Genomics 112 (2020) 114–126

Contents lists available at ScienceDirect

Genomics

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

Original Article

A TRIZ-inspired bat algorithm for gene selection in cancer classification

Mohammed Azmi Al-Betar[a,*], Osama Ahmad Alomari[b], Saeid M. Abu-Romman[c]

[a] Department of Information Technology, Al-Huson University College, Al-Balqa Applied University, P.O. Box 50, Al-Huson, Irbid, Jordan
[b] Department of Computer Engineering, Faculty of Engineering and Architecture, Istanbul Gelisim University, Istanbul, Turkey
[c] Department of Biotechnology, Faculty of Agricultural Technology, Al-Balqa Applied University, Jordan

ARTICLE INFO

Keywords:
Gene selection
Bat-inspired algorithm
Optimization
TRIZ
MRMR
SVM
Classification

ABSTRACT

Gene expression data are expected to make a great contribution in the producing of efficient cancer diagnosis and prognosis. Gene expression data are coded by large measured genes, and only of a few number of them carry precious information for different classes of samples. Recently, several researchers proposed gene selection methods based on metaheuristic algorithms for analysing and interpreting gene expression data. However, due to large number of selected genes with limited number of patient's samples and complex interaction between genes, many gene selection methods experienced challenges in order to approach the most relevant and reliable genes. Hence, in this paper, a hybrid filter/wrapper, called rMRMR-MBA is proposed for gene selection problem. In this method, robust Minimum Redundancy Maximum Relevancy (rMRMR) as filter to select the most promising genes and an modified bat algorithm (MBA) as search engine in wrapper approach is proposed to identify a small set of informative genes. The performance of the proposed method has been evaluated using ten gene expression datasets. For performance evaluation, MBA is evaluated by studying the convergence behaviour of MBA with and without TRIZ optimisation operators. For comparative evaluation, the results of the proposed rMRMR-MBA were compared against ten state-of-arts methods using the same datasets. The comparative study demonstrates that the proposed method produced better results in terms of classification accuracy and number of selected genes in two out of ten datasets and competitive results on the remaining datasets. In a nutshell, the proposed method is able to produce very promising results with high classification accuracy which can be considered a promising contribution for gene selection domain.

1. Introduction

DNA microarray provides useful information at the molecule level that can be used for detection and classification of cancer diseases [1]. From molecule biology point of view, DNA micrarray is valuable analytical tool that enables the biologists to analyze and monitor thousand of genes in one experiment. It finally generates gene expression data, which are considered as key marker for cancer diseases classification. However, this data is considered as high dimensional dataset since it contains a large number of genes (features) as well as relatively few number of patient samples [2,3]. In classification process, the aforementioned dataset characteristics pose a challenge for machine learning algorithm, and it's very critical to mitigate this issue prior performing classification task. Therefore, to overcome this challenge, gene selection is required. Gene selection is a process in data mining, which can reduce feature dimension by selecting a small set of genes that can achieve similar or better classification performance than using all genes [4]. The main advantages of reducing the number of genes from biological perspective [5] are 1) helping molecular biologists of identify the underlying molecular mechanism, related to gene expression of cancer diseases. 2) and interpret the pattern of the selected genes to discover new therapy targeted these genes. 3) reducing clinical cost.

Traditionally, gene selection methods are broadly divided into three groups: filter, wrapper, and hybrid methods [4]. Filter methods rely on interior characteristics of data in evaluating the quality or the relevancy of the genes to the target class. Widely used filter approaches are the ReliefF [6], Chi-square [7], Kullback-Leibler [8], Minimum Redundancy Maximum Relevancy (MRMR) [9], Robust MRMR (rMRMR) [10]. Wrapper methods are typically divided into two main components: search techniques and evaluation. Search techniques generate candidate gene subsets, and machine learning algorithms can be applied to evaluate their predictive accuracy. Compared to filter methods, wrapper methods often produce better classification results but they are more expensive in terms of computation time. Hybrid method combines the filter and wrapper methods, and complement the benefits of both methods [11]. To be more specific, the integration of both methods

[*] Corresponding author.
E-mail addresses: mohbetar@bau.edu.jo (M.A. Al-Betar), oalomari@gelisim.edu.tr (O.A. Alomari).

https://doi.org/10.1016/j.ygeno.2019.09.015
Received 17 June 2019; Received in revised form 5 September 2019; Accepted 17 September 2019
Available online 30 October 2019
0888-7543/ © 2019 Elsevier Inc. All rights reserved.

leads to obtain informative gene subset with satisfied classification accuracy results. However, the hybrid method is still in its infancy. Therefore, further investigations can be carried out in order to develop more sophisticated hybrid methods.

In wrapper-based approaches, the gene search space grows exponentially with the increase in number of genes. In such scenario, the introducing of exhaustive search algorithm to generate all possible subset of genes is impractical and entails expensive time consuming. Therefore, researchers frequently use metaheuristic approaches to obtain the desired solution while exploring the entire search space. Various metaheuristic approaches have been applied to solve the gene selection problem such as Binary Particle Swarm Optimization and Combat Genetic Algorithm (BPSO-CGA) [12], Harmony search with a Markov blanket (HSA-MB) [13], improvised Interval Value based Particle Swarm Optimization (IVPSO) [14], Genetic Algorithm with Artificial Bee Colony [15], Cellular learning automata-ant colony optimization feature selection (CLACOFS) [16], hybrid binary black hole algorithm and modified binary particle swarm optimization (BPSO (4–2)-BBHA) [17], and Correlation-based Feature Selection with improved-Binary Particle Swarm Optimization [18]. However, most of these approaches can easily reach the stagnation situation in local optima caused by complex interaction between the genes and huge gene search space [13,18]. Therefore, in order to address gene selection problems, a robust search-based approach relied on” efficient search operators” is required to optimize effectively the gene search space and to gain the near-optimal solution.

In specific, swarm-based intelligence methods are normally initiated with a population of random solutions. These solutions are iteratively reconstructed based on the idea of sharing knowledge by collective behavior whereby the fittest solutions to drive the swarm toward optimality. The main merits of swarm intelligence methods: highly scalable, self-organized, adaptable, flexible, Collective Robustness, and individual Simplicity. However, the main limitations of swarm intelligence methods are: time-Critical applications, parameter tuning, and Stagnate with a premature convergence [19]. Bat Algorithm (BA) is swarm-based intelligence method [20], in which its echolocation behaviour is intensively studied and modelled in optimization context. As an advantage, the optimization procedure of BA combines local search operator and global search operator, which are the key success factors of metaheuristic method. Owing to these merits, BA drawn a great attention as a promising method for solving a wide range of optimisation problems [21] such as clustering [22], scheduling [23,24], classifications [25], fault diagnosis [26], image processing [27], gene selection [28,29], and global optimization algorithms [30,31]. Many variants of BA have been produced to improve its performance in solving several optimization problems. In the context of gene selection problem, BA has been initially investigated for gene selection, however it should be more efficiently utilized according to characteristics of gene selection problem.

TRIZ is abbreviation of Teoriya Resheniya Izobretatelskikh Zadatch, also known as theory of inventive problem solving. This theory, introduced by Genrich Altshuller in 1985 [32], is constructed based on intensive and exhaustive analysis of one million patents. The TRIZ problem solving methodology consists of improving and worsening features, and inventive principles, which are considered as guideline for TRIZ research to solve design problems.

This paper proposes a hybrid filter/wrapper gene selection method based on rMRMR approach and modified BA algorithm (MBA). The proposed method is called rMRMR-MBA, in which rMRMR is operated first and then start ranking the genes according to it discriminative power, and thus the gene search space is defined by those highly ranked genes. This is to fine-tune the search space to the wrapper approach.

The main motivation behind this research is to modify the BA in wrapper approach to be more suitable and efficient according to the complexity of gene selection search space problem. Specifically, the main contribution of the proposed method lies on the incorporation of TRIZ inventive solution with basic BA to further optimize its search process and explore the interaction between genes, and thus reach and navigate the most promising search space regions. Extensive experiments were performed on ten popular microarray benchmark datasets to test the rMRMR-MBA method for the gene selection problem. The characteristics of datasets are varied in terms of number of genes, samples and classes. For the performance evaluation, MBA is evaluated by studying the convergence behavior of BA with and without incorporating TRIZ-inspired optimization operators. For the comparative evaluation, the results of the proposed rMRMR-MBA were compared with the results of previous gene selection methods using the same microarray datasets. The comparative results demonstrated that the proposed method is able to produce the best results in two out of ten datasets, and competitive results for the remaining datasets.

The remaining parts of this paper is organized as follows: related background is described in Section 2. The proposed method is illustrated in Section 3. The experimental results and comparative evaluation is analyzed in Section 5. Finally, Section 6 concludes the paper and recommends possible future enhancements.

2. Research background

The fundamentals of BA is provided here to show the original version. Thereafter, the theory of TRIZ innovation solution is described in detail.

