Knowledge-Based Systems 223 (2021) 107034

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Gene selection for microarray data classification based on Gray Wolf Optimizer enhanced with TRIZ-inspired operators

Osama Ahmad Alomari [a,*], Sharif Naser Makhadmeh [b,c,e], Mohammed Azmi Al-Betar [c,d], Zaid Abdi Alkareem Alyasseri [e,f], Iyad Abu Doush [g,h], Ammar Kamal Abasi [c,i], Mohammed A. Awadallah [j,c], Raed Abu Zitar [k]

[a] Department of Computer Engineering, Faculty of Engineering and Architecture, Istanbul Gelisim University, Istanbul, Turkey
[b] Faculty of Information Technology, Middle East University, Amman, Jordan
[c] Artificial Intelligence Research Center (AIRC), Ajman University, Ajman, United Arab Emirates
[d] Department of Information Technology, Al-Huson University College, Al-Balqa Applied University, P.O. Box 50, Al-Huson, Irbid, Jordan
[e] Center for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
[f] ECE Department-Faculty of Engineering, University of Kufa, P.O. Box 21, Najaf, Iraq
[g] Computing Department, American University of Kuwait, Salmiya, Kuwait
[h] Computer Science Department, Yarmouk University, Irbid, Jordan
[i] School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang, Malaysia
[j] Department of Computer Science, Al-Aqsa University, P.O. Box 4051, Gaza, Palestine
[k] Sorbonne University Center of Artificial Intelligence, Sorbonne University-Abu Dhabi, Abu Dhabi, United Arab Emirates

ARTICLE INFO

Article history:
Received 4 October 2020
Received in revised form 11 March 2021
Accepted 7 April 2021
Available online 15 April 2021

Keywords:
Gray Wolf Optimizer
Gene selection
Optimization
TRIZ
rMRMR
SVM
Classification

ABSTRACT

DNA microarray technology is the fabrication of a single chip to contain a thousand genetic codes. Each microarray experiment can analyze many thousands of genes in parallel. The outcomes of the DNA microarray is a table/matrix, called gene expression data. Pattern recognition algorithms are widely applied to gene expression data to differentiate between health and cancerous patient samples. However, gene expression data is characterized as a high dimensional data that typically encompassed of redundant, noisy, and irrelevant genes. Datasets with such characteristics pose a challenge to machine learning algorithms. This is because they impede the training and testing process and entail high resource computations that deteriorate the classification performance. In order to avoid these pitfalls, gene selection is needed. This paper proposes a new hybrid filter-wrapper approach using robust Minimum Redundancy Maximum Relevancy (rMRMR) as a filter approach to choose the top-ranked genes. Modified Gray Wolf Optimizer (MGWO) is used as a wrapper approach to seek further small sets of genes. In MGWO, new optimization operators inspired by the TRIZ-inventive solution are coupled with the original GWO to increase the diversity of the population. To evaluate the performance of the proposed method, nine well-known microarray datasets are tested. The support vector machine (SVM) is employed for the classification task to estimate the goodness of the selected subset of genes. The effectiveness of TRIZ optimization operators in MGWO is evaluated by investigating the convergence behavior of GWO with and without TRIZ optimization operators. Moreover, the results of MGWO are compared with seven state-of-art gene selection methods using the same datasets based on classification accuracy and the number of selected genes. The results show that the proposed method achieves the best results in four out of nine datasets and it obtains remarkable results on the remaining datasets. The experimental results demonstrated the effectiveness of the proposed method in searching the gene search space and it was able to find the best gene combinations.

1. Introduction

Feature selection is selecting/combining features to reduce the amount of data processing, and is considered an important

* Corresponding author.
E-mail addresses: oalomari@gelisim.edu.tr (O.A. Alomari), s.makhadmeh@ajman.ac.ae (S.N. Makhadmeh), mohbetar@bau.edu.jo (M.A. Al-Betar), zaid.alyasseri@uokufa.edu.iq (Z.A.A. Alyasseri), idoush@auk.edu.kw (I.A. Doush), ammar_abasi@student.usm.my (A.K. Abasi), ma.awadallah@alaqsa.edu.ps (M.A. Awadallah), raed.zitar@sorbonne.ae (R.A. Zitar).

https://doi.org/10.1016/j.knosys.2021.107034
0950-7051/

step when solving classification problems [1–3]. DNA microarray is a molecular analysis technology that allows us to study and analyze more than a thousand of genes in one experiment obtained from a large number of cells and tissues. The development of DNA microarray leads to generate high-dimensional data in fields such as clinical diagnosis, drug discovery, etc [4]. Gene expression data is high-dimensional data produced by DNA microarray experiments. It is widely used for the classification and detection of cancer diseases [5,6]. However, this data contains a large number of redundant, irrelevant, and noisy genes, which poses a challenge to the machine learning algorithm because the construction of a predictive model based on these unrelated genes leads to deterioration in classification performance. The most common way to address this challenge is by using gene selection. Gene selection is a pruning procedure that eliminates redundant and irrelevant genes and maintains the most relevant and meaningful ones [7]. Further insights can be gained by gene selection [5] such as (1) helping biologists researchers to figure out the molecular mechanism related to gene expression of cancer diseases; (2) a potential way of creating a new therapy through extensive analysis of the patterns of the selected genes, and (3) decreasing clinical cost. Generally, gene selection methods mainly fall into two categories [7]: filter-based approaches and wrapper-based approaches. Filter approaches evaluate the genes in short computation time because it performs filtering based on the inartistic properties of the training dataset without involving machine learning algorithms in the evaluation process. Examples of well-known filter approaches are Kullback–Leibler [8], Chi-square [9], ReliefF [10], Minimum Redundancy Maximum Relevancy (MRMR) [11], and Robust MRMR (rMRMR) [12].

Wrapper approaches formulate the gene selection as an optimization search problem [13–19], where candidate gene subsets produced by a search technique or machine learning algorithm are employed to evaluate these candidates. Wrapper approaches produce better classification accuracy than filter approaches but it is more expensive in computation time. Hybrid approaches (filter-wrapper) attract more attention from the researchers [20–23] especially in high dimensional data such as microarray data because it can leverage the benefits of both filter and wrapper approaches and it is proven to be more effective for microarray data classification.

However, having an effective and sophisticated hybrid gene selection approach still needs further investigations [20,21]. One of the challenges facing gene selection methods is the fact that the number of possible solutions grows exponentially when the number of genes increases. Therefore, many researchers put efforts to identify and select the near-optimal candidate gene subsets by modifying existing metaheuristic approaches. A metaheuristic technique is a problem independent algorithmic framework that embeds a set of strategies to develop optimization search algorithms [24].

Many metaheuristic algorithms are adapted or modified to yield better results for gene selection problems. Examples of the gene or feature selection methods using metaheuristic algorithms are Correlation-based Feature Selection with improved-Binary Particle Swarm Optimization [20], Harmony search with a Markov blanket (HSA-MB) [25], binary flower pollination algorithm (FPA) with β-hill climbing (called FPAβ-hc) [26], and rMRMR approach with modified BA algorithm (rMRMR-MBA) [6]. In addition to Binary JAYA Algorithm with Adaptive Mutation (BJAM) [27], and multi-objective salp Swarm Algorithm with dynamic locality (MODSSA) [14]. However, since the gene selection problem search space is comprised of a large set of permutations and the complex interaction between the genes, the above-mentioned gene selection methods are easily stuck in local optima.

The gray wolf optimizer (GWO) is a recent swarm-based intelligence that formulated according to the social hierarchy of gray packs and the hunting procedure [28]. It involves three main phases, the first phase is seeking for the prey, which represents the exploratory search mode. The second and third phases are encircling and attacking the prey, which represents the exploitative search mode. GWO has several merits over other swarm-based intelligence in literature as it is simple in adaptation, easy to use by the naive optimizer, parameter-free, and sound-and-complete. Therefore, GWO gains a tremendous interest from a wide variety of research communities as reported in [29]. To put a few examples of this application such as machine learning [30], networking [31], Security [32] , image processing [33] scheduling [34,35], engineering [36–38], and bioinformatics [39,40].

TRIZ is the abbreviation of Teoriya Resheniya Izobretatelskikh Zadatch, also well known as the theory of inventive problem-solving. This theory, proposed by Genrich Altshuller in 1985 [41], emerged from a deep analysis of one million patent records. The methodology of TRIZ inventive problem encompasses worsening and improving features, and inventive principles that are regarded as a set of instructions for TRIZ researchers in solving design problems. In practice, these principals are formulated as optimization operators and used to increase the diversity of the population that leads to enhance the metaheuristic algorithms performance in solving optimization problems in different areas [42–44]. In TRIZ principles, there are three main operations used to support metaheuristic convergence behavior: dynamization, segmentation, and local quality. The main role of dynamization and segmentation is to divide the solutions into multi groups and identify the length of each group. In each local quality, three operators will be used to update the solution, which are mutation, 2-opt, and swap operators. This is to scan effectively the interaction between the genes and increase the diversity of the evolved solutions to avoid the stagnation of local optima problem. Quite recently, the gene selection problem is solved by Triz-based bat-inspired algorithm [6]. The principle of Triz with three optimization operators inspired by Triz inventive solution has been employed in the main improvement loop of bat algorithm. The main function of these three operators is to diversify the current bat solution updated by original bat operators. This current solution is entered to the Triz optimization operators to implement the split, mutation, 2-opt, and swap operations every iteration. The results showed the impact of the Triz operators on the convergence behavior of bat algorithm for gene selection problem.

This research proposes a hybrid filter/wrapper gene selection method based on rMRMR to serve as a filter-based approach. The modified Gray wolf optimizer algorithm (MGWO) is used to serve as a wrapper-based approach, it is called rMRMR-MGWO. In rMRMR-MGWO, the TRIZ principles are incorporated within MGWO to maintain its diversity during the search of the promising genes from the search space. The main function of these Triz operators is to iteratively improve the diversity of the whole updated GWO population at the end of each improvement loop of GWO. This can be considered as a hybrid mechanism with inventive operators. These concepts are formulated as optimization operators, where each solution in the population is evolved using the three TRIZ principles: dynamization, segmentation, and local quality. The proposed method is evaluated using Microarray data including nine datasets of different sizes and characteristics. The results of the proposed algorithm are measured based on the number of genes and classification accuracy. The results produced by rMRMR-GWO and rMRMR-MGWO are compared to show the impact of the TRIZ operations on the convergence behavior of rMRMR-GWO. Finally, the proposed rMRMR-MGWO is compared against seven state-of-the-art methods using the same Microarray datasets. The comparative evaluation proves the viability and

efficiency of the TRIZ principle when used in GWO. The proposed method is able to outperform the state-of-the-art methods in four out of nine Microarray datasets. In sum, the proposed approach is a very efficient addition to the classification problem with huge amounts of data and can be applied to other optimization methods to tackle similar problems in different domains.

The remaining part of this paper consist of the following sections: Section 2 presents the related work, Section 3 elaborates on the research background about GWO and Triz concepts and principles. The proposed method which describes how the GWO is Incorporated with TRIZ operations for gene selection is presented in Section 4. The evaluation of the proposed method is conducted in the experiment and result section (Section 5). The paper is concluded and future work is recommended in Section 6.

