# A study on metaheuristics approaches for gene selection in microarray data: algorithms, applications and open challenges

Alok Kumar Shukla[1] · Diwakar Tripathi[2] · B. Ramachandra Reddy[2] · D. Chandramohan[2]

## Abstract

In the recent decades, researchers have introduced an abundance of feature selection methods many of which are studied and analyzed over the high dimensional datasets typically tiny number of instances and hundreds or thousands of genes. Feature selection methods provide a way of reducing computation cost, improving prediction performance and better understanding of the data structure. However, it is a challenging task due to two reasons such as the considerable solution space and feature interaction. A diversity of feature selection methods is established and applied on high dimensional datasets which includes the metaheuristic algorithms. In this paper, we focus on the basic algorithmic structures of metaheuristic for feature selection that reveals the predominate genes, called biomarkers in microarray gene expression data series with limited resources. In addition, more than hundred articles are carefully screened to prepare the up-to-date comprehensive work on the metaheuristic approach for feature selection and also discussed a range of open issue of recent metaheuristic approaches for feature selection. Furthermore, we have applied some metaheuristic techniques for feature selection on gene expression datasets to demonstrate the applicability of methods. Based on this comprehensive survey, this article suggest some crucial recommendations to researchers for choosing a suitable method from the repository of feature selection methods.

**Keywords** Biomarker · Feature selection · Gene expression · Machine learning · Metaheuristic

## 1 Introduction

DNA microarray-based emerging technologies have been applied in the fields of bioinformatics and biotechnology which exponentially increases data size concerning features and number of instances [1, 2]. Generally, genes that are relevant to a specific target annotation are known as biomarker. Biomarker discovery is a vital task for the researcher, as well as for the medical or pharmaceutical company with the goal of identifying genes that can be targeted by drugs. A typical microarray technology involves the hybridization of mRNA molecule to the DNA template. The microarray datasets are categorized as unlabelled, partially labeled and adequately labeled. It leads to the growth of gene selection techniques namely supervised and unsupervised to discover the organic patterns (such as tissues, cell lines, etc.) from the instances [3]. DNA microarray technologies have been applied to manage classification [4] and clustering [5] problems.

In data mining and bioinformatics domains, feature selection (FS) is an remarkable approach to reduce the dimensionality of the dataset (i.e., leukemia and colon) [6]. The basic principle of feature selection is to choose $m$ relevant feature subset from $n$ original DNA microarray dataset attributes on some criteria [7]. One of the major issue has been investigated and addressed by the researchers in DNA microarray as "curse of dimensionality". However, not all attributes are important, because many of them are redundant or even meaningless, which can decrease the classification accuracy of any learning algorithms. The construction of features [or Feature Extraction (FE)] is strictly related to the selection of attributes, which can also diminish the higher dimensionality [8, 9]. The main difference between FE and FS is as FS picks a subset of the candidate features whereas the feature extraction creates new features from candidate features.

In general, existing feature selection approaches overlook the fact that for a given cardinality, there are numerous

✉ Alok Kumar Shukla
alokjestshukla@gmail.com

1 Department of Computer Science and Engineering, G L Bajaj Institute of Technology and Management, Greater Noida, Uttar Pradesh, India

2 Madanapalle Institute of Technology and Science, Madanapalle 517325, India

subsets with similar information quality. It addresses the critical issues by removing irrelevant and noise data [10]. Generally, FS methods are categorized into four groups such as filter, wrapper, hybrid and embedded methods [11]. In most of the situations, a comprehensive search for choosing the optimal subset of features in a given data set is practically impossible. In the recent literature, various pursuit methods have been employed for the selection of features, such as iterative search, heuristic search and random search [12, 13]. However, most of the methods for selecting existing features still show stagnation in the excellent location and a high computational cost [14]. It is not possible to solve the enormous number of attributes using traditional methods. Therefore, researchers have employed an effective search methods for feature selection by using specific fitness function to reduce the dimensionality of microarrays, efficiently. In this article, our attention on metaheuristic based feature selection, because it is more superior to other well known filters or hybrid methods towards to readability and interpretability.

The major problems in non-metaheuristic algorithms for feature selection is trapped in local (or weak) solutions. To resolve this problem, in the mid-sixties, a wide range of algorithms have been widely applied and showing research interest in this domain [15]. The most popular metaheuristic algorithms are available in swarm intelligence (SI) [16] and evolutionary algorithm (EA) [17, 18] are: genetic algorithm (GA) [19], particle swarm optimization (PSO) [20], artificial bee colony (ABC), ant colony optimization (ACO), Bacterial Foraging Optimization (BFO), and Gravitational search algorithm (GSA), teaching learning based optimization (TLBO) [21], Sequential Forward Search (SFS) [22], and Sequential Backward Search (SBS) [23] which can applied to considerable number of features, but small supervised gene expression data sizes (ten to hundreds). Teaching learning based optimization (TLBO) for finding the near to optimal solution. Genetic algorithm (GA) usages the concept of Darwinian evolution based on the survival of the fittest [24], ACO imitates the behavior of scouring a bee [25], BFO method mimics the foraging strategy of *Escherichia coli* bacteria [26] and GSA works on the principle of the explosion of a mass [27]. The most informative set of genes can be picked by a corresponding fitness value of all features belonging to the datasets. Still, there are no comprehensive guidelines on the merits and demerits of different metaheuristic methods with their more appropriate areas of application. The approximate distribution of publications and proceeding last 8-year w.r.t. metaheuristic algorithms are depicted in Fig. 1. The articles used in this survey are obtained from all the central databases, such as Web of Science, Scopus, Google Scholar etc.

The main idea behind investigating the metaheuristic algorithms is to tackle complex optimization problems
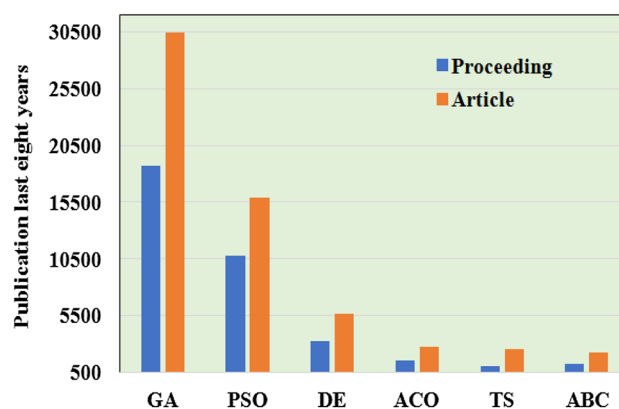


**Fig. 1** Number of publications on metaheuristics algorithms (January 2010–June 2019)

where classical optimization methods have to be failed. These methods are now accepted as some of the most practical approaches for solving many real world problems like gene selection [28]. There are several advantages of using metaheuristic algorithms for optimization as given below:

- Broad applicability: It can be applied to any problems that can be formulated as function optimization problems.
- Hybridization: It can be combined with more than one stochastic or classical optimization techniques.
- Ease of implementation: It is easy to implement and has a less complicated programming structure with less complicated operations.
- Efficiency and flexibility: It can be able to solve larger problems rapidly.
- It can easily handle multiple objective problems of stochastic nature [29].

However, there are some disadvantages of the metaheuristic methods that should be present here:

- In general, the optimization performance is highly dependent on control parameter tuning.
- It do not appear mathematical base, in compared to more traditional techniques [30].
- It cannot prove optimality.
- It cannot probably reduce the search space.
- Repeatability of optimization results obtained with the same initial condition settings is not guaranteed.

Since to our awareness, there is no broad discussion of the metaheuristic methods for feature selection. Ang et al. [2] have offered a comprehensive survey and provided the straightforward organization of gene selection and reviews filter, wrapper, and hybrid methods in the high dimensional

datasets. Author has not systematically reviewed the wrapper methods for gene selection and also less contribution in the description of data characteristics [31]. This paper presents a comprehensive review of the metaheuristic approaches for features selection in order to provide algorithms applicability with merits and demerits to the researchers involved in cutting-edge research.

The organization of the article described as follows. Section 2 summarize the Broadway concept of gene selection, which is used to improve solution diversity in the FS problems. Section 3 defines the taxonomy of metaheuristic techniques for gene selection that are used for high dimensional datasets. Section 4 describes the classification methods. Section 5 presents the study the experimental results based on metaheuristic approaches for feature selection. Section 6 represents a range of open challenges recommendations for future directions and followed by conclusions in Sect. 7.

## 2 Outline of gene relevancy

Gene selection aims to choose an optimal features subset from the $m$ cardinality of features which is more relevant and correlated to each other. Yu and Liu [32] have introduced the concept of features subset selection into four aspects: (a) fully irrelevant and noisy variables, (b) weakly relevant and redundant features, (c) weakly relevant and non-redundant variables, and (d) strongly relevant features as depicted in Fig. 2. An optimal subset contains all the features in the class (c) and (d), respectively.

For better accuracy and discriminative power, relevant features play a crucial role in gene selection. minimal Redundancy and Maximum Relevancy (mRMR) as an approach for feature selection is presented by the Ding and Peng [33] and successfully employed further in various
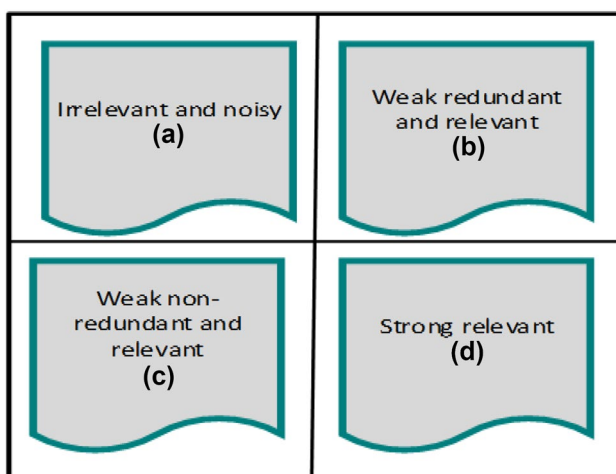
real-life applications [34–37]. As shown in Fig. 2, we can provide an analysis of features taxonomies based on their redundancy and relevancy.

The range of the variables $F = \{f_1, f_2, f_3, \ldots, f_n\}$ and instance space is defined as $s = \{s_1 * s_2 * s_3, \ldots, s_m\}$ respectively. Our objective function is represented as $f : s \Rightarrow l$ according to its meaningful features, where, $l$ present a space of labels.

**Definition 1** Gene selection: let, the original set of variables $F$ and $L(\cdot)$ be an assessment standard to be maximized and defined as $L : F' \subseteq F \Rightarrow R$. The original subset of features can be considered under the following concerns [38]:

1. Let, $|F| = m$ and $|F| = n$, then, $L(F')$ is maximized, where and $m \gg n$ and $F' \subset F$.
2. Set a threshold $\delta$ i.e., $L(F') > \delta$, to find a subset of features with the least number ($m \gg n$).
3. Discovering the Optimization method $L(F')$ with optimal subsets of features $|F|$.

There is a continuous problem of selecting variables in which each characteristic $f_k \in F$ assigns weights $w_{(k)}$ to preserve the theoretical significance of the variables. The allocation of binary weight is considered in the problem of selecting binary characteristics [39]. The subset of optimal features is considered one of the most optimal subsets. Therefore, the above definition does not guarantee that the subset of optimal features is distinctive. The subset of best features is defined in terms of precision of the induced classifier as below in Definition 2.

**Definition 2** Let, the dataset be defined by features $\{f_1, f_2, f_3, \ldots, f_n\}$ from a distribution $\rho$ over the labeled instance space and inducer $\aleph$. An optimal feature subset, $f_{opt}$, is a subset of the features such that the accuracy of the induced classifier $C = \aleph(D)$ is "maximal" [40]. The diversity of approaches have been applied to solve features selection problems, where filter-based gene selection approaches have recently extended attention and shown the efficient results. A common procedure for gene selection is shown as in Algorithm 1.



**Fig. 2** Feature taxonomies based on their redundancy and relevancy

---

**Algorithm 1** Overall process for feature selection

F: Set of features of dataset with n features     SG: Successor Generator operator     S: Evaluation criteria

$F_{opt}$ = Optimal features subset or weighted features

Initialization:

F'=Access point (F)

$F_{opt}$=select best F' from S

Repeat

F'= Strategy-search (F', SG(S), F)

$F_{opt}$ = Select the best F' according to S

If(S(F')>=S($F_{opt}$)$or(S(F') == S(F_{opt})and|F'| < \quad F_{opt})$

then $F_{opt}$ = F';

Until stopping criteria not meet.