2.1. BAT-inspired algorithm (BAT)

Bat-inspired algorithm (BA) is a nature inspired metaheuristic algorithm. It has been introduced by Xin-She Yang in 2010 [20], to imitate the echolocation behavior of bats. Bats share similar biological behavior in terms of navigating and hunting. They mainly rely on echolocation to seek for prey and/or avoid obstacles in the dark. While seeking for prey, bats emit pulses to the surrounding environment and listen for the echoes that bounced back from the surrounding object. By means of these echoes, bats can recognize and locate preys and obstacles. 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 Eq. (1). Determining the upper and lower bound frequencies is based on the domain of interest.

$$F_i = F_{min} + (F_{max} - F_{min}) \times \beta \tag{1}$$

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.

$$V_i(t+1) = V_i(t) + (X_i(t) - Gbest) \times F_i \tag{2}$$

where $Gbest$ is the best solution, F_i represents the frequency of the ith bat and the position X_i of each bat. Each bat's position is updated as shown in Eq. (3).

$$X_i(t+1) = X_i(t) + V_i(t+1) \tag{3}$$

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

$$x_{new} = x_{old} + \varepsilon A_t \tag{4}$$

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 Eqs. (5) and (6).

$$A_i(t+1) = \alpha A_i(t) \tag{5}$$

$$r_i(t+1) = r_i(0)(1 - e^{(-\gamma \times t)}) \tag{6}$$

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

Algorithm 1 bat-inspired algorithm

```
Initialize the bat population $X_i (i = 1, 2, .., n)$ and $V_i$.
Define pulse frequency $F_i$.
Initialize pulse rate $r_i$ and the loudness $A_i$.
while $t < Maxnumberofiterations$ do
    Generate new solutions by adjusting frequency.
    Updating velocities and positions [equations (3) to (2)].
    if $rand > r_i$ then
        Select a solution $X$ among the best solutions randomly.
        Generate a local solution around the selected best solution
    end if
    if $rand < A_i$ and $f(X_i) < f(x^*)$ then
        Accept the new solutions
        Increase $r_i$ and reduce $A_i$
    end if
    Rank the bats and find the current Gbest
end while
```

2.2. Theory of inventive problem solving (TRIZ)

The methodology of TRIZ problem solving method for technical contradictions demands TRIZ researchers to identify the improving and worsening features of a design problem. Thereafter, a mapping process to the improving and worsening features into TRIZ technical contradictions matrix is performed. Based on this process, a list of inventive principles is suggested to facilitate the designer task in solving design problem. In particular, the TRIZ designers can use 39 improving and worsening features, and 40 inventive principles to solve design problem, as shown in Table 1. However, the interpretation of each suggested inventive principle is very subjective, therefore, the solutions rely on the creativity of the designers. TRIZ-inspired solution has been widely applied in several fields, such as software development [33], service quality [34], and engineering [35–37]. Furthermore, several TRIZ-inspired metaheuristic algorithms have been proposed to address real life problems. For example, Duran-Novoa et al. (2011) [38] proposed evolutionary algorithm (EA) based on TRIZ-inventive solutions to tackle inventive problem based on dialectical negation. Specifically, a new conceptual framework that integrated EA with TRIZ is proposed to facilitate computer-aided problem solving. The main contributions of this research are the inversion of the traditional EA selection or survival of the fittest, and the coupling of EA with new dialectical operators inspired by TRIZ principles. The findings proved that effectiveness promoted by TRIZ inspired evolutionary algorithm can be interpreted, understood and developed systematically. Mei et al. (2010) [39] introduced bees algorithm coupled with extra optimization operators inspired from TRIZ-inventive solution in order to optimize the process of assemble sequences task of an assembly machine in particular of moving- board-with-time-delay (MBTD) type. Within framework of bees algorithm, TRIZ principles include Dynamisation, Segmentation and Local Quality, are formulated as optimization search operators, and then inserted after requiting bees for the selected sites step. After the current bee passed the latter step, it is further optimized by TRIZ-inspired optimization operators. The experimental results demonstrated that bees algorithm coupled with the TRIZ-inspired operator is dominant when compared with the previous algorithms and the original bees algorithm.

3. Proposed method

The proposed method for gene selection composes of two stages: the filter approach stage and wrapper approach stage, as depicted in Fig. 1. A full description of both stages will be presented in the following sections.

3.1. Stage I: filter approach

In this stage, a relatively recent modified version of Minimum Redundancy Maximum Relevancy (MRMR), known as robust MRMR (rMRMR) [10], was employed to perform filtering process that is bound to give rank score for all genes in the gene selection problems. This filtering process was carried out by rMRMR in order to reduce high dimensionality of the original dataset and feed the wrapper approach with discriminative genes. Similar to MRMR, rMRMR tries to find genes that have maximum relevancy with class and minimum redundancy between them. But in contrast to MRMR, the relevancy computation process relies in ensemble of filters metrics over various characteristics (distance, probability distribution, information theory, etc.), where MRMR proceed relevancy computation in only one filter metric (i.e., mutual information). MRMR has weakness related to high variability in the classification performance, which is also existing in any single filter approach. Therefore, rMRMR is proposed to overcome this weakness

Table 1
The 40 inventive principles of TRIZ.

1.Seqmentation	2.Taking Out	3.Local Quality	4.Asymmetry
5.Merging	6.Universality	7.”Nested Doll”	8.Anti-Weight
9.Preliminary Anti-Action	10.Preliminary Action	11.Beforehand Cushioning	12.Equipotentiality
13.”The Other way round”	14.Spheroidality-Curvature	15.”Dynamisation”	16.Partial or Excessive Actions
17.Another Dimension	18.Mechanical Vibration	19.Periodic Action	20.Continuity of Useful Action
21.Skipping	22.”Blessing in Disguise”	23.Feedback	24.”Intermediary”
25.Self-Service	26.Copying	27.Cheap Short-Living Objects	28.Mechanics Substitution
29.Pneumatics and Hydraulics	30.Flexible Shells and Tin Films	31.Porous Materials	32.Color Changes
33.Homogeneity	34.Discarding and Recovering	35.Parameter Changes	36.Phase Transitions
37.Thermal Expansion	38.Strong Oxidants	39.Inert Atmosphere	40.Composite Materials

and increase the robustness and stability of MRMR. The general mechanism of rMRMR is summarized as follows, pseudo-code is given in Algorithm 2.

Step 1: ***Initialization***.

In this step, three popular filters, ie ReliefF, Chi-Square and Kullback-Liebler, were chosen from a wide range of filters. Each single filter will be executed independently, and then all genes were evaluated and scored according to its discriminative power. The genes scores obtained from each filter are combined or aggregated into one gene ranking list by using ‘Mean’ of the scores (lines 6 to 12 in Algorithm 2).

Step 2: ***Hybridization***.

The ranking gene list obtained from ensemble of filter will be hybridized with filtering process of MRMR, as follows. Firstly, in the temporary evaluation (lines 16 to 18 in Algorithm 2), the initial gene relevancy scores are assigned in accordance with corresponding gene scores (i.e., Mean score) in the ranking gene list. Secondly, in gene relevancy score (lines 26 in Algorithm 2), the calculation of relevancy score relies on two metrics: mutual information $I(G_x, c)$ and ranking gene list $R(G_i)$. The main reason of this integration process is to introduced diversity in order to avoid the bias result of single filter and thus enhance the robustness and stability of MRMR.

Step 2: ***Filtering process outcomes***.

In this step, the top ranked genes that meet the predefined threshold assigned by the user, will be passed to subsequent stage to further select a small set of meaningful genes.