2. Related works

Recently many wrapper and hybrid filter-wrapper approaches are proposed to solve gene selection in order to improve the quality of cancer classification. Most of the aforementioned approaches depend on Metaheuristic algorithms due to its excellent performance in the selection of the most promising and informative gene subsets for cancer microarray data. In Hu et al. [45], the shuffled frog leaping algorithm (ISFLA) is enhanced by implementing an absolute balance group strategy and updating the strategy of chaos memory weight factor (CMWF). ISFLA is a wrapper-based feature selection approach that enhanced the SFLA search capabilities. High-dimensional biomedical datasets have been used to demonstrate that ISFLA can achieve remarkable quality. In particular, for Lung-Cancer-Ontario and Nervous-System datasets, ISFLA achieved an average accuracy of 75.06% and 81.67% with 14.33 and 32.33 features, respectively. Yan et al. [46] implemented an improved binary clonal flower pollination algorithm (IBCFPA) to improve the search capability of original BCFPA and achieved a higher accuracy of classification by integrating absolute balance group strategy and Gaussian Mutation. The superior capacity of IBCFPA was confirmed by the experimental results of six publicly biological datasets. Yan et al. [47] implemented the Coral Reefs Optimization algorithm with the Simulated Annealing and Tournament Selection (BCROSAT) to solve the feature selection problem. The proposed algorithm is applied for Nervous System datasets and Colon Tumor datasets and it achieved an average accuracy of 82.00% and 92.31% with 21.4 and 20.5 features, respectively. Preeja et al. [48] propose a Binary Krill Herd algorithm (BKH). The efficiency of the BKH is measured using time and accuracy when it selects features using one dataset from UCI datasets. For example, the BKH algorithm achieved 94% classification accuracy with ten features, while the FAST algorithm achieved only 56% classification accuracy using the same dataset and with the same number of features. Moreover, hybrid filter-wrapper approaches have attracted much attention from the researchers in solving the gene selection problem. These approaches can leverage the benefits of both filter and wrapper approaches. In practice, initially, the filter is running on the experimented data and start ranking the genes according to its own metrics. Thereafter, the highly ranked genes are passed to the wrapper approach to seek for further informative genes. Yang et al. [49] proposed a gene selection method for microarray data classification based on Information gain (IG) as a filter approach and genetic algorithm (GA) as a wrapper approach, namely (IG–GA). This hybrid approach starts its gene selection procedure by applying IG on the experimented data and measuring each gene relevancy score with a class label. The gene is sorted based on their relevancy and the top-ranked genes are subject to further pruning procedure driven by the GA, where the GA generates candidate gene subsets and evaluates them using KNN machine learning algorithm. The results demonstrate that IG–GA can produce promising results on some microarray data, such as brain-tumor. Alomari et al. [22] proposed MRMR-BA gene selection method for microarray data classification, which combined MRMR as a filtering approach and Bat algorithm as a wrapper approach. Experimental results show that the MRMR-BA is able to provide promising solutions for microarray data. Sahu [50] introduced an approach for gene selection problem using IG and Improved particle swarm optimization, called (IG-IPSO). Experimental results demonstrated the superiority of IG-IPSO in minimizing the number of genes and simultaneously maximizing the classification accuracy. For instance, IG-IPSO managed to achieve 98.37% classification accuracy with 34 genes on Leukemia1 dataset. Zhang et al. [51] introduced a novel hybrid strategy with the IG as the filter phase and the improved binary krill herd (MBKH) as a wrapper phase. The modifications applied to the MBHA led to deeper search and effective exploration of the gene search space. The results reveal that the proposed method can yield improvement in the convergence rate, classification accuracy, and a number of selected genes comparing to the BKH and other several new gene selection algorithms.

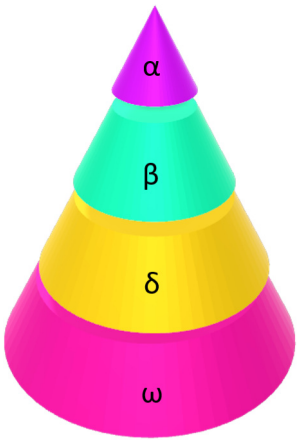

Fig. 1. Gray wolves domination levels.

3. Research background

3.1. Gray Wolf Optimizer (GWO)

GWO is a swarm-based optimization algorithm that mimics the lifestyle of gray wolves in nature, particularly in hunting and social leadership mechanism. GWO algorithm was proposed and mathematically formulated by Mirjalili in 2014 [28].

3.1.1. Gray wolves domination and inspiration

The pack of gray wolves contains four types of wolve members, including alpha (α), beta (β), delta (δ), and omega (ω) wolfs. These members are distributed based on their domination levels, where α is at the highest level and ω is at the lowest level of the wolves pack, as shown in Fig. 1.

The α wolf is the wisest in the pack, where it has high experiences in managing and making decisions for the pack, such as controlling the hunting mechanism and choosing a habitat. The β wolves are at the second domination level of the hierarchy. The β wolves usually stand behind the α wolf to support him in managing and controlling the pack. The third domination level of the hierarchy contains the δ wolves. The wolves belonging to this level are in charge of helping, supporting, and guarding the elderly and weak wolves. The ω wolves are the rest of the wolves in the pack.

Table 1
A list of 40 inventive principles extracted from TRIZ inventive solution.

1. Segmentation	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. Another Dimension	14. Mechanical Vibration	15. Periodic Action	16. Continuity of Useful Action
17. "The Other way round"	18. Spheroidality-Curvature	19. "Dynamisation"	20. Partial or Excessive Actions
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

The gray wolves hierarchy is the basis of their life style in the pack, where it manages their transactions, hunting mechanisms, and daily life. The primary benefit of this hierarchy is leading the hunting of prey. Once the prey found, the α wolf commands the others to encircle it and leads β and δ wolves to attack it.

3.2. Gray wolf optimizer algorithm

In GWO, the same inspiration and social hierarchy of gray wolves are used along with their domination levels. Each wolf is considered as a candidate solution for a particular optimization problem. Each level of the first three domination levels (i.e., α, β, and δ levels) contains only one solution, where the α level contains the best solution, and β, and δ levels contain the second and third best solutions, respectively. The rest of the solutions are placed in the ω level. The wolves in ω level must assist the wolves in α, β, and δ levels by encircling them using the formulation below.

$$\overrightarrow{D} = |\overrightarrow{C} \times \overrightarrow{X}_p(t) - \overrightarrow{X}(t)|, \tag{1}$$

$$\overrightarrow{X}(t+1) = \overrightarrow{X}_p(t) - \overrightarrow{A} \times \overrightarrow{D}, \tag{2}$$

$$\overrightarrow{A} = 2 \times \overrightarrow{a} \times \overrightarrow{r}_1 - \overrightarrow{a}, \tag{3}$$

$$\overrightarrow{C} = 2 \times \overrightarrow{r}_2, \tag{4}$$

$$a = 2 - t \times \frac{2}{I}, \tag{5}$$

where $\overrightarrow{X}_p(t)$ is the prey position at tth iteration, $\overrightarrow{X}(t)$ are the wolves position at tth and $(t+1)^{th}$ iterations, respectively, $\overrightarrow{r}_1$ and $\overrightarrow{r}_2$ are two random vectors, $\overrightarrow{A}$ and $\overrightarrow{C}$ are two coefficient vectors, and I is the maximum number of iterations. The primary goal of $\overrightarrow{C}$ and $\overrightarrow{A}$ is to avoid local optima stagnation and balance between exploration and exploitation, respectively. GWO is able to avoid stagnation in local optima by changing the value of $\overrightarrow{C}$ randomly, and also able to exploit and explore a particular search space if $|\overrightarrow{A}| < 1$ and $|\overrightarrow{A}| > 1$, respectively.

The solutions in ω level should be updated at each iteration based on the solution in α, β, and δ levels using the formulation below.

$$\overrightarrow{D}_\alpha = |\overrightarrow{C}_1 \times \overrightarrow{X}_\alpha - \overrightarrow{X}|, \tag{6}$$

$$\overrightarrow{D}_\beta = |\overrightarrow{C}_2 \times \overrightarrow{X}_\beta - \overrightarrow{X}|, \tag{7}$$

$$\overrightarrow{D}_\delta = |\overrightarrow{C}_3 \times \overrightarrow{X}_\delta - \overrightarrow{X}|, \tag{8}$$

$$\overrightarrow{X}_1 = \overrightarrow{X}_\alpha - \overrightarrow{A}_1 \times \overrightarrow{D}_\alpha, \tag{9}$$

$$\overrightarrow{X}_2 = \overrightarrow{X}_\beta - \overrightarrow{A}_2 \times \overrightarrow{D}_\beta, \tag{10}$$

$$\overrightarrow{X}_3 = \overrightarrow{X}_\delta - \overrightarrow{A}_3 \times \overrightarrow{D}_\delta, \tag{11}$$

$$\overrightarrow{X}(t+1) = \frac{\overrightarrow{X}_1 + \overrightarrow{X}_2 + \overrightarrow{X}_3}{3}, \tag{12}$$

3.3. Theory of Inventive Problem Solving (TRIZ)

Theory of Inventive Problem Solving (TRIZ) is a technique that relies on identifying and coding how creative innovations are processed to make them more predictable. Using this principle, a TRIZ technical contradictions matrix is developed with 39 features for improving and declining features, and 40 inventive principles are to make creativity an *exact science* [43]. This list is presented in Table 1.

The TRIZ principle is utilized to address problems in different domains such as software development [52], service quality [53], export packing of Persian Lime [54], and engineering [44,55,56]. In addition, several metaheuristic algorithms use principles of TRIZ to tackle different types of problems.

Duran-Novoa et al. [43] proposes an evolutionary algorithm (EA) that build solutions using TRIZ-inventive concept to solve inventive problems based on dialectical negation. The method presents new dialectical operators that stimulates the principles of TRIZ. The proposed method demonstrates the capability of using TRIZ toward better convergence.

Mei et al. [42] present a bees algorithm that is hybridized with TRIZ-based optimization operators to improve the steps of solving the problem of task sequencing of moving-board-with-time-delay (MBTD) assembly machine. The TRIZ pillars of Dynamisation, Segmentation, and Local Quality are developed to improve the search of bees algorithm, specifically after rewarding bees for picking the sites. The results show that this algorithm outperforms other comparable methods, including the original bees algorithm.

In another work Mei et al. [44] proposes updating the bees algorithm based on TRIZ principles by embedding new operators to solve the problem of the assembly of printed circuit boards (PCBs) using a machine of the moving-board-with-time-delay (MBTD) type. The proposed algorithm outperforms other comparable methods from the literature.

Recently, Al-Betar et al. [6] tackle the problem of gene selection. The proposed method used robust Minimum Redundancy Maximum Relevancy (rMRMR) as a filter to prune the genes and choose the favorable genes based on the interaction between them, along with a modified Bat Algorithm (MBA) as a wrapper technique to search for a small group of distinct genes. The method is named rMRMR-MBA. In MBA, new optimization operators based on TRIZ inventive solution are coupled with MBA to empower its searching process, and effectively explore the interaction between genes. The evaluation outcomes prove that rMRMR-MBA outperforms other ten comparative methods according to the classification accuracy and the number of selected genes in two out of ten datasets. It also generates competitive results for the remaining datasets.

