adjusting frequency.
6:   Updating velocities and positions [(5) to (7)].
7:   if  $rnd > r_i$  then
8:     Select a solution  $X$  among the best solutions randomly.
9:     Run  $\beta$ -hill climbing algorithm
10:    while  $t < Maxnumberofiterations$  do
11:       $X' = improve(N(X))$ 
12:      for  $i \in \{1, \dots, N\}$  do
13:        if  $rnd \leq \beta$  then
14:           $X''_i \in \{0, 1\}$ 
15:        end if
16:      end for
17:      if  $f(X'') \geq f(X)$  then
18:         $X = X''$ 
19:      end if
20:       $itr = itr + 1$ 
21:    end while
22:  end if
23:  if  $rnd < A_i$  and  $f(X_i) < f(x^*)$  then
24:    Accept the new solutions
25:    Increase  $r_i$  and reduce  $A_i$ 
26:  end if
27:  Rank the bats and find the current Gbest
28: end while

```

4 Experimental setup & results

In this section, the performance of the proposed method was evaluated using benchmark microarray datasets, which is established by Kent ridge biomedical data set repository.¹ These dataset are detailed in Section 4.1. The implementation of filter and wrapper approaches was programmed using two languages (i.e., Java and Matlab). In filter approach stage, ReliefF, Chi-Square and Kullback-Liebler are implemented in java using weka tool [52], while rMRMR is implemented using Matlab. In wrapper approach stage, BA, β -hill climbing and SVM are implemented using java. The SVM used in this method was based on the one prepared in libsvm [53]. RBF kernel is adopted for SVM classifier. As conventional, grid search algorithm was used for tuning parameters of SVM classifier [54]. The experiments are performed on an Intel Core Quad 2.66 GHz CPU with 4 GB of RAM in Microsoft Windows 7 platform.

As remembering, the proposed method has two stages: filter and wrapper. In stage I (filter approach), the rMRMR is proposed to improve the performance of the classical MRMR by means of combining the capability of the three other filtering approaches (ReliefF, Chi-Square and Kullback-Liebler) with the framework of MRMR. The effect of rMRMR on the classification accuracy is deeply studied in Section 4.2 in which the proposed rMRMR is compared against MRMR and other well-regarded filtering-based approaches. In stage II (wrapper approach): the β HC is hybridized with BA to improve its exploitation aspects. Therefore, the effect of the β HC on the convergence behavior of HBA is also studied in Section 4.3 where the BA with and without β HC are compared. Moreover, rMRMR with Genetic Algorithm [55] (rMRMR-GA) and rMRMR and Particle Swarm Optimization [56] (rMRMR-PSO) are re-implemented to gene selection problem, and their performance is compared to the performance of the proposed method (rMRMR-HBA). Finally, the comparative evaluation between the proposed rMRMR-HBA and other methods using the same datasets are fully discussed in Section 4.4.

For proper usage of stage I (filtering approach), SVM builds a classification model using top ranked 50 genes generated by rMRMR, MRMR, ReliefF, Chi-Square, Kullback-Liebler and ensemble method (i.e. ReliefF, Chi-Square, and Kullback-Liebler), Note that top ranked 50 genes for SVM is configured in accordance with the previous studies [14, 17, 55, 57]. Subsequently, the highly ranked genes resulting from filtering stage are passed into wrapper stage. It is worth recalling that in wrapper approach, SVM is used to validate and assess each candidate gene subset generated from HBA algorithms.

¹<http://sdmc.lit.org.sg/GEDatasets/Datasets>

In tables below (Tables 5, 6) is evaluated based on three measurements: the classification accuracy, the predictive gene subset length and fitness function values. while comparative table (Table 6), only the classification accuracy and the predictive gene subset length are used as done previously in the literature [2, 17]. Classification accuracy is calculated by (23), where gene subset length is number of selected genes.

$$\text{Classification Accuracy} = \frac{\text{\#of correctly predicted instances}}{\text{\# of instances}} \quad (23)$$

Table 1 shows the parameter setting values of the proposed method used in the experimental step. These parameters values are carefully selected based on some preliminary experiments and based on the previous parameter setting theory concluded by other studies using BA and β HC [36, 40]. Note that number of iteration is 100 for both algorithms.

4.1 Dataset used

To evaluate the proposed methods, we carried out our experiments using ten *de facto* datasets of gene expression profile [58, 59]. The datasets and their characteristics are summarized in Table 2. In table, the row represents the key of the dataset instances. For each row, four values are presented including #Gene, which refers to the number of genes in the dataset. The second column includes the #Samples, which refers to the number of patient samples selected for testing; In the third column, the #Class refers to the corresponding disease statuses; The last column shows the description of each cancer disease.

4.2 Effect rMRMR in the proposed method

In order to study the effect of the proposed rMRMR on the performance of classification process, the classification

Table 1 Parameter setting of the proposed method

Algorithm	Parameter	Selected value
BA	Number of artificial bats	100
	F_{min}	0.3
	F_{max}	1
	A	0.5
	r	0.5
	α	0.9
	γ	0.9
β -Hill climbing	Number of iterations	50
	β	0.005

Table 2 Datasets characteristic

Datasets	# Genes	# Samples	# Classes	# Description
Breast	24481	97	2	97 samples from breast cancer patients (46 patients developed distance metastases; the other 51 remained healthy after their initial diagnosis for an interval of at least five years).
MLL	12582	72	3	AML<comma> ALL<comma> and mixed-lineage leukemia (MLL).
Colon	2000	62	2	40 colon cancer biopsies vs.22 normal biopsies.
ALL-AML	7129	72	2	Two acute leukemia <comma> i.e.<comma> Acute Myelogenous Leukemia (AML) and Acute Lymphoblastic Leukemia (ALL).
ALL-AML-3C	7129	72	3	AML<comma> ALL B-cell<comma> and ALL T-Cell.