---

**Definition 3** Relevance to Object [41] : "A feature $x_i \in X$ is relevant to an object concept $C$; if there exists a pair of examples and in the instance space such that and differ only in their assignment to $x_i$ and $C(A) \neq C(B)$".

**Definition 4** Strongly Relevant to Instances [41]: "A feature $x_i \in X$ is strongly relevant to the instance $S$ if there exists a pair of examples $A, B \in S$ that only differ in their assignment to $x_i$ and $C(A) \neq C(B)$ or a feature $x_i \in X$ is strongly relevant to an objective $C$ in distribution of $P$, if there exists a pair of examples $A, B \in I$ with $P(A) \neq 0$ and $P(B) \neq 0$ that only differ in their assignment to $x_i$ and $C(A) \neq C(B)$".

**Definition 5** Weakly Relevant to Instances [41]: "A feature $x_i \in X$ is weakly relevant to instance $S$ if there exists at least a proper $X' \subset X(x_i \in X')$ where $x_i$ is strongly relevant with respect to $S$. Or, variable $x_i \in X$ is weakly relevant to objective $C \in distribution of P$ if there exists at least a proper $X' \subset X(x_i \in X')$, where $x_i$ is strongly relevant with respect to P″. The above definitions concentrate on which features are meaningful. Put in other words, it need to utilize relevance as a measure of unpredictability to indicate how "complicated" a function is.

**Definition 6** Relevance as a Complexity Measure [41]: "Given an instance of data and a set of concept $C$, let $r(S, C)$ be the number of variables relevant using Definition 3 to a concept $C$ in that, out of all those whose error over $S$ is the least, has the fewest relevant features". Otherwise, we will impose an optimal performance on $S$ with concept $C$ using the least number of features. The relevant concepts mentioned above are independent of the precise algorithm of learning. This means that a certain relevant function is not necessary for algorithms to learn.

**Definition 7** Incremental Usefulness [42]: "Given an instance of data $S$, a learning algorithm $L$, and a subset of variables $X'$, variable $x_i$ is incrementally useful to $L$ with respect to $X'$, if the accuracy of the hypothesis that

$L$ produces using the variable set $x_i \cup X'$ is better than the accuracy achieved using just the features subset $X''$.

**Definition 8** Entropy Relevance [43]: "Denoting mutual information $I(X;c) = H(c) - H(c|X)$ with Shannon Entropy $X$, the entropy relevance of $X$ to $c$ is defined as $r(X : c) = I(X : c)/H(c)$". Let $c$ be the objective seen as a feature and $X$ represent the original set of features, a subset $X' \subset X(x_i \in X')$ is sufficient if $I(X' : c) = I(X, c)$. For a sufficient subset, it must satisfy $r(X' : c) = r(X, c)$.

## 2.1 Basic progressive flowchart of feature selection

The way of picking a subset of relevant genes from the original datasets, is partitioned into five main steps as illustrates in Fig. 3. At each level, the decision is made which affects the gene selection performance [44].

**Stage 1** *Define search direction* This step defines the initial point and the search direction. According to the forward search process, the search starts with a null set and includes novel features in each successive iteration sequentially [45]. On the contrary, the search process starts with an original set of features and then features are removed successively in
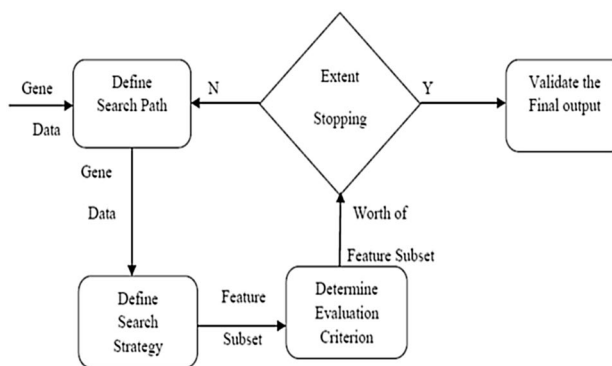


**Fig. 3** The overall progressive flowchart of gene selection

each successive iteration; is known as backward elimination search [46].

**Stage 2** *Define a search strategy* Many obsolete measures that evaluate features individually do not work well. Therefore, features subset must be evaluated in a group. To handle this issue, Gheyas and Smith [47] addressed the outlines of the search strategy. A high-quality search approach should present exceptional global search capability, high convergence ratio to get the nearest global optimal solution, acceptable local search solution, and high computational capability.

**Stage 3** *Define an evaluation criterion* Firstly, evaluation processes of gene selection are categorized into five different varieties such as filter, wrapper, ensemble, hybrid and embedded [48]. Filter method is also recognized as an open-loop method [49]. It is a simple, effective method which selects the features subsets regarding to underlying characteristics of features, additional information in terms of learning tasks. This approach mainly estimates the feature characteristics with four different types of evaluation criteria namely information theory, dependency, consistency and distance [40].

Wrapper method is known as a closed-loop method which encloses the gene selection around the learning algorithm and makes use of classification accuracy as a fitness function for features subset evaluation [19]. It picks the relevant or discriminative features subset by using the specific classifier concerning for minimizing the prediction error [40].

Embedded technique is a built-in feature selection tool that implement the features in the machine learning method and employs its properties through variable evaluation. It is more proficient and less computational cost, and more conformable than wrapper based method in terms of solution quality [22].

Hybrid method is created by merging two dissimilar methods for feature selection, e.g., filter and wrapper. Its activities inherit the advantages of individual methods to gain computational strengths [50, 51]. It employs dissimilar evaluation criteria in dissimilar search stages to get better proficiency and presentation for enhancing computational performance.

Ensemble method is one of the important processes for feature selection that intends to create a cluster of optimal variable subsets and then generate a collective outcome out of the group [52, 53]. For the comprehensive discussion of the ensemble-based gene selection can be found in [54]. It is deliberately planned to deal with the inconstancy and perturbation issues in the many feature selection algorithms. The performance of feature selection depends on a particularly selected subset. Thus, it should be relatively flexible and vigorous when dealing with high-dimensional datasets.

**Stage 4** *Describes the stopping criteria* When the FS methods achieves the optimal number of features, then the gene subset selection procedure should stop. An appropriate stopping criterion is prevaricated over-fitting and produced the more capable process to produce an optimal feature subsets with the less computational load. The decisions made in the prior stages may influence the preference of stopping criterion. The general stopping criteria are as follows:

- Describe the predetermined fixed number of features.
- Describe the predetermined fixed number of iterations.
- Determine the stall generation over two consecutive generations in percentage.
- Obtaining the most excellent variable subset according to accurate evaluation method.

**Stage 5** *Validate the optimum output* To measure the effectiveness of important feature subsets for better classification, a large number of judgment or validation methods have been presented in the previous literature [55].

In order to select the best subsets of features, from the literature, it is observed that there are two key aspects such as maximize the accuracy of the classification and minimize the attributes present in datasets. These are often contradictory goals. Therefore, selection of features can be solved by using multi-objective problems (MOPs) to invent a set of compromise results between two conflict objectives. In recent years, current research in this oversight it gained great attention, where metaheuristic techniques provide the evolutionary computation (EC) techniques using a population-based approach is particularly suitable for the optimization of multiple objectives.

## 2.2 Background information

### 2.2.1 Current literature on feature selection

In this subsection, we described the metaheuristic techniques in three characteristics: Search techniques (Exploration methods), assessment criteria, and some conflict objectives.

1.1 *Exploration methods* In the literature, various methods are available for FS that makes use of complete/exhaustive search [56]. This is on the grounds when the quantity of features is moderately little, as a rule, these techniques are excessively costly starting thereof view. Therefore, various evolutionary methods have been employed for subset selection, such as a heuristic search algorithm, in which typical cases are Sequential-Forward-Selection (SFS) [34], Sequential-Backward-Selection (SBS) [57]. But these approaches have nesting effect. To avoid the problem of the "nesting" effect

two techniques as Sequential Floating Forward Selection (SBFS) [14] and Sequential Floating Forward Selection (SFFS) [58] have addressed. As a comparison to a static method, it gives better performance concerning to computational cost and optimal feature subsets. Han et al. [59] have proposed an approach to investigate a subset of relevant characteristics using the BPSO coding scheme with the help of the ELM classifier. Zhang et al. [60] have presented a heuristic search and regression method, to select the features in high-dimensional data series. From experimental results show that heuristic methods achieved comparable performance as a comparison to the backtracking algorithm, with less computational time. Further, metaheuristic methods treated as active methods and useful to solve FS problems. These methods include GA, PSO, ACO.

1.2 *Assessment criteria* The classification performance is an assessment criterion for optimal feature subsets selection by metaheuristic approaches for feature evaluation. To evaluate the assessment of substantial learning algorithms such as support vector machine (SVM) [61], Naive Bayes (NB) [62], k-nearest neighbor (k-NN) [63], Decision Tree (DT) [64], LASSO [65], artificial neural network (ANN) [66], linear discriminant analysis (LDA) [67], etc., have been employed in metaheuristic for better classification of tumors and cancer [68] from the microarray datasets.

In case of filter method, some criteria to measure features importance/weight on datasets such as information theory, correlation, distance and consistency are utilized [22]. Various researchers have found the commonly used filter-based gene selection techniques i.e., Joint Mutual Information (JMI) [69], Information Gain [70], Relief-F [71], Chi-Square [72], F-statistic [73], Mutual Information (MI) [74], that work to reduce a considerable number of features but small supervised gene expression data size. One of the most popular filter method as mRMR [75], where MI is used to quantify the relevance of each attribute regarding object class. All the essential attributes are selected for better classification. As evaluation performance, many kinds of literature confirmed that filter based methods do not perform well to problems above tons of features [76].

1.3 *Number of conflict objectives* The excellent FS strategy intend to maximize classification performance only during the search process where, classification performance and some features included in a separate fitness function. To the extent our information, all the algorithms for the FS of the multi-objective characteristics depend on metaheuristic methods, since their population-based way that produces various solutions in single trails that particularly appropriate for Multi-Objective Optimization (MOO)[77].

### 2.2.2 Taxonomy of features selection approaches

This paper is courtyard on metaheuristic approaches for feature selection which characterized into various classifications as shown in Fig. 4, with three distinct criteria: search technique, assessment and objectives/problems. These criteria are the key segments of an FS strategy. The most popular metaheuristic method is genetic algorithm (GA), the optimal feature sets uses the respective learning categories for classification/regression with various approaches such as SVM, k-NN, and Lasso.

In literature, a wide range of metaheuristic based feature selection algorithms namely GA, Swarm based Algorithm, Artificial Bee Colony (ABC), Ant Colony Optimization (ACO), and Harmony Search (HS). GA is based on the Darwinian theory evolution to achieve the fittest or best features set [78]; PSO implements search for the behavior of a flock of birds or a fish school to look for food [79]; ABC imitates the practice of scouring a bee [80]; and last ACO is based on the behaviour of an ant looking for a destination from the source [81].

Based on assessment standards, filters as well as wrapper along with combination of both methods are evaluated. As per the objective, FS methods are characterized into single objective (SO), Multi-objective (MO) and Many-objective (MOB) methods, here multi-objective methodologies compare to techniques mean to discover a Pareto front of trade-off solutions. These objectives are dependent on fitness function.

## 3 Metaheuristic method for feature selection

### 3.1 Genetic algorithm for feature selection

GA [82], inspired by the process of natural selection and working in parallel heuristic research and it solves the problem of optimization based on the process of natural genetic schemes. GA method plays an energetic part in FS using best attributes with the help classification measures as a fitness evaluator with classifiers such as SVM, k-NN, and Lasso. Feature selection problem is binary optimization problem as a feature is selected or not, are represented by using 1 and 0 bits, respectively. But, optimization approaches are prepared for the optimizing the criteria in continuous search spaces. So, for optimizing the features subset, there is a need to covert the optimization approaches which can work in binary search space.

