Engineering Applications of Artificial Intelligence 100 (2021) 104210

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Survey paper

Review of swarm intelligence-based feature selection methods

Mehrdad Rostami [a,*], Kamal Berahmand [b], Elahe Nasiri [c], Saman Forouzandeh [d]

[a] Department of Computer Engineering, University of Kurdistan, Sanandaj, Iran
[b] Department of Science and Engineering, Queensland University of Technology, Brisbane, Australia
[c] Department of Information Technology and Communications, Azarbaijan Shahid Madani University, Tabriz, Iran
[d] Department of Computer Engineering University of Applied Science and Technology, Center of Tehran Municipality ICT org.Tehran, Iran

ARTICLE INFO

Keywords:
Machine learning
Dimensionality reduction
Feature selection
Evolutionary algorithms
Swarm intelligence

ABSTRACT

In the past decades, the rapid growth of computer and database technologies has led to the rapid growth of large-scale datasets. On the other hand, data mining applications with high dimensional datasets that require high speed and accuracy are rapidly increasing. An important issue with these applications is the curse of dimensionality, where the number of features is much higher than the number of patterns. One of the dimensionality reduction approaches is feature selection that can increase the accuracy of the data mining task and reduce its computational complexity. The feature selection method aims at selecting a subset of features with the lowest inner similarity and highest relevancy to the target class. It reduces the dimensionality of the data by eliminating irrelevant, redundant, or noisy data. In this paper, a comparative analysis of different feature selection methods is presented, and a general categorization of these methods is performed. Moreover, in this paper, state-of-the-art swarm intelligence is studied, and the recent feature selection methods based on these algorithms are reviewed. Furthermore, the strengths and weaknesses of the different studied swarm intelligence-based feature selection methods are evaluated.

1. Introduction

Pattern recognition is one of the most important applications of machine learning in different sciences. Machine learning methods utilized in the many areas of medical diagnosis (Liu et al., 2017; Li et al., 2019b), marketing (Huang and Tsai, 2009; Yuan et al., 2020), image processing (Liang et al., 2017; Zhou et al., 2017), text mining (Wang and Hong, 2019; Kou et al., 2020), information retrieval (Lin et al., 2014; Ji et al., 2019), Identification (Bi et al., 2019; Koide et al., 2020), etc. One of the significant goals of modeling and classification of data is to predict based on the train data and available features. Huge datasets with high dimensional feature space and a relatively smaller number of samples are critical issues for machine learning tasks (Forouzandeh et al., 2020). Once there are a number of irrelevant and redundant features among the initial feature set, dimensionality reduction is one of the essential techniques to eliminate these features. Chen et al. (2019). Dimensionality reduction can improve the performance of the machine learning algorithm and reduce the computational complexity by removing irrelevant and redundant features (Wang et al., 2019; Tang et al., 2019). In the previous years, two approaches for dimensional reduction were presented: feature selection and feature extraction (Farahat et al., 2013). In feature extraction, the primary feature space is mapped to a smaller space. In fact, in this technique, by combining existing features, fewer features are created so that these features contain all (or most of) the information contained in the primary features. Moreover, in feature selection, a subset of initial features is selected by removing the irrelevant and redundant feature.

The total search space to find the most relevant and non-redundant features, including all possible subsets, is 2^n, where n is the number of original features. Comprehensive search ensures that the most appropriate features are found, but usually, this is not computationally feasible, even for medium-sized datasets. Since the evaluation of all possible subsets is very costly, a solution must be searched that is both computationally feasible and useful in terms of quality. Many feature selection methods use metaheuristic algorithms to avoid increasing computational complexity (Welikala et al., 2015; Singh and Singh, 2019; Alshamlan et al., 2015). These algorithms will be able to optimize the problem of feature selection with appropriate accuracy within an acceptable time. Metaheuristic algorithms can be classified into two main categories: Swarm intelligence (SI) and Evolutionary Algorithms (EA). SI is a relatively new category of evolutionary computation comparing with EAs and other single-solution based approaches. SI algorithms utilized approximate and non-deterministic techniques to effectively and efficiently explore and exploit the search space in order

* Corresponding author.
E-mail addresses: m.rostami@eng.uok.ac.ir (M. Rostami), kamal.berahmand@hdr.qut.edu.au (K. Berahmand), el.nasiri@azaruniv.ac.ir (E. Nasiri), Saman.forouzandeh@gmail.com (S. Forouzandeh).

https://doi.org/10.1016/j.engappai.2021.104210
Received 8 September 2020; Received in revised form 21 January 2021; Accepted 15 February 2021
Available online 25 February 2021

to find near-optimal solutions The most popular nature-inspired meta-heuristic group is swarm-based techniques. Swarm Intelligence (SI) is a type of artificial intelligence that is based on collective behaviors in decentralized and self-organized systems. These systems usually consist of a population of simple actors who interact locally and with their environment.

Feature selection has been an active research area in data mining, pattern recognition, statistics communities, and other engineering applications (Liu et al., 2019b). Due to the wide applications of feature selection in most engineering issues, reviewing different feature selection methods can be effective in choosing a suitable approach for that application. In other words, the choice of feature selection methods differs among various engineering application areas. For example, in many engineering problems such as biomedical engineering and industrial engineering, the accuracy of the feature selection method is far more important than the computational complexity of that method. In fact, in these cases, we prefer to choose a feature selection method with maximum accuracy, even if this method has high computational complexity. On the other hand, in many engineering problems, such as image processing and machine vision, the computational complexity of the feature selection method is also very important, and in choosing the feature selection method in this area, computational complexity and execution time must also be considered. Given this, it is important to review different feature selection methods to choose the appropriate method for that application.

In recent decades, many SI-based algorithms have been employed to feature sections (Figueiredo et al., 2019). Despite the highly acceptable performance of these methods, only a few papers are reviewing SI based feature selection methods. In Basir and Ahmad (2014) a comparison of swarm intelligence-based feature selection method is conducted. The authors of this paper focused only on the well-known and traditional SI algorithm, and only the feature selection methods based on these algorithms studied. Also, in Brezočnik et al. (2018) a comprehensive literature review of SI algorithms for feature selection is performed. One of the limitations of this paper is that it lacks experiment results and only states the strengths and weaknesses of different SI-based feature selection methods. Moreover, in Nguyen et al. (2020) swarm intelligence approaches to feature selection in data mining are surveyed. One of the shortcomings of this paper is that it focuses on PSO, ABC, and ACO-based feature selection methods, and many of the other SI algorithms that have been proposed in recent years (i.e. Gravitational Search Algorithm, Gray Wolf Optimization, Whale Optimization Algorithm, Salp Swarm Algorithm, etc.) have not been discussed. Another drawback of this paper is that it has not performed any numerical analysis to compare different methods and only the different methods have been studied theoretically. Furthermore, Nayar et al. (2019) reviewed the problems encountered during the process of feature selection and how swarm intelligence has been used for selection of optimal set of features. This paper also gives a concise overview of various SI algorithms like PSO, ACO, ABC. One of the disadvantages of this paper is that the different methods have only been compared in terms of classification accuracy criteria, and no comparison has been made in terms of computational complexity, execution time, and a number of selected features. Another drawback of this paper is the lack of review of new SI algorithms such as Gray Wolf Optimization, Whale Optimization Algorithm, and Salp Swarm Algorithm

With the lack of a previous comprehensive review of SI-based feature selection methods, the main purpose of this paper is to fill the gap in coverage of SI algorithms for feature selection. This paper seeks to provide a comprehensive overview of SI-based feature selection methods and their categorization. Moreover, this paper tries to review the state-of-the-art and most well-known SI-based method used to feature selection. Also, in this paper, various experiments have been designed to compare the performance of different SI-based methods to allow a more accurate evaluation of these algorithms.

The remainder of this paper is organized as follows: Section 2 presents the introduction of the feature selection problem, Section 3 review different feature selection method and categories previous methods. Section 4 reviews SI based feature selection methods. Section 5 reports the experimental results of different SI-based methods. Finally, Section 6 presents the conclusion.

2. Introduction to feature selection

An essential issue with machine learning techniques is the high-dimensionality problem of a dataset where the feature subset size much greater than the pattern size (Rostami and Moradi, 2014; Berahmand et al., 2020; Rostami et al., 2020a). For example, in the medical applications that include very high-dimensional datasets, the classification parameters are also increased. Therefore, the performance of the classifier declines significantly (Saeys et al., 2007; Chandrashekar and Sahin, 2014; Liu and Yu, 2005). According to a general rule for a classification problem with n dimension and C class, at least $10 \times n \times C$ training data is required (Cadenas et al., 2013; Liu and Zheng, 2006; Jain et al., 2000). When it is not possible to provide this number of training data practically, reducing the feature subset size, reduces the number of required training data. As a result, the performance of the classification algorithm increases (Shu et al., 2020; Gokalp et al., 2020).

Dimensionality reduction is one popular technique to remove noise (i.e., irrelevant) and redundant features. It is an efficient method for improving accuracy performance, lowering computational complexity, building more generalized models, and decreasing the required storage (Wang et al., 2019; Tang et al., 2019). During the past few years, two major approaches have been proposed for dimension reduction: Feature extraction and feature selection (Farahat et al., 2013). Feature extraction involves a linear or nonlinear transformation from the original feature space to a new one with lower dimensionality. Feature selection, on the other hand, directly reduces the number of original features by selecting a subset of them that still retains sufficient information for classification.

Feature selection has been an active research area in data mining, pattern recognition, and statistics communities. The main idea of feature selection is choosing a subset of available features, by eliminating features with little or no predictive information, and also redundant features that are strongly correlated (Liu et al., 2019b). Feature selection is a fundamental research topic in statistical pattern recognition, machine learning, data mining with a long history started in the 1970s, and there have been several attempts to improve the performance of the feature selection methods (Saeys et al., 2007; Chandrashekar and Sahin, 2014; Liu and Yu, 2005). The main idea of feature selection is choosing a subset of available features, by eliminating features with little or no predictive information, and also redundant features that are strongly correlated (Liu et al., 2019b).

In general, the feature selection process consists of four main stages: subset generation, subset evaluation, stopping criteria, and validation of results. In each iteration of the search process, a subset of the candidate feature set is generated from the original features, and its appropriateness is measured by an evaluation criterion. The subset generation process and its evaluation are repeated until a predetermined stop criterion is reached. At the end of this process, the best subset of the selected feature is validated on the test dataset (Tadist et al., 2019; Renuka Devi and Sasikala, 2019).

From a general point of view, feature selection methods are divided into two categories supervised and unsupervised feature selection methods (Tang et al., 2018). In supervised methods, a set of train data is available, each of which is described by taking features values along with the class label, while in unsupervised methods, and train data lacks class tags. In general, it can be said that feature selection methods have better efficiency and more reliable performance in the supervised mode due to the use of class labels. Therefore, it is more difficult to select a feature in the unsupervised mode, and in many studies, this area has been considered (Ding et al., 2020; Zhang et al., 2019b).