ALL-AML-4C	7129	72	4	AML-Bone Marrow<comma> AML-Peripheral Blood<comma> ALL B-cell<comma> and T-Cell.
Lymphoma	4026	62	3	Three most prevalent adult lymphoid tumors.
CNS	7129	60	2	Outcome of the treatments for 60 central nervous system cancer patients (21 survivors and 39 failures).
Ovarian	15154	253	2	Proteomic spectra of 91 normal persons and 162 ovarian cancer patients.
SRBCT	2308	83	4	Small<comma> round blue cell tumors (SRBCT) from childhood.

behavior of rMRMR is compared with each basic MRMR, ReliefF, Chi-Square, Kullback-Liebler worked individually and an ensemble of filters (i.e. ReliefF, Chi-Square, Kullback-Liebler). For comparative purpose, we performed a 10-fold cross validation to estimate classification accuracy for all comparative methods.

In Table 3, the results of all comparative methods is summarized in terms average classification accuracy along with standard deviation values which represented as (average \pm standard deviation). The best results are highlighted in **bold** font. It can be noted that the rMRMR is able to obtain the best recorded results for four out of the 10 datasets.

Table 3 The classification accuracy of the proposed rMRMR as compared to the other gene selection methods by 10-fold cross validation on 10 microarray datasets

Dataset	rMRMR	MRMR	Reliff	Chi-square	Kullback-Liebler	Ensemble
Breast	83.505 \pm 11.78	78.351 \pm 14.47	79.381 \pm 11.20	75.258 \pm 12.05	65 \pm 9.45	76.29 \pm 12.26
MLL	97.222 \pm 5.37	100	95.833 \pm 6.02	95.83 \pm 6.29	95.833 \pm 6.29	95.833 \pm 6.02
Colon	79.032 \pm 19.56	80.65 \pm 19.54	79.032 \pm 18.16	80.645 \pm 12.56	83.871 \pm 14.45	85.48 \pm 15.23
ALL_AML	98.611 \pm 5.01	98.611 \pm 5.01	97.222 \pm 5.91	97.222 \pm 5.91	94.4 \pm 6.94	98.611 \pm 5.01
ALL_AML_3c	98.611 \pm 4.29	97.222 \pm 7.01	95.833 \pm 7.38	97.222 \pm 5.37	94.444 \pm 9.38	97.222 \pm 5.37
ALL_AML_4c	95.833 \pm 5.56	95.833 \pm 8.01	94.444 \pm 6.60	94.444 \pm 8.75	93.1 \pm 8.78	94.444 \pm 8.21
Lymphoma	100	100	100	100	100	100
CNS	88.333 \pm 9.77	80 \pm 16.01	75 \pm 10.72	75 \pm 15.18	60 \pm 16.82	83.33 \pm 12.42
Ovarian	100	100	98.814 \pm 1.83	100	99.605 \pm 1.20	100
SRBCT	100	100	100	98.795 \pm 3.33	96.386 \pm 6.30	100
AVG	94.114	93.066	91.55	91.13	88.26	93.12
WTL		2/4/4	1/2/7	1/2/7	1/1/8	1/4/5

Furthermore, it has similar best score results of four datasets (*i.e.*, *ALL_AML*, *Lymphoma*, *Ovarian*, and *SRBCT*) as obtained by others. Furthermore, rMRMR obtains the second and fourth best results on MLL and Colon datasets, respectively.

In terms of the average of all result averages (AVG) shown in the last row of Table 3, the proposed method clearly achieved the best AVG results for almost all tested datasets. Therefore, the rMRMR reveals an efficient outcomes in addressing the problem that consists large number of genes.

In addition, rMRMR and other methods are compared using the win/tie/loss (WIT) aspect shown in the last row of Table 3. The WTL row denotes the number of times that the method is win/tie/loss than rMRMR. As an example, the entry “2/4/4” in Table 3 indicates that MRMR losses four cases, tie four cases, and wins on only two in comparison with our method rMRMR. Notably, the proposed method is able to outperforms other filter-based gene selection methods in almost all tested datasets. In a nutshell, the proposed filter-based method archives promising improvement in the classifier performance. This outstanding results thanks to ensemble filter component that empower the robustness and stability of MRMR.

4.3 Effect of β HC on the performance of HBA

In this section, the effect of β HC on the convergence behavior of the proposed method is studied. Note that the top ranked 50 genes obtained by Stage I (filtering approach) are used in HBA in order to cut-of the scale of the search space by the most informative genes. The BA with β HC (*i.e.*, HBA), BA without β HC (*i.e.*, BA), GA, and PSO are compared together. Both methods were executed 30 independent runs. Recall, Table 1 shows the parameters setting of HBA. The experimental results obtained are summarized in Table 4 in terms of $|G|$, ACC , $|F|$, T and $'T - sig'$ which represent the average number of selected genes, the average classification accuracy, the average fitness value introduced in (22), average of execution time (minutes), and Wilcoxon signed-rank statistical test to shows whether there exits any statistically significant difference between HBA and other algorithms, respectively. In addition, the best results of $|G|$, ACC , T and $|F|$ are remarked in **bold** font.

In Table 4, $'T - sig'$ row, with the probability range of $\alpha \leq 0.05$, ‘*’ denotes that the results of the HBA are significantly better than the corresponding algorithm while ‘-’ denotes that the results of the HBA algorithm are not significantly better than the corresponding algorithm.