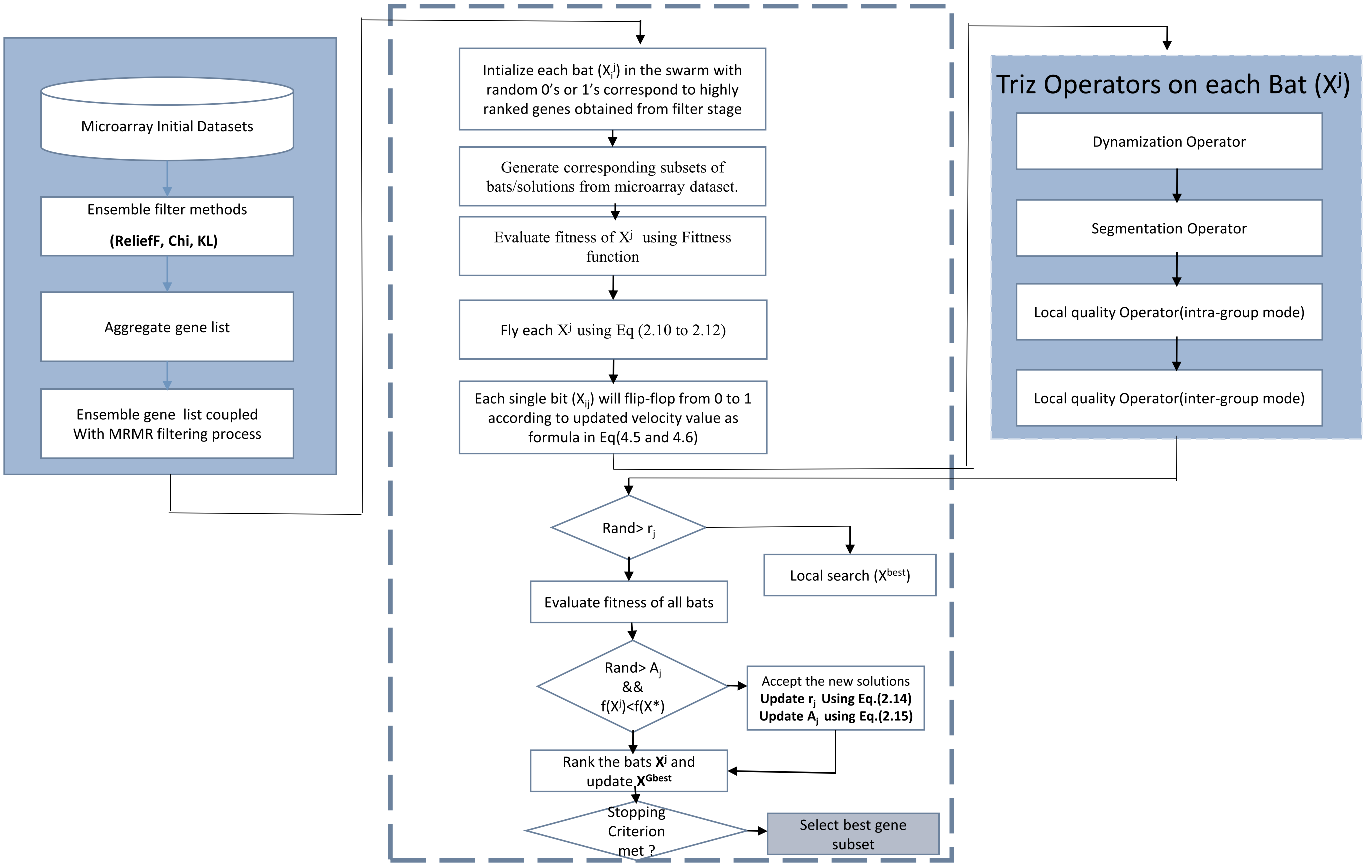
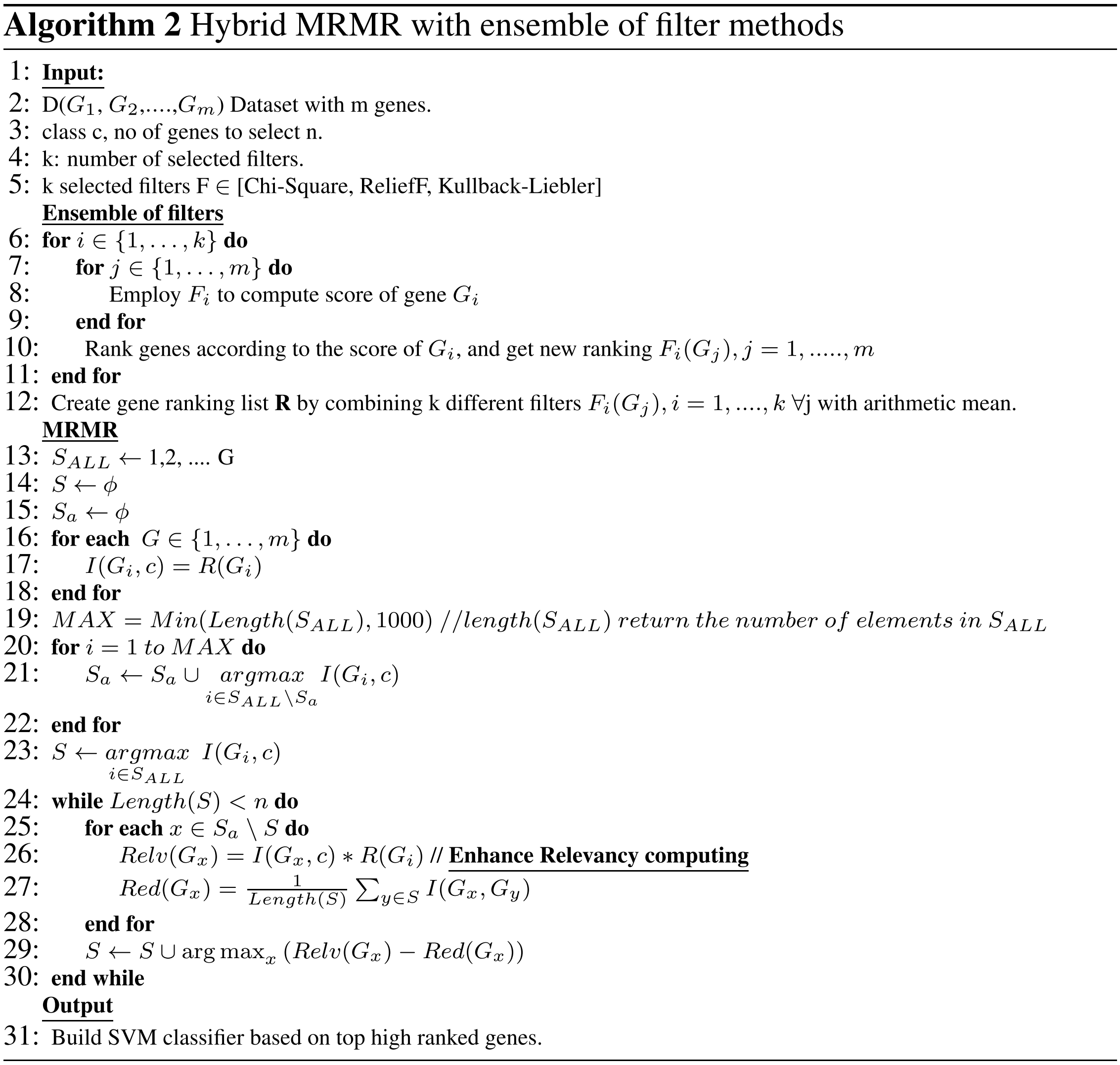

Microarray Initial Datasets
Ensemble filter methods
(ReliefF, Chi, KL)
Aggregate gene list
Ensemble gene list coupled
With MRMR filtering process
Intialize each bat (X_i^j) in the swarm with random 0’s or 1’s correspond to highly ranked genes obtained from filter stage
Generate corresponding subsets of bats/solutions from microarray dataset.
Evaluate fitness of X^j using Fittness function
Fly each X^j using Eq (2.10 to 2.12)
Each single bit (X_{ij}) will flip-flop from 0 to 1 according to updated velocity value as formula in Eq(4.5 and 4.6)
Rand> r_j
Local search (X^{best})
Evaluate fitness of all bats
Rand> A_j
&&
$f(X^j)<f(X^*)$
Accept the new solutions
Update r_j Using Eq.(2.14)
Update A_j using Eq.(2.15)
Rank the bats X^j and update X^{Gbest}
Stopping Criterion met ?
Select best gene subset
Triz Operators on each Bat (X^j)
Dynamization Operator
Segmentation Operator
Local quality Operator(intra-group mode)
Local quality Operator(inter-group mode)

Fig. 1. Flowchart of the proposed method (rMRMR-MBA) for gene selection problem.

Algorithm 2. Hybrid MRMR with ensemble of filter methods



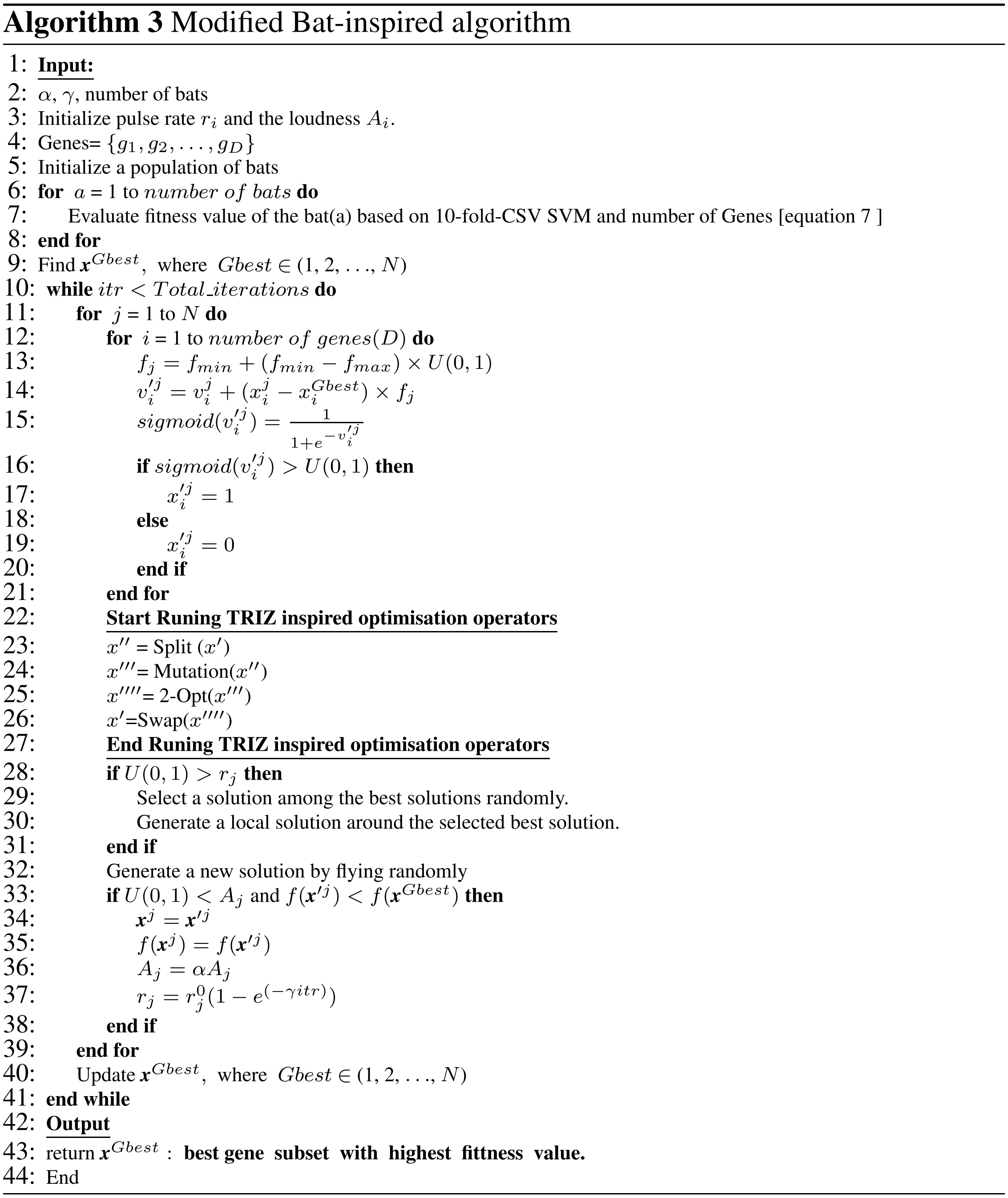
Algorithm 2 Hybrid MRMR with ensemble of filter methods

```
Input:
D($G_1$, $G_2$,....,$G_m$) Dataset with m genes.
class c, no of genes to select n.
k: number of selected filters.
k selected filters F ∈ [Chi-Square, ReliefF, Kullback-Liebler]
Ensemble of filters
for $i \in \{1, \dots, k\}$ do
    for $j \in \{1, \dots, m\}$ do
        Employ $F_i$ to compute score of gene $G_i$
    end for
    Rank genes according to the score of $G_i$, and get new ranking $F_i(G_j), j = 1, ....., m$
end for
Create gene ranking list R by combining k different filters $F_i(G_j), i = 1, ...., k$ ∀j with arithmetic mean.
MRMR
$S_{ALL} \leftarrow$ 1,2, .... G
$S \leftarrow \phi$
$S_a \leftarrow \phi$
for each $G \in \{1, \dots, m\}$ do
    $I(G_i, c) = R(G_i)$
end for
$MAX = Min(Length(S_{ALL}), 1000)$ $//length(S_{ALL})\ return\ the\ number\ of\ elements\ in\ S_{ALL}$
for $i = 1\ to\ MAX$ do
    $S_a \leftarrow S_a \cup \underset{i \in S_{ALL} \setminus S_a}{argmax}\ I(G_i, c)$
end for
$S \leftarrow \underset{i \in S_{ALL}}{argmax}\ I(G_i, c)$
while $Length(S) < n$ do
    for each $x \in S_a \setminus S$ do
        $Relv(G_x) = I(G_x, c) * R(G_i)$ // Enhance Relevancy computing
        $Red(G_x) = \frac{1}{Length(S)} \sum_{y \in S} I(G_x, G_y)$
    end for
    $S \leftarrow S \cup \arg\max_x (Relv(G_x) - Red(G_x))$
end while
Output
Build SVM classifier based on top high ranked genes.
```