In this review, the papers are collected by Google Scholar to get publications from several databases, including Elsevier, Springer, IEEE, etc. Papers that were published between 2010 and 2020 were primarily considered to ascertain the current status of study. The first search was done with the aim of finding review papers using the following keywords: review, dimensionality reduction, feature selection, evolutionary algorithm, swarm intelligence and meta-heuristics. In the second search, the following keywords are used for finding research papers focusing on the swarm intelligence-based feature selection methods: feature selection, swarm intelligence, dimensionality reduction, evolutionary algorithm, optimization, nature inspired metaheuristic, swarm-based techniques, metaheuristic optimization algorithm, population-based algorithms and evolutionary computation. In addition, more papers from Science Direct, Springer and IEEE Xplore recommender systems and references of the included papers also included. Then, the papers that focused on feature extraction and non-swarm intelligence-based feature selection methods were excluded. In other words, only relevant works that used only swarm intelligence algorithms for feature selection were included. This led to a total of 85 papers, including six review studies, to be thoroughly reviewed on the feature selection. The other references were related to other methods used in feature selection problem.

3. Backgrounds

Feature selection methods generally search through the solution space to optimize two conflicting objectives: maximizing the relevancy to the target class and minimizing the redundancy of selected features (Chen et al., 2020). Many search strategies are employed to optimize these objectives. These methods are generally categorized into single-objective and multi-objective methods (Hu et al., 2018; Senawi et al., 2017). In the single-objective methods, the population is optimized using only one objective in the fitness function. As a result, the choice of objective and definition of the fitness function will greatly affect the accuracy of the optimization algorithm. Also, there are usually several objectives in many optimization issues, and defining a fitness function with just one goal reduces optimization performance. One approach to overcome these challenges is to consider several different objectives in the fitness function of the feature selection problem. Modeling feature selection as a multi-objective problem can obtain a set of non-dominated feature subsets to meet different requirements in real-world applications. Most of the previously proposed feature selection methods are utilized single-objective function, and there are only a few methods of multi-objective function, such as MOPSO (Xue et al., 2013), MOGA (Morita et al., 2003), MOFSEF (Wu et al., 2020), MOFSEEG (Martín-Smith et al., 2017), and MOFSGADM (Li et al., 2020a). Moreover, some of the previous feature selection methods try to find minimum redundant features with maximum dependence on output labels. In Wang et al. (2018), a novel fuzzy multiple kernel learning model for classification is developed. In this model, a new criterion is proposed, which selects a subset of dissimilar and relevant features.

Moreover, recently, graph-based methods are used in machine learning techniques to extract the similarity relationships among the data (Berahmand et al., 2020; Yazdi et al., 2019; Majbouri Yazdi et al., 2020). In feature selection, graph-based methods provide an underlying manifold structure as a universal framework to reflect the relationships between features. Several research efforts employed graph-based methods for solving the feature selection problem. For example, in Bandyopadhyay et al. (2014), a dense subgraph finding approach is adopted for the unsupervised feature selection problem. Another clustering-based feature subset selection algorithm for high dimensional data is proposed in Song et al. (2013). This work utilizes a graph-theoretic clustering method for similar grouping features. In Zhang and Hancock (2012), a hypergraph-based method is proposed for feature selection. This work uses an information-theoretic criterion to evaluate the appropriateness of different features by considering the related class label of each sample. In Moradi and Rostami (2015a), the concept of graph clustering with the node centrality measure is integrated with the unsupervised feature selection process. This work is extended by the authors of Ghaemi and Feizi-Derakhshi (2016) to choose more informative features. In Henni et al. (2018), the authors employed Google's PageRank centrality measure to rank features based on their importance. Hashemi et al. (2020) propose another graph-based feature selection method for the multi-label high-dimensional dataset. The authors of this paper utilize the PageRank centrality measure to rank the features based on their value in the graph. Also, in this paper, the correlation distance criterion is used to remove irrelevant features. In Li et al. (2019a), an unsupervised graph-based feature selection method for high dimensional data is proposed. In this paper, the Laplacian graph and local geometrical structure are used for better representation of the features space. In Zhu et al. (2017), a subspace clustering guided unsupervised feature selection method is proposed. This work uses the subspace clustering to the learning of the clustering and then those features with good preservation ability of the cluster labels are selected.

Generally, previous methods for select optimal feature sets can be categorized into four groups consist of filter, wrapper, embedded, and hybrid models. These categories are explained in the following subsections.

3.1. Filter model

The filter model evaluates the relevance of features without using any learning algorithm. Therefore, the methods in this approach are typically fast. In this model, features are evaluated and ranked using the information-theoretic measures, and then those with the highest ranks are selected (Labani et al., 2018). From one point of view, the filter-based methods can be classified into univariate and multivariate methods. The univariate methods only consider the relevancy of features to the target class according to a specific criterion such as: Information Gain (IG) (Raileanu and Stoffel, 2004), Gain Ratio (GR) (Mitchell, 1997), Term Variance (TV) (Theodoridis and Koutroumbas, 2009), Mutual information (MI) (Xu et al., 2007), Gini Index (GI) (Raileanu and Stoffel, 2004), Laplacian score (LS) (He et al., 2005), and Fisher score (FS) (Gu et al., 2011). However, these methods ignore the dependency between features results in presenting similar features in the final feature set as well as building weak and complicated learning models. To solve this issue, a branch of the filter approach called multivariate methods handles both irrelevant and redundant features in their ranking strategies. In other words, multivariate filter methods were proposed, aiming at the consideration of feature dependencies to some degree. There are some multivariate methods, including minimal redundancy maximal relevance (mRMR) (Peng et al., 2005), Relevance redundancy feature selection (RRFS) (Ferreira and Figueiredo, 2012), MIFS (Battiti, 1994), Normalized mutual information feature selection (NMIFS) (Estévez et al., 2009), MIFS-U (Kwak and Choi, 2002), Unsupervised feature selection based on Ant Colony Optimization (UFSACO) (Tabakhi et al., 2014) and Hilbert–Schmidt independence criterion (HSIC) (Yamada et al., 2014).

3.2. Wrapper model

In the wrapper approach, a classifier is used and trained to evaluate a set of prominent features. In the wrapper model, a given learning model is used to evaluate a subset of features in their search processes in order to choose a set of features with the highest classification accuracy. Most of the wrapper methods employ iterative search processes wherein each iteration of the learning model is used to guide the population of solutions towards the best solution. However, due to a learning model being involved in the searching process of the wrapper approach, these methods often suffer from high computational cost and loss of generality. Filter methods are fast enough and their results do not rely on a specific classifier, and thus are appropriate for real-world applications (Chandrashekar and Sahin, 2014; Liu and Yu, 2005). Although the wrappers model may select a better feature subset, they are expensive to run and can break down with high dimensional medical dataset. This is due to the use of learning algorithms in the feature subset quality calculation.

3.3. Hybrid model

The hybrid approach is a combination of the filter and wrapper models which attempts to take advantage of both models. This approach mainly focuses on achieving the best possible performance by a particular learning algorithm and time complexity similar to the filter-based methods.

3.4. Embedded model

The embedded approach considers the feature selection problem as part of the machine learning method. In fact, in the embedded approach, a machine learning method is used for seeking the final feature subset (Zhang et al., 2015a).

In Table 1 the main characteristic of different feature selection methods is summarized. In this table, six attributes (i.e., the number of objectives, Type of method, search strategy, application, weakness, and strength) of different feature selection methods are reported.

4. Swarm intelligence-based feature selection

Feature selection methods based on how to evaluate the features are classified into two categories: feature ranking and subset selection methods. Feature ranking methods, based on a specified criterion, assign a score to each feature. Then the features that did not get enough scores are removed. But in subset selection methods, the space of possible subsets is searched to find the optimal subset. If the number of initial features is n, the search space for the optimal subset contains all the subsets of the features, which is equal to 2^n different states. In other words, in the property ranking methods, the value of each property is evaluated independently, and the relationship between features is not considered. In these methods, it is assumed that the features are independent of each other and that a possible dependence between the features is not considered. Although this simplistic assumption reduces the computational complexity of the feature selection method, in many cases, it may reduce the performance of the feature selection method.

Feature subset selection is an NP-Hard problem. In the simplest way, the best subset can be found by evaluating all possible subsets with an exhaustive search strategy. Although this method guarantees an optimal feature subset, finding the optimal solution is very time consuming and even impractical even for medium-sized datasets. Since it is very costly to evaluate all possible subsets, a feature subset must be searched, that, it is acceptable both in terms of computational complexity and in terms of appropriateness (Li et al., 2020b; Santucci et al., 2020). One approach to solving complex optimization and NP-Hard problems is meta-heuristics algorithms. Meta-heuristic algorithms are approximate approaches that can find satisfactory solutions over an acceptable time instead of finding the optimal solution (Zhang et al., 2015b). These algorithms are one of the categories of approximate optimization algorithms that have s strategies to escape from local optima and can be used in a wide range of optimization problems.

Many feature selection methods use meta-heuristics to avoid increasing computational complexity in the high dimensional dataset. These algorithms use primitive mechanisms and operations to solve an optimization problem and search for the optimal solution over a number of iterations (Barak et al., 2015). These algorithms often start with a population containing random solutions and try to improve the optimality of these solutions during each iteration step. At the beginning of most of the meta-heuristic algorithms, a number of initial solutions are randomly generated, and then a fitness function is utilized to calculate the optimality of the individual solutions of the generated population. If none of the termination criteria are met, production new generation will begin. This cycle is repeated until one of the termination criteria is met (Wang et al., 2020a; Hu et al., 2020a).

Meta-heuristic approaches can be classified into two categories: Evolutionary Algorithms (EA) and Swarm Intelligence (SI) (Zhang et al., 2015b). An EA uses mechanisms inspired by biological evolution, such as reproduction, mutation, recombination, and selection. Candidate solutions to the optimization problem play the role of individuals in a population, and the fitness function determines the quality of these solutions. After repetitions of the evolutionary algorithm, the initial population evolves and moves towards global optimization (Gong et al., 2020). Moreover, swarm-intelligence-based optimization methods usually consist of a number of uncomplicated members of artificial factors. The concept of SI optimization algorithms is taken from nature, and each factor Takes on a simple task, but The relationship of these factors with each other and Somewhat random reactions of these factors create global intelligent behavior, which cannot be done by any of the factors alone (Yong et al., 2016).

In the remainder of this section, SI-based feature selection methods, such as Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Artificial Bee Colony optimization (ABC), Differential Evolution (DE), Gravitational Search Algorithm (GSA), Firefly Algorithm (FA), Bat Algorithm (BA), Cuckoo Optimization Algorithm (COA), Gray Wolf Optimization (GWO), Whale Optimization Algorithm (WOA) and Salp Swarm Algorithm (SSA) are reviewed and outlined in Table 2.