Table 4 Comparison of the gene selection methods on ten selected datasets

Dataset	Criteria	GA	PSO	BA	HBA
Breast	ACC	88.17	92.16	90.34	95.40
	$ G $	22.72	20.86	15.63	12.63
	$ F $	81.45	85.39	86.02	91.26
	T	0:09	5:00	11:66	125:00
	T-Sig.	*	*	*	
MLL	ACC	100	100	100	100
	$ G $	9.80	11.20	10.67	8.00
	$ F $	96.08	95.52	95.73	96.80
	T	0:04	1:39	5:00	41:25
	T-Sig.	*	*	*	
Colon	ACC	89.94	93.76	93.33	97.85
	$ G $	10.30	19.60	10.16	12.27
	$ F $	87.84	87.17	90.60	93.37
	T	0:34	1:40	15:00	125:00
	T-Sig.	*	*	*	
ALL-AML	ACC	100	100	100	100
	$ G $	9.86	9.96	5.23	4.07
	$ F $	96.05	96.01	97.91	98.37
	T	0:03	1:30	1:66	16:60
	T-Sig.	*	*	*	
ALL-AML3c	ACC	99.53	100	100	100
	$ G $	10.66	13.13	7.17	5.33
	$ F $	95.36	94.75	97.13	97.87
	T	0:04	1:84	1:66	15:00
	T-Sig.	*	*	*	
ALL-AML4c	ACC	98.70	99.95	99.68	100
	$ G $	17.50	15.76	10.20	6.73
	$ F $	91.96	93.66	95.66	97.31
	T	0:04	0:83	1:66	30:00
	T-Sig.	*	*	*	
Lumphoma	ACC	100	100	100	100
	$ G $	11.03	10.80	13.97	8.13
	$ F $	95.59	95.68	94.41	96.75
	T	0:04	0:81	1:66	21:66
	T-Sig.	*	*	*	
CNS	ACC	94.27	98.77	96.94	100
	$ G $	25.60	20.20	19.23	11.20
	$ F $	85.18	90.94	89.86	95.52
	T	0:04	1:16	8:33	50:00
	T-Sig.	*	*	*	
Ovarian	ACC	100	100	100	100
	$ G $	5.30	7.66	3.90	3.07
	$ F $	97.88	96.93	98.44	98.77
	T	0:16	0:16	10:00	100:00
	T-Sig.	*	*	-	

Table 4 (continued)

Dataset	Criteria	GA	PSO	BA	HBA
SRBCT	ACC	100	100	100	100
	#G	14.83	13.30	11.06	9.13
	F	94.07	94.68	95.57	96.35
	T	0:08	5:00	3:33	50:00
	T-Sig.	*	*	*	

Wilcoxon signed-rank statistical test is performed on fitness value since its composed of classification accuracy and number of selected genes. Apparently in Table 4, HBA is able to obtain higher classification accuracy than BA and other algorithms on four datasets (i.e., Breast, Colon, ALL_AML_4c, and CNS) and similar classification accuracy to BA and other algorithms on six datasets (i.e., MLL, ALL_AML, ALL_AML_3c, Lymphoma, ovarian, and SRBCT). In terms of classification accuracy and number of selected genes, HBA is able to achieve a better trade-off between classification accuracy and number of selected genes than BA and other algorithms for all datasets except for ‘Colon’ dataset. In Colon dataset, HBA achieved higher classification accuracy but with slightly increasing number of selected genes than BA.

In term of fitness function, HBA exhibits higher fitness function value ($|F|$) than other algorithms on all datasets. Although the execution time (T) is not a major factor in gene selection domain since it is not a real time, the average execution time of 30 runs replicated for each version of the proposed method in addition to those recorded by GA and PSO are summarized in Table 4. It has to be noted that T recorded by HBA and BA are reasonably acceptable. GA has less T than PSO, BA, and HBA. Notably, HBA records the highest T value since this version is a hybrid version and therefore takes higher execution. Further, between the results of HBA and other algorithms, there exist a significant difference on four out of ten datasets. For instance, HBA gets significantly better results than BA and other algorithms on nine datasets and its not significantly better than BA on only one dataset (i.e., Ovarian). In sum, the experimental results exhibit that HBA is able to yield smaller set of reliable genes with higher/equivalent classification accuracy than BA and other algorithms on almost all datasets. This fruitful results is owing to the integration of β -hill climbing algorithm with BA that significantly improve the local search capability of BA, which enhance the exploitation aspect. Thereby, HBA is capable of eliminating the irrelevant and redundant genes that empower it to identify smaller set of reliable genes and impose it to obtain competitive or higher classification accuracy.

The convergence behavior trend of HBA against BA is shown in Fig. 3 for all experimented datasets. Notably for all figures shown, convergence rate of HBA is superior in comparison with that produced by BA.

4.4 Comparative evaluation

To further assess the effectiveness of the proposed method, rMRMR-HBA was compared with the state-of-the-art methods, as shown in Table 5. Table 6 shows the average criteria over multiple independent runs for each method, including the average of the classification accuracy (ACC) and the number of selected genes (#G). The best results are highlighted in **bold**.

As illustrated in Table 6, in terms of the classification accuracy, rMRMR-HBA obtained higher classification accuracy than other comparative methods for three out of ten experimented datasets (i.e., Breast, ALL_AML_4c, and CNS). Furthermore, the rMRMR-HBA is able to yield the best recorded classification accuracy as done by other comparative methods in six out of ten datasets (i.e., MLL, ALL_AML, ALL_AML_3c, Lymphoma, Ovarian, and SRBCT). The only ‘Colon’ dataset is solved by rMRMR-HBA with 4th rank which has a shallow search space. In general, the classification accuracy produced by the rMRMR-HBA is very promising for all dataset experimented in comparison with other comparative methods.

The number of genes is also recorded in the results due to the fact that there are a correlation between the classification accuracy and the number of genes. As an overall view, the proposed method obtained best ‘number of genes’ and ‘classification accuracy’ results for two out of ten datasets (i.e., ALL_AML_3c and Ovarian) and competitive results on the remainder once.

To elaborate, in Breast dataset, the MA-C method obtained (95.26%) classification accuracy with 183 genes, while the rMRMR-HBA approach achieves (95.4%) classification accuracy with only 12.63 genes. For MLL dataset, the rMRMR-HBA method selected 8 genes to achieve (100%) classification accuracy. The HSA-MB method selected fewer genes with only 6.6 to achieve (99.55%) classification accuracy on the same dataset. For Colon dataset, GPSO obtained the lowest of selected genes with only 2 and highest classification accuracy (100%). By contrast, the rMRMR-HBA method selected 12.27 genes and achieves (97.85%) classification accuracy. For ALI-AML dataset, the LDA-GA method select fewer genes with only 3 genes and achieves (99.5%) classification accuracy. In contrast, the rMRMR-HBA method selected 4.07 genes and achieves (100%) classification accuracy.