4. Proposed method for gene selection

The blueprint of the proposed gene selection method lies on two consecutive processes: (i) filtering process where the rMRMR is utilized to reduce the gene search space dimension and (ii) wrapper process where the Triz-inspired concepts are

incorporated with GWO coupled with SVM classifier to improve the final classification accuracy as well as minimize the number of Representative genes. The gene selection representation and the objective function are given in Sections 4.2 and 4.3, respectively.

The main motivation behind the hybridization of filter-wrapper is based on the success attributes of each method. To elaborate, the idea behind the utilizing rMRMR filtering approach in the proposed method is due to the fact that the rMRMR revealed the best filtering performance in comparison other filtering-based method in same dataset as studied in [12]. The wrapper approach is normally used optimization method coupled with efficient classifier for gene selection. As mentioned earlier, GWO is the most successful swarm-based optimization algorithm over other optimization methods due to its impressive advantages such as it is easy-to-use, simple in concepts, derivative free, flexible and scalable, robust, and sound-and-complete. Therefore, GWO has been growth exponentially over the years [29]. In spite of GWO advantages, there is a chronic premature convergence dilemma strike on its performance due to the lack of exploration power. Therefore, Triz-inspired optimization operator come to the fore to handle such issue. Due to its robustness classification performance in highly dimension datasets, SVM is almost used as a classifier for all methods used for microarray data [57–59].

The rMRMR-MGWO steps are thoroughly discussed subsequently. The flowchart of rMRMR-MGWO is provided in Fig. 2.

4.1. Robust minimum redundancy maximum relevancy

Robust minimum redundancy maximum relevancy is an improved version of MRMR [12], known as rMRMR. rMRMR is multivariate filter algorithm ranks gene subset by measuring gene-to-gene redundancy and gene-to-class relevancy. The main target of rMRMR filter algorithm is to provide a distinguished subset of genes that embeds insights and robust genes correlation able to classify class categories accurately. This is to overcome the computational barriers induced by using high dimensional data in classification. The relevancy and redundancy calculations in MRMR only rely on mutual information. However, rMRMR is different than basic MRMR in the metrics used for calculating the gene-to-target relevancy, as it involves ensembling of filters from various characteristics (information theory, distance, probability distribution, etc.) in calculating gene-to-target relevancy score. Regarding the redundancy computations, rMRMR follows the same computations of basic MRMR. The main purpose of the modification applied in relevancy computation is to subject the gene relevancy score to various filter methods from different characteristics. The reason is to overcome the high variability in the classification performance induced by applying single filter approach since it only relies on one metric, which makes it sensitive to the characteristics of the experimented dataset. In other words, rMRMR is designed to promote the robustness and the stability of MRMR. The main procedure of rMRMR is discussed subsequently. Moreover, the pseudo-code of rMRMR is provided in Algorithm 1.

Step1: ***Initialization***.

In the initialization step, three well-regarded filter algorithms are selected to form the ensemble method which are ReliefF, Chi-Square and Kullback–Liebler. In practical implementation, the three filter algorithms are carried out independently, they evaluate all genes in each experimented dataset based on its discriminative power. For each experimented dataset, the genes scores estimated by each single filter are aggregated into one ranked gene list through "Mean of the scores" (lines 6 to 12 in Algorithm 1).

Step 2: ***Hybridization***.

The hybridization in this step is done by combining the gene ranking list obtained from the previous step with relevancy computations in MRMR, as follows. Firstly, iteratively for each gene in the experimented dataset, two relevancy scores will be calculated based on mutual information (i.e, the current gene and the target) and the mean score for the same gene in the gene ranking list (lines 16 to 18 in Algorithm 1). Secondly, a new factor (i.e, gene mean score) has been added to the relevancy computation, where the final gene score is estimated by multiplying gene relevancy score $I(G_x, c)$ with mean gene score $R(G_i)$ (lines 26 in Algorithm 1). The main purpose of this hybridization is to avoid bias results of single-based results and to introduce diversity in the measurements used for calculating the gene relevancy score which eventually leads to improve the robustness and stability of MRMR.

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

Eventually, according to the predefined threshold for the number of selected genes, the filtered gene subset will be passed to the wrapper approach to find a further smaller set of high relevancy and informative genes.

Algorithm 1 Hybrid MRMR with ensemble of filter methods

1: **Input:**
2: D(G_1, G_2,....,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 ∈ [Chi-Square, ReliefF, Kullback–Liebler]
Ensemble of filters
6: **for** $i \in \{1, \ldots, k\}$ **do**
7: **for** $j \in \{1, \ldots, m\}$ **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 $F_i(G_j), j = 1, \ldots..., m$
11: **end for**
12: Create gene ranking list **R** by combining k different filters $F_i(G_j), i = 1, \ldots., k$ ∀j with arithmetic mean.
MRMR
13: $S_{ALL} \leftarrow$ 1,2, ... G
14: $S \leftarrow \phi$
15: $S_a \leftarrow \phi$
16: **for each** $G \in \{1, \ldots, m\}$ **do**
17: $I(G_i, c) = R(G_i)$
18: **end for**
19: $MAX = Min(Length(S_{ALL}), 1000)$ //*length*(S_{ALL}) *return the number of elements in* S_{ALL}
20: **for** $i = 1$ *to MAX* **do**
21: $S_a \leftarrow S_a \cup \underset{i \in S_{ALL} \setminus Sa}{argmax}\ I(G_i, c)$
22: **end for**
23: $S \leftarrow \underset{i \in S_{ALL}}{argmax}\ I(G_i, c)$
24: **while** $Length(S) < n$ **do**
25: **for each** $x \in S_a \setminus S$ **do**
26: $Relv(G_x) = I(G_x, c) * R(G_i)$ // **Enhance Relevancy computing**
27: $Red(G_x) = \frac{1}{Length(S)} \sum_{y \in S} I(G_x, G_y)$
28: **end for**
29: $S \leftarrow S \cup \arg\max_x (Relv(G_x) - Red(G_x))$
30: **end while**
Output
31: The resulting gene subset is fed to subsequent stage (i.e, wrapper approach).

4.2. Solution representation

The gene selection is classified as a combinatorial optimization problem. The solution representation in this problem is comprised of candidate gene subset [60,61]. The navigation process in gene search space for finding a desired and acceptable candidate gene subset becomes more challenging by increasing the number

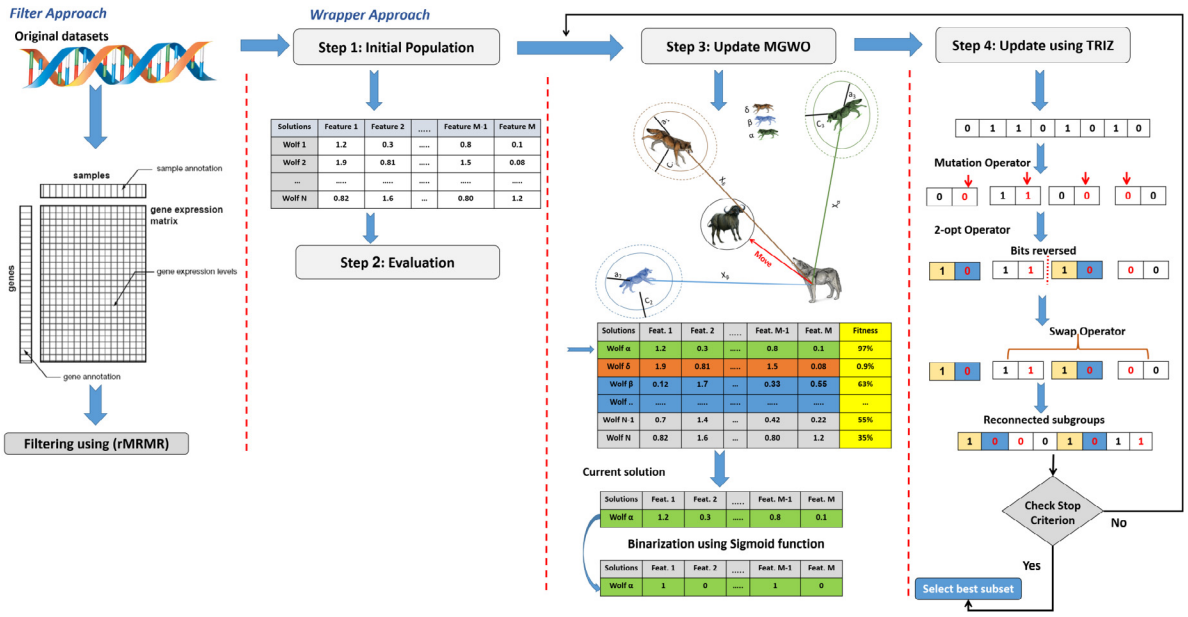

Fig. 2. Flowchart of rMRMR-MGWO method.

of selected genes in the experimented data. In order to formulate the problem mathematically, the number of genes is (symbolized as N), and the number of possible candidate gene subsets is (symbolized as $[2^N]$). Each candidate solution is symbolized as $\boldsymbol{x}$, and it is composed of binary strings $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$, where the N represents the length of a string or the size of gene subset. Moreover, in the binary string, the bit '1' implies that the gene is selected, while '0' implied that the gene is ignored.

4.3. Fitness function

The main purpose of gene selection methods is to reduce the number of genes and simultaneously boost the classification accuracy. To achieve this purpose, many studies combined classification accuracy and the size of gene subset into a single weighted function and use it as fitness function to assess each candidate gene subset as in Eq. (13) below:

$$fitness = \alpha \times acc(classifier) + \beta \times (1 - \frac{s}{p}) \tag{13}$$

where p stands for the number of genes in the experimented dataset, and s stands for the size of the candidate gene subset. The value of the two weighting factors related to classification accuracy (α) and the size of the candidate gene subset (β) are 1 and 0.001, respectively [45,51]. In this study, the classification accuracy is estimated by performing 10-fold cross-validation with SVM classifier. Note that the 10-fold cross-validation is widely used in the gene selection domain to validate the classification due to its consistency, and less results variability with regard to input data [62]. Note that the almost all comparative methods used k-fold CV for validation purpose.

4.4. Implementation process of the proposed rMRMR-MGWO method

A new gene selection method rMRMR-MGWO is introduced in this section. In this method, rMRMR is played a vital role by pruning the original genes in the cancer microarray data and selected the most biologically relevant genes. In each experimented data, the genes by rMRMR are used as inputs for further gene selection optimization process driven by MGWO. MGWO is composed of four main steps, which will be thoroughly discussed below. The flow chart and pseudocode of rMRMR-MGWO are illustrated in Fig. 2 and Algorithm 3.

Step 1: Initialization.

In this step, the number of iterations (Max_{itr}) is initialized, and each wolf is denoted as a solution of the gene selection problem, where each solution is a binary vector of length D, as shown in Eq. (14). In other words, the decision variable in the solution is only accepting 0 or 1, called position in GWO. GWO searching processes are lunched by generating n wolves as random binary vectors.

$$\boldsymbol{GWOP} = \begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_D^1 \\ x_1^2 & x_2^2 & \cdots & x_D^2 \\ \vdots & \vdots & \cdots & \vdots \\ x_1^n & x_2^n & \cdots & x_D^n \end{bmatrix} \tag{14}$$

subject to:

$x \in \{0, 1\}$

Step2: Evaluation.