3.2. Stage II: wrapper approach

In this stage, wrapper approach is performed for further filtering process and seek for the most informative subset of genes from the top ranked genes, which are obtained by rMRMR. BA is modified by inserting new operators inspired from TRIZ to refine the search space effectively. As BA is population based algorithm which has the capability of managing a population of solutions at the same time, however, it provides a wide coverage in the search space but without operating extensive local search on single region in the search space. Therefore, a couple of TRIZ principles are formulated as optimization operators to exploit each single search space region effectively. The ultimate goal of our proposed method is to maximize classification accuracy while minimizing the number of selected genes. The pseudocode of the proposed method is presented in Algorithm 3. The components and process of the proposed method in Stage II are illustrated in the following sections:

Algorithm 3. Modified Bat-inspired algorithm



```
Algorithm 3 Modified Bat-inspired algorithm
Input:
$\alpha$, $\gamma$, number of bats
Initialize pulse rate $r_i$ and the loudness $A_i$.
Genes= $\{g_1, g_2, \ldots, g_D\}$
Initialize a population of bats
for $a$ = 1 to $number\ of\ bats$ do
    Evaluate fitness value of the bat(a) based on 10-fold-CSV SVM and number of Genes [equation 7 ]
end for
Find $\mathbf{x}^{Gbest}$, where $Gbest \in (1, 2, \ldots, N)$
while $itr < Total\_iterations$ do
    for $j$ = 1 to $N$ do
        for $i$ = 1 to $number\ of\ genes(D)$ do
            $f_j = f_{min} + (f_{min} - f_{max}) \times U(0,1)$
            $v_i^{\prime j} = v_i^j + (x_i^j - x_i^{Gbest}) \times f_j$
            $sigmoid(v_i^{\prime j}) = \frac{1}{1+e^{-v_i^{\prime j}}}$
            if $sigmoid(v_i^{\prime j}) > U(0,1)$ then
                $x_i^{\prime j} = 1$
            else
                $x_i^{\prime j} = 0$
            end if
        end for
        Start Runing TRIZ inspired optimisation operators
        $x''$ = Split ($x'$)
        $x'''$= Mutation($x''$)
        $x''''$= 2-Opt($x'''$)
        $x'$=Swap($x''''$)
        End Runing TRIZ inspired optimisation operators
        if $U(0,1) > r_j$ then
            Select a solution among the best solutions randomly.
            Generate a local solution around the selected best solution.
        end if
        Generate a new solution by flying randomly
        if $U(0,1) < A_j$ and $f(\mathbf{x}^{\prime j}) < f(\mathbf{x}^{Gbest})$ then
            $\mathbf{x}^j = \mathbf{x}^{\prime j}$
            $f(\mathbf{x}^j) = f(\mathbf{x}^{\prime j})$
            $A_j = \alpha A_j$
            $r_j = r_j^0(1 - e^{(-\gamma itr)})$
        end if
    end for
    Update $\mathbf{x}^{Gbest}$, where $Gbest \in (1, 2, \ldots, N)$
end while
Output
return $\mathbf{x}^{Gbest}$ : best gene subset with highest fitness value.
End
```

3.2.1. Solution representation

In the context of optimization, gene selection problems are types of combinatorial problems, in which search space formed by candidate gene subsets [40,41]. The gene search space is expanded and become complicated to be addressed by increasing the number of genes. To introduce mathematically, if N represents the number of genes, there are [2^N] candidate subsets of genes.

Each candidate solution (i.e., solution x) advance to gene selection problem is represented by a binary string of length N, $x = (x_1, x_2, \ldots, x_N)$, where the gene is preserved if the corresponding bit x_i in the candidate solution equals 1. Otherwise, the gene is discarded.

3.2.2. Fitness function

As aforementioned, each candidate gene subset is a series of 0's and 1's bits. Genes coded in ones will only be considered in the evaluation. The fitness function used to evaluate the effectiveness of each individual solution is presented in the Eq. (7) [42].

$$\alpha \times R(D) + \beta \times \frac{|C| - |R|}{|C|} \tag{7}$$

where $\alpha R(D)$ stands for the classification accuracy estimated by running ten multiple cross-validation with SVM classifiers, on the training dataset on the basis of gene subset R to decision D. The gene subset size is denoted by $|R|$. $|C|$ is the total number of genes. The two weighting

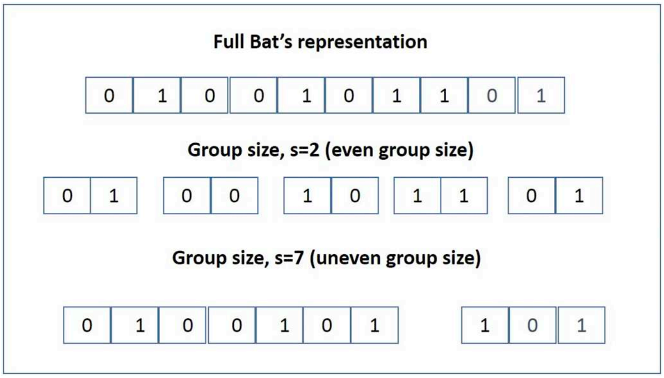

Full Bat's representation
0 1 0 0 1 0 1 1 0 1
Group size, s=2 (even group size)
0 1 | 0 0 | 1 0 | 1 1 | 0 1
Group size, s=7 (uneven group size)
0 1 0 0 1 0 1 | 1 0 1

Fig. 2. The segmentation operator.

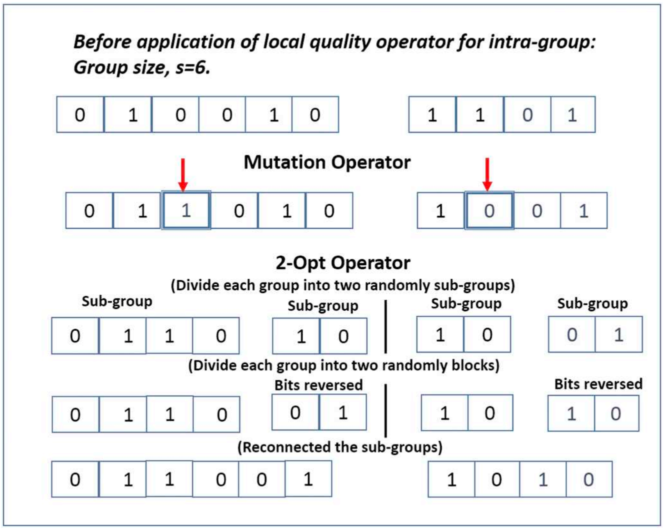


Fig. 3. The Local Quality operator in the intra-group mode.

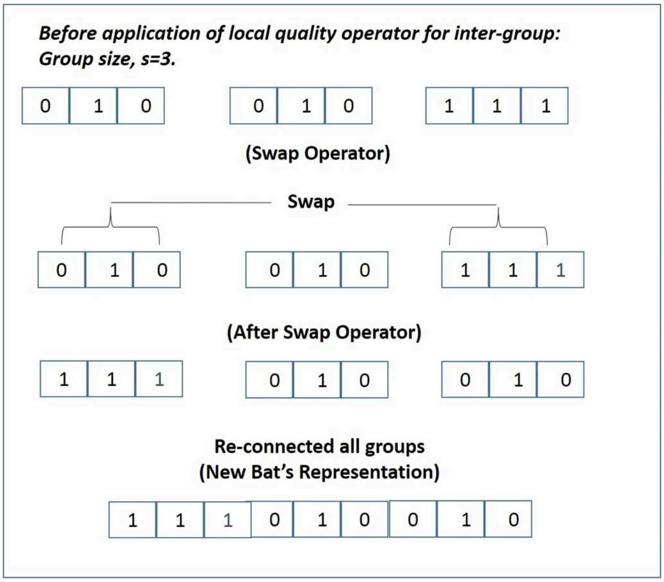


Fig. 4. The Local Quality operator in the inter-group mode.

factors α and β are related to classification accuracy and gene subset length, respectively. $\alpha \in [0,1]$ and $\beta = (1\text{-}\alpha)$. The classification accuracy is more important than the subset length. In this paper, α is set to 0.8 [3,42].