For ALL_AML_3c dataset, rMRMR-HBA method obtained the lowest of selected genes with only 5.33 genes and highest classification accuracy (100%). For ALL_AML_4c dataset, the rMRMR-HBA method selected 6.73 genes and achieves (100%) classification accuracy. In comparison, HSA-MB method selected slightly fewer genes with 6.37 genes; however, its classification accuracy is smaller. For Lymphoma dataset, the HSA-MB approach

selected fewer genes. The HSA-MB method selected 3.75 genes and achieves (99.99%) classification accuracy. In contrast, rMRMR-HBA method selected 8.13 genes and archives (100%) classification accuracy.

For CNS dataset, the rMRMR-HBA method selected quiet higher genes with 11.2 genes but achieved the highest classification accuracy (100%). By contrast, LDA-GA method obtained the lowest of selected genes with only

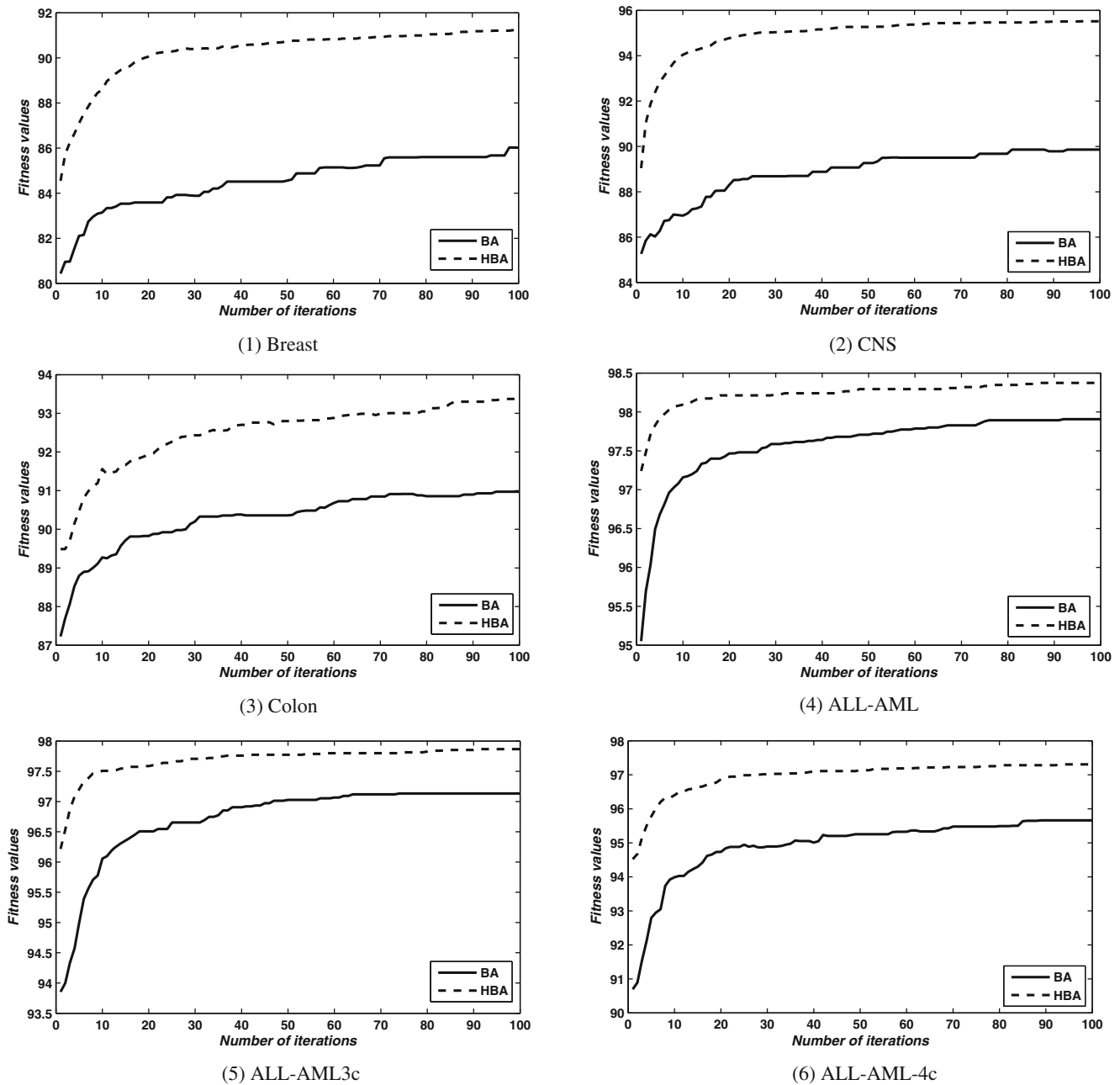


Fig. 3 The convergence behavior of BA and HBA for 10 datasets

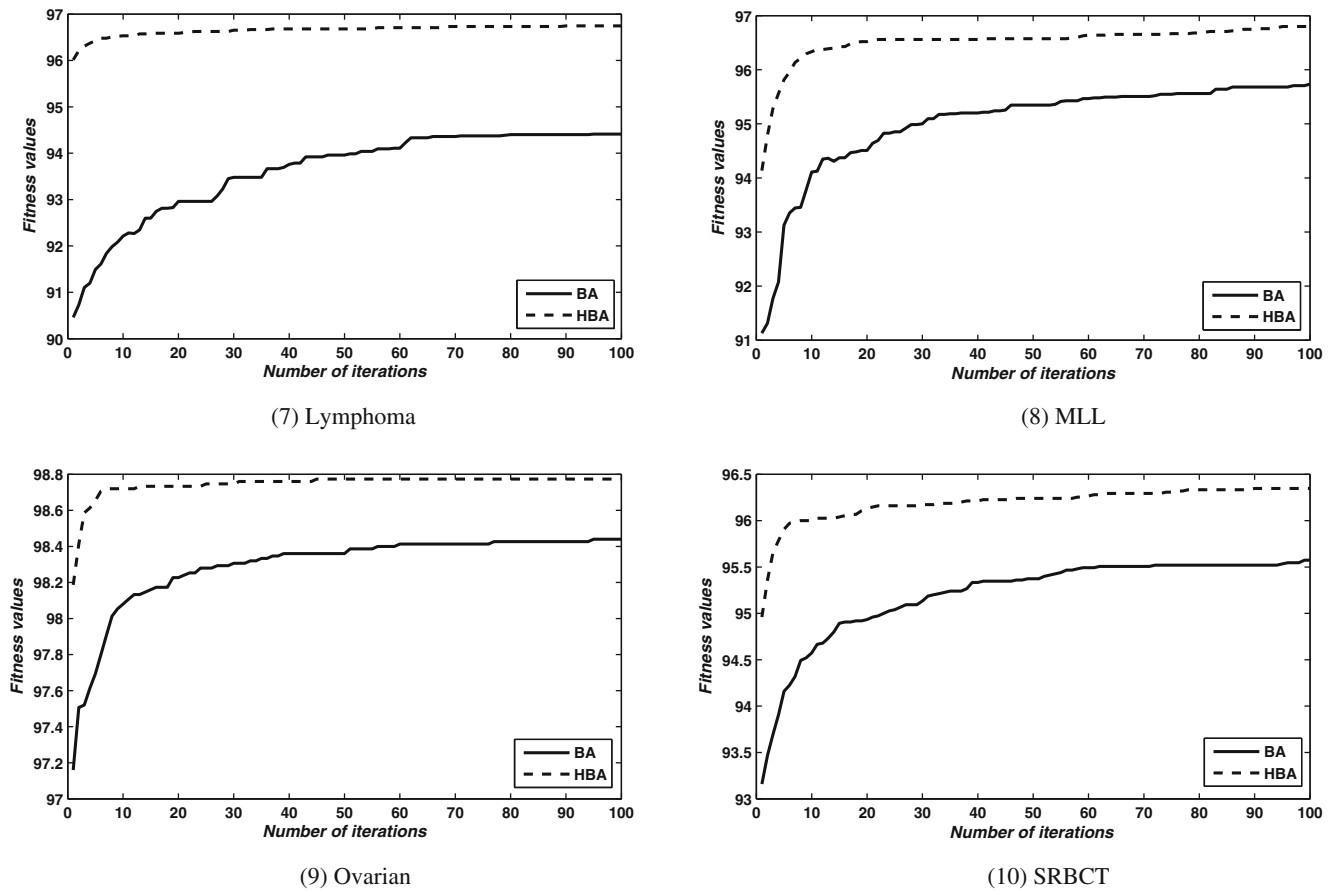


Fig. 3 (continued)

4 and achieves (99.3%) classification accuracy. Although there are many existing methods that achieve (100%) classification accuracy for Ovarian dataset, the rMRMR-HBA approach selected the smallest number of predictive genes. The rMRMR-HBA approach selected only 3.07 genes to

achieve (100%) classification accuracy. For SRBCT dataset, the HSA-MB method select fewer genes with only 3 genes and achieves (99.57%) classification accuracy. In contrast, the rMRMR-HBA method selected slightly higher 9.13 genes and achieves (100%) classification accuracy.

Table 5 Key to comparative methods

Key	Method name	Reference
IG-SGA	Information Gain and Standard Genetic Algorithm	[60]
CFC-iBPSO	Correlation-based Feature Selection with improved-Binary Particle Swarm Optimization	[56]
HSA-MB	Hybridizing harmony search with a Markov blanket for gene selection problems	[17]
MBEGA	Markov Blanket-Embedded Genetic Algorithm for gene selection	[55]
MRMR-GA	Minimum Redundancy Maximum Relevancy with a Genetic Algorithm	[2]
MA-C	Correlation-based memetic framework	[61]
BPSO-CGA	Binary Particle Swarm Optimization and a Combat Genetic Algorithm	[18]
GPSO	Geometric Particle Swarm Optimization	[62]
BIRSW	Best Incremental Ranked Subset	[63]
LDA-GA	Fisher Linear Discriminate Analysis-based Genetic Algorithm	[64]