To estimate the goodness of features subsets, various researchers have integrated various classification methods such as SVM, KNN, ANN, DT, NB, multiple linear regression [83] and extreme learning machine (ELM) [84] as a
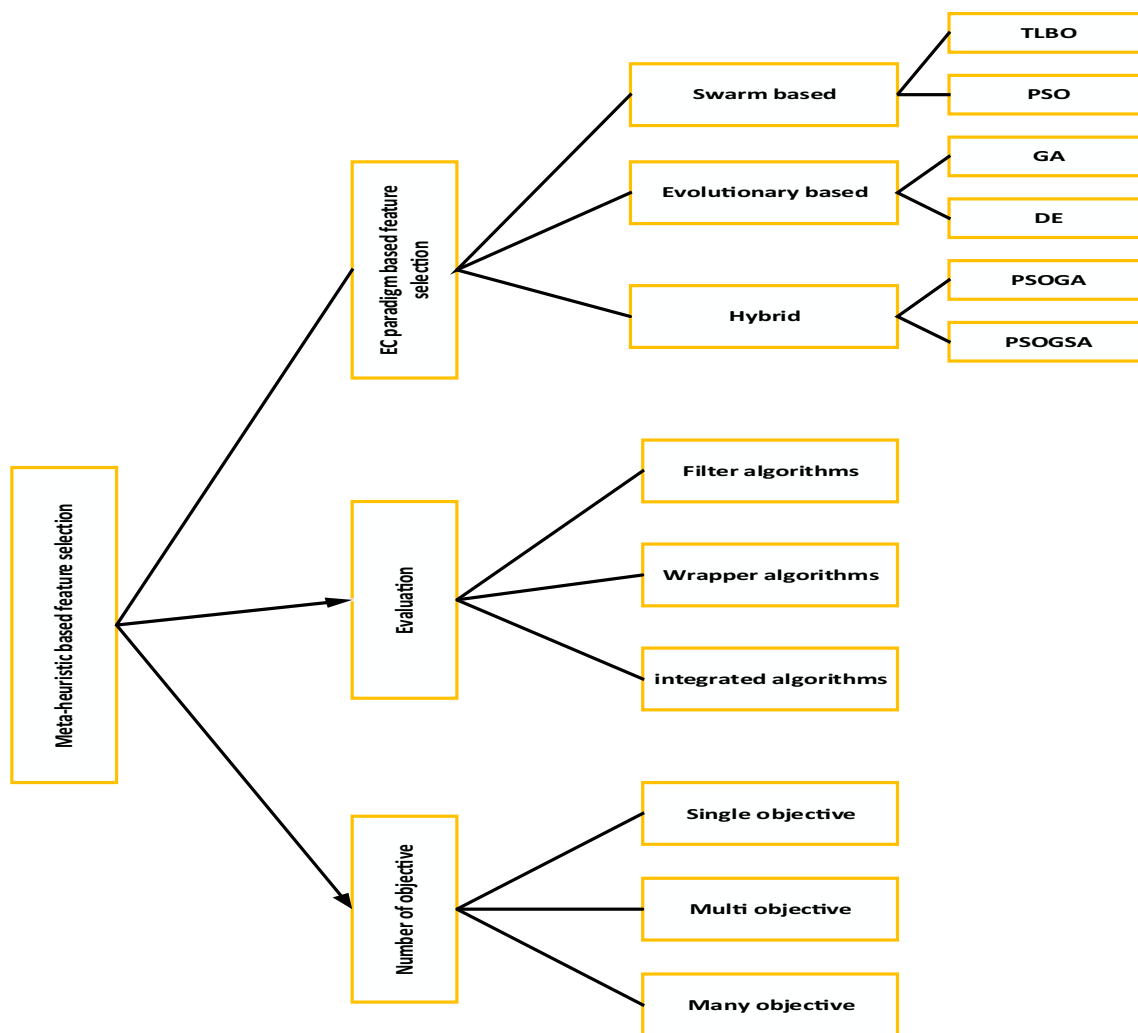
**Fig. 4** Overall categories of metaheuristic for feature selection

wrapper for metaheuristic approaches. The most popular classification techniques are SVM and k-NN and have better classification performances and effortlessness. Using filter criteria such as information theory [85], consistency measures [86] and fuzzy set theory [87] have been employed with GA for feature selection.

Various enhancement on GA have been introduced for getting optimal subset using crossover and mutation variations. Srinivas and Patnaik [88] have proposed an approach to adjust both the crossover and mutation and to remove the local minima from the search space. Dugan and Erkoç [89] have introduced an extended GA concept as self-adaptive genetic algorithm (SAGA) to search the level of adaptation iteratively. Similarly, GA method is used in a two-stage filtering method, in first stage features ranking is evaluated and selected the top-most features which were passed in GA for optimal feature selection [90]. In contrast, Ghamisi and Benediktsson [91] introduced a different feature selection

scheme that is based on the amalgamation of a GA and PSO were proposed. Similarly, Cho et al. [50] have suggested the Quantum GA combined with an improved self-adaptive (SA) scheme that is used for solving Electromagnetic optimization problems.

In [92] proposed a novel Markov blanket-embedded genetic algorithm (MBEGA) for gene selection problem. In particular, embedded Markov blanket-based memetic operators add or delete features (or genes) from genetic algorithm (GA) solution so as to quickly improve the solution and fine-tune the search. Empirical results on synthetic and microarray benchmark datasets suggested that MBEGA was effective and efficient to eliminate irrelevant and redundant features based on both Markov blanket and predictive power in classifier model. Similarly, A new approach for predicting drug effectiveness was presented by [93]. The approach was based on machine learning and genetic algorithms. A global search mechanism, weighted decision

tree, decision-tree-based wrapper, correlation-based heuristic, and identification of intersecting feature sets were employed for selecting significant genes. The feature selection approach has resulted in 85% reduction of number of features. The relative increase in accuracy and specificity for the significant gene/SNP set was 10% and 3.2%.

Feature selection with GA by utilizing multi-objective methodologies began much consideration as a contrast with single objective feature determination technique. The vast majority of the multi-objective methods depend on non-dominated based (NSGA-II) or its variations [94, 95]. Author [96] aim was to preserve global diversity better by implementing Localized IMGA (LIMGA) and Dual Dynamic Migration Policy (DDMP). LIMGA creates unique evolution trends by using different kind of GA for each island. DDMP was a new migration policy which rules the individual migration. DDMP determines the state of an island according to its diversity and attractivity level. By determining its states, DDMP ensures the individual migrating to the correct island dynamically.

Even though there are more takes a report at multi-objective-based feature selection utilizing GA than utilizing other wrapper approaches, the capability of GA for MOFS has still not been thoroughly researched since attribute/feature selection is an intricate task that necessitates mainly composed MOGA to scan for the non-dominated solutions. Traditional genetic algorithm uses the two crucial tuning operators for feature selection, e.g., crossover and mutation, and provides chances to classify good feature or to discover the most exquisite feature sets, but this is a challenging task. In GA, the crucial, vital problem is when and how to apply the adjustment operators and parameter settings which influence their performance in feature selection.

## 3.2 PSO for feature selection

The combination of coding schemes, such as continuous and binary has been used PSO as single objective and multi-objective for feature selection in the filter as well as wrapper method [97]. The illustration of each swarm in the PSO for the FS is generally a string of bits. So, the dimension is equivalent to the complete attributes existent in the datasets. The representation of bit string as binary and real number in binary PSO and continuous PSO, respectively. At the point, when the binary characterization is utilized, one indicates that the relating feature is chosen and 0 infers that it is not chosen. At the point when continuous representation is used, a threshold as $\theta$ is generally decided the feature subsets from the data series, that is, if the value is higher than as $\theta$, the corresponding characteristic is selected, otherwise, discarded.

As we can see in mentioned literature [98], a large number of research work have been done on PSO for single

objective than multi-objective, and more article on wrapper than filter determination. For metaheuristic approaches, distinctive learning methods have been utilized with PSO to assess the decency of the chosen features, e.g., SVM, KNN, DT, RF [99] and ensemble approach [100]. To intensification the performance of FS problems, the researcher has been familiarized a large number of new PSO algorithms, containing initialization methods, a way of representation, fitness functions and search mechanisms. For example, Cervante et al. [101] established an inventive generation method for FS with PSO, which show that generation and expressively increased the performance of PSO. In three different variants of encoding schemes of PSO are continuous encoding [102], binary encoding [103] and mixture of both encoding [104].

In PSO, the best subset is evaluated with the assistance of fitness function. PSO as wrapper approaches, numerous current works utilized only the classification performance as a fitness function [105], which generally prompted expansive feature subsets. The process to solve the problem by simultaneously optimizing, two conflict objectives such as feature subsets and fitness solutions [106]. Research on multi-objective for FS, PSO [14], and is the focused to optimize the performance (for example, accuracy) and the number of features as two separate objectives. Usually, PSO of being easy to implement and structure of updating solution is easier as compared to GA algorithm. Nowadays, still, an open issue for emergent fresh PSO algorithms is mainly new search methods and parameter control strategies for large-scale feature selection. In [107], author have combined the modified discrete particle swarm optimization (PSO) and support vector machines (SVM) for tumor classification. The modified discrete PSO was applied to select genes, while SVM was used as the classifier. The proposed approach was used to the microarray data of 22 normal and 40 colon tumor tissues and shown good prediction performance.

## 3.3 Ant colony for feature selection

As the previous work on ACO has introduced for a variety of optimization applications and display that more mechanism on the ensemble as compare filter and hybrid methods [108]. In [109], a new undersampling method ACOSampling called based on the idea of ant colony optimization (ACO) to address this problem. The algorithm starts with feature selection technology to eliminate noisy genes in data. Then author randomly and repeatedly divided the original training set into two groups: training set and validation set. So far, most of the attention on single objective-based feature selection, and there are only a few methods has been investigated in multi-objective methods. Shunmugapriya and Kanmani [110] have proposed ACO method for feature selection using

the values of pheromone in ACO for the preferences the features and also recommended the modernized the pheromone traces of the edges that link every two different characteristics of the best solution so far. The experimental result has found a better solution regarding the accuracy of the classification by using a proposed method than the GA and PSO method. Similarly, in [111] author proposed a new modal using the ACO to concurrently select the characteristics and optimize the SVM parameters, in addition, weight optimizer method is also introduced for determining the probabilities of a specific feature through ACO. Similarly, for the selection of features, from [112] have used the two optimizers such as (ACO and DE) methods for integration, where DE optimization method was applied to find the subset of optimal features based on the solutions achieved from ACO.

The structure of ACO for node selection usually is a graph, where attributes represent as nodes to create a chart model. All ants signifies a subset of characteristics, in which the selected features are the visited nodes. Most of the cases, ACO based algorithms, the symptoms are entirely related to each other in the graph, but in [113] each characteristic was related to only two functions. Similarly, for feature selection by Aghdam et al. [114] have suggested a new way of presentation pattern to decrease the search space, during this only selected and not selected feature is connected using two edges. In most ACO-based methodologies, the order of execution has employed as the fitness evaluation. In articles [111, 112], the suitability of ants (features subset) has assessed by utilizing the average classification accuracy. However, the performance of individual features was also considered to enhance the performance additionally. The extraordinary procedure in [114] included both the classification accuracy and the number of highlights. Afterward, by expanding the effort on SO-ACO and a fuzzy classifier for including feature selection [113], Vieira et al. [115] offered an approach based on multi-objective and aimed to minimize the classification error and the number of features. In general, the volume of filter approaches is much sophisticated in ACO for feature selection, comparably as GA, PSO and DE methods.

## 3.4 Hybrid techniques for feature selection

A large number of hybrid metaheuristic methods has available for feature selection, including GA-PSO, PSO-GSA, TLBOSA, HS-GA, Hybrid Gravitational Search Algorithm (HGSA), the Hybrid Genetic Algorithm (HGA), and WOA-SA, CSPSO and TLBOGSA [116–118] have used for the multi-objective problems. DE has introduced to solve feature selection problems [119]. Most of the work focused on refining the DE search mechanisms, in addition, representation pattern has also been presented. Definitely, several relevant documents using hybrid techniques based on conventional

DE methods for feature selection has been presented. For example, Hancer et al. [120] investigated the DE with filter method, where DE algorithm found the optimal subsets of feature using filter technique. From the experimental result, it has observed that it has achieved the good result as compared to existing wrapper methods. Other technique such as memetic algorithms for feature selection integration with a single solution based optimization algorithm such as local search strategy and it provides a good combination of the wrapper and filter methods [121]. Furthermore, a large variety of optimization methods have also been employed to solve the complex problems (i.e., feature selection) including MS-PSO, FSS-EBNA, GAPSO, and MMSDE. A new hybrid wrapper approach which was based on cellular learning automata (CLA) with ant colony method (ACO) used to find the set of features which improve the classification accuracy [122]. CLA has applied due to its capability to learn and model complicated relationships. The selected features from last phase were evaluated using ROC curve and the most effective while smallest feature subset was determined.