3.2.3. Incorporating TRIZ operators with bat algorithm

In this stage, BA is modified, called MBA, by adding extra optimization search operators inspired from TRIZ inventive solution to promote the searching process into basic BA and effectively explore the interaction among the genes. According to the description of TRIZ inventive principles, three principles are selected to be formulated as optimization search operators. These principles include dynamization, segmentation and local quality. They divided each object/solution into several groups and subsequently alter each group to examine the interaction among genes in order to produce an inventive solution for a given problem. In the context of gene selection optimization problem, the behavior of these principles could be formulated as optimization operators to optimize the gene search space, which are namely TRIZ-inspired optimization operators. The incorporation of TRIZ-inspired optimization operators with BA algorithm is presented in flow chart in Fig. 1 and pseduocoded in Algorithm 3.

Based on the flow chart, MBA start running its searching process by establish an initial population, and then they evaluated using fitness function, as shown in Eq. 7. After that bat motion step, each bat in the population is manipulated by TRIZ-inspired optimisation operators (dynamization, segmentation, and local quality). The optimization function of each operator are discussed as follows:

Dynamization Operator In this operator, the solution representation of the current bat is maintained, but it will be divided into groups in the subsequent operator. The main function of dynamization operator is to determine the number of elements in the groups at each iteration. Notable, the number of elements in the groups is generated randomly and control the number of divided groups in the subsequent operator.

Segmentation Operator In this operator, the current bat is divided into several groups according to the number of elements in each group (N) that assigned by dynamization operator. If the solution size of the current bat is divisible by (N), then the size of each group is even. Otherwise, the size of each group is uneven. Fig. 2 illustrate the process of segmentation operator.

Local Quality Operator The main task of local quality operator is to apply variation in each group in the solution. This task can be formulated as optimization task by means of using neighborhood search operators to give variation or diversity in each part in the solution. In this research, the local quality operator is applied into two modes: (1) intra-group mode (within a group); and (2) inter-group mode (between groups). For Intra-group mode, two neighborhood search operators are applied in each group, which are mutation operator and 2-Opt operation. In mutation operator, a random position in each group is generated and its value is flipped to "1" if it "0" and vice versa. While in the 2-Opt operation, each group is further divided into two sub-groups. One of the two sub-groups is reversed and the other is maintained. This process is ended by recombining the sub-groups. The entire process of intra-group mode is illustrated in Fig. 3. For the inter-group mode, new arrangement in the sequence of the groups is taken place by using swap operator. In practices, two random groups are selected and then their positions are swapped in the groups sequence. Fig. 4 shows the process of the inter-group mode.

In the remaining process, MBA completed its search procedure as similar as basic BA, where Pulse rate and Loudness are operated to further optimize the current bat/solution. Eventually, the entire search procedure of MBA is iterated until the termination condition is met. If it is met, then MBA output best fitness bat represents the most predictive gene subset with minimum number of genes.

4. Time complexity of the proposed rMRMR-MBA

The time complexity required to execute the proposed rMRMR-MBA method is depend on the complexity of the operator of rMRMR and MBA methods. The rMRMR as pseudocoded in Algorithm 2 requires $\mathcal{O}(n \times m)$ where n is the number of selected genes while m is the total number of genes. On the other hand, the time complexity of the MBA depends on two parts: the BA operators and the proposed Triz-inspired operators. According to the pseudocode of MBA shown in Algorithm 3 (Lines 10–41), the time complexity is $\mathcal{O}(Total_{i}terations \times N \times D)$. Note that the Triz-inspired operators required time complexity as follows: the *Split* function requires $\mathcal{O}(D/2)$, the *Mutation* function requires $\mathcal{O}(D/2)$, the 2-Opt function requires $\mathcal{O}(D/2)$, and the Swap function requires $\mathcal{O}(2)$. In a nutshell, the overall time complexity for the proposed rMRMR-MBA method is $\mathcal{O}(Total_{i}terations \times N \times D)$ where $Total_{i}terations$ refers to the maximum number of iterations, N refers to the number of bats (i.e., solutions), while D is the total number of genes (i.e., solution dimensionality). It is worth mentioning that the time complexity is also depends on the time complexity required to calculate the objective function by SVM. Therefore, the time complexity is calculated regardless the computational time required for calculating the objective function value at each iteration for each generated solution.

5. Experimental setup and results

The performance of the proposed method rMRMR-MBA is evaluated using 10 gene expression datasets, which were taken from the `http://csse.szu.edu.cn/staff/zhuzx/Datasets.html`. In our experiments, the involved algorithms in both stages were programmed using two languages (i.e., Java and Matlab). In filter stage, rMRMR is implemented using Matlab, while other filters (i.e., ReliefF, Chi-Square, and Kullback-Liebler) are implemented in java using weka tool [43]. In wrapper stage, modified BA and SVM are implemented using java. In particular, SVM was implemented using LIBSVM [44]. All the experiments are performed on an Intel Core Quad 2.66 GHz CPU with 4 GB of RAM.

5.1. Dataset used

The selected benchmark datasets include "Breast", "MLL", "Colon", "ALLAML", "ALLAML-3C", "ALLAML-4C", "Lymphoma", "CNS", "Ovarian" and "SRBCT" datasets. These datasets are commonly used in many studies and cover the example of small, medium, and large dimensional datasets. The characteristics of the selected datasets are summarized in Table 2. Table 2 contains the number of genes (# Genes), number of samples of each dataset (# Samples) and the number of classes (# Classes).

5.2. Parameter settings

In both stages (i.e., filter and wrapper), the parameter setting values are assigned based on some preliminary experiments and based on previous parameters from the referenced papers. In filter approach, the threshold for the top ranked genes in rMRMR is assigned to 50, which is in accordance to previous studies [5,13,45,46]. In wrapper approach, for MBA, the parameters used in the algorithm are number of artificial bats, minimum frequency (F_{min}), maximum frequency (F_{max}), loudness (A), pulse rate (r), α and γ. 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 [20,28,47]. For SVM, Radial Basis Function (RBF) kernel is selected to perform classification task, as well as traditional grid search algorithm was experimented to give the best parameter values for RBF kernel [48]. The parameter setting values of the MBA are shown in Table 3.

Table 2
Datasets characteristic.

Datasets	# Genes	# Samples	# Classes
Breast	24,481	97	2
MLL	12,582	72	3
Colon	2000	62	2
ALL-AML	7129	72	2
ALL-AML-3C	7129	72	3
ALL-AML-4C	7129	72	4
Lymphoma	4026	62	3
CNS	7129	60	2
Ovarian	15,154	253	2
SRBCT	2308	83	4

Table 3
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

Table 4
Comparsion between BA and MBA.

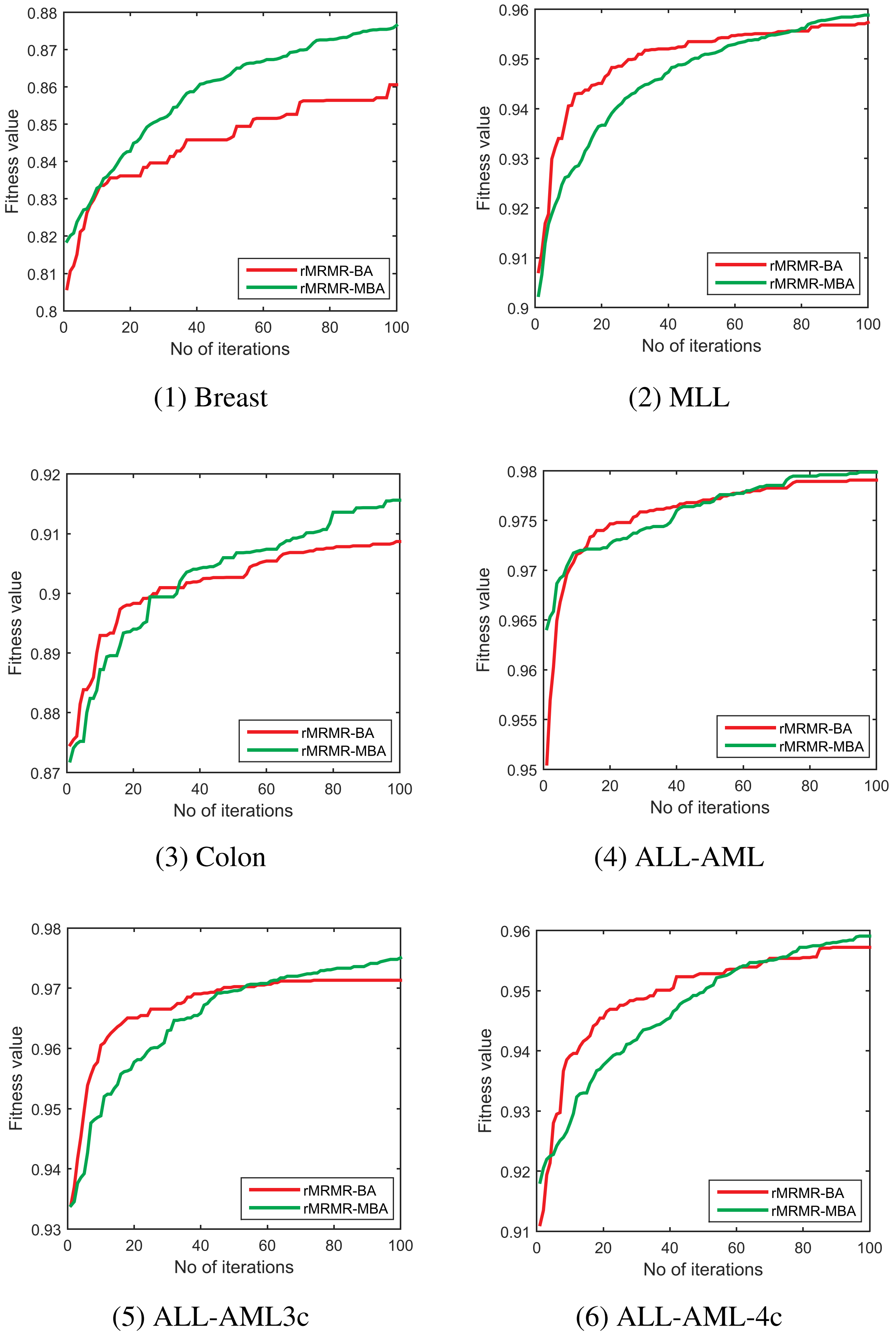
Algorithm		Dataset				
		Breast	MLL	Colon	ALL-AML	ALL-AML3c
BA	$\vert \# G\vert$	**15.63**	10.67	10.16	5.23	7.17
	ACC	90.34	100	93.33	100	100
	$(\vert \# F\vert)$	86.02	95.73	90.60	97.91	97.13
MBA	$\vert \# G\vert$	18.37	**10.3**	**9.7**	**5.03**	**6.23**
	ACC	**93.75**	**100**	**94.30**	**100**	**100**
	$(\vert \# F\vert)$	**87.65**	**95.88**	**91.56**	**97.99**	**97.51**
	T–Sig.	*	–	–	–	*
		ALL-AML-4c	Lymphoma	CNS	Ovarian	SRBCT
BA	$\vert \# G\vert$	**10.2**	13.97	19.23	**3.9**	11.06
	ACC	99.68	100	96.94	**100**	100
	$(\vert \# F\vert)$	95.67	94.41	89.86	**98.44**	95.57
MBA	$\vert \# G\vert$	10.23	**8.83**	**16.43**	3.93	**10.47**
	ACC	**100**	**100**	**99.61**	100	**100**
	$(\vert \# F\vert)$	**95.91**	**96.47**	**93.12**	98.43	**95.81**
	T–Sig.	–	*	*	–	*