In this step, the wolves are mainly evaluated according to their positions. For example, if the current wolf has 5 positions equal to 1 and the remaining positions are equal to 0 then the biological genes data mapped to the 1's positions (i.e., selected genes) will be fetched and represented as a new reduced data. Afterward, 10-fold-cross-validation will only be applied using SVM on this reduced data. Practically, the reduced data will be divided into 10 folds, where the first fold is testing data, and the other 9 folds are represented as the training data. An SVM model is constructed based on the training data and the model is used on evaluating the testing data. This process is repeated 10 times and on each time classification accuracy is estimated and a different fold is assigned to the subsequent evaluation. Eventually, the average classification accuracy is computed over the ten testing folds. In the GWO, the evaluation of each wolf is mainly depending on average classification accuracy and the size of the gene subset as introduced in Eq. (13). The fittest solution is assigned to X_α, and the second and third best solutions are assigned to X_β, and X_δ, respectively. Algorithm 2 exhibits the pseudo-code for the social hierarchy of the proposed MGWO.

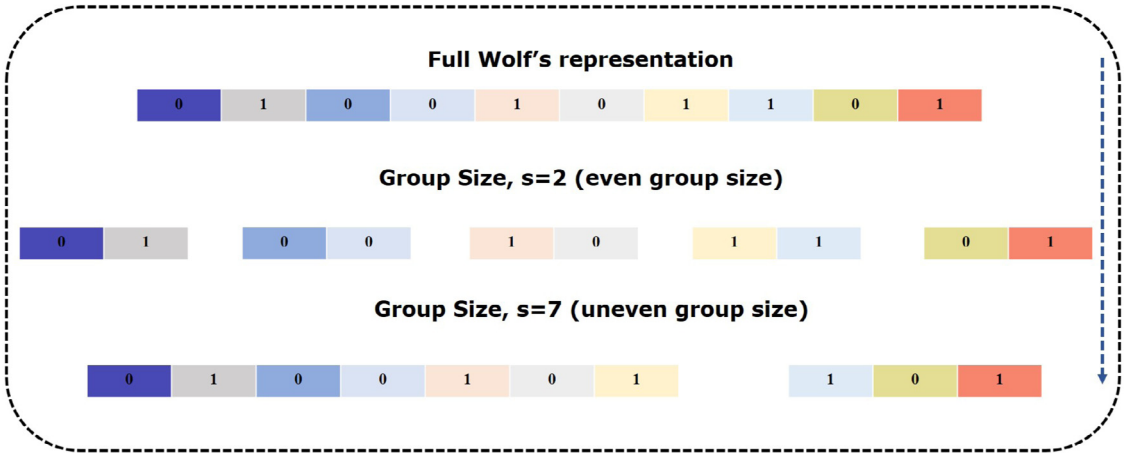

Fig. 3. The segmentation operator.

Algorithm 2 Pseudo code for the social hierarchy of the proposed MGWO

```
while (itr < Max_itr) do
  for each solution (j) do
    Calculate the fitness of the solution
    X_α = the fittest solution
    X_β = the second best solution
    X_δ = the third best solution
  end for
  t = t + 1
end while
Return X_α, X_β, X_δ;
```

Step3: Update MGWO population

In this step, MGWO follows the same operations as that of the original GWO; i.e. encircling prey, searching for prey (exploration), attacking prey (exploitation), and hunting mechanism. This is implemented for navigating the gene search space and updating the MGWO population, as discussed in Section 3.1. Updating the solutions in the MGWO population can be done through Equations from (1) to (12). Conventionally, the updating mechanism is operated by measuring the distance between each wolf/solution in the population and social hierarchy-based solutions (i.e., X_α, X_β, and X_δ). The current wolf/solution updates its positions/decision variables according to $\overrightarrow{X_\alpha}$, $\overrightarrow{X_\beta}$, and $\overrightarrow{X_\delta}$. Consequently, a new solution is generated ($\overrightarrow{X_1}$) on the basis of $\overrightarrow{X_\alpha}$ using Eqs. (6), (3), (4), and (9). These steps are repeated to generate two new solutions $\overrightarrow{X_2}$ and $\overrightarrow{X_3}$, where $\overrightarrow{X_2}$ is generated based on $\overrightarrow{X_\beta}$ using Eqs. (7), (3), (4), and (10), and X_3 is generated based on X_δ using Eqs. (8), (3), (4), and (11). Eventually, the solutions $\overrightarrow{X_1}$, $\overrightarrow{X_2}$, and $\overrightarrow{X_3}$ are aggregated using the mean to produce a new solution $\overrightarrow{X}(t+1)$. It should be noted that since the $\overrightarrow{X}(t+1)$ positions have real values, they are transferred as binary vector using Eqs. (15) and (16).

$$sigmoid(\overrightarrow{X}(t+1)) = \frac{1}{1+e^{-\overrightarrow{X}(t+1)}} \tag{15}$$

$$\overrightarrow{X}(t+1) = \begin{cases} 1 & \text{if } sigmoid(\overrightarrow{X}(t+1)) > U(0,1), \\ 0 & Otherwise \end{cases} \tag{16}$$

Moreover, the new solution $\overrightarrow{X(t+1)}$ is produced at each iteration and evaluated using the fitness function as given in Eq. (13). In the subsequent step, the new solution $\overrightarrow{X(t+1)}$ is passed for further optimization of search operators to produce more quality solutions.

Step4: Update GWO population using TRIZ-inspired operators

In this step, new optimization operators inspired by TRIZ inventive solution are incorporated in the basic GWO search process to maintain the population's diversity in order to explore the interactions between the genes effectively for the purpose of accessing and defining the most promising regions in the gene search space. This, in turn, improves the gene selection outcomes. TRIZ is a problem-solving method that is comprised of 40 inventive principles (as shown in Table 1) extracted from analyzing 1 million patents records. Due to its value and significance, this research is motivated to formulate some principles as its optimization search operators. In this work, the descriptions of these principles are investigated to check its feasibility and suitability to be formulated as optimization search operators. As a result, three principles (i.e., dynamization, segmentation, and local quality) are selected and used as extra optimization operators within the GWO search procedure. Based on the MGWO method flowchart, after the current solutions are updated by the original mechanism of GWO (as discussed in the previous step), these solutions will be prone to other updating mechanism driven by dynamization, segmentation, and local quality optimization operators as discussed below.

Dynamization Operator In this operator, there is no alternation to the positions (i.e., decision variables) in the current wolf. However, the solution representation of the gray wolf will be split into smaller groups by the segmentation operator. In the meantime, dynamization plays the role of identifying the number of elements in each group. It should be noted that the number of elements in each individual is generated arbitrarily based on this number and the total number of groups is determined. These settings will be mainly used by the segmentation operator to reorganize the wolf's solution representation.

Segmentation Operator Within this operator, the solution vector of the current wolf is reorganized into groups and the number of elements (N) in each group

is already set by the dynamization operator. In case the length of the solution is dividable by (N) then the all group size is even, otherwise; it is uneven. For further clarification, the process of segmentation operator is illustrated in Fig. 3.

Local Quality Operator In the previous operations, the current wolf position vector is divided into smaller groups without any permutations or alterations applied to the wolf's position vector. However, the local quality operator will be altered using neighborhood search operators including three operators: mutation, 2-opt operation, and swapping to provide diversity in each group. Note that the local quality operator in this research is proceeded into two modes: (1) intra-group mode (changes are applied inside the groups); and (2) inter-group mode (changes are applied in the order of the groups). For Intra-group mode, mutation and 2-opt operations are applied for each group as follows. Initially, the mutation operator generates a random position between 0 and $N-1$, and its value is flipped to "1" in case the original position value is "0" and visa versa. While in the 2-Opt operation, every single group is split into two sub-groups. One of the two sub-groups will be maintained and the other one is selected to be altered by the 2-opt operation. The 2-opt operation is applied to the selected sub-group by reversing the order of the positions' values. For example, "0010101" to "1010100". Afterward, the separated sub-groups are merged to become one group again. Fig. 4 illustrates the entire process of intra-group mode. In the inter-group mode, a swap operator is applied in the solution groups by generating two random group indexes and their contents are swapped in the group sequence. The entire process is applied in inter-group mode and it is presented in Fig. 5.

Step 5: Check the stop criterion The steps 3 and 4 are iteratively carried out until the stop criterion (i.e., the maximum number of iterations) is satisfied. In case the condition is satisfied, the MGWO would have delivered the best fitness wolf for the gene selection problem which carries the most relevant biomarker-genes able to classify cancer diseases accurately. Algorithm 3 presents the pseudo-code of the five steps of the proposed MGWO.

Algorithm 3 Pseudo code of the proposed MGWO

```
Step1: Initialization.
Genes= {g_1, g_2, ..., g_D}
Initialize MGWO parameters (n, Max_itr).
while (itr ≤ Max_itr) do
   for each solution (j) do
      Step2: Evaluation.
      Calculate the fitness of the solution
      X_α = the fittest solution
      X_β = the second best solution
      X_δ = the third best solution
   end for
   Step3: Update MGWO population
   for each solution (j) do
      Update r_1, r_2
      Update the value of A_1 (Eq. (3))
      Update the value of C_1 (Eq. (4))
      Calculate X_1 (Eqs. (6), (9))
      Update r_1, r_2
      Update the value of A_2 (Eq. (3))
      Update the value of C_2 (Eq. (4))
      Calculate X_2 (Eqs. (7), (10))
      Update r_1, r_2
      Update the value of A_3 (Eq. (3))
      Update the value of C_3 (Eq. (4))
      Calculate X_3 (Eqs. (8), (11))
      Generate a new solution X(t + 1) (Eq. (12))
      Transfer X(t + 1) to binary vector using sigmoid function (Eq. (15))
   end for
   Start Running TRIZ inspired optimization operators
   x'' = Split (X(t + 1))
   x''' = Mutation(x'')
   x'''' = 2-Opt(x''')
   x' = Swap(x'''')
   End Running TRIZ inspired optimization operators
   Step 5: Check the stop criterion
   if The maximum number of the iteration is not reached then
      itr = itr + 1
   end if
end while
Return X_α
```

Table 2
Datasets characteristic.

Datasets	# Genes	# Samples	# Classes
Colon Tumor	2000	62	2
CNS	7129	60	2
ALL-AML	7129	72	2
Ovarian Cancer	15154	253	2
Lung Cancer	12601	203	5
ALL-AML-3C	7129	72	3
ALL-AML-4C	7129	72	4
MLL	12582	72	3
SRBCT	2308	83	4

5. Experimental setup and results

The features of the PC used in these experiments are : Intel Core Quad 2,66 GHz CPU with 4 GB of RAM, Win 10 operating system. In our experiments, filter approaches (i.e., ReliefF, Chi-Square, and Kullback–Liebler) were implemented using an open-source machine learning platform, called Weka [63], while rMRMR was implemented based on Matlab. In the wrapper approach, GWO, modified GWO, and SVM were implemented based on Java. The SVM classifier was implemented in the open-source LIBSVM [64].

5.1. Dataset

The MGWO was tested on nine public microarray benchmark datasets. These *de facto* datasets are widely used in pattern recognition applications that are composed of evolutionary algorithms and machine learning to recognize patterns in genes to classify cancer samples from healthy samples [51]. These experimented datasets are different in dimension. The number of genes between 2000 to 15154 including irrelevant or poor prognostic genes. While the number of patient samples is between 60 to 235. Table 2 shows the main characteristics of the experimented datasets in terms of the number of samples, the number of genes, and class categories.