Table 6 Results of comparison between rMRMR-HBA and the state-of-art methods

Datasets	M	RMRMR-HBA	HSA-MB [17]	MBEGA [55]	CFC-iBPSO [56]	IG-SGA [60]	MRMR-GA [2]	MA-C [61]	BPSO-CGA [18]	GPSO [62]	BIRSW [63]	LDA-GA [64]
Breast	#G	12.63	5.06	14.5	32.7	-	-	183	-	4	-	-
	ACC	95.4	80.06	80.74	92.75	-	-	95.26	-	86.35	-	-
MLL	#G	8	6.6	32.1	30.08	-	-	108	-	-	-	-
	ACC	100	99.55	94.33	100	-	-	100	-	-	-	-
Colon	#G	12.27	4.16	24.5	4.2	60	15	-	214	2	3.5	7
	ACC	97.85	90.27	85.66	94.89	85.48	98.39	-	96.7	100	85.48	99.83
ALL_AML	#G	4.07	5	12.8	4.3	3	15	387	300	3	2.5	3
	ACC	100	99.34	95.89	100	97.06	100	99.56	100	97.38	93.04	99.5
ALL_AML_3c	#G	5.33	5.84	18.1	6	-	-	394	-	-	-	-
	ACC	100	99.18	96.64	100	-	-	99.53	-	-	-	-
ALL_AML_4c	#G	6.73	6.37	26.2	20.7	-	-	386	-	-	-	-
	ACC	100	96.79	91.93	97.63	-	-	98.61	-	-	-	-
Lymphoma	#G	8.13	3.75	34.3	24	-	15	-	196	-	10.3	-
	ACC	100	99.99	97.68	100	-	98.96	-	100	-	82.14	-
CNS	#G	11.2	7.43	20.5	10.5	38	-	374	-	-	-	4
	ACC	100	84.17	72.21	95.84	86.76	-	97.78	-	-	-	99.3
Ovarian	#G	3.07	5.73	9	3.3	-	-	247	-	4	-	6
	ACC	100	99.81	99.71	100	-	-	100	-	99.4	-	97.4
SRBCT	#G	9.13	8.9	60.7	34.1	-	-	526	880	-	-	-
	ACC	100	99.57	99.23	100	-	-	100	100	-	-	-

5 Conclusion and future work

In this paper, an efficient method was proposed to solve gene selection problem, which combine filter and wrapper methods, namely, rMRMR, HBA, and SVM classifier. In filter stage, rMRMR method is proposed by combining an ensemble of filter methods (i.e., ReliefF, Chi-Square, Kullback-Liebler) with MRMR to improve its robustness and stability. This combination mechanism is done as follows: firstly, the ensemble methods (i.e., ReliefF, Chi-Square, Kullback-Liebler) generate single gene ranking list by aggregate the results from each individual filter, subsequently, this ranked list is used by MRMR. Secondly, the ensemble scores of genes in coupled with the calculation of relevancy score in MRMR. In wrapper stage, we propose hybrid BA algorithm with β -hill climbing algorithm integrated with SVM to seek for optimal gene subset from top informative genes obtained by rMRMR method. This hybridization process of HBA is proposed to strike a right balance between exploration and exploitation capabilities and improve its local exploitation aspects.