4.1. PSO-based methods

Particle Swarm Optimization is a powerful SI-based optimization method, developed by Kennedy and Eberhart in 1995. This optimization method, inspired by the collective conduct of birds and fish, has recently been used in a number of researches to optimize the selection of the final feature subset. Unler et al. (2011) introduced a new search strategy to select the final feature set, which integrated the filter approach with the PSO-based wrapper approach. Moreover, Inbarani et al. (2014), developed a hybrid feature selection method for the medical application. The authors of this paper, proposed a hybrid feature selection approach by combining the wrapper PSO and rough sets theory to improve the classification accuracy of disease diagnosis. Furthermore, in Huang and Dun (2008) another hybrid approach for feature subset selection and parameter optimization is proposed. In this method the PSO optimization method, and SVM in integrated to improve the performance of feature selection. The authors of this paper combined PSO–SVM with a distributed parallel architecture to overcome the high computational complexity for the high dimensional datasets. Moreover, the authors of Xue et al. (2014) developed novel initialization methods and best particle updating strategies to improve the performance of optimization tasks in feature subset selection. This method aims to select a subset of relevant and non-redundant features for computational complexity reduction. Banka and Dara (2015), proposed an integration model of binary PSO (BPSO) algorithm and Hamming distance to select the final feature set in the classification task. In this method, a hamming distance is used to improve the velocity update process in a BPSO algorithm and then this improved BPSO is utilized for feature selection problems. In Yong et al. (2016), and improved multi-objective PSO algorithm is proposed for unreliable data classification. In this paper, two new operators of the reinforced memory and the hybrid mutation are introduced to improve the search ability of the PSO algorithm. In Moradi and Gholampour (2016), and efficient PSO-based feature selection method is proposed by the integration of filter and wrapper approaches. The proposed method, called HPSO-LS, introduced a new local search to select the subset of non-redundant and relevant features. In Zhang et al. (2017a) a multi-Objective particle swarm optimization approach is developed for cost-based feature selection in classification. The main goal of this method is to produce a Pareto front of Non-dominated solutions, that is, feature set, to meet different prerequisites

Table 1
The main characteristics of different feature selection methods. The synonyms used in this Table are: **MOP**: Multi-Objective Optimization, **SOP**: Single-Objective Optimization, **RB**: Ranking-Based, **SSB**: Subset selection-based, **ACO:** Ant Colony Optimization, **GA:** Genetic Algorithm, **ABC:** Artificial Bee Colony, **DE:** Differential Evolution, **NSGA:** Nondominated Sorting Genetic Algorithm.

Methods	Number of objectives	Type	Search strategy	Application	Weakness	Strength
RDC (Rehman et al., 2015)	SOP	Filter-RB	Univariate	Textual	Ignoring feature dependency, Low performance	Fast, Independent of classifier
DFS (Uysal and Gunal, 2012)	SOP	Filter-RB	Univariate	Textual		
NDM (Rehman et al., 2017)	SOP	Filter-RB	Univariate	Textual		
FS (Gu et al., 2011)	SOP	Filter-RB	Univariate	Textual/Numeric		
GI (Raileanu and Stoffel, 2004)	SOP	Filter-RB	Univariate	Textual/Numeric		
MI (Xu et al., 2007)	SOP	Filter-RB	Univariate	Textual/Numeric		
LS (He et al., 2005)	SOP	Filter-RB	Univariate	Numeric		
IG (Raileanu and Stoffel, 2004)	SOP	Filter-RB	Univariate	Textual/Numeric		
CHI (Sebastiani, 2002)	SOP	Filter-RB	Univariate	Textual/Numeric		
GR (Mitchell, 1997)	SOP	Filter-RB	Univariate	Numeric		
TV (Theodoridis and Koutroumbas, 2009)	SOP	Filter-RB	Univariate	Numeric		
RRFS (Ferreira and Figueiredo, 2012)	SOP	Filter	MultivariateGreedy	Numeric	Using a greedy search strategy and easily trapping into the local optima.	Having higher accuracy than univariate methods
mRMR (Peng et al., 2005)	SOP	Filter	MultivariateGreedy	Numeric		
RSM (Lai et al., 2006)	SOP	Filter	MultivariateGreedy	Numeric		
GUFS (Farahat et al., 2013)	SOP	Filter	MultivariateGreedy	Numeric		
MECY_FS (Wang et al., 2015)	MOP	Filter	GA	Numeric	Only applicable to numeric datasets, Having lower performance in text datasets.	Using information-theoretic methods to evaluate the objective function that has made them fast.
RRFSACO (Tabakhi and Moradi, 2015)	SOP	Filter/SSB	MultivariateACO	Numeric	Being single-objective, Forcing the evolving population to form a particular feature set due to the use of a single quality function.	Considering the feature dependency
GCACO (Moradi and Rostami, 2015b)	SOP	Filter/SSB	MultivariateACO	Numeric		
MGCACO (Ghimatgar et al., 2018)	SOP	Filter/SSB	MultivariateACO	Numeric		
FASTFS (Song et al., 2013)	SOP	Filter/SSB	MultivariateGraph-based	Textual/Numeric		
GCNC (Moradi and Rostami, 2015a)	SOP	Filter/SSB	MultivariateGraph-based	Numeric		
FSGA (Yang et al., 2011)	SOP	Filter/SSB	MultivariateGA	Foreign Fiber		
MICGSOFS (Lyu et al., 2017)	SOP	Filter/SSB	MultivariateSequential	Biomedical Datasets		
MBFFS (Hua et al., 2020)	SOP	Filter/SSB	Markov Blanket	Numeric		
BT-SFS (Yan et al., 2018)	SOP	Wrapper	Sequential	Fault Detection	Depending on the initial solutions and easily trapping into the local optima	having low computational complexity
MOABC (Hancer et al., 2015a)	MOP	Wrapper	ABC	Numeric	Inefficient for high-dimensional datasets due to high computational complexity.	Low probability of trapping into the local optimum, High classification accuracy
TMABC-FS (Zhang et al., 2019a)	MOP	Wrapper	ABC	Numeric		
MOPSOFS (Xue et al., 2013)	MOP	Wrapper	PSO	Numeric		
MOACO (Ke et al., 2010)	MOP	Wrapper	ACO	Numeric		
MOWGA (Vignolo et al., 2013)	MOP	Wrapper	GA	Face Recognition		
MODEFS (Mlakar et al., 2017)	MOP	Wrapper	DE	Face Recognition		
MONSGA (González et al., 2019)	MOP	Wrapper	NSGA-II	Motor imagery		
Nested GA (Sayed et al., 2019)	MOP	Wrapper	NSGA-II	Microarray		

of decision-makers in real-world applications. In Jain et al. (2018), integration of correlation feature selection with modified binary PSO algorithm is used for gene selection and cancer classification. This method selects a high relevant feature subset by eliminating the irrelevant and redundant features. In Zhang et al. (2018a), a combination method of improved mRMR and Shuffled Frog Leaping Algorithm is developed to improve the acoustic defect detection accuracy. In this method, to reduce the dimensions of the original features, a single-objective function is defined using the MRMR criterion and then this function is optimized using the PSO algorithm. Finally, after selecting the final feature subset, the final acoustic defect detection is made using the neural network classifier. In Qasim and Algamal (2018) a novel PSO-based feature selection method is proposed by a combination of the regression model and Bayesian information criterion to improve the performance of disease diagnosis. In Prasad et al. (2018), a recursive particle swarm optimization strategy is developed to select a subset of relevant and non-redundant features in DNA microarray datasets. The authors of this paper combined the various filter-based feature selection measures with the developed wrapper-based recursive PSO method. In Pashaei et al. (2019) a feature selection method is developed in cancer classification using binary black hole algorithm and improved BPSO algorithm. In this method, by combining the advantages of wrapper and filter approach, a hybrid feature selection method is introduced. This hybrid feature selection method, embedded the binary black hole algorithm is in the PSO algorithm to make this more effective and to facilitate the exploration and exploitation of the PSO to improve the performance further. Moreover, Gunasundari et al. (2018), developed an improved BPSO method for feature selection. This method uses a new Win–Win strategy to increase the accuracy of feature selection in liver and kidney disease diagnosis. Furthermore, the authors of Yan et al. (2019), introduced a novel hybrid PSO-based feature selection method to the analysis of Laser-induced breakdown spectroscopy. In this method, an attempt has been made to use the advantages of coating and filter methods simultaneously. In Xue et al. (2020) a PSO-based feature selection with multiple classifiers is proposed to improve

Table 2
Outlining the reviewed swarm intelligence based feature selection methods.