5.2. Parameter settings

The setting of parameters for both filter and wrapper approaches are illustrated in this section. In filter approach, for parameter top-ranked genes, M = 50 has been adopted according to previous studies [5,25,57,65]. In the wrapper approach for MGWO, the number of wolves and the maximum amount of iterations are set to 100 based on experimental evaluation and the related studies in the literature [5,12,22,57]. The other parameters related to the machine learning algorithm are set as follows: 10-fold-CV schema, which has been widely adopted to evaluate the performance of the machine learning algorithm, is

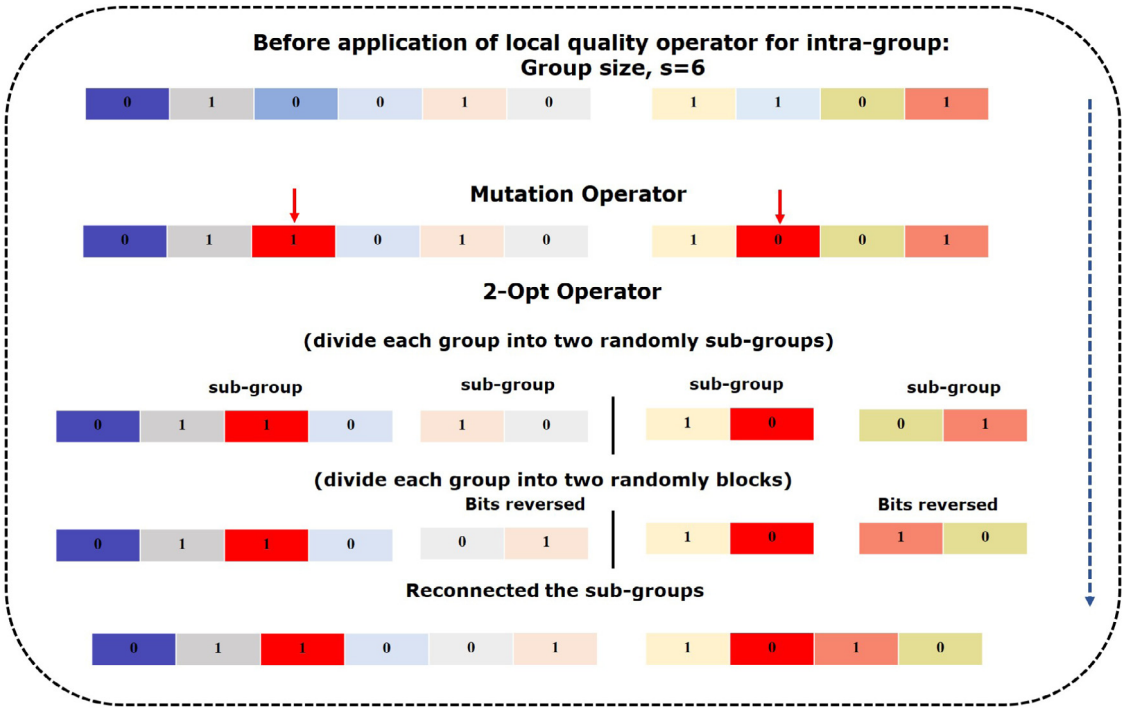

Fig. 4. The local quality operator in the intra-group mode.

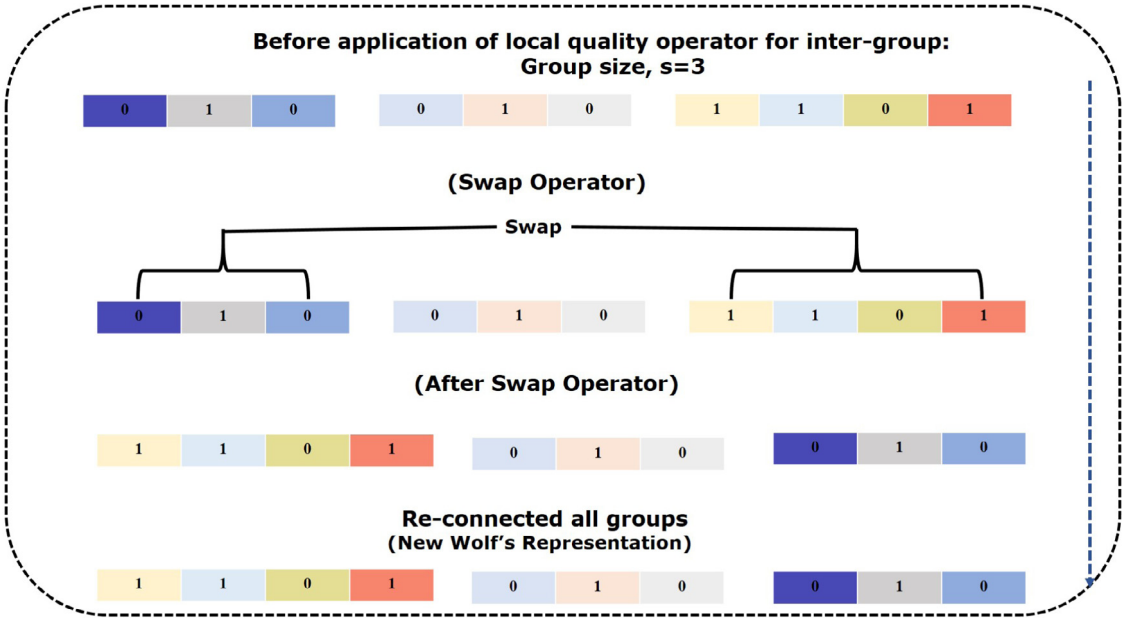

Fig. 5. The local quality operator in the inter-group mode.

used in this research. Moreover, for the SVM kernel and the values of its parameters selection, Radial Basis Function (RBF) kernel is chosen to carry out the classification task, as well as, the grid search is harnessed to optimize the plenty parameter C and the kernel parameter Gamma.

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

The performance of basic GWO and GWO coupled with TRIZ-inspired operators (MGWO) is examined in this section. Since both algorithms are stochastic algorithms, they were performed in 30 independent runs. The results of classification accuracy (ACC), precision, recall, F1score, MCC, number of selected genes ($|\#G|$), and Wilcoxon signed-rank statistical test are exposed in Table 3. Note: the best results of ACC, $|\#G|$ are marked in bold. The interpretation of the results of Wilcoxon signed-rank statistic is that if the probability range of $\alpha \leq 0.05$, the results of MGWO are significantly better than GWO, and the symbol '$*$' is displayed in the corresponding dataset. In case of the probability range α condition is not satisfied, it implies that there is no significant statistical results were found.

From Table 3, it can be observed that MGWO achieves better or similar results on most evaluation measurements (i.e., classification accuracy, Precision, Recall, F1score, and MCC) on all

Table 3
The performance of the proposed Triz-based GWO using different evaluation measures.

	Dataset	Colon	CNS	AL-AML	Ovarian	Lung	All-Aml-3c	All-AML-4c	MLL	SRBCT
rMRMR-MGWO	ACC	0.9586	0.9938	1	1	0.9791	1	0.9990	1	1
	Precision	0.9683	0.9974	1	1	1	1	0.9989	1	1
	Recall	0.9409	0.9873	1	1	1	1	0.9996	1	1
	F1score	0.968	0.9953	1	1	1	1	0.9991	1	1
	MCC	0.9052	0.9757	1	1	1	1	0.9962	1	1
	No_Of_genes	9.8	17.46	5.06	3.56	15.8	6.7	11.36	8.4	12.3
GWO	ACC	0.9413	0.9738	1	1	0.9752	0.9986	0.9884	0.9990	1
	Precision	0.9533	0.9829	1	1	0.9445	0.9982	0.9911	0.9989	1
	Recall	0.9197	0.9571	1	1	0.9882	0.9994	0.9962	0.9995	1
	F1score	0.9546	0.98	1	1	0.9579	0.9977	0.9863	0.999	1
	MCC	0.86397	0.9611	1	1	0.92	0.92	0.93	0.98	1
	No_Of_genes	5.5	16.1	5.5	16.76	4.033	10.06	10.8	11.466	14.53
	T-Sig.	*	–	–	–	*	–	*	*	–

datasets compared to GWO. Conventionally, both methods tend to find fewer number of genes with high classification accuracy. In comparison, MGWO resulted in better classification accuracy with a lower number of selected genes for six datasets (i.e., ALL-AML, Ovarian Cancer, Lung Cancer, ALL-AML-3c, MLL, and SRBCT). In term of number of selected genes, MGWO managed to select fewer number of genes than GWO on six datasets (i.e., AL-AML, Ovarian, Lung, All-Aml-3c, MLL, and SRBCT), while GWO select fewer number of genes compared to MGWO on three datasets (i.e., Colon, CNS, and ALL-AML-4c). Table 3 also demonstrates the results of Wilcoxon signed-rank statistical test between the MGWO and the MGWO. It can be inferred that MGWO achieves statistically significant results on four datasets (i.e., Colon, Lung, All-AML-3c, and MLL).

Moreover, the convergence curve is plotted in Fig. 6 to investigate the search process of GWO and MGWO while optimizing the gene search space. The graph shows that MGWO converges higher (i.e., better) than GWO for almost all datasets. However, for ALL-AML, Ovarian, and SRBCT, GWO and MGWO have similar convergence behavior.

To further evaluate the effectiveness of the proposed method, the results of the proposed methods are compared against embedded gene/feature selection methods, as depicted in Table 4. Please note that the best results are highlighted in bold. The embedded methods involved in this comparison are : LASSO [66], Random Forest [67], Elastic Nets [68], and Decision tree [69]. It can be seen from the table that the proposed method (rMRMR-MGWO) outperform all other embedded gene selection methods, except Elastic Nets. rMRMR-MGWO outperforms Elastic Nets on four datasets (i.e, Colon Tumor, ALL-AML-3c, ALL-AML-4c, and MLL) , while Elastic produces better results than rMRMR-MGWO on two datasets (i.e., CNS and Lung Cancer). Both methods achieve similar result on the remaining datasets.

In sum, the experimental results proved that the MGWO is able to produce the best compromise in the trade-off between the classification accuracy and number of selected genes. This outstanding result is a credit of the incorporation of TRIZ-inspired optimization operators with MGWO that helps in exploring the interactions between the genes effectively by maintaining the diversity of the population, accessing and defining the most promising regions in the gene search space.

According to the study [23], the average number of the selected genes obtained by optimization algorithms cannot be utilized for biological analysis. Because of that, the results of proposed method in multiple independent runs were ordered by the fitness function that involved classification accuracy and number of selected genes, as shown in Table 3, and in previous methods as seen in Table 7. In this study, for getting insight on the biological genes related to the cancer, the obtained genes for each dataset is listed in Table 5. The description of each gene can be found from the Gene database of National Center for Biotechnology Information (NCBI).

5.4. Comparative evaluations

In this section and to further assess the performance of the proposed method, its results are compared with state-of-art gene selection methods in the literature. These methods are tabulated in Table 6. The performance measures are used for the evaluation including: the average classification accuracy and the average of selected genes, which appear between parentheses, as shown in Table 6.