Extensive experiments were conducted into ten microarray benchmark datasets to test the performance of the proposed method. These data sets are high dimensional with high complex. The evaluation process involves three stages, (i) evaluating the filtering-based approach, (ii) evaluating the wrapper-based approach, and (iii) evaluating the proposed rMRMR-HBA method. In the first evaluation stage, the proposed rMRMR is evaluated against original MRMR and other classical filtering-based approach. The results show that the proposed rMRMR outperforms those produced by other comparative filtering-based approach for seven out ten datasets. In the second evaluation stage, the wrapper-based approach (i.e., HBA) is evaluated as follows: bat-inspired algorithm with and without β -hill climbing algorithm, as well as, the proposed method (HBA) is evaluated against other well-known methods (i.e., GA and PSO). The evaluation is conducted based on three main criteria: number of selected genes, classification accuracy, execution time, and fitness function value. The proposed HBA is able to exhibit better compromise among number of selected genes and classification accuracy than BA for all datasets except for 'Colon' dataset. In Colon dataset, HBA achieved higher classification accuracy but with slightly increasing number of selected genes than BA. In term of fitness function value, HBA obtained higher fitness function value ($|F|$) than BA on all datasets.

In the third evaluation stage, a comparison with state-of-art methods shows that our proposed method obtained an equivalent or higher classification accuracy in comparison with those yielded by other competitors in nine out of ten datasets. Furthermore, the best overall results in two out of ten datasets (i.e., ALL_AML_3c and Ovarian) and

competitive results on the remainder datasets, in terms of number of genes and classification accuracy, are obtained. This finding proves that rMRMR-HBA approach is capable to eliminate the irrelevant and redundant genes effectively due to the integration of rMRMR filter and HBA wrapper, that result in the identification of small set of informative genes. Therefore, these high impressive achievements can be inspired other researchers to contribute in the rMRMR-HBA which is a superior method pregnant with many discoveries and capabilities for gene selection domain.

As the proposed rMRMR-HBA revealed a very successful outcomes for gene selection, improving the performance of each individual filter-based gene selection method (i.e. MRMR) by means of homogeneous/heterogeneous ensembles of filtering-based gene selection methods are required in the future. In addition, rMRMR-HBA can be expanded to other high-dimensional datasets, including image and text mining data.

References

1. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA et al (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286(5439):531–537
2. El Akadi A, Amine A, El Ouardighi A, Aboutajdine D (2011) A two-stage gene selection scheme utilizing mrmr filter and ga wrapper. *Knowl Inf Syst* 26(3):487–500
3. Lai C-M, Yeh W-C, Chang C-Y Gene selection using information gain and improved simplified swarm optimization, *Neurocomputing*
4. Jain A, Zongker D (1997) Feature selection: evaluation, application, and small sample performance. *IEEE Trans Pattern Anal Mach Intell* 19(2):153–158
5. Kohavi R, John GH (1997) Wrappers for feature subset selection. *Artif Intell* 97(1):273–324
6. Ding C, Peng H (2005) Minimum redundancy feature selection from microarray gene expression data. *J Bioinforma Comput Biol* 3(02):185–205
7. Kononenko I (1994) Estimating attributes: analysis and extensions of relief. In: *European conference on machine learning*. Springer, pp 171–182
8. Su C-T, Hsu J-H (2005) An extended chi2 algorithm for discretization of real value attributes. *IEEE Trans Knowl Data Eng* 17(3):437–441
9. Kullback S, Leibler RA (1951) On information and sufficiency. *Ann Math Stat* 22(1):79–86
10. Li B-Q, Hu L-L, Chen L, Feng K-Y, Cai Y-D, Chou K-C (2012) Prediction of protein domain with mrmr feature selection and analysis. *PLoS One* 7(6):e39308
11. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *J Mach Learn Res* 3:1157–1182
12. Bolón-Canedo V, Sánchez-Marcoño N, Alonso-Betanzos A (2012) An ensemble of filters and classifiers for microarray data classification. *Pattern Recogn* 45(1):531–539
13. Seijo-Pardo B, Porto-díaz I, Bolón-canedo V, Alonso-betanzos A (2017) Ensemble feature selection: Homogeneous and heterogeneous approaches. *Knowl-Based Syst* 118:124–139

14. Ebrahimpour MK, Eftekhari M (2017) Ensemble of feature selection methods: a hesitant fuzzy sets approach. *Appl Soft Comput* 50:300–312
15. Narendra PM, Fukunaga K (1977) A branch and bound algorithm for feature subset selection. *IEEE Trans Comput* 26(9):917–922
16. Alshamlan HM, Badr GH, Alohalo YA (2015) Genetic bee colony (gbc) algorithm: a new gene selection method for microarray cancer classification. *Comput Biol Chem* 56:49–60
17. Shreem SS, Abdullah S, Nazri MZA (2014) Hybridising harmony search with a markov blanket for gene selection problems. *Inf Sci* 258:108–121