Reference	SI method	Number of objectives	Type	Application
Unler et al. (2011)	PSO	Single objective	Hybrid	Numerical
Inbarani et al. (2014)	PSO	Single objective	Hybrid	Medical
Huang and Dun (2008)	PSO	Single objective	Wrapper	Numerical
Xue et al. (2014)	PSO	Single objective	Wrapper	Numerical
Banka and Dara (2015)	PSO	Multi objective	Wrapper	Medical
Yong et al. (2016)	PSO	Multi objective	Wrapper	Numerical
Moradi and Gholampour (2016)	PSO	Single objective	Wrapper	Numerical/Medical
Zhang et al. (2017a)	PSO	Single objective	Wrapper	Numerical
Jain et al. (2018)	PSO	Single objective	Hybrid	DNA microarray
Zhang et al. (2018a)	PSO	Single objective	Filter	Signal
Qasim and Algamal (2018)	PSO	Single objective	Wrapper	Medical
Prasad et al. (2018)	PSO	Single objective	Wrapper	DNA microarray
Pashaei et al. (2019)	PSO	Single objective	Wrapper	Medical
Gunasundari et al. (2018)	PSO	Single objective	Wrapper	Medical
Yan et al. (2019)	PSO	Single objective	Wrapper	spectroscopy
Xue et al. (2020)	PSO	Single objective	Wrapper	Numerical
Rostami et al. (2020b)	PSO	Multi objective	Filter	Medical
Kabir et al. (2012)	ACO	Single objective	Hybrid	Numerical
Li et al. (2013)	ACO	Single objective	Wrapper	DNA microarray
Chen et al. (2013)	ACO	Single objective	Wrapper	Image
Forsati et al. (2014)	ACO	Single objective	Wrapper	Numerical
Ke et al. (2008)	ACO	Single objective	Filter	Numerical
Tabakhi et al. (2014)	ACO	Single objective	Filter	Numerical/Medical
Moradi and Rostami (2015b)	ACO	Single objective	Filter	Numerical/Medical
Dadaneh et al. (2016)	ACO	Single objective	Filter	Numerical/Medical
Liu et al. (2019a)	ACO	Multi objective	Wrapper	Numerical
Schiezaro and Pedrini (2013)	ABC	Single objective	Wrapper	Numerical/Medical
Hancer et al. (2015b)	ABC	Single objective	Wrapper	Numerical/Medical
Shunmugapriya and Kanmani (2017)	ABC	Single objective	Wrapper	Numerical/Medical
Hancer et al. (2018b)	ABC	Single objective	Wrapper	Numerical
Arslan and Ozturk (2019)	ABC	Single objective	Wrapper	Numerical
Zhang et al. (2019a)	ABC	Multi-Objective	Wrapper	Numerical
Wang et al. (2020b)	ABC	Multi-Objective	Wrapper	Numerical
Al-Ani et al. (2013)	DE	Single objective	Wrapper	Numerical/Medical
Hancer et al. (2018a)	DE	Multi objective	Filter	Numerical/Medical
Zhang et al. (2020)	DE	Multi objective	Wrapper	Numerical/Medical
Hancer (2020)	DE	Single objective	Wrapper	Numerical/Medical
Han et al. (2014)	GSA	Single objective	Wrapper	Numerical/Medical
Xiang et al. (2015)	GSA	Single objective	Wrapper	Numerical/Medical
Taradeh et al. (2019)	GSA	Single objective	Wrapper	Numerical/Medical
Zhang et al. (2017b)	FA	Multi-Objective	Wrapper	Numerical
Zhang et al. (2018b)	FA	Single objective	Wrapper	Facial expression
Larabi Marie-Sainte and Alalyani (2020)	FA	Single objective	Wrapper	Text
Selvakumar and Muneeswaran (2019)	FA	Single objective	Hybrid	Network Intrusion
Tawhid and Dsouza (2018)	BA	Single objective	Wrapper	Numerical/Medical
Liu et al. (2020)	BA	Single objective	Wrapper	Image Steganalysis
Al-Betar et al. (2020)	BA	Single objective	Hybrid	Medical
Elyasigomari et al. (2017)	COA	Single objective	Wrapper	Numerical/Medical
Jayaraman and Sultana (2019)	COA	Single objective	Wrapper	Medical
Prabukumar et al. (2019)	COA	Single objective	Wrapper	Gene expression
Emary et al. (2015)	GWO	Multi-Objective	Hybrid	Numerical/Medical
Emary et al. (2016)	GWO	Single objective	Wrapper	Numerical/Medical
Tu et al. (2019)	GWO	Single objective	Wrapper	Numerical/Medical
Abdel-Basset et al. (2020)	GWO	Single objective	Wrapper	Numerical/Medical
Mafarja and Mirjalili (2017)	WOA	Single objective	Wrapper	Numerical/Medical
Mafarja and Mirjalili (2018)	WOA	Single objective	Wrapper	Numerical/Medical
Nematzadeh et al. (2019)	WOA	Single objective	Filter	Medical
Faris et al. (2018)	SSA	Single objective	Wrapper	Numerical/Medical
Ibrahim et al. (2019)	SSA	Single objective	Wrapper	Numerical
Tubishat et al. (2020)	SSA	Single objective	Wrapper	Numerical/Medical
Al-Zoubi et al. (2020)	SSA	Single objective	Wrapper	Medical
Hegazy et al. (2020)	SSA	Multi-Objective	Wrapper	Numerical/Medical
Neggaz et al. (2020)	SSA	Single objective	Wrapper	Numerical/Medical

for increasing the classification accuracy and reducing computational complexity. In this paper, a new Self-Adaptive Parameter and Strategy are used to deal with the issue of feature selection in a high-dimensional dataset. The reported results showed that the use of these mechanisms greatly increased the search ability of particle optimization algorithms for high-dimensional datasets. Moreover, in Rostami et al. (2020b), a novel graph-based feature selection method is developed to increase disease diagnosis accuracy. In this method, using the node centrality criterion, a new mechanism for initializing the particles is proposed. Then, by defining a multi-objective fitness function, a subset of the final features that are least similar to each other and most relevant to the target class are selected. Finally, based on the selected features, the disease is diagnosed.

4.2. ACO-based methods

About three decades ago, a new heuristic-based optimization method called the Ant System (AS) is proposed by Dorigo and his colleagues to solve complicated optimization problems. AS was first utilized for the traveling salesman problem. Then, Dorigo and Caro (1999) developed the Ant Colony Optimization (ACO) algorithm to improve the AS algorithm. This improved algorithm can be able to search the global optimal by representing the problem as a graph model. Moreover, ACO has been successfully utilized in many types of research to select the final feature set. In Kabir et al. (2012) an ACO-based feature selection approach is developed by combining the neural network and IG method to remove redundant and irrelevant features. Moreover, the authors of Li et al. (2013) proposed a new ACO-based gene selection method in the DNA microarray dataset. This proposed method includes two phases of essential features by improved AS algorithm and then final features are selected using a modified ACO algorithm. Furthermore, in Chen et al. (2013) a new ACO-based feature selection approach is developed. In this approach, unlike the previous ant feature selection methods base and ACO, the feature selection problem is modeled as a directed graph. In Forsati et al. (2014), a new version of ACO is proposed to remove the irrelevant and redundant features. This method selects the final feature set by utilizing a local search strategy to avoid getting stuck in the local optimal. Moreover, in Ke et al. (2008), a rough set theoretic-based ACO feature selection method is developed. In Tabakhi et al. (2014) a novel ACO-based feature selection method is proposed for unsupervised mode. The authors of this paper selected the most non-redundant features that have the least similarity with each other. Moreover, Moradi and Moradi and Rostami (2015b), developed a filter-based feature selection approach utilizing the ACO algorithm and graph clustering. This approach represented the feature space as a clustered graph. Then, according to the similarity between the features and by defining a filter criterion, it selects a dissimilar and related subset of the features. In Dadaneh et al. (2016) proposed an unsupervised ACO-based feature selection method to remove redundant and irrelevant features. This method tries to select an optimal subset of features in a hierarchical process, by considering the similarity between features. In Liu et al. (2019a) combination of feature selection and ant colony optimization is proposed for improve the classification accuracy of imbalanced data. In this method, instead of using a single-objective fitness function, a multi-objective ant colony optimization algorithm is used to improve the performance feature selection. The reported results showed acceptable performance of the proposed method in classifying imbalanced and high-dimensional datasets.

4.3. ABC-based methods

The Artificial Bee Colony algorithm (ABC) is a SI-based optimization method inspired by the lifestyle of the bee population. This algorithm tries to imitate the food search conduct of the bee population. This optimization technique integrated a local search strategy with a random search strategy that can be utilized for complex optimization problems. This swarm intelligence-based optimization technique has been applied in many references to select a final feature set. In Schiezaro and Pedrini (2013), a novel feature selection method based on an artificial bee colony is developed to improve the machine learning task. In this method, the feature space is modeled as a binary vector and the classification accuracy is utilized as a fitness function. The authors of Hancer et al. (2015b), developed an improved ABC-based feature selection approach. This approach combined the similarity search strategy and a binary version of ABC to improve the performance of feature selection. Shunmugapriya and Kanmani (2017), proposed a novel feature selection method by integrating the ABC algorithm and ACO algorithm. This method tries to utilize the advantages of an artificial bee colony and an ant colony optimization algorithm, simultaneously. In Hancer et al. (2018b) a new feature selection approach is developed by the combination of the ABC algorithm and the pareto-optimal front surface. The authors of this paper used a multi-objective fitness function and genetic operators to improve the accuracy of the feature selection method and the convergence of the ABC algorithm. In Arslan and Ozturk (2019) a Multi Hive ABC Programming is developed to select the final feature set in high dimensional datasets. This approach utilized the ability of an automatic programming algorithm to remove irrelevant and redundant features. The authors of Zhang et al. (2019a), developed a multi-objective ABC-based feature selection approach. In this method, two new operators are used to improve its search capability and convergence of the ABC search strategy. In Wang et al. (2020b), an ABC-based feature selection is proposed by integrating of multi-objective optimization algorithm with a sample reduction strategy. This proposed method has both increased classification accuracy and reduced computational complexity.

4.4. DE-based methods

The Differential Evolution (DE) algorithm is an SI-based search strategy technique that has been developed to solve complex optimization problems. This optimization method is introduced to overcome the main weakness of the genetic algorithm, namely the lack of local search in this algorithm. The main difference between DE algorithm and the genetic algorithms is in the genetic selection operators. DE algorithm is used in different applications of machine learning tasks and feature selection. For example, Al-Ani et al. (2013) developed a wrapper-based feature subset selection using DE algorithm and a wheel-based search mechanism. Hancer et al. (2018a) proposed a multi-objective DE-based approach to remove the irrelevant and redundant features and improving classification accuracy. In Zhang et al. (2020), a multi-objective DE-based feature selection approach is developed. In this approach, a novel mutation operator is defined to escape the local optimal. Furthermore, in Hancer (2020), another new multi-objective DE-based feature selection approach is developed to feature selection and improve the performance of the clustering algorithm simultaneously.

4.5. GSA-based methods

Moreover, some SI-based optimization methods have been inspired by physical laws. Gravitational Search Algorithm (GSA) is (Rashedi et al., 2009) a physics-based optimization technique inspired by Newton's law of universal gravitation. GSA, a popular SI-based algorithm, has been widely applied in machine learning, and recently, many feature selection methods have been developed using this optimization algorithm. Han et al. (2014) proposed a GSA-based feature selection approach using a linear chaotic map to improve classification accuracy. In Xiang et al. (2015), a hybrid system for feature selection based on an enhanced GSA and K-nearest neighbor classifier is developed. In this method, a piecewise linear chaotic map for exploration, and sequential quadratic programming for exploitation are employed. In Taradeh et al. (2019), an efficient hybrid feature selection method is proposed to utilized the advantages of the SI and the genetic algorithm to improve the performance of the GSA algorithm.

4.6. FA-based methods

The Firefly Algorithm (FA) was introduced by Yang, in 2010a, the main idea was inspired by the optical connection between fireflies. The Firefly Algorithm is a sensible example of swarm intelligence in which low-performance agents can work together to achieve great results with high performance. In Zhang et al. (2017b) a novel FA-based feature selection method, called return-cost-based binary FFA is developed. The authors of this paper provide a variety of strategies to prevent premature convergence of the FA algorithm and thus improve the accuracy of feature selection. In Zhang et al. (2018b), a feature selection method is developed using the firefly optimization algorithm to increase the accuracy of classification and regression models. In this method, to improve convergence and prevent trapping in local optimization, some variation is added to standard FA, which improves the accuracy of final feature selection. Moreover, in Larabi Marie-Sainte and Alalyani (2020) FA-based feature selection method is proposed. In this method, after selecting the final features, these features are utilized to classify Arabic texts using an SVM classifier. Furthermore, in Selvakumar and Muneeswaran (2019), another FA-based feature selection method is presented for network intrusion detection. In this method, a combination of filter-based feature selection method (i.e., Mutual Information) and wrapper-based feature selection method (i.e., C4.5 and Bayesian network) has been utilized to select the final features.