It can be seen from Table 7 and Fig. 7 that the results of rMRMR-MGWO produced a similar or superior classification performance on six datasets. Meanwhile, the proposed rMRMR-MGWO managed to efficiently minimize the number of genes for each experimented dataset and simultaneously obtain high classification accuracy. In terms of classification accuracy and number of selected genes solely, rMRMR-MGWO achieved the best results on 2 datasets (i.e., ALL-AML-4c, and MLL). In term of number of the selected genes, IG-MBKH achieves the best result, where it manage to select the lowest number of selected genes on three datasets (CNS, ALL-AML, and Ovarian). rMRMR-MGWO selected the lowest number of selected genes on two datasets (i.e., ALL-AML-4c and MLL). Furthermore, each of SARA-SVM and ABCD selected the lowest number of selected genes on two dataset. SARA-SVM achieves the best result on Lung and SRBCT datasets, while ABCD provides the best result for Colon and ALL-AML-3c datasets. In sum, rMRMR-MGWO appears to be competitive and in some cases superior against state-of-arts in gene selection problem.

6. Conclusions an further directions

Building a classification system using gene expression data is an active research area in bioinformatics. It has been proposed to be used as an early tool for diagnosing cancer as it can differentiate between healthy and cancerous patients samples. In this paper, an efficient filter-wrapper gene selection method was proposed to identify the biological-relevant genes that work collaboratively for cancer distinction. The proposed method combines robust Minimum Redundancy Maximum Relevancy (rMRMR) as a filter approach and Modified Gray Wolf Optimizer (MGWO) as a wrapper approach. In MGWO, the performance of GWO, which is used as the search engine, is improved by incorporating the TRIZ-inspired concepts as a new operator to the evolution loop of GWO. Utilizing this operator, the diversity control of MGWO is enriched where the generated solution from the original GWO operator is entered to TRIZ operator for further improvement. This is to help MGWO in exploring interesting regions in the search space. Nine high dimensional benchmark datasets are used in the experiments to test the performance of the proposed method. These benchmark datasets vary in terms of the number of genes,

Table 4
The results of the selected genes by embedded gene selection methods and rMRMR-MGWO.

Dataset	Measure	rMRMR-MGWO	LASSO	Random Forest	Elastic Nets	Decision tree
Colon Tumor	ACC	**0.9414**	0.5694	0.7903	0.8548	0.7903
	Precision	**0.9533**	0.54206	0.85	0.9	0.775
	Recall	**0.9197**	0.78307	0.6818	0.7727	0.8181
	F1score	**0.9546**	0.50997	0.83950	0.8888	0.8266
	MCC	**0.8639**	0.3370	0.5376	0.68010	0.5724
CNS	ACC	97.3889	0.6666	0.7166	**1**	0.6666
	Precision	0.9829	0.74358	0.8461	**1**	0.7692
	Recall	0.9571	0.5238	0.4761	**1**	0.47619
	F1score	0.98	0.7435	0.79518	**1**	0.75
	MCC	0.9611	0.2673	0.3476	**1**	0.5016
All-AML	ACC	**1**	0.7083	0.9722	**1**	0.9305
	Precision	**1**	0.9574	0.9787	**1**	0.9787
	Recall	**1**	0.24	0.96	**1**	0.84
	F1score	**1**	0.8108	0.9787	**1**	0.9484
	MCC	**1**	0.2991	0.9387	**1**	0.8461
Ovarian Cancer	ACC	**1**	0.9841	0.9960	**1**	0.9960
	Precision	**1**	**1**	**1**	**1**	**1**
	Recall	**1**	0.9560	0.9890	**1**	0.9890
	F1score	**1**	0.9878	0.9969	**1**	0.9969
	MCC	**1**	0.9659	0.99143	**1**	0.99143
Lung Cancer	ACC	0.9752	0.7931	0.9162	**0.9753**	0.8768
	Precision	0.9445	0.5064	0.8461	**0.9537**	0.6674
	Recall	0.9882	0.8810	0.9632	**0.9883**	0.9502
	F1score	0.9579	0.5402	0.8678	**0.9665**	0.6626
	MCC	0.92	0.5	0.8357	**0.9564**	0.6566
All-AML-3c	ACC	**0.9986**	0.4861	87.5	0.9861	0.8333
	Precision	**0.9982**	0.3298	0.8225	0.9629	0.6357
	Recall	**0.9994**	0.6670	0.9469	0.9929	0.9121
	F1score	**0.9977**	0.2940	0.8034	0.9738	0.5866
	MCC	**0.97**	0.3	0.7477	0.9685	0.6088
All-AML-4c	ACC	**0.9884**	0.5278	0.7777	0.9167	0.7639
	Precision	**0.9911**	0.3991	0.7856	0.8886	0.8137
	Recall	**0.9962**	0.7954	0.9068	0.9641	0.8933
	F1score	**0.9863**	0.3876	0.81532	0.9044	0.8022
	MCC	**0.93**	0.2522	0.7254	0.8729	0.7032
MLL	ACC	**0.9990**	0.5694	0.9027	0.9861	0.6388
	Precision	**0.9989**	0.5420	0.8908	0.9833	0.5833
	Recall	**0.9995**	0.7830	0.9499	0.9924	0.8030
	F1score	**0.999**	0.5099	0.8947	0.9856	0.5133
	MCC	**0.98**	0.3370	0.8492	0.9789	0.5522
SRBCT	ACC	**1**	0.3493	0.9879	**1**	0.3614
	Precision	**1**	0.25	0.9913	**1**	0.26
	Recall	**1**	0.75	0.9956	**1**	0.7546
	F1score	**1**	0.1294	0.9907	**1**	0.1498
	MCC	**1**	0.2	0.9864	**1**	0.2033

Table 5
Genes selected by the proposed method.

Dataset	Ngenes	Gene ID in dataset
Colon Tumor	10	H08393, M76378, M76378, M59807, T47377, M80815, X61118, M82919, Z50753, U09564
CNS	17	U43747_s_at, M27492_at, S71824_at, U41737_at, M13194_at, M63962_rna1_at, U17566_at, L40396_at, L48513_at, M96739_at, U21936_at, M13207_at, S78296_at, S66541_at, L33799_at, J04760_at, U49928_at
ALL-AML	5	Y07604_at, M31303_rna_at, U77604_at, M84371_rna1_s_at, U82759_at
Ovarian Cancer	4	MZ2.7921478, MZ435.07512, MZ2.8234234, MZ435.46452
Lung Cancer	16	gene9134, gene12097, gene9170, gene4452, gene8766, gene9733, gene10892, gene8472, gene8882, gene9672, gene4788, gene9840, gene3227, gene10139, gene11300, gene4244
ALL-AML-3c	7	M31303_rna1_at, J03473_at, X58072_at, M27891_at, D00749_s_at, X76223_s_at, X00274_at
ALL-AML-4c	11	D00749_s_at, M23197_at, X69398_at, X77094_at, U70063_at, M80254_at, M63138_at, D30756_at, X66401_cds1_at, M96326_rna1_at, U57721_at
MLL	8	36553_at ,37710_at, 37933_at, 1535_at, 34699_at, 1065_at, 35588_at, 32541_at
SRBCT	12	gene1955, gene509, gene1386, gene545, gene1613, gene2046, gene246, gene2162, gene153, gene1327, gene1389, gene1003

samples, and classes. Two well-known measurements in the gene selection research field are used for the evaluations, which are the number of selected genes and classification accuracy. Initially, the proposed method is evaluated using two phases. In the first

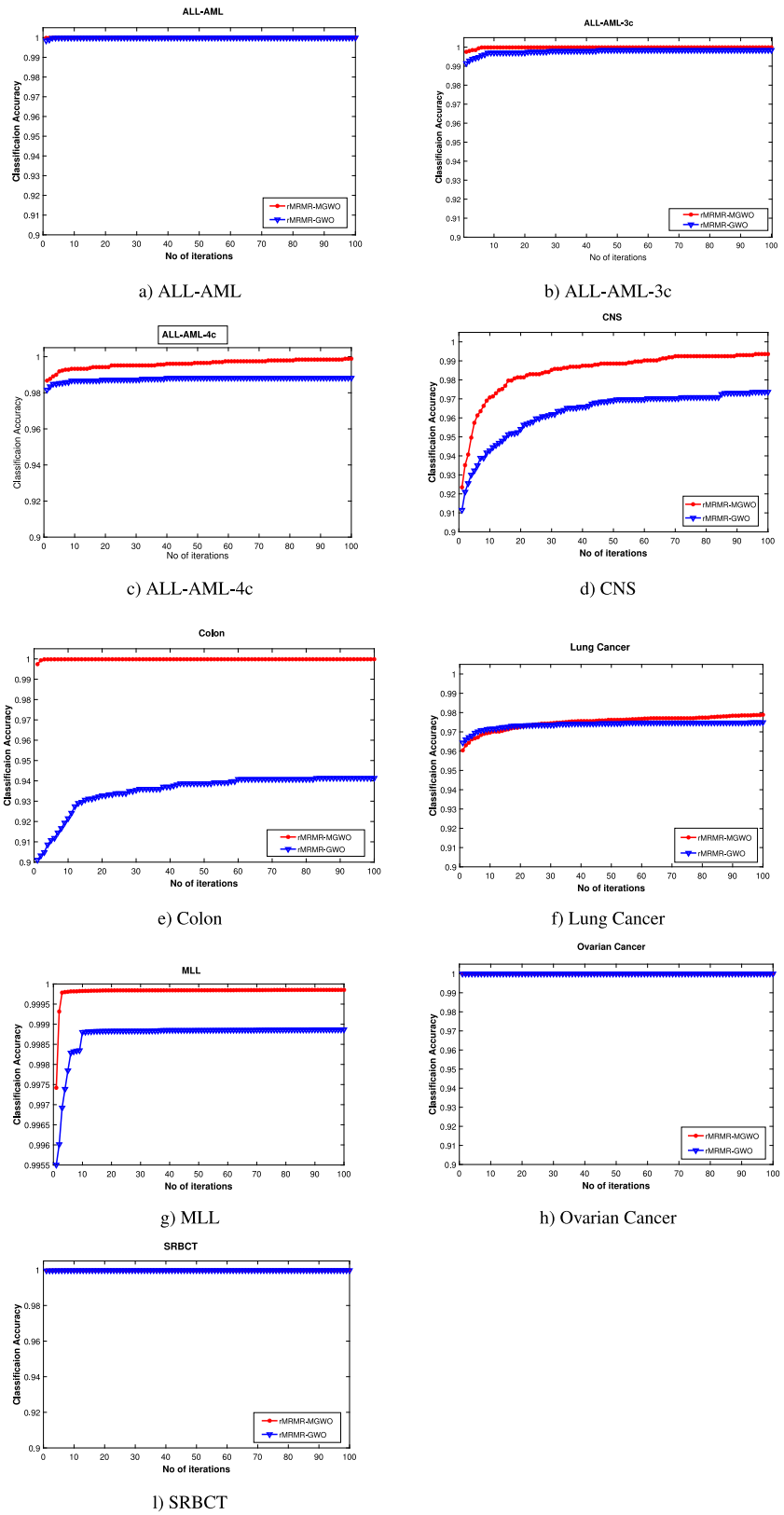

Fig. 6. The convergence behavior of the proposed method rMRMR-MGWO for all datasets.

phase, the results of MGWO (i.e., with TRIZ-inspired optimization operators) and GWO (without TRIZ-inspired optimization operators) are compared. According to the classification accuracy results, the proposed method rMRMR-MGWO achieved higher or similar classification results on all datasets against rMRMR-GWO. To elaborate, rMRMR-MGWO and rMRMR-GWO have the same classification accuracy results for ALL-AML, Lung Cancer, and SRBCT datasets. Other datasets (i.e., Colon, CNS, Ovarian, ALL-AML-3c, ALL-AML-4c, and MLL), the rMRMR-MGWO can achieve