5.3. Effect of Triz-inspired operators on the performance of MBA

In this section, the effect of TRIZ-inspired operators on the convergence behavior of the MBA is studied. Furthermore, the MBA is compared with basic BA. Both methods were executed 30 independent runs. The experimented results in terms of average classification accuracy (ACC), the average number of selected genes ($\vert \# G\vert$), and the average fitness value ($\vert \# F\vert$) were presented in Table 4. In order to show significant statistical difference between both methods, Wilcoxon signed-rank statistical test was used. Moreover, the best results of ACC, $\vert \# G\vert$ and $\vert \# F\vert$ are highlighted in bold font.

In Table 4, $T - sig$ row, with the probability range of $\alpha \leq 0.05$, '$*$'implies that the results achieved by MBA method are significantly better than the BA, while '$-$'implies that the results achieved by MBA method are not significantly better than the BA. Wilcoxon signed-rank statistical test is considered only fitness to process statistical calculation due to fitness value composed of classification accuracy and number of selected genes.

As shown in Table 4, MBA achieved higher classification accuracy



(1) Breast

(2) MLL

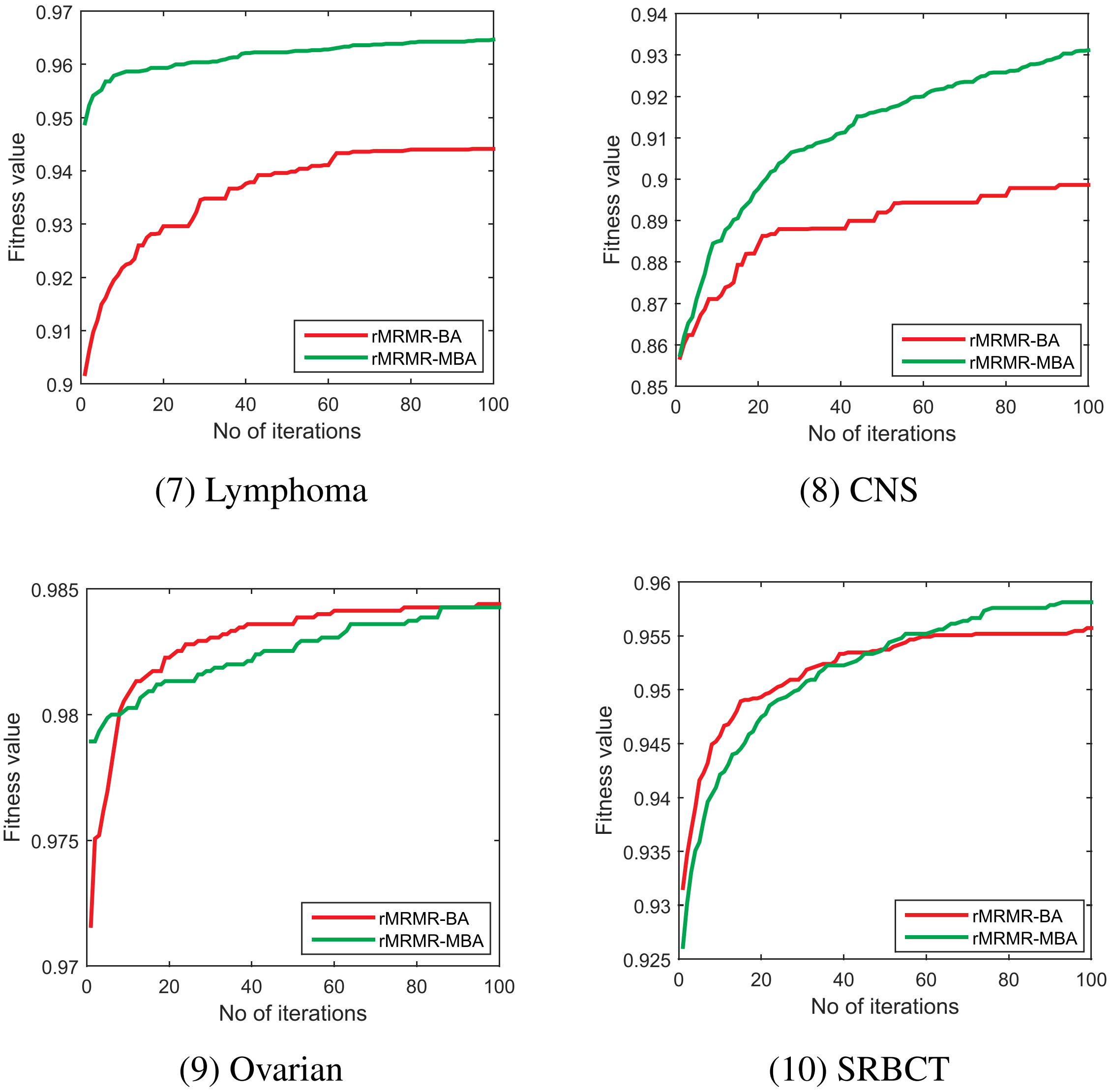
(3) Colon

(4) ALL-AML

(5) ALL-AML3c

(6) ALL-AML-4c

Fig. 5. The convergence behavior of BA and MBA for 10 datasets.



(7) Lymphoma

(8) CNS

(9) Ovarian

(10) SRBCT

Fig. 5. (*continued*)

than BA on nine datasets (i.e., Breast, MLL, Colon, ALLAML, ALLAML-3C, ALLAML-4C, Lymphoma, CNS and SRBCT). In other hand, BA achieved higher classification accuracy than MBA in only one dataset (i.e., Ovarian). In term of classification and number of selected genes, MBA has resulted in higher classification accuracy and smaller number of selected genes on seven datasets (i.e., MLL, Colon, ALLAML, ALLAML-3C, Lymphoma, CNS and SRBCT). In Ovarian dataset, BA showed slightly better results than MBA. In the remaining two datasets (Breast and ALLAML-4C), none of the both methods can overcome each other in both measurements. In term of fitness function, MBA yields higher fitness value (| # F|) than BA on all datasets except Ovarian dataset. Table 4 also shows the results of wilcoxon signed-rank statistical test between MBA and BA. It can be inferred that there are significant differences in favor of MBA in five datasets (i.e., Breast, ALLAML-3C, Lymphoma, CNS, and SRBCT).