18. Chuang L-Y, Yang C-H, Li J-C, Yang C-H (2012) A hybrid bpsoga approach for gene selection and classification of microarray data. *J Comput Biol* 19(1):68–82
19. Li S, Wu X, Tan M (2008) Gene selection using hybrid particle swarm optimization and genetic algorithm. *Soft Comput-A Fusion Found Methodol Appl* 12(11):1039–1048
20. Yang X-S (2010) A new metaheuristic bat-inspired algorithm. In: *Nature inspired cooperative strategies for optimization (NICSO 2010)*. Springer, pp 65–74
21. Ramesh B, Mohan VCJ, Reddy VV (2013) Application of bat algorithm for combined economic load and emission dispatch. *Int J Electric Eng Telecommun* 2(1):1–9
22. Musikapun P, Pongcharoen P (2012) Solving multi-stage multi-machine multi-product scheduling problem using bat algorithm. In: *2nd international conference on management and artificial intelligence*, vol 35. IACSIT Press Singapore, pp 98–102
23. Yang X-S, Hossein Gandomi A (2012) Bat algorithm: a novel approach for global engineering optimization. *Eng Comput* 29(5):464–483
24. Lin J-H, Chou C-W, Yang C-H, Tsai H-L et al (2012) A chaotic levy flight bat algorithm for parameter estimation in nonlinear dynamic biological systems. Source: *J Comput Inf Technol* 2(2):56–63
25. Mishra S, Shaw K, Mishra D (2012) A new meta-heuristic bat inspired classification approach for microarray data. *Procedia Technol* 4:802–806
26. Alomari OA, Khader AT, Al-Betar MA, Abualigah LM (2017) Gene selection for cancer classification by combining minimum redundancy maximum relevancy and bat-inspired algorithm. *Int J Data Mining Bioinforma* 19(1):32–51
27. Komarasamy G, Wahi A (2012) An optimized k-means clustering technique using bat algorithm. *Eur J Sci Res* 84(2):26–273
28. Al-Betar MA, Awadallah MA, Faris H, Yang X-S, Khader AT, Alomari OA (2018) Bat-inspired algorithms with natural selection mechanisms for global optimization. *Neurocomputing* 273:448–465
29. Akhtar S, Ahmad A, Abdel-Rahman E (2012) A metaheuristic bat-inspired algorithm for full body human pose estimation. In: *2012 Ninth Conference on Computer and robot vision (CRV)*. IEEE, pp 369–375
30. Tosun Ö, Marichelvam M (2016) Hybrid bat algorithm for flow shop scheduling problems. *Int J Math Oper Res* 9(1):125–138
31. Roeva ON, Fidanova SS (2013) Hybrid bat algorithm for parameter identification of an e. coli cultivation process model. *Biotechnol Biotechnol Equip* 27(6):4323–4326
32. Afrabandpey H, Ghaffari M, Mirzaei A, Safayani M (2014) A novel bat algorithm based on chaos for optimization tasks. In: *2014 Iranian Conference on Intelligent systems (ICIS)*. IEEE, pp 1–6
33. Wang G, Guo L, Duan H, Liu L, Wang H (2012) A bat algorithm with mutation for ucav path planning. *The Scientific World Journal*
34. Kira K, Rendell LA The feature selection problem: Traditional methods and a new algorithm. In: *AAAI*, Vol 2, 1992, pp 129–134
35. Jović A, Brkić K, Bogunović N A review of feature selection methods with applications. In: *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, IEEE, 2015, pp 1200–1205
36. Nakamura RY, Pereira LA, Costa K, Rodrigues D, Papa JP, Yang X-S (2012) Bba: a binary bat algorithm for feature selection. In: *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, pp 291–297
37. Davis L Bit-climbing, representational bias, and test suite design. In: *ICGA*, 1991, pp 18–23
38. Ingber L, Rosen B (1992) Genetic algorithms and very fast simulated reannealing: A comparison. *Math Comput Modell* 16(11):87–100
39. Dueck G (1993) New optimization heuristics: The great deluge algorithm and the record-to-record travel. *J Comput Phys* 104(1):86–92
40. Al-betar MA (2017) β -hill climbing: an exploratory local search. *Neural Comput Appl* 28(1):153–168
41. Alyasseri ZAA, Khader AT, Al-Betar MA, Awadallah MA (2018) Hybridizing β -hill climbing with wavelet transform for denoising ecg signals. *Inf Sci* 429:229–246
42. Holland JH (1992) *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, NJ
43. Vapnik VN (1999) *An overview of statistical learning theory*. *IEEE Trans Neural Netw* 10(5):988–999
44. Peng H, Long F, Ding C (2005) Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell* 27(8):1226–1238. <https://doi.org/10.1109/TPAMI.2005.159>
45. Şen B, Peker M, Çavuşoğlu A, Çelebi FV (2014) A comparative study on classification of sleep stage based on eeg signals using feature selection and classification algorithms. *J Med Syst* 38(3):1–21
46. Dietterich TG (2000) Ensemble methods in machine learning. In: *International workshop on multiple classifier systems*. Springer, pp 1–15
47. Duval B, Hao J-K, Hernandez Hernandez JC (2009) A memetic algorithm for gene selection and molecular classification of cancer. In: *Proceedings of the 11th Annual conference on Genetic and evolutionary computation*. ACM, pp 201–208
48. Dash M, Liu H (1997) Feature selection for classification. *Intell Data Anal* 1(3):131–156
49. Amaldi E, Kann V (1998) On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. *Theor Comput Sci* 209(1):237–260
50. Li X, Yin M (2013) Multiobjective binary biogeography based optimization for feature selection using gene expression data. *IEEE Trans NanoBiosci* 12(4):343–353