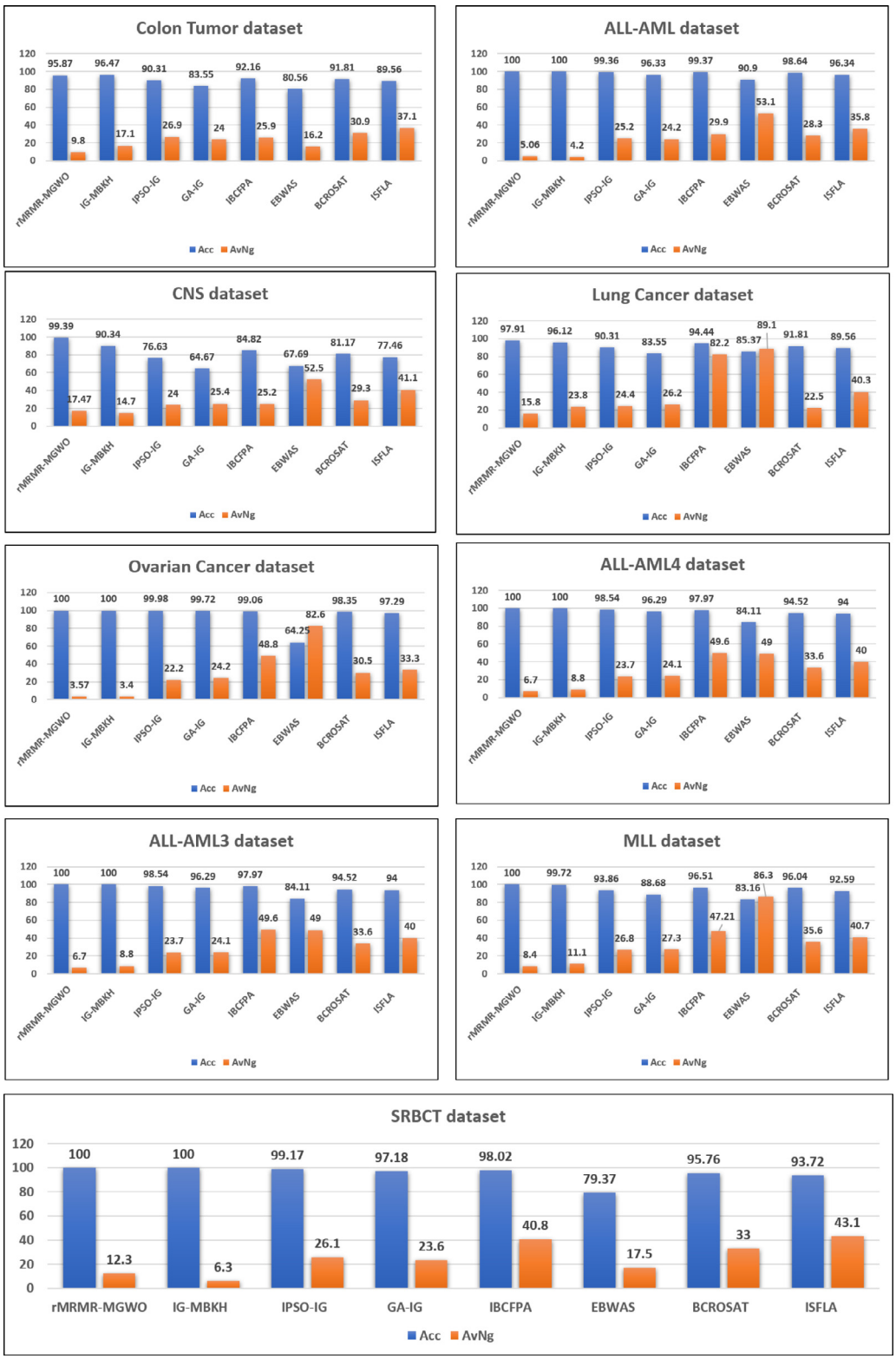

Fig. 7. Compared the experimental results of rMRMR-MGWO with other methods.

better results. According to both, the number of genes and classification accuracy results, the rMRMR-MGWO can yield better results than GWO in 6 out of 9 datasets while for the other three datasets, the number of genes yielded by rMRMR-GWO are less than the rMRMR-MGWO, although the classification accuracy results are better in favor of rMRMR-MGWO. In the second phase, rMRMR-MGWO is also compared against nine state-of-the-art methods using the same benchmark datasets. In terms of classification accuracy results, the rMRMR-MGWO is able to outperform the comparative methods in three out of nine comparative results while it is able to achieve five better results for other datasets achieved by other comparative methods. In terms of both the number of genes and classification accuracy results, rMRMR-MGWO achieves the best results on 4 datasets (i.e., Lung Cancer,

Table 6
Key to comparative methods.

Key	Method name	Reference
IG-MBKH	Information Gain and a Modified Binary Krill Herd Algorithm	[51]
GA-IG	Information gain and search strategy based on genetic algorithm (GA)	[49]
IPSO-IG	Information Gain and improved binary particle swarm optimization	[50]
IBCFPA	Absolute balance group strategy to the original binary clonal flower pollination algorithm	[46]
EBWSA	Elite Binary Wolf Search Algorithm	[70]
BCROSAT	Simulated Annealing and Tournament Selection to Coral Reefs Optimization algorithm	[47]
ISFLA	improved shuffled frog leaping algorithm	[45]
MIM-mMFA	Mutual Information Maximization-modified Moth Flame Algorithm	[71]
$DE - Fs^{pm}$	Permutational-based Differential Evolution algorithm	[71]
SARA-SVM	Simulated Annealing and Rao algorithm	[59]
ABCD	Artificial Bee Colony based on Dominance	[23]
BGWOCMALOL	GWO variant enhanced with a covariance matrix adaptation evolution strategy CMAES), levy flight mechanism, and orthogonal learning (OL) strategy	[72]

Table 7
Comparison results of the competitive algorithms.

Algorithm	Colon Tumor	CNS	ALL-AML	Lung Cancer	Ovarian Cancer
rMRMR-MGWO	95.87(9.80)	99.39(17.47)	100 (5.06)	97.9(15.80)	**100.00**(3.57)
IG-MBKH	96.47(17.10)	90.34(**14.70**)	**100.00**(**4.20**)	96.12(23.80)	**100.00**(**3.40**)
IPSO-IG	90.31(26.90)	76.63(24.00)	99.36(25.20)	90.31(24.40)	99.98(22.20)
GA-IG	83.55(24.00)	64.67(25.40)	96.33(24.20)	83.55(26.20)	99.72(24.20)
IBCFPA	92.16(25.90)	84.82(25.20)	99.37(29.90)	94.44(82.20)	99.06(48.80)
EBWAS	80.56(16.20)	67.69(52.50)	90.90(53.10)	85.37(89.10)	64.25(82.60)
BCROSAT	91.81(30.90)	81.17(29.30)	98.64(28.30)	91.81(22.50)	98.35(30.50)
ISFLA	89.56(37.10)	77.46(41.10)	96.34(35.80)	89.56(40.30)	97.29(33.30)
MIM-mMFA	100(31.00)	**100**(24.70)	100(7.50)	**100**(35.30)	98.42(35.90)
$DE - Fs^{pm}$	97.2(27.1)	–	100 (7.50)	–	100(90)
SARA-SVM	97.02(9)	–	97.65 (7)	90.22 (**5**)	99.15(6)
ABCD	99.23(**6.32**)	–	–	**100**(11.28)	–
BGWOCMALOL	**100**(29.5)	**100**(210)	100(96.5)	**100**(380)	–
Algorithm	ALL_AML_3c	ALL_AML_4c	MLL	SRBCT	
rMRMR-MGWO	**100**(6.70)	**99.91**(**11.37**)	**100**(**8.40**)	**100**(12.30)	
IG-MBKH	**100.00**(8.80)	99.44(15.80)	99.72(11.10)	**100**(6.30)	
IPSO-IG	98.54(23.70)	94.07(25.30)	93.86(26.80)	99.17(26.10)	
GA-IG	96.29(24.10)	89.89(25.00)	88.68(27.30)	97.18(23.60)	
IBCFPA	97.97(49.60)	94.35(45.60)	96.51(47.21)	98.02(40.80)	
EBWAS	84.11(49.00)	78.54(47.20)	83.16(86.30)	79.37(17.50)	
BCROSAT	94.52(33.60)	88.94(38.20)	96.04(35.60)	95.76(33.00)	
ISFLA	94.00(40.00)	90.91(32.20)	92.59(40.70)	93.72(43.10)	
MIM-mMFA	**100**(18.70)	–	100(33)	**100**(27.30)	
$DE - Fs^{pm}$	95.7(212.3)	–	–	99.9(821.4)	
SARA-SVM	98.02(7)	–	–	99.81(**5**)	
ABCD	**100**(**3.12**)	–	–	**100**(5.04)	
BGWOCMALOL	–	–	–	**100**(37.5)	

ALL-AML-3c, ALL-AML-4c, and MLL). In a nutshell, the proposed method reveals a high efficiency by being able to maximize classification accuracy while minimizing the number of genes. This can be considered as an outstanding contribution pregnant with a plethora of future developments for research communities. As rMRMR-MGWO provides very successful outcomes for the gene selection problem, in the future, it can be further enhanced in different perspectives:

1. Hybridization with other local search-based algorithms to empower its exploitation.
2. Application to different realistic datasets or bioinformatics problems such as protein tertiary structure prediction [73].
3. Using other advanced machine learning algorithms in the classification part such as conventional neural networks. (CNN) [74].

CRediT authorship contribution statement

Osama Ahmad Alomari: Suggested, Represented, Coded the contributions, Accomplished the experimental design, Executed the program, Experimental scenarios on the server, Develop the mathematical models of the proposed approach, Supervision. **Sharif Naser Makhadmeh:** Writing - original draft, Data curation, Validation, Formal analysis, Visualization. **Mohammed Azmi Al-Betar:** Writing - original draft, Data curation, Validation, Formal analysis, Visualization. **Zaid Abdi Alkareem Alyasseri:** Writing - original draft, Data curation, Validation, Formal analysis, Visualization. **Iyad Abu Doush:** Writing - original draft, Writing - review & editing, Validation. **Ammar Kamal Abasi:** Writing - original draft, Data curation, Validation, Formal analysis, Visualization. **Mohammed A. Awadallah:** Writing - original draft, Data curation, Validation, Formal analysis, Visualization. **Raed Abu Zitar:** review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] H. Wahsheh, I.A. Doush, M. Al-Kabi, I. Alsmadi, E. Al-Shawakfa, Using machine learning algorithms to detect content-based Arabic web spam, J. Inf. Assur. Secur. 7 (1) (2012).
[2] R. Sawalha, I.A. Doush, Face recognition using harmony search-based selected features, Int. J. Hybrid Inf. Technol. 5 (2) (2012) 1–16.
[3] I. Abu Doush, M.I. Al-Saleh, Can genetic algorithms help virus writers reshape their creations and avoid detection? J. Exp. Theoret. Artif. Intell. 29 (6) (2017) 1297–1310.

[4] S. Wang, Q. Cheng, Microarray analysis in drug discovery and clinical applications, in: Bioinformatics and Drug Discovery, Springer, 2006, pp. 49–65.