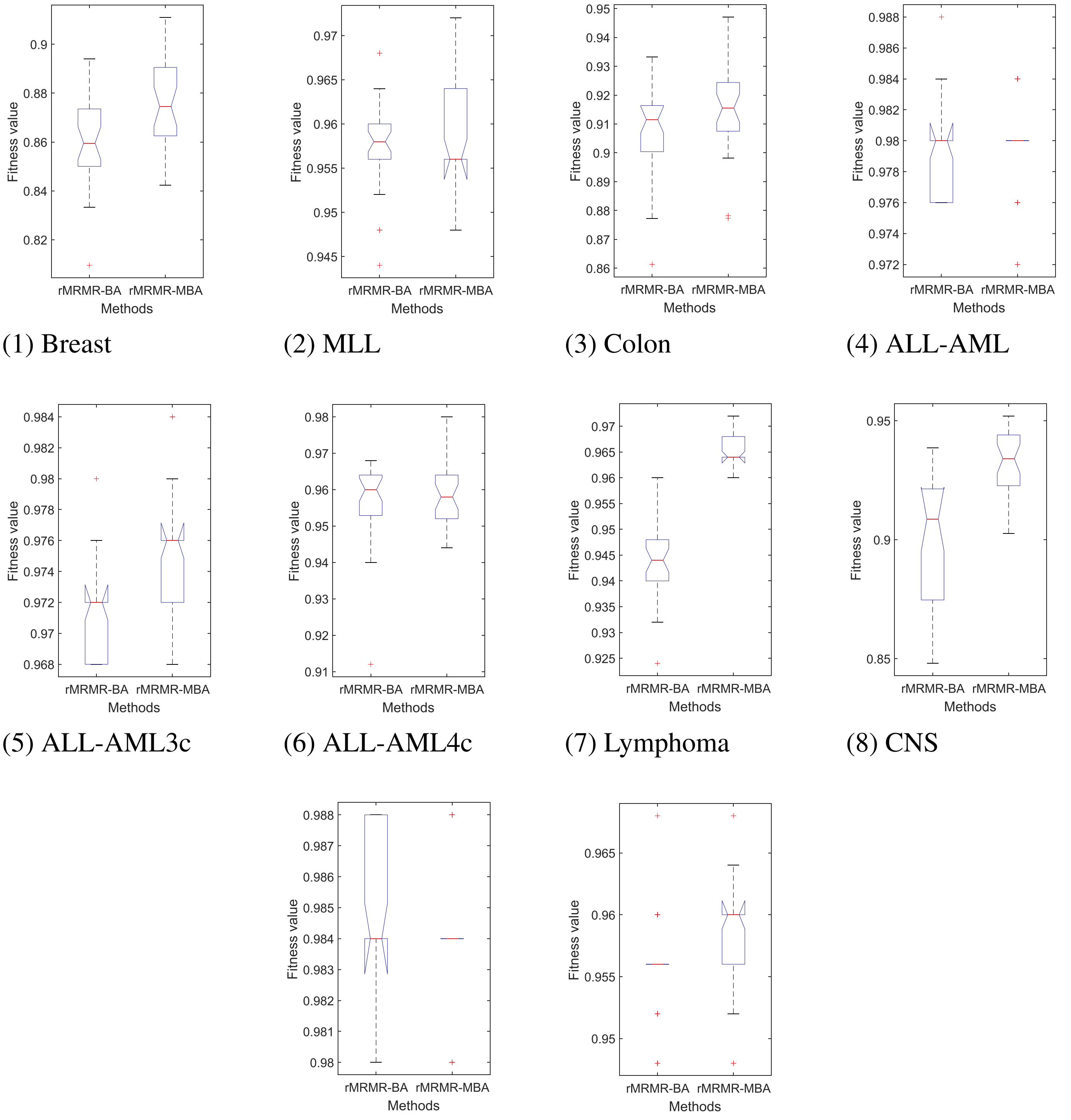
The convergence behaviour for both methods is plotted on all experimented datasets, as shown in Fig. 5. The convergence behaviour trend of MBA is significant better than BA on three datasets (i.e., Breast, Lymphoma, and CNS). For (MLL, Colon, ALLAML, ALLAML-3C, ALLAML-4C, and SRBCT), even though MBA perform slightly worse than BA in these datasets in early stage of evolution, but in late stage of evolution the MBA is able to converge better than BA. On the other hand, for only one dataset(i.e., Ovarian), the convergence trend of BA is slightly better than MBA. In summary, the experimental results demonstrated that MBA produced best compromise between the classification accuracy and number of selected genes on most of experimented datasets. The outstanding results of TRIZ-inspired optimization operators that assist BA further explore the interaction between the genes that managed it to access the most interesting region in the search space which consists of small set of relevant and informative genes.

In terms of diversity behavior of the proposed rMRMR-MBA, as shown in the boxplots in Fig. 6 for all dataset used, the proposed method is able to maintain the diversity during the search and converge to the highly accurate results.

5.4. Comparative evaluations

In this section, the effectiveness of the proposed method is further assessed by comparing rMRMR-MBA with the state-of-the-art methods, as shown in Table 5. The results reported in Table 6 are based on 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.

According to the Table 6, rMRMR-MBA has resulted higher or similar classification accuracy than other comparative methods on nine out of ten datasets. In only ‘Colon’ dataset, rMRMR-MBA is ranked third



(1) Breast

(2) MLL

(3) Colon

(4) ALL-AML

(5) ALL-AML3c

(6) ALL-AML4c

(7) Lymphoma

(8) CNS

(9) Ovarian

(10) SRBCT

Fig. 6. Boxplots for all datasets to show the diversity behaviour of the proposed rMRMR-MBA.

after LDA-GA and MRMR-GA. In term of classification accuracy and number of selected genes, the proposed method obtained the lowest number of selected genes with highest classification accuracy results on two out of ten datasets (i.e., ALL_AML_3c, Ovarian), and competitive results on the remaining datasets.

6. Conclusion and future work

In this paper, an efficient filter/wrapper method was proposed to address gene selection problem. In the proposed method, rMRMR is used as filter approach while MBA and SVM are used as wrapper approach. In wrapper approach, BA algorithm, which represents the search engine in wrapper approach, is improved by incorporating within the searching procedure further optimisation search operators. This is to promote a wide coverage in the gene search space and in particular explore the interaction between the genes effectively.

Ten high dimensional well-known gene expression dataset are experimented to test the performance of the proposed method. The

Table 5
Key to comparative methods.

Key	Method name	Reference
CFC-iBPSO	Correlation-based Feature Selection with improved-Binary Particle Swarm Optimization	[18]
IG-SGA	Information Gain and Standard Genetic Algorithm	[49]
MRMR-BA	Minimum Redundancy Maximum Relevancy with a Bat-Inspired Algorithm	[28]
HSA-MB	Hybridising harmony search with a Markov blanket for gene selection problems	[13]
MBEGA	Markov Blanket-Embedded Genetic Algorithm for gene selection	[45]
MRMR-GA	Minimum Redundancy Maximum Relevancy with a Genetic Algorithm	[2]
MA-C	Correlation-based memetic framework	[50]
BPSO-CGA	Binary Particle Swarm Optimisation and a Combat Genetic Algorithm	[12]
BIRSW	Best Incremental Ranked Subset	[51]
LDA-GA	Fisher Linear Discriminate Analysis-based Genetic Algorithm	[52]

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

Datasets	M	rMRMR-MBA	HSA-MB [13]	MBEGA [45]	CFC-iBPSO [18]	MRMR-BA [28]	IG-SGA [49]	MRMR-GA [2]	MA-C [50]	BPSO-CGA [12]	BIRSW [51]	LDA-GA [52]
Breast	$\|\#G\|$	12.63	**5.06**	14.5	32.7	18.3	–	–	183	–	–	–
	ACC	**95.4**	80.06	80.74	92.75	88.8	–	–	95.26	–	–	–
MLL	$\|\#G\|$	8	**6.6**	32.1	30.08	19.03	–	–	108	–	–	–
	ACC	**100**	99.55	94.33	**100**	79.86	–	–	**100**	–	–	–
Colon	$\|\#G\|$	12.27	4.16	24.5	4.2	8.13	60	15	–	214	**3.5**	7
	ACC	97.85	90.27	85.66	94.89	93.12	85.48	98.39	–	96.7	85.48	**99.83**
ALL_AML	$\|\#G\|$	4.07	5	12.8	4.3	5.23	3	15	387	300	**2.5**	3
	ACC	**100**	99.34	95.89	**100**	100	97.06	**100**	99.56	**100**	**93.04**	99.5
ALL_AML_3c	$\|\#G\|$	**5.33**	5.84	18.1	6	7.27	–	–	394	–	–	–
	ACC	**100**	99.18	96.64	**100**	100	–	–	99.53	–	–	–
ALL_AML_4c	$\|\#G\|$	6.73	**6.37**	26.2	20.7	11.03	–	–	386	–	–	–
	ACC	**100**	96.79	91.93	97.63	99.77	–	–	98.61	–	–	–
Lymphoma	$\|\#G\|$	8.13	**3.75**	34.3	24	37.73	–	15	–	196	10.3	–
	ACC	**100**	99.99	97.68	100	86.36	–	98.96	–	**100**	82.14	–
CNS	$\|\#G\|$	11.2	7.43	20.5	10.5	19.2	38	–	374	–	–	**4**
	ACC	**100**	84.17	72.21	95.84	94.22	86.76	–	97.78	–	–	**99.3**
Ovarian	$\|\#G\|$	**3.07**	5.73	9	3.3	3.83	–	–	247	–	–	6
	ACC	**100**	99.81	99.71	**100**	100	–	–	**100**	–	–	97.4
SRBCT	$\|\#G\|$	9.13	**8.9**	60.7	34.1	12.83	–	–	526	880	–	–
	ACC	**100**	99.57	99.23	**100**	100	–	–	**100**	**100**	–	–

characteristics of these datasets varied in terms of number of genes, samples, and classes. The evaluation process involved two phases. In the first phase, BA is evaluated with and without TRIZ-inspired optimization operators. The evaluation is achieved based on three criteria: number of selected genes, classification accuracy, and fitness function value. In term of classification accuracy and number of selected genes, the proposed MBA have resulted better compromise on most of experimented datasets. In term of fitness function value, MBA produced higher fitness function value ($|F|$) than BA on all datasets except 'Ovarian' dataset. In the second phase, the proposed rMRMR-MBA is compared against the state-of-art gene selection methods. rMRMR-MBA achieved equivalent or higher classification accuracy than other comparative methods on nine out of ten datasets. In other hand, in terms of classification accuracy and number of selected genes, rMRMR-MBA achieved the best overall results on two out of ten datasets (i.e., ALL-AML3c, Ovarian) and competitive results on the remaining datasets.

The formulation of TRIZ-inventive solution as optimisation search operators and integrated them with BA algorithm, result in practical and effective gene selection tool that capable to produce the most biological relevant genes related to disease. For future work, the performance of the proposed method can be investigated using new class of machine learning algorithm (i.e., Deep learning). Furthermore, rMRMR-MBA can be expanded to other high-dimensional datasets, including image and test data.