In bioinformatics, data mining and machine learning domains, GA is another effective feature selection algorithm that extracts useful information from datasets. And, multiple extensions of conventional GA are proposed in recent decades [123]. Several significant articles are presented to solve the feature selection problem using hybrid techniques using conventional GA methods. Recently a diversity of metaheuristic methods have been applied to solve features selection problems. Since all metaheuristic algorithms have their strength and weakness, these concerns are beneficial for the further potential investigation to address different new challenges in the domain of feature selection.

## 3.5 Other feature selection techniques

In the unlabeled dataset, unsupervised learning plays a vital role in finding a hidden pattern. A primary example of unsupervised learning is clustering techniques [124], which tries to discover natural groupings in a set of objects without knowledge of class labels. The selection of functionalities that use unsupervised learning techniques are beyond the scope of this document and will not be analyzed in detail. But, in this section, we refereed some articles that perform a selection of unsupervised functions. Selecting functions that use unsupervised learning can provide better description and reliability of data than non-supervised learning [125]. Several documents attempting to resolve feature selection using unsupervised learning can be found in [126].

To address the relevant genes, Kalousis et al. [127] proposed a gene (feature) selection method using GA for grouping functions, in which a GA has been applied for finding the best cluster center value of a grouping method to group entities into different clusters. The characteristics of each

group were classified according to their distance values at the center of the group. Similarly, Khatami et al. [128] have applied PSO and sample pixels of an image are used to obtain the conversion matrix weights for color differentiation, while *K*-medoids provide a measure of the fitness conditions for the PSO procedure.

## 4 Measures in filter methods

To deal with the curse of dimensionality problem, we need to perform a dimensionality reduction task, i.e., feature selection, to assess the goodness of *m* features from the *n*-cardinality datasets. Generally, FS methods are classified into two basic methods (i.e., Subset evaluation and Subset generation) [129]. A subset generation procedure is a search technique that selects feature subsets based on specific search strategies, namely sequential search and random search. By using sets of evaluation methods, estimate the feature subset based on some criteria. The criterion used for feature selection is based on their dependency on algorithms. These are categorized in two ways, namely dependent criteria and independent criteria. The independent standards are correlation measures, distance measures, information measures, precision measures, and consistency measures. The four types of filter procedures in metaheuristic for feature selection can be seen as follows:

In the feature selection field using filter approach, mutual information based methods are most popular than all other. The use of information measures is mainly in four traditions.

1. Before using a metaheuristic technique, apply the information criterion to the rank of individual attributes. One of the examples of such kind of methods is Symmetrical Uncertainty (SU) and Mutual Information (MI) is evaluated based on filter rank, and then the top-most features are passed in ACO or GA-based feature selection [130].
2. One of the best examples of local search optimization algorithm is a memetic algorithm, information criterion is Mutual information [131] and symmetrical uncertainty [111] is applied as filter method to refine the poor solution quality obtained by a GA or PSO for feature selection.
3. Another approach as includes an information criterion for updating or search operator. MI has been covered the position vector by using the PSO optimization algorithm with the help of wrapper scheme as SVM [132].
4. Lastly, MI used as an objective function in a metaheuristic algorithm. It is the most proficient approach to feature selection. Based on the clue of "max-relevance and min-redundancy", MI method is used to quantify the redundancy within a subset of attributes and the relevance between characteristics and the labels. Different

metaheuristic algorithm objectives have maximized the significance and minimize the redundancy.

Correlation is a measurement of how strong two variables are linearly related. Authors in article [117], two correlation methods have proposed to assess the relevancy and redundancy of EA [133] and NSGA-II [134] and in [11] for the selection of features in two credit approval datasets. Similarly, Hu et al. [135] proposed a multi-population GA for selection of variables and also create a relationship between variable and labels, which were used as a filtering measure to assess the GA performance.

Distance measurements (DM) are also recognized as measures of separability, divergence, and discrimination. Signal to Noise (S2N) ratio has used for the selection of top-ranked features, and GA method used top-ranked features for the good classification [136]. To assess the goodness of each agent in PSO, the S2N ratio was applied for feature selection.

The measures of consistency are based on the fact that two samples, which have the unique features values, with the same label. Arauzo-Azofra et al. [137] introduced the first method based on filters for the selection of features based on the measurement of coherence. Regarding evolutionary calculation, GA was the first EA algorithm to use the coherence measurement for the selection of selection [138].

Fuzzy logic gives the degree of membership to the feature. It is also able to quantify the imprecision and uncertainty using a membership function, which can be used to assess the excellence of features. Using fuzzy fitness function two optimization algorithm such as PSO and GA have been applied for feature selection in SO [139] and MO approaches [140].

In recent years, the researcher has been investigated the used of more than one evaluation measures simultaneously in a unique FS algorithm which has become popular because every method has its advantages and disadvantages. Emmanouilidis et al. [141] have studied five different filtering measures in NSGA-II for the selection of features, including pairs of inconsistent examples as a measure of coherence, correlation of the attribute class as a measure of dependence/correlation, measurement of the inter-class distance, and entropy representation as for information measure.

In summary, various filter methods have been implemented in metaheuristic for selection of features. Most popular methods such as information measures, correlation measures and distance measures are relatively inexpensive from computational views, while coherence, the approximate set and measures based on silenced theories can handle better with noisy data. Compared to EA methods, generally, the performance of filters methods is gives poor results concerning classification accuracy. But, it can be less complex than

wrapper approaches [142] regarding computational time in large datasets. Therefore, developed a filter specific measures based on the characteristics of a metaheuristic technique can increase efficiency and effectiveness, which offers an important direction for future research.

### 4.1 Classification and regression method

This subsection provides a brief overview of two classifiers ant their accuracy is utilized as fitness function for feature selection. It is a basic approach to data mining that involves the construction of classifier. We present the support vector Machine (SVM) [143], and k-nearest neighbor (k-NN) [144] and LASSO methods.

#### 4.1.1 Support vector machine

Separating the feature vector and predict the correct class label is the major challenging task for the data mining classification algorithm. To resolve separation a set of feature vector which has different class memberships by the supervised machine-learning SVM model that analyses data for classification and constructing optimal decision planes classifier [143]. It got the popularity amongst the other machine learning classification. The main objective is drawing a hyper plane to split the dense feature vector dataset and margins between the sets of feature vectors maximum. To constructing an optimal decision plane, an iterative inductive learning model is used to minimize an error function $\wedge(w)$ defined in Eqs. 1 and 2.

$$\wedge(w) = \frac{1}{2}(WW^T) + C \sum_{i=m} \phi_i. \tag{1}$$

Subject to be constraints

$$Y_i[w^T k + \rho] \geq 1 - \lambda_i \quad and \quad i = 1, 2, 3, \ldots, m \tag{2}$$

where, c represents the capacity constant, $w$ as the vector of coefficients, and represents parameters for handling non-separable data (inputs), non-negative slack variables $\phi_i$ to the representation the deviations from the margin, $\rho$ a constant, $\lambda_i, with, i = 1, \ldots, m$ are the parameters to use noisy. For each training samples $I$, $x_i$ are the independent variables represented by actual class labels $y_i$. SVM accomplishes non-linear problem solved by kernel function k' which transforms data into higher space.

#### 4.1.2 k-nearest neighbour

The recent trends of classification problems, nearest neighbors (NN) is applied for distinguishes the classification unknown data point by its most intimate neighbor whose class is already known. It can solve real-world problems with the availability of inexpensive platform. Cover and Hart [145] investigated the purposed k-nearest neighbor (k-NN), and it concerns to find a group of $k$ nearest objects in the training samples which is nearest to the test object, and bases on the label of the majority of particular dataset [146] in this neighborhood.

#### 4.1.3 LASSO

Lasso a regularization technique that's useful for feature selection and to prevent over-fitting training data [65]. It works by penalizing the sum of absolute value of weights found by the regression. Lasso is great for reducing the feature space, because when $\lambda$ is sufficiently large, then many of the weights wi are driven to zero have been widely considered for the high-dimensional data analysis. Given a data set that consists of n observations $\{(x_i, l_i)|1 \leq i \leq n\}$. where $x_i = (x_{i1}, \ldots, x_{ip})$ is a p-dimensional vector of predictors and $l_i$ is a response variable, regression model is written as:

$$l_i = \beta x_i + \epsilon_i, \quad i = 1, \ldots, n,$$

where $\beta = (\beta 1, \ldots, \beta p)$ is a p-dimensional vector of regression coefficients and $\epsilon_i$ is a random error term which is assumed to be independently and identically normally distributed with mean of zero and variance of $\sigma^2$. It assumes that the response is mean-corrected and the predictors are standardized, so the intercept term is not included in the model. LASSO is a FS process based on a regression model with L1-norm regularization as:

$$\min_{\beta} \sum_{i=1}^{n} \left( l_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} \left| \beta_j \right|,$$

where $\lambda$ is non-negative hyper-parameter. Although LASSO has been successfully used in high-dimensional data.

### 4.2 Microarray datasets description

In this section, we demonstrate the comparative experimental study of the four feature selection method in five commonly used biomedical gene expression datasets namely Lung Cancer [147], Colon Cancer [148], Diffuse Large

**Table 1** Dataset description

| S. no. | Dataset | Number of sample | Number of genes | Number of classes |
|--------|---------|------------------|-----------------|-------------------|
| 1. | Lung cancer | 203 | 12,600 | 5 |
| 2. | Colon cancer | 62 | 2000 | 2 |
| 3. | DLBCL | 47 | 4026 | 2 |
| 4. | Leukemia | 72 | 7129 | 2 |
| 5. | SBRCT | 83 | 2308 | 4 |

B-cell Lymphoma (DLBCL) [149], Leukemia [150] and Small-Blue-Round-Cell Tumor (SBRCT) [151] which was downloaded from http://www.gems-system.org. A different set of features with the cancer classes are present in each dataset. The dataset descriptions in terms of sample number, number of genes, and labeled classes are summarized in Table 1.

## 4.3 Validation methods

To measure the acceptability of feature subset for classification, different error estimation strategies have been suggested. It is also vital to choose a validation method (classifier accuracy) for the selected classifier. Most study achieves the validation using either CV or bootstrap techniques [152]. In this paper, we use tenfold cross-validation, which is performed in all classifier for classification performance. It randomly splits the dataset into training and testing samples. In the training dataset that consists of 90% of the data samples and other testing subset consisting of 10% of the data samples to estimate the performance based on the confusion matrix.

## 4.4 Performance measures

We measure the classification performance with the help of two classifiers SVM and KNN with four performance measures accuracy, sensitivity, precision and F-measure. These performance measures are defined as follows.

1. Accuracy: To predict the percentage of correctly classified samples, it is formulated as in Eq. 3.

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} * 100 \quad (3)$$

2. Sensitivity: Percentage of positive instances that are predicted as positive. It is also called True Positive Rate (TPR) or Recall. It is formulated as in Eq. 4.

$$Sensitivity(Recall(Re)) = \frac{TP}{TP + FN} * 100 \quad (4)$$

3. Precision: It is the percentage of positive predictions that are correct. This is also called positive predicted value. It is formulated as in Eq. 5.

$$Precision(Pr) = \frac{TP}{TP + FP} * 100 \quad (5)$$

4. F-measure: It is a composite measure which favors algorithms with higher sensitivity and challenges those with higher specificity as in Eq. 6.

$$F\text{-}measure = \frac{2 * Pr * Re}{Pr + Re} * 100 \quad (6)$$

Here, TP, TN, FP, and FN are true positive, true negative, false positive and false negative in the independent datasets.