51. Ahmad Alomari O, Tajudin Khader A, Azmi Al-Betar M, Mohammad Abualigah L Mrrm ba: A hybrid gene selection algorithm for cancer classification. *J Theor Appl Inform Technol* 95(12):2610–2618
52. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The weka data mining software: an update. *ACM SIGKDD Explor Newslett* 11(1):10–18
53. Chang C-C, Lin C-J (2011) Libsvm: a library for support vector machines. *ACM Trans Intell Syst Technol (TIST)* 2(3):27
54. Huang C-L, Wang C-J (2006) A ga-based feature selection and parameters optimization for support vector machines. *Expert Syst Appl* 31(2):231–240
55. Zhu Z, Ong Y-S, Dash M (2007) Markov blanket-embedded genetic algorithm for gene selection. *Pattern Recogn* 40(11):3236–3248
56. Jain I, Jain VK, Jain R Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification. *Applied Soft Computing*
57. Bolón-Canedo V, Sánchez-Marroño N, Alonso-Betanzos A, Benítez JM, Herrera F (2014) A review of microarray datasets and applied feature selection methods. *Inf Sci* 282:111–135

58. Li T, Zhang C, Ogihara M (2004) A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. *Bioinformatics* 20(15):2429–2437
59. Li J, Liu H Kent ridge bio-medical data set repository, Institute for Infocomm Research. <http://sdmc.lit.org.sg/GEDatasets/Datasets.html>
60. Salem H, Attiya G, El-Fishawy N (2017) Classification of human cancer diseases by gene expression profiles. *Appl Soft Comput* 50:124–134
61. Kannan SS, Ramaraj N (2010) A novel hybrid feature selection via symmetrical uncertainty ranking based local memetic search algorithm. *Knowl-Based Syst* 23(6):580–585
62. Talbi E-G, Jourdan L, Garcia-Nieto J, Alba E (2008) Comparison of population based metaheuristics for feature selection: Application to microarray data classification. In: 2008 IEEE/ACS International Conference on Computer Systems and Applications. IEEE, pp 45–52
63. Ruiz R, Riquelme JC, Aguilar-Ruiz JS (2006) Incremental wrapper-based gene selection from microarray data for cancer classification. *Pattern Recogn* 39(12):2383–2392
64. Bonilla-Huerta E, Duval B, Hernández JCH, Hao J-K, Morales-Caporal R (2011) Hybrid filter-wrapper with a specialized random multi-parent crossover operator for gene selection and classification problems. In: International Conference on Intelligent Computing. Springer, pp 453–461



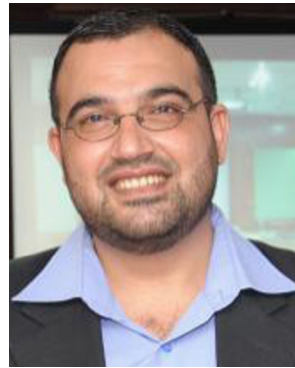
Osama Ahmad Alomari received his M.Sc. in computer science from the Universiti Kebangsaan Malaysia (UKM), Malaysia in 2014. He is currently a PhD Candidate in the school of Computer Sciences at Universiti Sains Malaysia. His research interests include evolutionary algorithms, nature-inspired computation, and their applications to optimization problems.



Ahamad Tajudin Khader received the BSc. and M.Sc. in Mathematics from the University of Ohio, USA, in 1982 and 1983, respectively. He obtained his Ph.D. in Computer Science from University of Strathclyde, U.K., in 1993. He is currently a Professor in the School of Computer Sciences, Universiti Sains Malaysia. His current position is a Deputy Dean of Graduate Studies and Research which responsible for at least 200 researchers and Ph.D students. The author published

hundreds of high quality papers in international Journals, conferences, and book chapters.

He is a Member of the Institute of Electrical and Electronics Engineers (IEEE) as an international professor, Association of Computing Machinery (ACM), Special Interest Group for Genetic and Evolutionary Computation (SIGEVO), and Computational Intelligence Society. His research interest Mainly focuses on Optimisation and Scheduling.



Mohammed Azmi Al-Betar is an associated professor in the Department of Computer Science at Al-Huson University College, Al-Balqa Applied University, and a senior member of computational intelligence research group in the school of computer science at University Sains Malaysia (USM). He received his B.Sc and M.Sc from Computer Science Department at Yarmouk University, Irbid-Jordan in 2001 and 2003, respectively.

He obtained his PhD from the School of Computer Sciences at University Sains Malaysia (USM), Pulau Penang - Malaysia in Oct. 2010. The author published a number of high quality papers in international Journals and conferences. His research interests are mainly directed to metaheuristic optimisation methods and hard combinatorial optimisation problems including Scheduling and Timetabling.



Mohammed A. Awadallah received the graduate degree in computer science from the Islamic University, Gaza, Palestine, in 2002, the M.Sc. degree in Computer Information System (CIS) from the Arab Academy for Banking and Financial sciences, Amman, Jordan, in 2005, and PhD from the School of Computer Sciences at the Universiti Sains Malaysia (USM), Pulau Penang, Malaysia, in 2014.

He is currently an Assistant Professor with the Department of Computer Science, Al-Aqsa University, P. O. Box 4051, Gaza, Palestine. He has authored many publications in international journals, conferences, and book chapters. His research interests on optimization methods (Genetic Algorithm, Artificial Bee Colony, Harmony Search, Bat Algorithm, Hill Climbing, etc.) and combinatorial optimization problems like scheduling, rostering and timetabling problems.