4.7. BA-based methods

The Bat Algorithm (BA) (Yang, 2010b) is an algorithm inspired by the collective behavior of bats in the natural environment, introduced by Yang in 2010. The bat algorithm is a kind of swarm intelligence-based algorithm that is inspired by the echolocation behavior of bats. Bats find the exact path and location of their prey by sending sound waves and receiving reflections. When the sound waves return to the transmitter of the bat waves, the bird can draw an audio image of the obstacles in front of its surroundings and see the surroundings well. In Tawhid and Dsouza (2018), a hybrid variant of bat algorithm and improved PSO algorithm is to improve the feature selection performance. In this proposed method, the PSO algorithm is used to improve the convergence power of the hybrid algorithm. Moreover, in Liu et al. (2020), a binary BA-based feature selection method for image steganalysis is proposed. This method selects the most relevant feature from raw features extracted to improve the final detection accuracy. Furthermore, Al-Betar et al. (2020) used the bat algorithm to search optimal features subset to increase the accuracy of cancer classification. In this method, a combination of filter and wrapper approach was used to improve the performance of feature selection. In this method, robust mRMR as a filter to select the most relevant features and an improved BA algorithm as a search strategy in the wrapper approach is presented to select the final feature subset.

4.8. COA-based methods

The Cuckoo Optimization Algorithm (COA) is another algorithm based on swarm intelligence, inspired by the special lifestyle of a bird called the cuckoo (Rajabioun, 2011). The specific habitude of laying eggs and the reproduction of this bird has been the basis for the formation of this optimization algorithm. Like other evolutionary algorithms, the cuckoo optimization algorithm begins with an initial population of cuckoos. This early population of cuckoos has a number of eggs that are placed in the nest of the host bird. Some eggs, which are more similar to host bird eggs, have a better chance of growing into adult birds. The other eggs are identified and destroyed by the host butterfly. Grown eggs show that the nest is a better place in the search space, and the usefulness of that area is higher. The goal of the cuckoo optimization algorithm, which is the optimization function, is to find the place where most eggs have a more chance to survive. In Elyasigomari et al. (2017), a COA-based feature selection method is developed to improve the cancer classification accuracy. In this method, first, the irrelevant features are removed using a simple and fast filter-based feature selection. Then, from these relevant features, the final features are selected by integration wrapper-based feature selection and COA algorithm. In Jayaraman and Sultana (2019), a combination method of cuckoo search algorithm and neural network is developed for feature selection. In this method, after selecting a feature subset of non-redundant and relevant features, the final chosen features are sent to the classifier, and heart disease is classified. Moreover, in Prabukumar et al. (2019), another cuckoo search-based feature selection method is proposed for improving the disease diagnosis accuracy. In this method, the process of feature subset selection is optimized using the cuckoo search optimization, and then these selected features are sent to the SVM classifier to identify lung cancer.

4.9. GWO-based methods

Gray Wolf Optimization (GWO) is a new evolutionary algorithm-based optimization technique inspired by gray wolves (Mirjalili et al., 2014). This optimization method is one of the latest bio-inspired techniques, which imitates the hunting process of a pack of gray wolves in nature. Recently, some GWO-based method has been utilized as a tool for feature selection in machine learning applications. In Emary et al. (2015), a Multi-Objective GWO was applied to search for the most relevant and non-redundant features. This proposed method utilized the low computational complexity of the filter model to improve the accuracy of the wrapper model. In Emary et al. (2016), a binary version of the GWO is developed to select the optimal feature subset for classification tasks. In the wrapper-based feature selection method, classification accuracy, and the number of selected features are utilized for the fitness function. In Tu et al. (2019), to enhance the previous GWO-based feature selection approach, a multi-strategy ensemble GWO is developed for feature selection. Moreover, Abdel-Basset et al. (2020) developed a GWO-based wrapper feature selection approach that integrated with a mutation operator for data classification. The mutation operator proposed in this paper tries to remove the redundant and irrelevant features.

4.10. WOA-based methods

Whale Optimization Algorithm (WOA) is another SI-based optimization method inspired by the hunting behavior of humpback whales (Mirjalili and Lewis, 2016). This optimization method consists of three operators to imitate the search for prey, encircling prey, and bubble-net foraging behavior of humpback whales. Recently, the WOA algorithm is successfully applied in many different optimization problems and feature selection. In Mafarja and Mirjalili (2017), a hybrid WOA with a simulated annealing algorithm is developed for feature selection. The aim of using simulated annealing in this hybrid method is to improve the exploitation by searching the most promising regions located by the WOA algorithm. In Mafarja and Mirjalili (2018), a WOA-based wrapper feature selection algorithm is developed. In this study, tournament and roulette wheel selection mechanism and also crossover and mutation operators are used to improve the exploration and exploitation of the search process of WOA algorithm. Nematzadeh et al. (2019) introduced a frequency-based filter feature selection method using a whale algorithm on high-dimensional medical datasets. In this method, a filter criterion is utilized to discard the irrelevant features using the WOA. Then, the reminder features are ranked based on another filtering method, namely, Mutual Congestion.

4.11. SSA-based methods

Salp Swarm Algorithm (SSA) (Mirjalili et al., 2017) another bio-inspired algorithm based on swarm intelligence is proposed for solving optimization problems. This algorithm is inspired by the swarming behavior of salps when moving and forage in the oceans. In Faris et al. (2018) an efficient binary Salp Swarm Algorithm with a crossover scheme is proposed to improve the accuracy of feature selection. Moreover, in Ibrahim et al. (2019) a hybrid optimization method for the feature selection problem is proposed by the integration of the slap swarm algorithm with particle swarm optimization. This combination between both swarm optimization algorithms improved the efficacy of the exploration and the exploitation steps. In Tubishat et al. (2020), the integration of Improved SSA and novel local search algorithm is developed for the feature selection problem. In this method, the accuracy of the classification accuracy of KNN classifier on the training data is used as a fitness function, and the final features are selected in a repetitive process. Also, in this method, to create the initial population, a method based on Opposition Based Learning is presented, which has caused diversity in the initial solutions. Furthermore, in Al-Zoubi et al. (2020) a Salp Swarm Algorithm-based feature weighting method is developed to predict the presence of liver disorder, heart, and Parkinson's disease. Also, in Hegazy et al. (2020) solution accuracy, reliability, and convergence speed of basic SSA is improved by adding a new control parameter and inertia weight. Then this improved algorithm is tested in the feature selection problem. Moreover, in Neggaz et al. (2020), a new variant of SSA optimization algorithm is proposed for feature selection. In this method, an additional phase is utilized to overcome the problem of being stuck in local optima by encouraging the exploration of the search.

5. Experimental results

In this section, the performances of the different swarm intelligence-based feature selection methods are evaluated. The results are shown in terms of the number of selected features and the classification accuracy (ACC). The classification accuracy calculated as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Where TP, TN, FP, and FN stand for the number of true positives, true negatives, false positives, and false negatives, respectively.

In each experiment, each feature selection method is run ten times, and the mean and standard deviation (square root of its variance) of ten different runs is used to compare different methods. Moreover, in each run, each dataset is normalized, and it is randomly divided into a training set (66% of the dataset) and a test set (34% of the dataset). The training set is utilized for the feature selection process, while the test set is applied for evaluating the proposed feature selection method. To fulfill fair experiments, all evaluated methods are carrying out on the same train/test dataset. Due to the randomness of the train and test set, both the average and the standard deviation of the result are reported. In the remainder of this section, used datasets, utilized classifiers, evaluated methods, results, and discussion are explained in the following subsections.

5.1. Datasets

In this study, several datasets, with different specifications, were utilized to evaluate different SI-based feature selection methods and compare their performance with each other. These datasets include SpamBase, Sonar, Arrhythmia, Madelon, Isolet, and Colon taken from the UCI repository (Asuncion and Newman, 2007) and have been extensively used in the literature. The basic characteristics of these datasets are summarized in Table 3. These datasets have been chosen considering diverse characteristics such as the number of features and the number of different classes. For example, Colon is a very high dimensional dataset with a small sample size, while SpamBase is an example of a low dimensional, with a large sample size dataset. Again Isolet is a multi-class dataset that has 26 different kinds of classes.

In some of these datasets, different features have different values. In this situation, features with a larger value range may dominate the features with a smaller value range, and maybe more likely to be selected. To overcome this challenge, before the feature selection process, all different datasets are normalized using the Max–Min normalization. Using this normalization method, the range of values of all used datasets is changed to [0, 1] range.

Moreover, in some of the used datasets, there are several missing values. To overcome this difficulty, the miss values in these features are inserted by averaging the available data corresponding to the available features.

Table 3
Characteristics of the used medical datasets.

Dataset	Features	Classes	Patterns
SpamBase	57	2	4601
Sonar	60	2	208
Arrhythmia	279	16	351
Madelon	500	2	4400
Isolet	617	26	1559
Colon	2000	2	62

5.2. The utilized classifier

To evaluate the generalizability of the proposed methods in different classifiers, in these experiments, three classifiers, including Support Vector Machine (SVM), Naïve Bayes (NB), and AdaBoost (AB) are utilized.

Support vector machine SVM is one of the supervised learning algorithms that was proposed by Vapnik. The goal of SVM is to maximize the margin between data samples, and in recent years it has shown good performance for classification and regression problems. Naïve Bayes (NB) is a group of simple probabilistic classification algorithms based on the probability that classifies data by assuming the independence of random variables and using the base theorem. AdaBoost (AB), short for "Adaptive Boosting", is a machine learning meta-algorithm formulated by Yoav Freund and Robert Schapire. An AdaBoost classifier is a meta-estimator that begins by fitting a classifier and then fits additional copies of it on the same dataset, and then the weights of incorrectly classified instances are adjusted such that subsequent classifiers focus more on severe cases.

The experimental workbench is Weka (Waikato environment for knowledge analysis) (Hall et al., 0000), which is a collection of data mining methods. In this work, SMO, Naïve Bayes, and AdaBoostM1 as the WEKA implementation of SVM, NB, and AB were used.

5.3. The evaluated methods

In these experiments, to compare the performance of different methods of selecting a feature based on, from each SI-based algorithm, one feature selection method was chosen and evaluated in the experimental result. For a fair evaluation, all of the methods examined in this section were selected from among wrapper-based methods. Due to the inherent differences between filter-based and wrapper-based feature selection methods, which usually filter-based model have lower computational complexity and wrapper-based models have higher accuracy, it is not possible to compare these methods. For a fairer comparison, the performed experiments in this section are divided into two separate parts, in the first part the wrapper-based methods and in the second part, the filter-based methods were compared with each other. These wrapper-based methods include PSO-based (Xue et al., 2020), ACO-based (Liu et al., 2019a), ABC-based (Wang et al., 2020b), DE-based (Hancer, 2020), GSA-based (Taradeh et al., 2019), FA-based (Selvakumar and Muneeswaran, 2019), BA-based (Tawhid and Dsouza, 2018), COA-based (Elyasigomari et al., 2017), GWO-based (Abdel-Basset et al., 2020), WOA-based (Mafarja and Mirjalili, 2018) and SSA-based (Neggaz et al., 2020). Furthermore, filter-based methods include PSO-based (Zhang et al., 2018a), ACO-based (Moradi and Rostami, 2015b), DE-based. Hancer et al. (2018a) and WOA-based (Nematzadeh et al., 2019).