## 5 Experimental results

This section, demonstrates the experimental results obtained by four metaheuristic approaches based feature selections on five gene datasets with two classifiers SVM and KNN and one regression Lasso approach. We have implemented four metaheuristic methods such as ACO, PSO, DE, and GA with the performance of the two classifiers and Lasso regression as the objective function. All metaheuristic methods are performed on the identical machine, and MATLAB environment on 2.4 GHz Pentium Core i7 with 8 GB RAM running the Windows 8 operating system. For a fair comparison on the same computing environment, algorithms such as ACO, PSO, DE, and GA nature-inspired algorithms are run with a

**Table 3** Performance using SVM classification algorithm on dataset with the help of with four feature selection methods

| Dataset | Measures | GA | PSO | DE | ACO |
|---|---|---|---|---|---|
| Colon cancer | Acc | 88.26 | 87.89 | 89.81 | 84.1 |
| | #feat | 38 | 47 | 37 | 51 |
| | ETime | 521.95 | 759.42 | 785.75 | 860.23 |
| DLBCL | Acc | 87.09 | 85.22 | 76.56 | 75.61 |
| | #feat | 31 | 37 | 42 | 48 |
| | ETime | 978.52 | 725.45 | 787.43 | 865.32 |
| Prostate | Acc | 93.66 | 82.96 | 78.01 | 88.89 |
| | #feat | 2 | 27 | 22 | 30 |
| | ETime | 892.42 | 850.54 | 701.41 | 649.51 |
| SBRCT | Acc | 85.03 | 87.53 | 80.89 | 82.91 |
| | #feat | 33 | 36 | 41 | 49 |
| | ETime | 765.54 | 902.15 | 878.17 | 885.03 |
| Lung cancer | Acc | 91.03 | 90.09 | 83.89 | 81.37 |
| | #feat | 29 | 31 | 36 | 40 |
| | ETime | 689.21 | 958.14 | 785.43 | 896.32 |

**Table 2** Parameter setting

| S. no. | Parameters | GA | PSO | DE | ACO |
|---|---|---|---|---|---|
| 1. | Population size | 20 | 20 | 20 | 20 |
| 2. | Number of generations | 100 | 100 | 100 | 100 |
| 3. | Number of runs | 10 | 10 | 10 | 10 |
| 4. | Performance | Accuracy | Accuracy | Accuracy | Accuracy |

**Table 4** Performance using k-NN classification algorithm on dataset with the help of with four feature selection methods

| Dataset | Measures | GA | PSO | DE | ACO |
|---|---|---|---|---|---|
| Colon cancer | Acc | 82.09 | 81.89 | 79.81 | 80.18 |
| | #feat | 39 | 42 | 44 | 48 |
| | ETime | 926.95 | 681.42 | 885.75 | 920.23 |
| DLBCL | Acc | 80.89 | 82.09 | 78.96 | 77.71 |
| | #feat | 37 | 40 | 44 | 39 |
| | ETime | 828.35 | 925.15 | 887.54 | 925.32 |
| Prostate | Acc | 92.86 | 80.36 | 83.01 | 84.79 |
| | #feat | 27 | 31 | 33 | 35 |
| | ETime | 692.42 | 1050.25 | 954.41 | 836.01 |
| SBRCT | Acc | 81.03 | 82.03 | 78.89 | 74.75 |
| | #feat | 33 | 31 | 37 | 41 |
| | ETime | 505.54 | 549.15 | 652.17 | 754.03 |
| Lung cancer | Acc | 88.97 | 89.64 | 83.64 | 82.27 |
| | #feat | 34 | 38 | 36 | 41 |
| | ETime | 1089.21 | 1038.14 | 1205.43 | 996.32 |

**Table 5** Performance using LASSO algorithm on dataset with the help of with four feature selection methods

| Dataset | Measures | GA | PSO | DE | ACO |
|---|---|---|---|---|---|
| Colon cancer | Acc | 81.53 | 80.48 | 76.53 | 77.69 |
| | #feat | 48 | 89 | 75 | 66 |
| | ETime | 752.52 | 981.85 | 898.85 | 1020.53 |
| DLBCL | Acc | 78.85 | 79.89 | 76.42 | 76.85 |
| | #feat | 51 | 59 | 66 | 68 |
| | ETime | 712.30 | 856.93 | 915.37 | 987.58 |
| Prostate | Acc | 88.89 | 87.52 | 85.63 | 83.12 |
| | #feat | 37 | 30 | 36 | 34 |
| | ETime | 763.42 | 942.25 | 968.52 | 1025.36 |
| SBRCT | Acc | 77.36 | 75.12 | 74.37 | 75.36 |
| | #feat | 45 | 40 | 39 | 61 |
| | ETime | 735.12 | 731.61 | 732.22 | 786.90 |
| Lung cancer | Acc | 84.22 | 87.53 | 85.32 | 84.51 |
| | #feat | 52 | 59 | 62 | 68 |
| | ETime | 901.25 | 997.23 | 1045.43 | 1236.36 |

suitable parameter setting which can be seen in Table 2. Furthermore, selected the top-most ranked subset is 200 on gene datasets. To complete the process, the procedure is repetitive as ten iterations to allow each part of data to become as test data. As shown in Tables 3 and 4, the accuracy of the classifier and the corresponding mean values of number of features selected by respective approach in respective gene datasets. Here, mean of tenfold-cross-validation approach is used for comparative analysis with two measures such as accuracy (acc) and mean of number of features selected in each approach (#feat) and execution time (in s).

As wrapper approach such as SVM uses a fitness function in Table 3. It shows the performance of metaheuristic methods on small gene datasets. As obtained results, GA method indicates the best classification accuracy achieved by the classifier i.e., SVM as 93.56% in DLBCL dataset with 24 genes and lowest classification accuracy as 76.91% in Colon Cancer Dataset using ACO wrapper with 48 genes. The best results among all metaheuristic for feature selection have been highlighted (Table 5).

Table 4 shows the performance of wrapper method over other approaches using k-NN. The GA method indicates the best classification accuracy achieved by the classifier as 92.86% in DLBCL dataset with 27 features and lowest classification accuracy as 74.75% in Leukaemia Dataset using ACO wrapper with 41 features. The best results among all metaheuristic for feature selection have been highlighted.

As one of the most attractive regression methods such as ridge, Lasso and Elastic-Net [153] can popularly employed in both machine learning and biomedicine. The comparison of classification accuracies and gene selection of four methods on five biological data over 10 runs are summarized in Table 6. The average classification accuracy of GA is 81.53 percent, which is 1.3, 3.3, 5.6 percent higher compared to PSO, DE and ACO, respectively in colon cancer data.

In addition to the above cases, the convergence to an optimal solution is an essential issue in the feature selection problems, and comparison between the metaheuristics algorithm has shown in Fig. 5a, b with respective classifiers. Figures 5, 6, 7, 8 and 9 shows convergence curves of the

**Fig. 5** Convergence curve for feature selection on Leukemia dataset with **a** KNN, **b** SVM
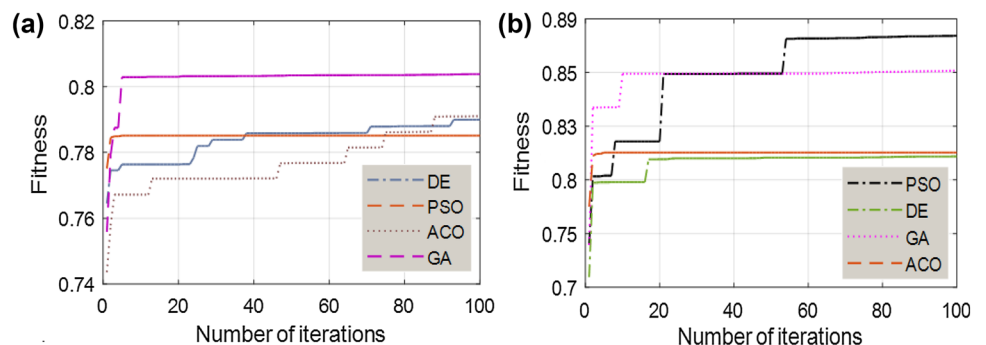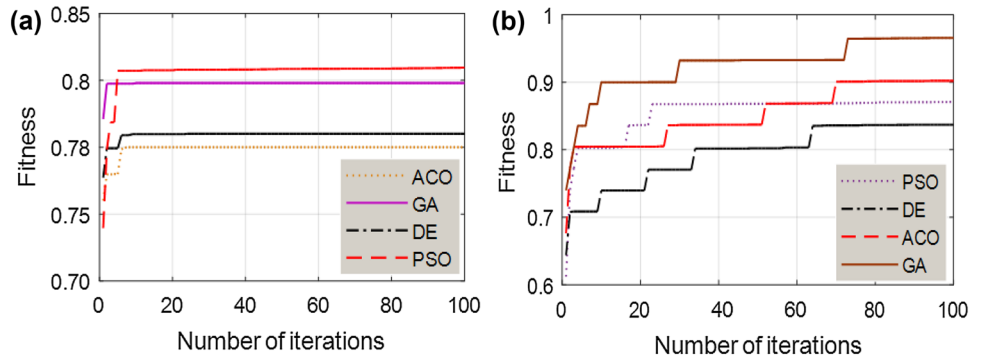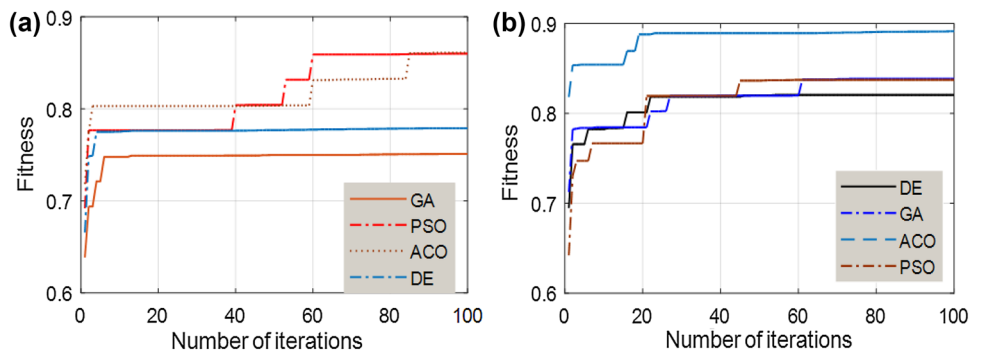
**Fig. 6** Convergence curve for feature selection on SBRCT dataset with **a** KNN, **b** SVM
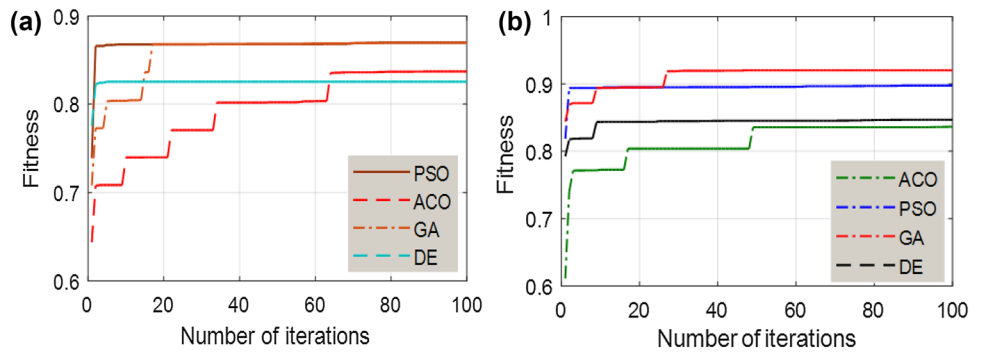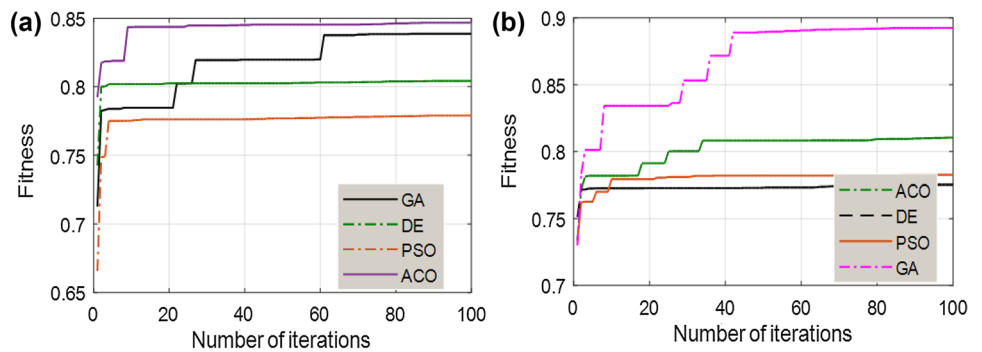


**Fig. 7** Convergence curve for feature selection on DLBCL dataset with **a** KNN, **b** SVM



**Fig. 8** Convergence curve for feature selection on lung cancer dataset with **a** KNN, **b** SVM



**Fig. 9** Convergence curve for feature selection on Colon dataset with **a** KNN, **b** SVM



metaheuristic methods, where, the x-axis presents the number of iterations, and the y-axis presents the performance of the subset containing the specified number of iterations.