Acknowledgment

This work was supported by grant from the Deanship of Scientific Research in Al-Balqa Applied University, Al-Salt, Jordan (Grant No. DSR-2019-#30). The authors are pleased to acknowledge the editor and reviewers for their valuable comments and substantial efforts which improve the readability and presentation of this paper.

References

[1] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (5439) (1999) 531–537.

[2] 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.

[3] C.-M. Lai, W.-C. Yeh, C.-Y. Chang, Gene selection using information gain and improved simplified swarm optimization, Neurocomputing.

[4] A. Jain, D. Zongker, Feature selection: evaluation, application, and small sample performance, Pattern Analysis and Machine Intelligence, IEEE Transactions on 19 (2) (1997) 153–158.

[5] V. Bolón-Canedo, N. Sánchez-Maroño, A. Alonso-Betanzos, J.M. Benítez, F. Herrera, A review of microarray datasets and applied feature selection methods, Inf. Sci. 282 (2014) 111–135.

[6] I. Kononenko, Estimating attributes: analysis and extensions of relief, European Conference on Machine Learning, Springer, 1994, pp. 171–182.

[7] C.-T. Su, J.-H. Hsu, An extended chi2 algorithm for discretization of real value attributes, IEEE Trans. Knowl. Data Eng. 17 (3) (2005) 437–441.

[8] S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Stat. 22 (1) (1951) 79–86.

[9] C. Ding, H. Peng, Minimum redundancy feature selection from microarray gene expression data, J. Bioinforma. Comput. Biol. 3 (2) (2005) 185–205.

[10] O.A. Alomari, A.T. Khader, M.A. Al-Betar, M.A. Awadallah, A novel gene selection method using modified mrmr and hybrid bat-inspired algorithm with β-hill climbing, Appl. Intell. 48 (11) (2018) 4429–4447.

[11] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, The Journal of Machine Learning Research 3 (2003) 1157–1182.

[12] L.-Y. Chuang, C.-H. Yang, J.-C. Li, C.-H. Yang, A hybrid bpso-cga approach for gene selection and classification of microarray data, J. Comput. Biol. 19 (1) (2012) 68–82.

[13] S.S. Shreem, S. Abdullah, M.Z.A. Nazri, Hybridising harmony search with a markov blanket for gene selection problems, Inf. Sci. 258 (2014) 108–121.

[14] D. Ramyachitra, M. Sofia, P. Manikandan, Interval-value based particle swarm optimization algorithm for cancer-type specific gene selection and sample classification, Genomics data 5 (2015) 46–50.

[15] H.M. Alshamlan, G.H. Badr, Y.A. Alohali, Genetic bee colony (gbc) algorithm: a new gene selection method for microarray cancer classification, Comput. Biol. Chem. 56 (2015) 49–60.

[16] F.V. Sharbaf, S. Mosafer, M.H. Moattar, A hybrid gene selection approach for microarray data classification using cellular learning automata and ant colony optimization, Genomics 107 (6) (2016) 231–238.

[17] E. Pashaei, E. Pashaei, N. Aydin, Gene selection using hybrid binary black hole algorithm and modified binary particle swarm optimization, Genomics.

[18] I. Jain, V.K. Jain, R. Jain, Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification, Appl. Soft Comput. 62 (2018) 203–215.

[19] H. Ahmed, J. Glasgow, Swarm intelligence: concepts, models and applications, School of Computing, Queens University Technical Report.

[20] X.-S. Yang, A new metaheuristic bat-inspired algorithm, Nature Inspired Cooperative Strategies for Optimization (NICSO 2010), Springer, 2010, pp. 65–74.

[21] M. Chawla, M. Duhan, Bat algorithm: a survey of the state-of-the-art, Appl. Artif. Intell. 29 (6) (2015) 617–634.

[22] G. Komarasamy, A. Wahi, An optimized k-means clustering technique using bat algorithm, Eur. J. Sci. Res. 84 (2) (2012) 26–273.

[23] B. Ramesh, V.C.J. Mohan, V.V. Reddy, Application of bat algorithm for combined economic load and emission dispatch, Int, J. of Electricl Engineering and Telecommunications 2 (1) (2013) 1–9.

[24] P. Musikapun, P. Pongcharoen, 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, 2012, pp. 98–102.

[25] S. Mishra, K. Shaw, D. Mishra, A new meta-heuristic bat inspired classification approach for microarray data, Procedia Technology 4 (2012) 802–806.

[26] M. Kang, J. Kim, J.-M. Kim, Reliable fault diagnosis for incipient low-speed bearings using fault feature analysis based on a binary bat algorithm, Inf. Sci. 294 (2015) 423–438.

[27] C. Karri, U. Jena, Fast vector quantization using a bat algorithm for image compression, Engineering Science and Technology, an International Journal 19 (2) (2016) 769–781.

[28] O.A. Alomari, A.T. Khader, M.A. Al-Betar, L.M. Abualigah, Gene selection for cancer classification by combining minimum redundancy maximum relevancy and bat-inspired algorithm, International Journal of Data Mining and Bioinformatics 19 (1) (2017) 32–51.

[29] M. Dashtban, M. Balafar, P. Suravajhala, Gene selection for tumor classification using a novel bio-inspired multi-objective approach, Genomics 110 (1) (2018) 10–17.

[30] M.A. Al-Betar, M.A. Awadallah, H. Faris, X.-S. Yang, A.T. Khader, O.A. Alomari, Bat-inspired algorithms with natural selection mechanisms for global optimization, Neurocomputing 273 (2018) 448–465.

[31] M.A. Al-Betar, M.A. Awadallah, Island bat algorithm for optimization, Expert Syst. Appl. 107 (2018) 126–145.

[32] G. Altshuller, 40 Principles: TRIZ Keys to Innovation, 1 Technical Innovation Center, Inc., 2002.

[33] D.L. Mann, B. Maizlish, Systematic (Software) Innovation, IFR Press, 2008.

[34] J. Kim, J. Kim, Y. Lee, W. Lim, I. Moon, Application of triz creativity intensification approach to chemical process safety, J. Loss Prev. Process Ind. 22 (6) (2009) 1039–1043.

[35] D. Russo, C. Rizzi, G. Montelisciani, Inventive guidelines for a triz-based eco-design matrix, J. Clean. Prod. 76 (2014) 95–105.

[36] M. Ang, D. Pham, K. Ng, Application of the bees algorithm with triz-inspired operators for pcb assembly planning, Proceedings of 5th Virtual International Conference on Intelligent Production Machines and Systems (IPROMS2006), 2009, pp. 454–459.

[37] M. Li, X. Ming, L. He, M. Zheng, Z. Xu, A triz-based trimming method for patent design around, Comput. Aided Des. 62 (2015) 20–30.

[38] R. Duran-Novoa, N. Leon-Rovira, H. Aguayo-Tellez, D. Said, Inventive problem solving based on dialectical negation, using evolutionary algorithms and triz heuristics, Comput. Ind. 62 (4) (2011) 437–445.

[39] C.A. Mei, D. Pham, J.S. Anthony, W.N. Kok, Pcb assembly optimisation using the bees algorithm enhanced with triz operators, IECON 2010-36th Annual Conference on IEEE Industrial Electronics Society, IEEE, 2010, pp. 2708–2713.

[40] B. Duval, J.-K. Hao, J.C. Hernandez Hernandez, A memetic algorithm for gene selection and molecular classification of cancer, Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation, ACM, 2009, pp. 201–208.

[41] M. Dash, H. Liu, Feature selection for classification, Intelligent data analysis 1 (3) (1997) 131–156.

[42] X. Li, M. Yin, Multiobjective binary biogeography based optimization for feature selection using gene expression data, IEEE Transactions on NanoBioscience 12 (4) (2013) 343–353.

[43] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The weka data mining software: an update, ACM SIGKDD Explorations Newsletter 11 (1) (2009) 10–18.

[44] C.-C. Chang, C.-J. Lin, Libsvm: a library for support vector machines, ACM Transactions on Intelligent Systems and Technology (TIST) 2 (3) (2011) 27.

[45] Z. Zhu, Y.-S. Ong, M. Dash, Markov blanket-embedded genetic algorithm for gene selection, Pattern Recogn. 40 (11) (2007) 3236–3248.

[46] M.K. Ebrahimpour, M. Eftekhari, Ensemble of feature selection methods: a hesitant fuzzy sets approach, Appl. Soft Comput. 50 (2017) 300–312.

[47] O.A. Alomari, A.T. Khader, M.A. Al-Betar, L.M. Abualigah, Mrmr ba: a hybrid gene selection algorithm for cancer classification, J. Theor. Appl. Inf. Technol. 95 (12) (2017) 2610–2618.

[48] C.-L. Huang, C.-J. Wang, A ga-based feature selection and parameters optimizationfor support vector machines, Expert Syst. Appl. 31 (2) (2006) 231–240.

[49] H. Salem, G. Attiya, N. El-Fishawy, Classification of human cancer diseases by gene expression profiles, Appl. Soft Comput. 50 (2017) 124–134.

[50] S.S. Kannan, N. Ramaraj, A novel hybrid feature selection via symmetrical uncertainty ranking based local memetic search algorithm, Knowl.-Based Syst. 23 (6) (2010) 580–585.

[51] R. Ruiz, J.C. Riquelme, J.S. Aguilar-Ruiz, Incremental wrapper-based gene selection from microarray data for cancer classification, Pattern Recogn. 39 (12) (2006) 2383–2392.

[52] E. Bonilla-Huerta, B. Duval, J.C.H. Hernández, J.-K. Hao, R. Morales-Caporal, Hybrid filter-wrapper with a specialized random multi-parent crossover operator for gene selection and classification problems, International Conference on Intelligent Computing, Springer, 2011, pp. 453–461.