Moreover, all these methods are implemented using C# on an Intel Core-i7 CPU with 8 GB of RAM.[1]

[1] The datasets and source code are available from the corresponding author (Mehrdad Rostami, E-mail: m.rostami@eng.uok.ac.ir) on reasonable request.

5.4. Parameters of different methods

Similar to all feather selection methods, the swarm intelligence-based method has a number of parameters, such as population size, number of iterations, etc. These parameters are important for feature selection methods because they directly control the behaviors of the learning model and have a considerable impact on the performance of final accuracy. To optimally choose these parameters, it is necessary to repeatedly set parameters and generate a number of predictions with different combinations of values, and then evaluate the prediction accuracy to select the best parameter values. As a result, choosing the best values for the parameters is an optimization problem. One way to optimize the selection of parameter values is to use an exhaustive search algorithm. Given that the accuracy of the learning model must be calculated to evaluate each combination of parameter values, this approach will not be applicable in situations where the construction of the learning model has high computational complexity.

In this paper, to implement different methods and adjust the parameters of each method, the parameter optimization method proposed in Wu et al. (2019) is used for choosing the best values for their parameters. In this parameter optimization algorithm, the Bayesian theory-based optimization algorithm is used to solve the problem.

5.5. Results

In these experiments, the classification accuracy and feature subset size are used as the performance evaluation criteria. In the experiments, first, the performances of different wrapper SI-based feature selection methods are compared over different classifiers. Table 4 summarize the average classification accuracy (in %) over ten independent runs of the different SI-based wrapper feature selection methods using SVM, NB, and AB classifiers. Each entry of Table 4 denotes the mean value as well as standard deviation (shown in parenthesis) of ten independent runs. The best result is indicated in bold face and underlined, and the second-best is in bold face. Table 4 reveals that, in most cases, the PSO-based method performs better than the other SI-based feature selection method. For example, in SpamBase dataset on the SVM classifier, PSO-based method obtained a 92.84% classification accuracy. In contrast, for ACO, ABC, DE, GSA, FA, BA, COA, GWO, WOA, and SSA-based method, these values were reported.91.87, 90.32, 88.78, 88.09, 87.91, 90. 83, 91.16, 91.73, 88.64 and 90.19, correspondingly.

Moreover, Figs. 1 to 3 show the average classification accuracy over all datasets on the SVM, Naive Bayes, and AdaBoost classifiers, respectively. As can be seen in these figures, on SVM and AB classifiers, the PSO-based method had the highest average classification accuracy, and on the Naive Bayes classifier, COA-based method won the highest rank. The results of Fig. 1 show that the PSO-based obtained 89.35% average classification accuracy and achieved the first rank with a margin of 0.50 percent compared to the ACO-based method, which obtained the second-best average classification accuracy. Moreover, from Fig. 2 results, it can be seen that the differences between the obtained classification accuracy of the COA-based method and the second-best ones (PSO-based) and third-best ones (ACO-based) on Naive Bayes classifier were reported 0.16 (i.e., 89.24–89.08) and 0.88 (89.24–88.36) percent. Furthermore, on the AB classifier, the PSO-based feature selection method gained the first rank with an average classification accuracy of 88.92%, and the ACO-based and ABC-based feature selection methods were ranked second and third with an average classification accuracy of 88.28% and 87.73%, respectively.

Table 5 records the number of selected features of the eleven wrapper SI-based feature selection methods for each dataset. It can be observed that, generally, all the eleven methods achieve a significant reduction of dimensionality by selecting only a small portion of the original features. Among the various methods, in the Arrhythmia, Madelon, Isolet, and Colon datasets, the PSO-based method has the best performance among the other SI-based, selecting only 7.21, 14.87,

Table 4
Average classification accuracy rate and as standard deviation (shown in parenthesis) over ten runs of the wrapper-based feature selection methods using SVM, Naive Bayes, and AdaBoost classifier. The best result is indicated in bold face and underlined, and the second-best is in bold face.

Dataset	Method	Classifier		
		SVM	Naive Bayes	AdaBoost
	PSO	**92.84 (2.83)**	**92.31 (1.56)**	**92.75 (2.35)**
	ACO	**91.87 (1.83)**	89.76 (2.82)	90.66 (2.91)
	ABC	90.32 (2.09)	89.91 (1.93)	**91.91 (3.13)**
	DE	88.78 (2.54)	88.32 (2.19)	89.94 (2.87)
	GSA	88.09 (1.96)	87.98 (2.02)	88.27 (2.62)
SpamBase	FA	87.91 (2.17)	**92.04 (1.82)**	90.45 (1.93)
	BA	90. 83 (2.94)	89.73 (2.37)	90.65 (1.76)
	COA	91.16 (1.65)	90.83 (2.74)	90.77 (1.84)
	GWO	91.73 (1.87)	91.85 (3.09)	91.84 (2.90)
	WOA	88.64 (2.37)	88.93 (2.18)	90.38 (1.45)
	SSA	90.19 (2.26)	90.06 (1.78)	90.19 (2.89)
	PSO	**88.14 (2.47)**	**87.91 (2.29)**	**86.93 (3.11)**
	ACO	**88.78 (2.87)**	87.32 (3.42)	85.81 (2.63)
	ABC	87.93 (1.82)	87.12 (2.81)	**86.78 (1.15)**
	DE	85.15 (2.78)	85.19 (2.22)	84.93 (2.81)
	GSA	84.71 (1.69)	85.11 (1.34)	85.62 (1.83)
Sonar	FA	84.92 (2.38)	84.19 (2.41)	86.62 (3.19)
	BA	84.03 (1.83)	84.66 (3.83)	85.92 (2.58)
	COA	84.73 (2.48)	**89.84 (2.14)**	85.28 (1.94)
	GWO	85.42 (1.19)	85.31 (2.27)	85.73 (2.08)
	WOA	86.11 (2.78)	85.65 (2.48)	85.98 (2.69)
	SSA	87.04 (3.02)	87.26 (2.19)	86.12 (2.73)
	PSO	**86.41 (2.71)**	86.05 (2.61)	**85.94 (2.37)**
	ACO	84.92 (2.38)	**86.23 (2.75)**	**85.71 (2.83)**
	ABC	**85.14 (2.83)**	85.72 (1.85)	84.31 (1.86)
	DE	82.28 (3.07)	84.78 (2.39)	83.81 (2.35)
	GSA	81.79 (2.47)	84.18 (2.61)	84.21 (2.66)
Arrhythmia	FA	82.29 (1.74)	83.28 (1.95)	82.18 (1.97)
	BA	83.36 (2.85)	84.73 (2.91)	83.46 (2.74)
	COA	83.91 (2.79)	**88.91 (2.63)**	82.31 (2.81)
	GWO	82.05 (1.96)	85.34 (1.79)	84.92 (1.70)
	WOA	83.38 (1.45)	83.61 (2.94)	83.38 (2.44)
	SSA	81.84 (2.93)	82.29 (2.09)	82.86 (2.39)
	PSO	**86.67 (3.17)**	86.63 (2.47)	**86.16 (1.74)**
	ACO	86.19 (2.28)	85.95 (1.38)	86.03 (2.18)
	ABC	**87.54 (2.73)**	**87.18 (1.81)**	**86.08 (2.91)**
	DE	85.14 (2.16)	85.05 (2.81)	85.92 (3.15)
	GSA	84.38 (3.19)	84.88 (1.92)	84.67 (2.72)
Madelon	FA	85.08 (2.82)	**86.95 (1.19)**	85.81 (2.38)
	BA	84.93 (1.84)	85.07 (2.39)	84.91 (1.96)
	COA	85.26 (2.62)	85.19 (1.83)	85.83 (2.67)
	GWO	84.27 (2.81)	85.19 (1.92)	85.37 (2.49)
	WOA	84.93 (2.37)	84.64 (2.37)	85.04 (2.38)
	SSA	85.09 (1.98)	84.94 (1.64)	84.77 (1.63)
	PSO	**85.61 (1.33)**	**85.37 (1.92)**	**85.41 (2.31)**
	ACO	**85.21 (1.91)**	**85.18 (1.82)**	**85.31 (2.39)**
	ABC	84.32 (2.93)	84.91 (2.36)	84.51 (1.77)
	DE	83.81 (2.92)	83.04 (1.73)	83.71 (2.37)
	GSA	84.04 (1.86)	84.23 (2.28)	83.81 (1.73)
Isolet	FA	84.78 (2.33)	84.01 (1.82)	84.66 (2.53)
	BA	83.91 (1.87)	83.19 (1.69)	83.81 (1.48)
	COA	85.01 (1.82)	84.81 (2.83)	85.01 (2.91)
	GWO	84.32 (2.76)	84.48 (1.91)	84.39 (2.71)
	WOA	84.51 (1.64)	84.01 (2.19)	83.92 (3.38)
	SSA	85.19 (2.81)	85.09 (2.31)	84.93 (2.61)
	PSO	**96.41 (2.81)**	**96.19 (2.31)**	**96.33 (2.39)**
	ACO	96.11 (1.72)	95.72 (1.81)	96.15 (1.81)
	ABC	93.71 (2.82)	92.81 (3.92)	92.81 (2.62)
	DE	92.29 (2.38)	91.18 (2,73)	90.91 (1.88)
	GSA	93.72 (1.84)	92.19 (1.92)	92.01 (1.03)
Colon	FA	90.37 (1.39)	90.09 (2.71)	91.26 (1.81)
	BA	92.28 (1.67)	91.26 (1.77)	91.25 (2.70)
	COA	**96.57 (3.19)**	**95.88 (2.98)**	**96.42 (1.78)**
	GWO	93.33 (2.82)	92.85 (2.71)	92.91 (3.11)
	WOA	91.19 (3.26)	90.88 (1.61)	91.27 (2.66)
	SSA	95.91 (2.18)	94.81 (2.60)	95.81 (2.36)

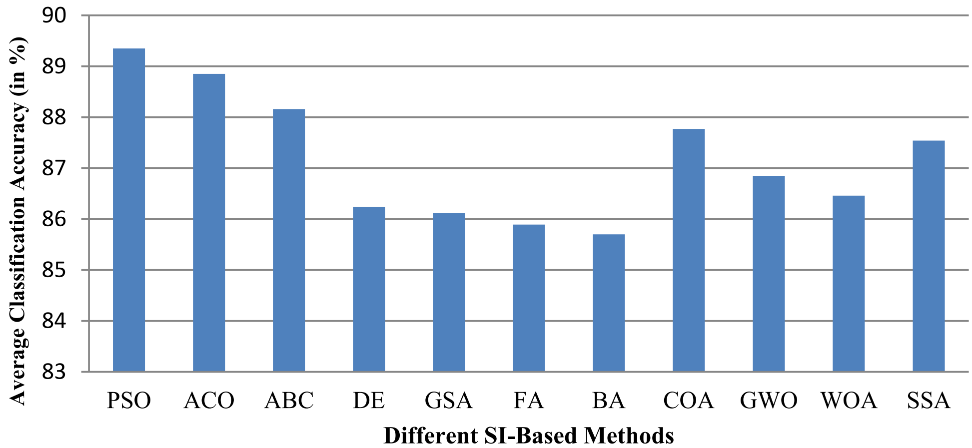

Fig. 1. Average classification accuracy over all datasets on the SVM classifier (wrapper-based methods).