## 5.1 Applications

In this section, we have shown the metaheuristic for feature selection in Table 7. It can be seen that metaheuristic

**Table 6** Solution quality of each metaheuristic on selected gene subsets using SVM classifier as the fitness function

| Dataset | Measure | GA | PSO | ACO | DE |
|---|---|---|---|---|---|
| Colon cancer | Sensitivity | 83.52 | 79.17 | 71.51 | 75.48 |
| | Precision | 82.24 | 78.81 | 69.97 | 74.01 |
| | F-measure | 83.07 | 78.84 | 71.57 | 72.19 |
| SRBCT | Sensitivity | 90.85 | 89.71 | 80.19 | 82.15 |
| | Precision | 89.75 | 88.16 | 79.79 | 78.75 |
| | F-measure | 92.10 | 89.42 | 80.14 | 81.56 |
| Lung cancer | Sensitivity | 87.12 | 85.78 | 82.89 | 86.74 |
| | Precision | 86.84 | 85.16 | 82.06 | 88.79 |
| | F-measure | 85.89 | 84.72 | 79.96 | 87.15 |
| DLBCL | Sensitivity | 88.00 | 81.38 | 85.37 | 76.36 |
| | Precision | 90.11 | 81.39 | 85.52 | 77.91 |
| | F-measure | 88.89 | 80.46 | 95.78 | 79.86 |
| Leukaemia | Sensitivity | 84.52 | 85.23 | 81.46 | 79.89 |
| | Precision | 83.05 | 86.94 | 80.36 | 79.15 |
| | F-measure | 82.85 | 87.14 | 80.31 | 80.05 |

approaches have been applied to a variety of areas. As, results are depicted in Tables 3 and 4 in terms of classification performance and number of features, SVM achieve the better accuracy as compared to KNN. So, for more analysis on SVM based feature selection, to evaluate the classification performance of the selected subset using four metaheuristic algorithms, i.e., GA, PSO, ACO, and DE, respectively on five microarray datasets. The classification sensitivity, specificity, and F-measure are presented in this Table 6. According to Table 6, support vector machine as a fitness function gives the highest performance with GA method almost all datasets. This result shows that the genetic algorithm (GA) is a robust metaheuristic as compare to other metaheuristic used in this experiment.

**Table 7** Application of metaheuristic algorithm used in different applications

| Sub areas and details | Types of metaheuristic applied and references |
|---|---|
| Electrical power systems | |
| Economic dispatch | Oppositional TLBO [80], Hybrid GA-MGA [81] |
| Optimal power flow | Multi-objective teaching-learning-based optimisation algorithm [142] |
| Micro-grid operation | Fuzzy-based clustering and teaching-learning based optimization [82] |
| Control systems and robotics | |
| System identification | Elitist teaching learning opposition based algorithm (ETLOBA)[83] |
| Controller design and tuning | Hybrid PSO-GS [84] |
| Bioinformatics | |
| Gene regulatory networks | GA/PSO [85] |
| Micro-array data analysis | Multimedia teaching and learning [86] |
| Protein folding | PSO-AB [87] |
| Operon prediction | Teaching-learning-based optimization [88] |
| Chemical engineering | |
| Fault diagnosis of chemical process | Wavelet LS+SVM integrated improved PSO [89] |
| Parameter estimation of chemical process | MIMO ANFIS-PSO [90] |
| Pattern recognition and image processing | |
| Data clustering | ABC [154], auto teaching learning-based optimization [94] |
| Pixel clustering and region based image segmentation | ABC with the neural system [95] |
| Image registration and enhancement | Artificial bee colony [99] |
| Artificial neural networks | |
| Training of wavelet neural networks | Fuzzy wavelet neural network (FWNN) [100] |
| Signal processing | |
| Digital filter design | Bat algorithm [101] |
| Others | |
| Structure designing and machining processes | Cuckoo search and EFS [102, 155] |
| Network Security | Cuckoo search (CS) [103] |
| Text mining | Differential evolution [104] |
| Multi-class cooperative | Teaching-based optimization algorithm with simulated annealing [105] |
| Hyperspectral image processing | Immune clonal strategy (ICS)-based effective band selection [156] |

# 6 Open challenges

There is no single feature selection algorithm suitable for all classification problems. The problem of selection of features depends on what exactly the task is, and so the issue of classification [157]. What seems to be a useful feature of a problem can be tremendous for another. Despising the suitability, the success and the promise of metaheuristics for the selection of the features, there are still some difficulties and challenges, which is analyzing here.

## 6.1 Scalability

Due to the tendency of large data [106, 151], getting significant features is the extremely difficult and risky problem in the FS process. The selection of the features of a data set above 300 features called as large-scale dataset for feature selection [108]. However, today the number of features in numerous domains, for example, gene analysis have hundreds or more than hundreds number of features without much stretch. This expands the computational cost and requires innovative search methods. However, the two aspects have their issues, so issue cannot be settled by expanding the computational power. To overcome the issues mentioned above, a large number of metaheuristic algorithms have been investigated by researchers to solve high dimensional problems [110] for feature selection. While, hybrid subset selections based algorithms have been introduced that investigated the trade-off between original features set and classification performance of the model with selected feature subsets [135]. The first phase of hybrid model, in general, is a filter part. With the help of the filtering approach, evaluates the importance of each feature and screen out an irrelevant subset of feature from the datasets. It gives much informative subset as compared to the original data set. By picking features from the candidate set of features which makes a trade-off between the predictive power of the candidate feature (relevance to the class vector) and its independence from all features previously selected. While, the subsequent method (wrapper) accelerates the search with filtered genes, and to optimize classification performance.

## 6.2 Computational cost

The major issue in the feature selection method is a computational expansion for solving the high dimensional problem. In metaheuristic-based feature selection problem, the computational cost is a thoughtful issue, as they often involve a large number of assessments [158]. In general,

the performance of filters methods such as mutual information [159, 160] has significant improved results toward to the classification accuracy but still it needs improvements and can be less complex than wrapper approaches regarding computational time in large datasets. To overcome the current challenge like a computational cost, two significant and effective factors must be considered: an efficient search technique and a rapid assessment measure [161]. It is emphasized that the parallel nature of metaheuristics is adequate such as grid computing, a graphics processing unit, and cloud computing that can be used to accelerate the process.

## 6.3 Search mechanisms

As we know, feature selection (FS) problem is a computationally complex problem which is a NP-hard problem. It has a large composite solution space [162]. To solve this problem, it requires a robust optimization search method, but recent valuable algorithms still have a necessary great improvement in finding a potential solution. For improved results, newly developed metaheuristic has the capability to explore the complete search space and also be able to exploit the local regions when needed. The new search scheme may include a local search (to form new memetic algorithms), the hybridization of different search schemes, the hybridization of metaheuristic with conventional algorithms [121]. In nature, metaheuristic algorithms are stochastic and approximate for optimizing the problems. It can have generated diversified optimal fitness when the different local solution is considered. When optimal fitness value is the same, then we select the smallest number of selected bits. In addition, the newly proposed search algorithms with high stability is also an important task.

## 6.4 Measures

The fitness evaluation as a function for suitability analysis is significant aspects in the metaheuristic for the features selection. It effectively influences the computational period, the performance of prediction and the search space landscape. In general, most of the computational period is used in the assessment procedure for optimization of metaheuristic algorithms and also for many filter-based approaches [69, 71]. Although there are some speedy assessment measures, such as information gain [163] to evaluate the characteristics individually rather than a group of features. Ignoring interactions between characteristics produces subgroups with redundancy and lack complementary features [164].

## 6.5 Dataset structure

The number of variables and instances in a data set is significant influences the work and design of experiments in feature selection problems [165]. A considerable portion of the current features selection methods are intended for actual data and depend on the suspicion that features have no express correlations to each other. As such, they overlook the inherent structures between the attributes. For instance, these FS methods can choose a similar subset of features even if the features have been reorganized. In some data mining applications, features indicate different kinds of structures, such as spatial or temporal softness, disjointed groups, stacked groups, trees, and graphics. When applying FS algorithms in data sets with structured characteristics, is useful to explicitly incorporate this prior knowledge, which could improve post-learning activities such as classification and clustering.

## 7 Conclusion

Although metaheuristic algorithms for feature selection have accomplished various success but they still face challenges and also their potential has not been fully recognized. In the metaheuristic algorithm, one of the main difficulties is scalability, since both the number of functions and the number of instances is increasing in the microarray datasets. In literature, a variety of conventional metaheuristic algorithms have been applied on microarray datasets based on the following attention: simplicity, stability, robustness, and computational requirements. The metaheuristic methods always provide benefits such as giving insight into the data, better classifier model, enhance generalization, and identification of irrelevant variables for feature selection. This survey shows a series of metaheuristic algorithms for addressing feature selection tasks and focused on key factors such as representation, search mechanisms, performance measures, and structure. In addition, experimental work has also conducted using metaheuristic approaches on large number of datasets for example, Colon, Leukemia, etc. Furthermore, recent examples of metaheuristic algorithms for feature selection from the literature have been presented with a summary of some noticeable applications, and also some issues have examined. Finally, a few proposals prescribed that will help to develop novel and effective metaheuristic approaches and solve the different kinds of problems.

## References

1. Dwivedi S, Vardhan M, Tripathi S, Shukla AK (2019) Implementation of adaptive scheme in evolutionary technique for anomaly-based intrusion detection. Evol Intell 1–15
2. Ang JC, Mirzal A, Haron H, Hamed HNA (2016) Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection. IEEE/ACM Trans Comput Biol Bioinform 13(5):971–989
3. Feng C-M, Xu Y, Liu J-X, Gao Y-L, Zheng C-H (2019) Supervised discriminative sparse PCA for com-characteristic gene selection and tumor classification on multiview biological data. IEEE Trans Neural Netw Learn Syst
4. Shukla AK, Singh P, Vardhan M (2019) Medical diagnosis of Parkinson disease driven by multiple preprocessing technique with scarce Lee Silverman voice treatment data. In: Engineering vibration, communication and information processing. Springer, Berlin, pp 407–421
5. Chen X, Huang JZ, Wu Q, Yang M (2017) Subspace weighting co-clustering of gene expression data. IEEE/ACM Trans Comput Biol Bioinform 16:352–364
6. Wang H, Jing X, Niu B (2017) A discrete bacterial algorithm for feature selection in classification of microarray gene expression cancer data. Knowl-Based Syst 126:8–19
7. Shukla AK, Singh P, Vardhan M (2018) A hybrid gene selection method for microarray recognition. Biocybern Biomed Eng 38(4):975–991
8. Yang W-H, Dai D-Q, Yan H (2008) Feature extraction and uncorrelated discriminant analysis for high-dimensional data. IEEE Trans Knowl Data Eng 20(5):601–614
9. Tan X, Deng L, Yang Y, Qu Q, Wen L (2019) Optimized regularized linear discriminant analysis for feature extraction in face recognition. Evol Intell 12(1):73–82
10. Cao P, Liu X, Yang J, Zhao D, Li W, Huang M, Zaiane O (2017) A multi-kernel based framework for heterogeneous feature selection and over-sampling for computer-aided detection of pulmonary nodules. Pattern Recogn 64:327–346
11. Tripathi D, Edla DR, Cheruku R, Kuppili V (2019) A novel hybrid credit scoring model based on ensemble feature selection and multilayer ensemble classification. Comput Intell 35(2):371–394
12. Shukla AK, Singh P, Vardhan M (2020) An adaptive inertia weight teaching-learning-based optimization algorithm and its applications. Appl Math Model 77:309–326
13. Loughran R, Agapitos A, Kattan A, Brabazon A, O'Neill M (2017) Feature selection for speaker verification using genetic programming. Evol Intell 10(1–2):1–21
14. Xue B, Zhang M, Browne WN (2013) Particle swarm optimization for feature selection in classification: a multi-objective approach. IEEE Trans Cybern 43(6):1656–1671
15. Baykasoğlu A, Hamzadayi A, Köse SY (2014) Testing the performance of teaching-learning based optimization (TLBO) algorithm on combinatorial problems: flow shop and job shop scheduling cases. Inf Sci 276:204–218
16. Karaboga D, Gorkemli B, Ozturk C, Karaboga N (2014) A comprehensive survey: artificial bee colony (ABC) algorithm and applications. Artif Intell Rev 42(1):21–57
17. Kim K-J, Cho S-B (2008) An evolutionary algorithm approach to optimal ensemble classifiers for DNA microarray data analysis. IEEE Trans Evol Comput 12(3):377–388
18. Shukla AK, Singh P, Vardhan M (2019) A hybrid framework for optimal feature subset selection. J Intell Fuzzy Syst 36(3):2247–2259