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

[6] M.A. Al-Betar, O.A. Alomari, S.M. Abu-Romman, A TRIZ-inspired bat algorithm for gene selection in cancer classification, Genomics 112 (1) (2020) 114–126.

[7] A. Jain, D. Zongker, Feature selection: Evaluation, application, and small sample performance, IEEE Trans. Pattern Anal. Mach. Intell. 19 (2) (1997) 153–158.

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

[9] 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.

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

[11] C. Ding, H. Peng, Minimum redundancy feature selection from microarray gene expression data, J. Bioinform. Comput. Biol. 3 (02) (2005) 185–205.

[12] 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.

[13] A.I. Hammouri, M. Mafarja, M.A. Al-Betar, M.A. Awadallah, I. Abu-Doush, An improved Dragonfly Algorithm for feature selection, Knowl.-Based Syst. 203 (2020) 106131.

[14] I. Aljarah, M. Habib, H. Faris, N. Al-Madi, A.A. Heidari, M. Mafarja, M. Abd Elaziz, S. Mirjalili, A dynamic locality multi-objective salp swarm algorithm for feature selection, Comput. Ind. Eng. 147 (2020) 106628.

[15] 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.

[16] A.I. Hammouri, M. Mafarja, M.A. Al-Betar, M.A. Awadallah, I. Abu-Doush, An improved dragonfly algorithm for feature selection, Knowl.-Based Syst. 203 (2020) 106131.

[17] M.A. Al-Betar, A.I. Hammouri, M.A. Awadallah, I.A. Doush, Binary β -hill climbing optimizer with S-shape transfer function for feature selection, J. Amb. Intell. Hum. Comput. (2020) 1–29.

[18] R.Z. Al-Abdallah, A.S. Jaradat, I.A. Doush, Y.A. Jaradat, A binary classifier based on firefly algorithm, Jordanian J. Comput. Inf. Technol. (JJCIT) 3 (3) (2017).

[19] H. Wang, X. Jing, B. Niu, A discrete bacterial algorithm for feature selection in classification of microarray gene expression cancer data, Knowl.-Based Syst. 126 (2017) 8–19.

[20] 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.

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

[22] 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, Int. J. Data Min. Bioinform. 19 (1) (2017) 32–51.

[23] V. Coleto-Alcudia, M.A. Vega-Rodríguez, Artificial bee colony algorithm based on dominance (ABCD) for a hybrid gene selection method, Knowl.-Based Syst. 205 (2020) 106323.

[24] M. Nasr, O. Farouk, A. Mohamedeen, A. Elrafie, M. Bedeir, A. Khaled, Benchmarking meta-heuristic optimization, 2020, arXiv preprint arXiv: 2007.13476.

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

[26] Z.A.A. Alyasseri, A.T. Khader, M.A. Al-Betar, O.A. Alomari, Person identification using EEG channel selection with hybrid flower pollination algorithm, Pattern Recognit. (2020) 107393.

[27] M.A. Awadallah, M.A. Al-Betar, A.I. Hammouri, O.A. Alomari, Binary JAYA algorithm with adaptive mutation for feature selection, Arab. J. Sci. Eng. (2020) 1–16.

[28] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, Adv. Eng. Softw. 69 (2014) 46–61.

[29] H. Faris, I. Aljarah, M.A. Al-Betar, S. Mirjalili, Grey wolf optimizer: a review of recent variants and applications, Neural Comput. Appl. 30 (2) (2018) 413–435.

[30] Q. Al-Tashi, H.M. Rais, S.J. Abdulkadir, S. Mirjalili, H. Alhussian, A review of grey wolf optimizer-based feature selection methods for classification, in: Evolutionary Machine Learning Techniques, Springer, 2020, pp. 273–286.

[31] X. Zhao, S. Ren, H. Quan, Q. Gao, Routing protocol for heterogeneous wireless sensor networks based on a modified grey wolf optimizer, Sensors 20 (3) (2020) 820.

[32] Q.M. Alzubi, M. Anbar, Z.N. Alqattan, M.A. Al-Betar, R. Abdullah, Intrusion detection system based on a modified binary grey wolf optimisation, Neural Comput. Appl. (2019) 1–13.

[33] X. Yan, Y. Zhang, D. Zhang, N. Hou, Multimodal image registration using histogram of oriented gradient distance and data-driven grey wolf optimizer, Neurocomputing (2020).

[34] C. Li, W. Wang, D. Chen, Multi-objective complementary scheduling of hydro-thermal-RE power system via a multi-objective hybrid grey wolf optimizer, Energy 171 (2019) 241–255.

[35] C. Lu, L. Gao, Q. Pan, X. Li, J. Zheng, A multi-objective cellular grey wolf optimizer for hybrid flowshop scheduling problem considering noise pollution, Appl. Soft Comput. 75 (2019) 728–749.

[36] X. Ma, X. Mei, W. Wu, X. Wu, B. Zeng, A novel fractional time delayed grey model with Grey Wolf Optimizer and its applications in forecasting the natural gas and coal consumption in Chongqing China, Energy 178 (2019) 487–507.

[37] X. Li, K.M. Luk, The grey wolf optimizer and its applications in electromagnetics, IEEE Trans. Antennas and Propagation (2019).

[38] M.A. Al-Betar, M.A. Awadallah, M.M. Krishan, A non-convex economic load dispatch problem with valve loading effect using a hybrid grey wolf optimizer, Neural Comput. Appl. (2019) 1–28.

[39] M. Vosooghifard, H. Ebrahimpour, Applying grey wolf optimizer-based decision tree classifer for cancer classification on gene expression data, in: Computer and Knowledge Engineering (ICCKE), 2015 5th International Conference on, IEEE, 2015, pp. 147–151.

[40] J.-Y. An, Z.-H. You, Y. Zhou, D.-F. Wang, Sequence-based prediction of protein-protein interactions using gray wolf optimizer–based relevance vector machine, Evol. Bioinform. 15 (2019) 1176934319844522.

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

[42] C.A. Mei, D. Pham, J.S. Anthony, W.N. Kok, PCB assembly optimisation using the Bees Algorithm enhanced with TRIZ operators, in: IECON 2010-36th Annual Conference on IEEE Industrial Electronics Society, IEEE, 2010, pp. 2708–2713.

[43] 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.

[44] M. Ang, D. Pham, K. Ng, Application of the Bees Algorithm with TRIZ-inspired operators for PCB assembly planning, in: Proceedings of 5th Virtual International Conference on Intelligent Production Machines and Systems (IPROMS2006), 2009, pp. 454–459.

[45] B. Hu, Y. Dai, Y. Su, P. Moore, X. Zhang, C. Mao, J. Chen, L. Xu, Feature selection for optimized high-dimensional biomedical data using an improved shuffled frog leaping algorithm, IEEE/ACM Trans. Comput. Biol. Bioinform. 15 (6) (2016) 1765–1773.

[46] C. Yan, J. Ma, H. Luo, G. Zhang, J. Luo, A novel feature selection method for high-dimensional biomedical data based on an improved binary clonal flower pollination algorithm, Hum. Hered. 84 (1) (2019) 1–13.

[47] C. Yan, J. Ma, H. Luo, A. Patel, Hybrid binary coral reefs optimization algorithm with simulated annealing for feature selection in high-dimensional biomedical datasets, Chemometr. Intell. Lab. Syst. 184 (2019) 102–111.

[48] V. Preeja, A. Shahana, A binary krill herd approach based feature selection for high dimensional data, in: 2016 International Conference on Inventive Computation Technologies (ICICT), vol. 2, IEEE, 2016, pp. 1–6.

[49] C.-H. Yang, L.-Y. Chuang, C.H. Yang, et al., IG-GA: a hybrid filter/wrapper method for feature selection of microarray data, J. Med. Biol. Eng. 30 (1) (2010) 23–28.

[50] B. Sahu, A combo feature selection method (filter+ wrapper) for microarray gene classification, Int. J. Pure Appl. Math. 118 (16) (2018) 389–401.

[51] G. Zhang, J. Hou, J. Wang, C. Yan, J. Luo, Feature selection for microarray data classification using hybrid information gain and a modified binary Krill Herd algorithm, Interdiscip. Sci. Comput. Life Sci. (2020).

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

[53] 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.

[54] A.A. Aguilar-Lasserre, V.E. Torres-Sánchez, G. Fernández-Lambert, C. Azzaro-Pantel, G. Cortes-Robles, M.A. Román-del Valle, Functional optimization of a Persian lime packing using TRIZ and multi-objective genetic algorithms, Comput. Ind. Eng. 139 (2020) 105558.

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

[56] 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.

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

[58] C.-M. Lai, W.-C. Yeh, C.-Y. Chang, Gene selection using information gain and improved simplified swarm optimization, Neurocomputing (2016).

[59] S.K. Baliarsingh, K. Muhammad, S. Bakshi, SARA: A memetic algorithm for high-dimensional biomedical data, Appl. Soft Comput. 101 (2021) 107009.

[60] B. Duval, J.-K. Hao, J.C. Hernandez Hernandez, A memetic algorithm for gene selection and molecular classification of cancer, in: Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation, ACM, 2009, pp. 201–208.
[61] M. Dash, H. Liu, Feature selection for classification, Intell. Data Anal. 1 (3) (1997) 131–156.
[62] C. Ambroise, G.J. McLachlan, Selection bias in gene extraction on the basis of microarray gene-expression data, Proc. Nat. Acad. Sci. 99 (10) (2002) 6562–6566.
[63] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software: an update, ACM SIGKDD Explor. Newsl. 11 (1) (2009) 10–18.
[64] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. (TIST) 2 (3) (2011) 27.
[65] Z. Zhu, Y.-S. Ong, M. Dash, Markov blanket-embedded genetic algorithm for gene selection, Pattern Recognit. 40 (11) (2007) 3236–3248.
[66] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1) (1996) 267–288.
[67] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[68] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. R. Stat. Soc. Ser. B Stat. Methodol. 67 (2) (2005) 301–320.
[69] V. Sugumaran, V. Muralidharan, K. Ramachandran, Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing, Mech. Syst. Signal Process. 21 (2) (2007) 930–942.
[70] J. Li, S. Fong, R.K. Wong, R. Millham, K.K. Wong, Elitist binary wolf search algorithm for heuristic feature selection in high-dimensional bioinformatics datasets, Sci. Rep. 7 (1) (2017) 1–14.
[71] R. Rivera-López, E. Mezura-Montes, J. Canul-Reich, M.A. Cruz-Chávez, A permutational-based differential evolution algorithm for feature subset selection, Pattern Recognit. Lett. 133 (2020) 86–93.
[72] J. Hu, H. Chen, A.A. Heidari, M. Wang, X. Zhang, Y. Chen, Z. Pan, Orthogonal learning covariance matrix for defects of grey wolf optimizer: Insights, balance, diversity, and feature selection, Knowl.-Based Syst. 213 (2021) 106684.
[73] M.S. Abual-Rub, M.A. Al-Betar, R. Abdullah, A.T. Khader, A hybrid harmony search algorithm for ab initio protein tertiary structure prediction, Netw. Model. Anal. Health Inf. Bioinform. 1 (3) (2012) 69–85.
[74] H. Dia, An object-oriented neural network approach to short-term traffic forecasting, European J. Oper. Res. 131 (2) (2001) 253–261.