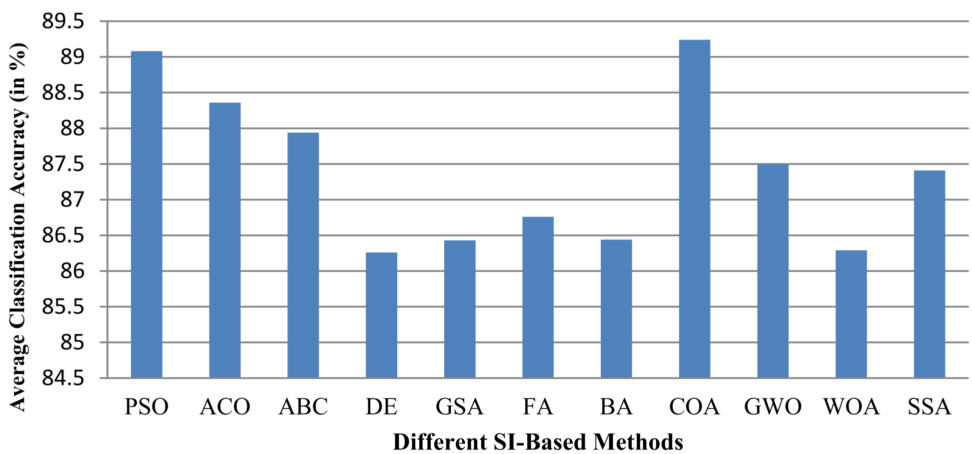

Fig. 2. Average classification accuracy over all datasets on the Naive Bayes classifier (wrapper-based methods).

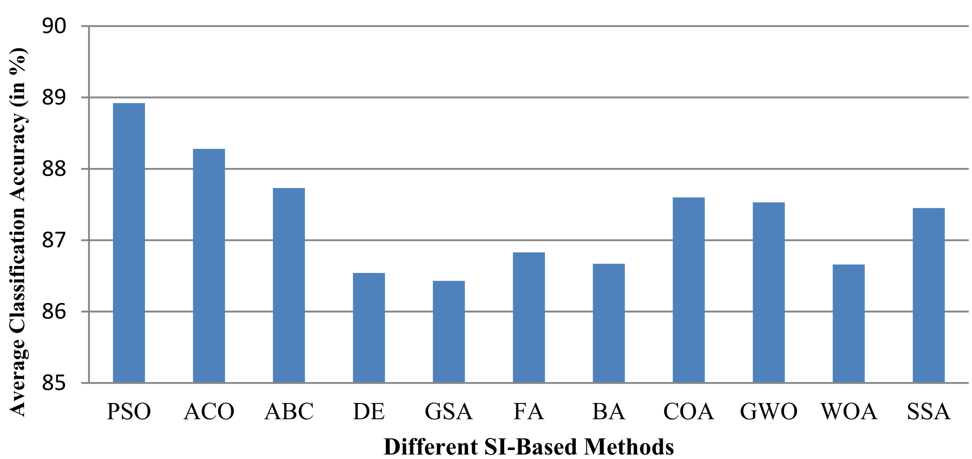

Fig. 3. Average classification accuracy over all datasets on the AdaBoost classifier (wrapper-based methods).

22.95, and 0.58%, respectively. Moreover, in the SpamBase and Sonar datasets, the ACO-base method selected an average of 8.18 and 7.11 features, respectively.

Also, several experiments were conducted to compare the execution time of different wrapper SI-based feature selection methods. In these experiments, corresponding execution times (in second) for each method, were reported in Table 6. Due to the fact that the feature selection process and the final classification process are independent, only the execution time for feature selection is reported in the data in this Table It can be seen from the results that generally the single objective SI-based feature selection methods are much faster than the multi objective SI-based feature selection methods (i.e., ACO-based and ABC-based). This is due to the fact that in multi-objective methods, several different criteria are usually considered to calculate the fitness of solutions; thus, these methods can be computationally more expensive than the single objective methods. Moreover, the reported results

Table 5
Average number of selected features of the different wrapper SI-based methods. (Minimum number of selected features is indicated in bold face and underlined and the second best is in bold face).

Dataset	Number of the original feature	Method	Number of selected features	The ratio of the selected features to the original features (in %)
SpamBase	57	PSO	**8.91**	**15.63**
		ACO	<u>**8.18**</u>	<u>**14.35**</u>
		ABC	8.98	15.75
		DE	9.32	16.35
		GSA	10.12	17.75
		FA	10.03	17.60
		BA	10.78	18.91
		COA	9.64	16.91
		GWO	10.73	18.82
		WOA	11.62	20.39
		SSA	9.17	16.09
Sonar	60	PSO	7.32	12.20
		ACO	<u>**7.11**</u>	<u>**11.85**</u>
		ABC	7.89	13.15
		DE	8.19	13.65
		GSA	8.28	13.80
		FA	7.98	13.30
		BA	8.08	13.47
		COA	**7.26**	**12.10**
		GWO	9.71	16.18
		WOA	8.48	14.13
		SSA	8.31	13.85
Arrhythmia	279	PSO	**20.12**	**7.21**
		ACO	21.88	7.84
		ABC	20.96	7.51
		DE	23.14	8.29
		GSA	22.61	8.1
		FA	24.08	8.63
		BA	22.83	8.18
		COA	21.39	7.67
		GWO	23.94	8.58
		WOA	22.06	7.91
		SSA	23.70	8.49
Madelon	500	PSO	<u>**74.35**</u>	<u>**14.87**</u>
		ACO	68.91	13.78
		ABC	**75.19**	**15.04**
		DE	77.84	15.57
		GSA	81.56	16.31
		FA	82.93	16.59
		BA	88.48	17.70
		COA	82.38	16.48
		GWO	79.19	15.84
		WOA	85.73	17.15
		SSA	78.03	15.61
Isolet	617	PSO	<u>**141.63**</u>	<u>**22.95**</u>
		ACO	**142.56**	**23.10**
		ABC	172.78	28.00
		DE	163.98	26.58
		GSA	159.62	25.87
		FA	152.83	24.77
		BA	143.48	23.25
		COA	145.85	23.64
		GWO	158.61	25.71
		WOA	164.92	26.73
		SSA	147.67	23.93
Colon	2000	PSO	<u>**11.63**</u>	<u>**0.58**</u>
		ACO	12.98	0.64
		ABC	13.26	0.66
		DE	13.64	0.68
		GSA	12.76	0.64
		FA	12.29	0.61
		BA	13.61	0.68
		COA	**12.06**	**0.60**
		GWO	14.98	0.75
		WOA	13.21	0.66
		SSA	12.45	0.62

revealed that the PSO-based feature selection method has the lowest average execution time overall datasets among all other methods. After the PSO-based method, WOA-based and GSA-based methods ranked second and third, respectively.

All of the methods evaluated in the first part of this subsection were in the category of wrapper-based feature selection methods. In the remainder of this subsection, some SI-based methods that use

Table 6
Average execution time (in second) of wrapper feature selection methods over ten independent runs.

Dataset	Wrapper-based feature selection method										
	PSO	ACO	ABC	DE	GSA	FA	BA	COA	GWO	WOA	SSA
SpamBase	6.78	8.41	8.93	9.44	6.82	7.68	7.83	9.12	8.73	7.11	8.64
Sonar	4.19	7.81	8.27	7.93	6.54	6.08	7.18	7.19	8.63	5.09	7.62
Arrhythmia	21.93	27.81	29.98	27.18	22.38	24.71	23.49	25.62	26.03	24.78	26.61
Madelon	89.51	98.32	108.67	109.67	99.32	101.56	105.84	99.53	104.86	88.73	101.78
Isolet	48.18	51.4	58.90	55.09	48.91	55.78	54.32	48.18	56.71	48.36	52.07
Colon	59.81	78.42	61.76	60.31	54.78	58.17	59.46	53.77	59.14	58.92	54.05
Average	38.4	45.36	46.09	44.94	39.79	42.33	43.02	40.57	44.02	38.83	41.8

Table 7
Average classification accuracy rate and as standard deviation (shown in parenthesis) over ten runs of the filter-based feature selection methods using SVM, Naive Bayes, and AdaBoost classifier. The best result is indicated in bold face and underlined, and the second-best is in bold face.

Dataset	Method	Classifier		
		SVM	Naive Bayes	AdaBoost
SpamBase	PSO	87.32 (1.83)	86.51 (2.35)	**87.98 (3.24)**
	ACO	**88.54 (2.11)**	**87.67 (2.29)**	**88.20 (1.78)**
	DE	**88.13 (1.64)**	**88.01 (2.81)**	87.62 (2.71)
	WOA	87.06 (1.82)	86.69 (2.95)	86.17 (1.94)
Sonar	PSO	86.03 (2.13)	**86.79 (2.43)**	**85.93 (3.72)**
	ACO	**87.74 (1.67)**	**86.28 (3.17)**	85.28 (2.39)
	DE	**86.21 (2.28)**	85.63 (2.38)	**85.89 (2.25)**
	WOA	86.13 (3.15)	86.12 (2.92)	84.93 (2.59)
Arrhythmia	PSO	**61.32 (1.98)**	**61.84 (2.39)**	**60.95 (1.71)**
	ACO	60.73 (2.93)	60.08 (2.19)	**61.38 (2.48)**
	DE	**60.85 (4.86)**	**60.13 (2.36)**	59.22 (2.70)
	WOA	60.01 (2.72)	59.94 (2.73)	60.03 (1.94)
Madelon	PSO	**75.61 (2.84)**	**74.58 (2.16)**	**74.86 (2.66)**
	ACO	**75.32 (3.06)**	**75.12 (2.81)**	**74.93 (2.87)**
	DE	73.28 (2.93)	72.98 (3.09)	73.38 (1.72)
	WOA	73.71 (1.69)	73.14 (2.62)	72.86 (2.39)
Isolet	PSO	**81.23 (2.75)**	**80.72 (2.31)**	**81.43 (2.38)**
	ACO	80.28 (2.94)	**80.46 (1.98)**	80.98 (1.16)
	DE	**81.09 (2.98)**	80.02 (2.72)	**81.16 (2.61)**
	WOA	80.14 (3.11)	79.81 (1.08)	80.02 (1.09)
Colon	PSO	**81.42 (2.19)**	**80.98 (3.81)**	**81.38 (1.98)**
	ACO	**84.53 (1.92)**	**84.13 (2.63)**	**84.77 (2.13)**
	DE	79.37 (2.83)	79.14 (3.17)	79.06 (2.41)
	WOA	78.81 (1.08)	79.56 (2.62)	78.12 (1.81)

filter approaches to search the final feature subset will be evaluated. These methods include PSO-based (Zhang et al., 2018a), ACO-based (Moradi and Rostami, 2015b), DE-based (Hancer et al., 2018a) and WOA-based (Nematzadeh et al., 2019).