19. Shukla A, Singh P, Vardhan M (2018) A two-stage gene selection method for biomarker discovery from microarray data for cancer classification. Chemom Intell Lab Syst 183:47–58

20. Chuang L-Y, Chang H-W, Tu C-J, Yang C-H (2008) Improved binary PSO for feature selection using gene expression data. Comput Biol Chem 32(1):29–38

21. Shukla AK, Singh P, Vardhan M (2018) Neighbour teaching learning based optimization for global optimization problems. J Intell Fuzzy Syst 34(3):1583–1594

22. Liu C, Wang W, Zhao Q, Shen X, Konan M (2017) A new feature selection method based on a validity index of feature subset. Pattern Recogn Lett 92:1–8

23. Ghaemi M, Feizi-Derakhshi M-R (2016) Feature selection using forest optimization algorithm. Pattern Recogn 60:121–129

24. Aguilar-Rivera R, Valenzuela-Rendón M, Rodríguez-Ortiz J (2015) Genetic algorithms and darwinian approaches in financial applications: a survey. Expert Syst Appl 42(21):7684–7697

25. Nikolić M, Teodorović D (2013) Empirical study of the bee colony optimization (BCO) algorithm. Expert Syst Appl 40(11):4609–4620

26. Chen Y-P, Li Y, Wang G, Zheng Y-F, Xu Q, Fan J-H, Cui X-T (2017) A novel bacterial foraging optimization algorithm for feature selection. Expert Syst Appl 83:1–17

27. Rashedi E, Nezamabadi-Pour H, Saryazdi S (2009) GSA: a gravitational search algorithm. Inf Sci 179(13):2232–2248

28. Shukla AK, Singh P, Vardhan M (2019) A new hybrid wrapper TLBO and SA with SVM approach for gene expression datasets. Inf Sci 503:238–254

29. Boussaïd I, Lepagnot J, Siarry P (2013) A survey on optimization metaheuristics. Inf Sci 237:82–117

30. Rao RV, Savsani VJ, Vakharia D (2012) Teaching-learning-based optimization: an optimization method for continuous non-linear large scale problems. Inf Sci 183(1):1–15

31. Kohavi R, John GH (1997) Wrappers for feature subset selection. Artif Intell 97(1–2):273–324

32. Yu L, Liu H (2004) Efficient feature selection via analysis of relevance and redundancy. J Mach Learn Res 5(Oct):1205–1224

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

34. Naghibi T, Hoffmann S, Pfister B (2015) A semidefinite programming based search strategy for feature selection with mutual information measure. IEEE Trans Pattern Anal Mach Intell 37(8):1529–1541

35. Maji P (2012) Mutual information-based supervised attribute clustering for microarray sample classification. IEEE Trans Knowl Data Eng 24(1):127–140

36. Liu H, Sun J, Liu L, Zhang H (2009) Feature selection with dynamic mutual information. Pattern Recogn 42(7):1330–1339

37. Shukla AK, Singh P, Vardhan M (2019) DNA gene expression analysis on diffuse large b-cell lymphoma (DLBCL) based on filter selection method with supervised classification method. In: Computational intelligence in data mining. Springer, Berlin, pp 783–792

38. Shukla AK, Tripathi D (2019) Identification of potential biomarkers on microarray data using distributed gene selection approach. Math Biosci 315:108230

39. Pashaei E, Aydin N (2017) Binary black hole algorithm for feature selection and classification on biological data. Appl Soft Comput 56:94–106

40. Wang A, An N, Yang J, Chen G, Li L, Alterovitz G (2017) Wrapper-based gene selection with markov blanket. Comput Biol Med 81:11–23

41. Blum AL, Langley P (1997) Selection of relevant features and examples in machine learning. Artif Intell 97(1–2):245–271

42. Caruana R, Freitag D (1994) Greedy attribute selection. In: Machine learning proceedings 1994. Elsevier, Amsterdam, pp 28–36

43. Wang H, Bell D, Murtagh F (1999) Axiomatic approach to feature subset selection based on relevance. IEEE Trans Pattern Anal Mach Intell 21(3):271–277

44. Wang Y, Yang X-G, Lu Y (2019) Informative gene selection for microarray classification via adaptive elastic net with conditional mutual information. Appl Math Model 71:286–297

45. Mao KZ, Tang W (2011) Recursive Mahalanobis separability measure for gene subset selection. IEEE/ACM Trans Comput Biology Bioinform 8(1):266–272

46. Tang J, Alelyani S, Liu H (2014) Feature selection for classification: a review. In: Data classification: algorithms and applications, p 37

47. Gheyas IA, Smith LS (2010) Feature subset selection in large dimensionality domains. Pattern Recogn 43(1):5–13

48. Tang J, Zhou S (2016) A new approach for feature selection from microarray data based on mutual information. IEEE/ACM Trans Comput Biol Bioinform 13(6):1004–1015

49. Shukla AK, Singh P, Vardhan M (2018) Predicting alcohol consumption behaviours of the secondary level students. In: Proceedings of 3rd international conference on internet of things and connected technologies (ICIoTCT), pp 8–14

50. Cho JH, Lee D-J, Park J-I, Chun M-G (2013) Hybrid feature selection using genetic algorithm and information theory. Int J Fuzzy Log Intell Syst 13(1):73–82

51. Singh P, Shukla A, Vardhan M (2017) Hybrid approach for gene selection and classification using filter and genetic algorithm. In: 2017 International conference on inventive computing and informatics (ICICI). IEEE, pp 832–837

52. Yang P, Hwa Yang Y, Zhou BB, Zomaya AY (2010) A review of ensemble methods in bioinformatics. Curr Bioinform 5(4):296–308

53. Edla DR, Tripathi D, Cheruku R, Kuppili V (2018) An efficient multi-layer ensemble framework with BPSOGSA-based feature selection for credit scoring data analysis. Arabian J Sci Eng 43(12):6909–6928

54. Osareh A, Shadgar B (2013) An efficient ensemble learning method for gene microarray classification. BioMed Res Int

55. Blattman C, Jamison J, Koroknay-Palicz T, Rodrigues K, Sheridan M (2016) Measuring the measurement error: a method to qualitatively validate survey data. J Dev Econ 120:99–112

56. Narendra PM, Fukunaga K (1977) A branch and bound algorithm for feature subset selection. IEEE Trans Comput 9:917–922

57. Butler-Yeoman T, Xue B, Zhang M (2015) Particle swarm optimisation for feature selection: a hybrid filter-wrapper approach. In: CEC, pp 2428–2435

58. Pudil P, Novovičová J, Kittler J (1994) Floating search methods in feature selection. Pattern Recogn Lett 15(11):1119–1125

59. Han F, Yang C, Wu Y-Q, Zhu J-S, Ling Q-H, Song Y-Q, Huang D-S (2017) A gene selection method for microarray data based on binary PSO encoding gene-to-class sensitivity information. IEEE/ACM Trans Comput Biol Bioinform 14(1):85–96

60. Zhang Y, Gong D-W, Cheng J (2017) Multi-objective particle swarm optimization approach for cost-based feature selection in classification. IEEE/ACM Trans Comput Biol Bioinform (TCBB) 14(1):64–75

61. Suykens JA, Vandewalle J (1999) Least squares support vector machine classifiers. Neural Process Lett 9(3):293–300

62. Friedman N, Geiger D, Goldszmidt M (1997) Bayesian network classifiers. Mach Learn 29(2–3):131–163

63. Kovács ZM, Guerrieri R (1991) A generalization technique for nearest-neighbor classifiers. In: Neural networks. 1991 IEEE international joint conference on. IEEE, pp 1782–1788

64. Learning M (2009) Decision-tree learning

65. Tibshirani R (1996) Regression shrinkage and selection via the Lasso. J R Stat Soc Ser B (Methodol) 58(1):267–288

66. Zhang Y, Ding X, Liu Y, Griffin P (1996) An artificial neural network approach to transformer fault diagnosis. IEEE Power Eng Rev 16(10):55–55

67. Sharma A, Paliwal KK, Imoto S, Miyano S (2014) A feature selection method using improved regularized linear discriminant analysis. Mach Vis Appl 25(3):775–786

68. Frohlich H, Chapelle O, Scholkopf B (2003) Feature selection for support vector machines by means of genetic algorithm. In: Tools with artificial intelligence, 2003. Proceedings. 15th IEEE international conference on. IEEE, pp 142–148

69. Bennasar M, Hicks Y, Setchi R (2015) Feature selection using joint mutual information maximisation. Expert Syst Appl 42(22):8520–8532

70. Roobaert D, Karakoulas G, Chawla NV (2006) Information gain, correlation and support vector machines. In: Feature extraction. Springer, Berlin, pp 463–470

71. Sun Y (2007) Iterative RELIEF for feature weighting: algorithms, theories, and applications. IEEE Trans Pattern Anal Mach Intell 29(6)

72. Jin X, Xu A, Bie R, Guo P (2006) Machine learning techniques and chi-square feature selection for cancer classification using sage gene expression profiles. In: International workshop on data mining for biomedical applications. Springer, Berlin, pp 106–115

73. Seijo-Pardo B, Porto-Díaz I, Bolón-Canedo V, Alonso-Betanzos A (2017) Ensemble feature selection: homogeneous and heterogeneous approaches. Knowl-Based Syst 118:124–139

74. Cover TM, Thomas JA (1991) Entropy, relative entropy and mutual information. Elem Inf Theory 2:1–55

75. Fan X, Li X (2017) Minimizing probing cost with MRMR feature selection in network monitoring. IEEE Commun Lett 21(11):2400–2403

76. Liu J, Lin Y, Lin M, Wu S, Zhang J (2017) Feature selection based on quality of information. Neurocomputing 225:11–22

77. Shukla AK, Singh P, Vardhan M (2018) An empirical study on multi-objective swarm algorithm for standard multi-objective benchmark problems. In: Proceedings of 3rd international conference on internet of things and connected technologies (ICIoTCT), pp 832–837

78. Jones G (1998) Genetic and evolutionary algorithms. In: von Rague P (ed) Encyclopedia of computational chemistry

79. Van den Bergh F, Engelbrecht AP (2004) A cooperative approach to particle swarm optimization. IEEE Trans Evol Comput 8(3):225–239

80. Karaboga D, Basturk B (2008) On the performance of artificial bee colony (ABC) algorithm. Appl Soft Comput 8(1):687–697

81. Gajjar S, Sarkar M, Dasgupta K (2016) FAMACROW: fuzzy and ant colony optimization based combined mac, routing, and unequal clustering cross-layer protocol for wireless sensor networks. Appl Soft Comput 43:235–247

82. Holland J, Goldberg D (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley, Reading

83. Jiang S, Chin K-S, Wang L, Qu G, Tsui KL (2017) Modified genetic algorithm-based feature selection combined with pretrained deep neural network for demand forecasting in outpatient department. Expert Syst Appl 82:216–230

84. Chyzhyk D, Savio A, Graña M (2014) Evolutionary ELM wrapper feature selection for Alzheimer's disease cad on anatomical brain MRI. Neurocomputing 128:73–80

85. Li Y, Zhang S, Zeng X (2009) Research of multi-population agent genetic algorithm for feature selection. Expert Syst Appl 36(9):11570–11581

86. De Stefano C, Fontanella F, Marrocco C, Di Freca AS (2014) A GA-based feature selection approach with an application to handwritten character recognition. Pattern Recogn Lett 35:130–141

87. Herrera F, Lozano M (2009) Fuzzy evolutionary algorithms and genetic fuzzy systems: a positive collaboration between evolutionary algorithms and fuzzy systems. In: Computational intelligence. Springer, Berlin, pp 83–130

88. Srinivas M, Patnaik LM (1994) Adaptive probabilities of crossover and mutation in genetic algorithms. IEEE Trans Syst Man Cybern 24(4):656–667

89. Dugan N, Erkoç Ş (2009) Genetic algorithms in application to the geometry optimization of nanoparticles. Algorithms 2(1):410–428

90. El Akadi A, Amine A, El Ouardighi A, Aboutajdine D (2011) A two-stage gene selection scheme utilizing MRMR filter and ga wrapper. Knowl Inf Syst 26(3):487–500

91. Ghamisi P, Benediktsson JA (2015) Feature selection based on hybridization of genetic algorithm and particle swarm optimization. IEEE Geosci Remote Sens Lett 12(2):309–313

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

93. Shah SC, Kusiak A (2004) Data mining and genetic algorithm based gene/SNP selection. Artif Intell Med 31(3):183–196

94. Tavana M, Li Z, Mobin M, Komaki M, Teymourian E (2016) Multi-objective control chart design optimization using NSGA-III and MOPSO enhanced with DEA and TOPSIS. Expert Syst Appl 50:17–39

95. Li H, Zhang Q (2009) Multiobjective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II. IEEE Trans Evol Comput 13(2):284–302

96. Gozali AA, Fujimura S (2019) DM-LIMGA: dual migration localized island model genetic algorithm—a better diversity preserver island model. Evol Intell 1–13

97. Elbes M, Alzubi S, Kanan T, Al-Fuqaha A, Hawashin B (2019) A survey on particle swarm optimization with emphasis on engineering and network applications. Evol Intell 1–17

98. Wei Y, Qiqiang L (2004) Survey on particle swarm optimization algorithm. Eng Sci 5(5):87–94

99. Zainudin M, Sulaiman M, Mustapha N, Perumal T, Nazri A, Mohamed R, Manaf S (2017) Feature selection optimization using hybrid relief-f with self-adaptive differential evolution. Int J Intell Eng Syst 10(3):21–29

100. Dara S, Banka H (2014) A binary PSO feature selection algorithm for gene expression data. In: Advances in communication and computing technologies (ICACACT), 2014 international conference on. IEEE, pp 1–6

101. Cervante L, Xue B, Zhang M, Shang L (2012) "Binary particle swarm optimisation for feature selection: a filter based approach. In: Evolutionary computation (CEC), 2012 IEEE Congress on. IEEE, pp 1–8

102. Lin T-L, Horng S-J, Kao T-W, Chen Y-H, Run R-S, Chen R-J, Lai J-L, Kuo I-H (2010) An efficient job-shop scheduling algorithm based on particle swarm optimization. Expert Syst Appl 37(3):2629–2636

103. Banks A, Vincent J, Anyakoha C (2007) A review of particle swarm optimization. Part I: background and development. Nat Comput 6(4):467–484

104. Huang C-L, Dun J-F (2008) A distributed PSO-SVM hybrid system with feature selection and parameter optimization. Appl Soft Comput 8(4):1381–1391

105. Mandal M, Mondal J, Mukhopadhyay A (2015) A PSO-based approach for pathway marker identification from gene expression data. IEEE Trans Nanobiosci 14(6):591–597

106. Reyes-Sierra M, Coello CC et al (2006) Multi-objective particle swarm optimizers: a survey of the state-of-the-art. Int J Comput Intell Res 2(3):287–308

107. Shen Q, Shi W-M, Kong W, Ye B-X (2007) A combination of modified particle swarm optimization algorithm and support vector machine for gene selection and tumor classification. Talanta 71(4):1679–1683

108. Jensen R, Shen Q (2003) Finding rough set reducts with ant colony optimization. In: Proceedings of the 2003 UK workshop on computational intelligence, vol 1(2), pp 15–22

109. Yu H, Ni J, Zhao J (2013) Acosampling: an ant colony optimization-based undersampling method for classifying imbalanced DNA microarray data. Neurocomputing 101:309–318

110. Shunmugapriya P, Kanmani S (2017) A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC hybrid). Swarm Evol Comput 36:27–36

111. Tabakhi S, Moradi P (2015) Relevance-redundancy feature selection based on ant colony optimization. Pattern Recogn 48(9):2798–2811

112. Moradi P, Rostami M (2015) Integration of graph clustering with ant colony optimization for feature selection. Knowl-Based Syst 84:144–161

113. Chen Y-C, Pal NR, Chung I-F (2012) An integrated mechanism for feature selection and fuzzy rule extraction for classification. IEEE Trans Fuzzy Syst 20(4):683–698

114. Aghdam MH, Ghasem-Aghaee N, Basiri ME (2009) Text feature selection using ant colony optimization. Expert Syst Appl 36(3):6843–6853

115. Vieira SM, da Costa Sousa JM, Kaymak U, Dubois D, Sousa J, Carvalho J (2009) Feature selection using fuzzy objective functions. In: IFSA/EUSFLAT conference, pp 1673–1678

116. Ibrahim AM, Tawhid MA (2019) A hybridization of cuckoo search and particle swarm optimization for solving nonlinear systems. Evol Intell 1–21

117. Shukla AK, Singh P, Vardhan M (2018) Hybrid TLBO-GSA strategy for constrained and unconstrained engineering optimization functions. Hybrid Metaheuristics Res Appl 84:41

118. Hsu H-H, Hsieh C-W, Lu M-D (2011) Hybrid feature selection by combining filters and wrappers. Expert Syst Appl 38(7):8144–8150

119. Khushaba RN, Al-Ani A, Al-Jumaily A (2008) Differential evolution based feature subset selection. In: Pattern recognition, 2008. ICPR 2008. 19th international conference on. IEEE, pp 1–4

120. Hancer E, Xue B, Zhang M (2018) Differential evolution for filter feature selection based on information theory and feature ranking. Knowl-Based Syst 140:103–119

121. Zhu Z, Ong Y-S, Dash M (2007) Wrapper-filter feature selection algorithm using a memetic framework. IEEE Trans Syst Man Cybern Part B (Cybern) 37(1):70–76

122. Sharbaf FV, Mosafer S, Moattar MH (2016) A hybrid gene selection approach for microarray data classification using cellular learning automata and ant colony optimization. Genomics 107(6):231–238

123. Priya RD, Sivaraj R (2017) Dynamic genetic algorithm-based feature selection and incomplete value imputation for microarray classification. Curr Sci (00113891) 112(1):126

124. Armano G, Farmani MR (2016) Multiobjective clustering analysis using particle swarm optimization. Expert Syst Appl 55:184–193

125. Boutemedjet S, Bouguila N, Ziou D (2009) A hybrid feature extraction selection approach for high-dimensional non-Gaussian data clustering. IEEE Trans Pattern Anal Mach Intell 31(8):1429–1443

126. Pudil P, Novovičová J, Choakjarernwanit N, Kittler J (1995) Feature selection based on the approximation of class densities by finite mixtures of special type. Pattern Recogn 28(9):1389–1398

127. Kalousis A, Prados J, Hilario M (2007) Stability of feature selection algorithms: a study on high-dimensional spaces. Knowl Inf Syst 12(1):95–116

128. Khatami A, Mirghasemi S, Khosravi A, Lim CP, Nahavandi S (2017) A new PSO-based approach to fire flame detection using k-Medoids clustering. Expert Syst Appl 68:69–80

129. Dash M, Liu H (1997) Feature selection for classification. Intell Data Anal 1(3):131–156

130. Kumar A, Kumar A (2016) Adaptive management of multimodal biometrics fusion using ant colony optimization. Inf Fusion 32:49–63

131. Sehhati M, Mehridehnavi A, Rabbani H, Pourhossein M (2015) Stable gene signature selection for prediction of breast cancer recurrence using joint mutual information. IEEE/ACM Trans Comput Biol Bioinform 12(6):1440–1448

132. Xing EP, Karp RM (2001) Cliff: Clustering of high-dimensional microarray data via iterative feature filtering using normalized cuts. Bioinformatics 17(suppl_1):S306–S315

133. Zhang H, Zhou A, Song S, Zhang Q, Gao X-Z, Zhang J (2016) A self-organizing multiobjective evolutionary algorithm. IEEE Trans Evol Comput 20(5):792–806

134. Deb K, Agrawal S, Pratap A, Meyarivan T (2000) A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In: International conference on parallel problem solving from nature. Springer, Berlin, pp 849–858

135. Hu Z, Bao Y, Xiong T, Chiong R (2015) Hybrid filter-wrapper feature selection for short-term load forecasting. Eng Appl Artif Intell 40:17–27

136. Grozdic DT, Jovicic ST (2017) Whispered speech recognition using deep denoising autoencoder and inverse filtering. IEEE/ACM Trans Audio Speech Lang Process (TASLP) 25(12):2313–2322

137. Arauzo-Azofra A, Benitez JM, Castro JL (2008) Consistency measures for feature selection. J Intell Inf Syst 30(3):273–292

138. Guerra-Salcedo C, Chen S, Whitley D, Smith S (1999) Fast and accurate feature selection using hybrid genetic strategies. In: Evolutionary computation, 1999. CEC 99. Proceedings of the 1999 congress on, vol 1. IEEE, pp 177–184

139. Chakraborty B (2002) Genetic algorithm with fuzzy fitness function for feature selection. In: IEEE international symposium on industrial electronics (ISIE'02), vol 1, pp 315–319

140. Zhou A, Qu B-Y, Li H, Zhao S-Z, Suganthan PN, Zhang Q (2011) Multiobjective evolutionary algorithms: a survey of the state of the art. Swarm Evol Comput 1(1):32–49

141. Emmanouilidis C, Hunter A, Macintyre J, Cox C et al (2001) A multi-objective genetic algorithm approach to feature selection in neural and fuzzy modeling. Evol Optim 3(1):1–26

142. Tran B, Zhang M, Xue B (2016) A PSO based hybrid feature selection algorithm for high-dimensional classification. In: Evolutionary computation (CEC), 2016 IEEE congress on. IEEE, pp 3801–3808

143. Vapnik VN (1999) An overview of statistical learning theory. IEEE Trans Neural Netw 10(5):988–999

144. Weinberger KQ, Blitzer J, Saul LK (2006) Distance metric learning for large margin nearest neighbor classification. In: Advances in neural information processing systems, pp 1473–1480

145. Cover T, Hart P (1967) Nearest neighbor pattern classification. IEEE Trans Inf Theory 13(1):21–27

146. Mining WID (2006) Data mining: concepts and techniques. Morgan Kaufmann, Los Altos

147. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M et al (2001) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci 98(24):13790–13795

148. Chandra B, Gupta M (2011) An efficient statistical feature selection approach for classification of gene expression data. J Biomed Inf 44(4):529–535

149. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, Gaasenbeek M, Angelo M, Reich M, Pinkus GS et al (2002) Diffuse large b-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med 8(1):68

150. Apolloni J, Leguizamón G, Alba E (2016) Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments. Appl Soft Comput 38:922–932

151. Singh P, Shukla A, Vardhan M (2017) A novel filter approach for efficient selection and small round blue-cell tumor cancer detection using microarray gene expression data. In: 2017 international conference on inventive computing and informatics (ICICI). IEEE, pp 827–831

152. Hastie T, Tibshirani R, Friedman J (2009) Unsupervised learning. In: The elements of statistical learning. Springer, Berlin, pp 485–585

153. Li J, Dong W, Meng D (2017) Grouped gene selection of cancer via adaptive sparse group Lasso based on conditional mutual information. IEEE/ACM Trans Comput Biol Bioinform

154. Karaboga D, Ozturk C (2011) A novel clustering approach: artificial bee colony (ABC) algorithm. Appl Soft Comput 11(1):652–657

155. Shukla AK (2019) Building an effective approach toward intrusion detection using ensemble feature selection. Int J Inf Secur Priv 13(3):601–614

156. Yin J, Wang Y, Hu J (2012) A new dimensionality reduction algorithm for hyperspectral image using evolutionary strategy. IEEE Trans Ind Inf 8(4):935–943

157. Reddy BR, Ojha A (2017) Performance of maintainability index prediction models: a feature selection based study. Evolv Syst 1–26

158. Shukla AK, Singh P, Vardhan M (2019) A new hybrid feature subset selection framework-based on binary genetic algorithm and information theory. Int J Comput Intell Appl 18:1950020

159. Sluga D, Lotrič U (2017) Quadratic mutual information feature selection. Entropy 19(4):157

160. Hoque N, Bhattacharyya D, Kalita JK (2014) MIFS-ND: a mutual information-based feature selection method. Expert Syst Appl 41(14):6371–6385

161. Sharma A, Imoto S, Miyano S (2012) A top-r feature selection algorithm for microarray gene expression data. IEEE/ACM Trans Comput Biol Bioinform 9(3):754–764

162. Çaliş B, Bulkan S (2015) A research survey: review of AI solution strategies of job shop scheduling problem. J Intell Manuf 26(5):961–973

163. Meyer PE, Schretter C, Bontempi G (2008) Information-theoretic feature selection in microarray data using variable complementarity. IEEE J Sel Top Signal Process 2(3):261–274

164. Hall MA, Smith LA (1999) Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper. In: FLAIRS conference 1999, pp 235–239

165. Collins LM, Dziak JJ, Li R (2009) Design of experiments with multiple independent variables: a resource management perspective on complete and reduced factorial designs. Psychol Methods 14(3):202

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.