In the first experiments, the performances of different filter-based methods are evaluated over different classifiers. Table 7 record the average classification accuracy (in %) over ten independent runs of the different filter SI-based feature selection methods using SVM, NB, and AB classifiers. Each entry of Table 7 shows the mean value as well as standard deviation (shown in parenthesis) of ten independent runs. The reported results of this Table shows that, in most cases, the ACO-based method performs better than the other filter SI-based feature selection method. For example, in the colon dataset on the SVM classifier, ACO-based method obtained 84.53% classification accuracy. In contrast, for PSO, DE, and WOA-based methods, these values were reported as 81.42, 79.37, and 78.81, respectively.

Also, Fig. 4 indicates the average classification accuracy over all datasets on the SVM, Naive Bayes, and AdaBoost classifiers. As can be seen in these reported results, on all classifiers, the ACO-based method had the highest average classification accuracy. For example, this figure shows that the ACO-based obtained 79.52% average classification accuracy on the SVM classifier and achieved the first rank with a margin of 0.70 percent compared to the PSO-based method, which obtained the second-best average classification accuracy. Furthermore, on the AB classifiers, the ACO-based feature selection method gained the first rank with an average classification accuracy of 79.26%, and the PSO-based and DE-based feature selection methods were ranked second and third with an average classification accuracy of 78.76% and 77.72%, respectively.

Table 8
Average execution time (in second) of filter feature selection methods over ten independent runs.

Dataset	Filter-based feature selection method			
	PSO	ACO	DE	WOA
SpamBase	3.47	4.19	4.97	4.32
Sonar	0.17	0.43	0.79	0.36
Arrhythmia	5.42	7.81	11.47	9.54
Madelon	11.34	14.87	12.38	12.98
Isolet	9.65	10.32	11.94	10.62
Colon	16.81	19.41	22.67	17.42
Average	7.81	9.50	10.70	9.20

Moreover, in Table 8 corresponding execution times (in second) for each filter SI-based feature selection method, were reported. Similar to the previous execution time table, only the time of the feature selection process is considered as the execution time, in Table 8. It can be seen that the multi-objective wrapper feature selection method (i.e. DE-based method) has a higher execution time than other single objective methods. Moreover, the reported results revealed that the PSO-based feature selection method has the lowest average execution time overall datasets among all other methods. After the PSO-based method, ACO-based methods ranked second.

5.6. Statistical analysis

In this subsection, the Friedman test (Friedman, 1940) is applied to the statistical analysis of the reported results. The Friedman test is a nonparametric test utilized to compare the performance of different feature selection on various datasets. For this purpose, each feature selection method is ranked on each dataset. To this end, the SPSS statistics acquired by IBM is used. In the Statistical test results, it is not possible to say that if the level of significance is less than the level of error, the difference between at least a pair of specimens is deducted. Since the test errors are considered at 5%, the level of significance must be lower than 0.05 to satisfy this constraint. Tables 12 and 13 present the average calculated ranking for different wrapper-based and filter-based swarm intelligence feature selection methods on each classifier. The results of Table 9 show that the PSO-based method has the best average ranking. Table 11 shows that the Friedman test has reported a p-value of 0.000084, 0.000089, and 0.000517 in the wrapper-based methods on SVM, NB, and AB classifiers, respectively. Since these values are below 0.05, it can be claimed that the results of the PSO-based method are significantly different from those of other wrapper-based methods. Moreover, Table 10 shows that PSO-based and ACO-based methods have almost equal average rank between different filter-based methods and better than other methods. The reported results of Table 11 showed that the Friedman test has reported a p-value of 0.060184, 0.085801, and 0.008101 in the filter-based methods on

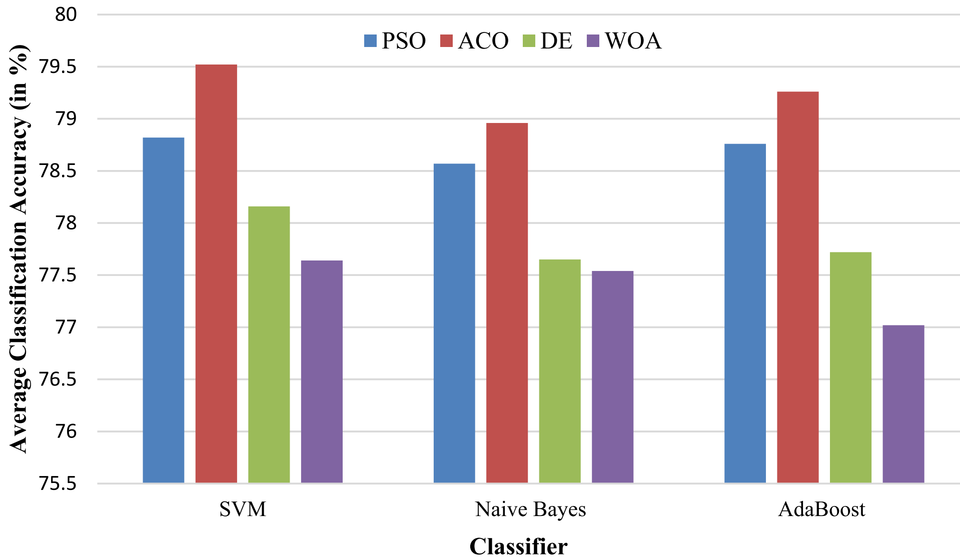

Fig. 4. Average classification accuracy over all datasets on different classifiers (wrapper-based methods).

Table 9
Average ranks of the different Wrapper-based feature selection methods on SVM, NB, and AB classifier.

Dataset	Wrapper-based feature selection method										
	PSO	ACO	ABC	DE	GSA	FA	BA	COA	GWO	WOA	SSA
SVM	1.50	2.33	4.17	7.83	9.17	8.17	8.17	4.33	7.17	7.17	5.67
Naive Bayes	1.83	3.50	4.33	8.67	8.67	7.33	8.33	3.00	5.17	8.83	6.00
AdaBoost	1.17	3.67	3.67	8.83	8.67	6.83	7.83	5.5	5.5	7.5	6.67

Table 10
Average ranks of the different Filter-based feature selection methods on SVM, NB, and AB classifier.

Dataset	Filter-based feature selection method			
	PSO	ACO	DE	WOA
SVM	2	1.83	2.5	3.67
Naive Bayes	1.83	1.83	3	3.33
AdaBoost	1.67	1.67	2.83	3.83

SVM, NB, and AB classifiers, respectively. Since these values are over 0.05, it can be claimed that the results of the best methods (i.e. PSO-based and ACO-based methods) are not significantly different from those of other filter-based methods.

5.7. Discussion

In this section, the strengths and weaknesses of the methods studied will be evaluated, and the factors that can lead to the superiority of a feature selection method will be analyzed.

1- Since each classifier has certain properties, the single classifiers are usually less accurate and generalization than the wrapper-based feature selection model that uses a combination of multiple classifiers. In other words, multiple-classifiers are usually highly accurate due to the diversity of used classifiers and the prediction performance of every single classifier. Unlike other compared wrapper SI-based feature selection methods that utilize a single classifier in their feature selection process to calculate the quality of a generated feature set, in the wrapper PSO-based feature selection method (Xue et al., 2020), four different classifiers include SVM, LDA, KNN, and ELM, are used as evaluation functions.

2- Given that in a large dataset, the number of unrelated and redundant features has also increased, evolutionary algorithms may be stuck in the local optimal. Moreover, many SI-based methods that use wrapper approaches to search for the optimal features subset are usually highly computationally complex and inefficient in high-dimensional datasets. Among the evaluated wrapper SI-based feature selection methods in this section, the PSO-based (Xue et al., 2020), ACO-based (Liu et al., 2019a) COA-based (Elyasigomari et al., 2017), and SSA-based (Neggaz et al., 2020) feature selection methods, highly accurate in the high-dimensional dataset (i.e., colon dataset with 2000 features), while other evaluated SI-based methods were only effective in low-dimensional datasets.

3- One of the main goals of an efficient feature selection method is to identify the optimal number of required features for the machine learning task and prevent the selection of too many or too few features during their feature selection process. If too many features are selected in a feature selection method, the probability of selecting redundant and irrelevant features will be increased; as a result, the prediction accuracy will be decreased. On the other hand, if too few features are selected, they will not be able to represent all the information of original features. Among the studied methods in this paper, multi-objective methods and methods that took into account the number of selected features in their fitness function showed better performance, and the number of final selected features by these methods were fewer.

4- Exploration of the search space and exploitation of the best solutions found are two conflicting objectives that must be taken into account when using a swarm intelligence-based method. Exploration means to generate diverse solutions to explore the search space on a global scale, while exploitation means to focus the search on a good solution region. A good balance between these two objectives will improve the performance of the searching method. Good SI-based feature selection methods should employ different strategies for their search processes, in order to develop a powerful method with a better balance

Table 11
The results of the Friedman statistics test.

	Wrapper-based feature selection			Filter-based feature selection		
	SVM	NB	AB	SVM	NB	AB
Chi-Square	36.009	35.857	31.334	7.400	6.600	11.800
df	10	10	10	3	3	3
Asymp.Sig (*p*-value)	0.000084	0.000089	0.000517	0.060184	0.085801	0.081010

between exploration and exploitation capabilities and better convergence speed. Among the studied wrapper methods in this paper, PSO-based (Xue et al., 2020), ACO-based (Liu et al., 2019a), ABC-based, SSA-based (Neggaz et al., 2020) method, demonstrated better performance in balancing the factors of exploration and exploitation. Furthermore, among the evaluated filter-based methods, the ACO-based method (Moradi and Rostami, 2015b) showed the best performance for the balance between exploration and exploitation.

5- A feature selection method can be evaluated from two aspects: efficiency and effectiveness. The efficiency of a feature selection method depends on the required time to find the final feature subset. While effectiveness depends on the quality of the selected feature subset. These two criteria have conflicted with each other, and usually, the improvement of one of them leads to the reduction of the other. Therefore, balancing these two criteria is an important and necessary issue in feature selection. Wrapper-based feature selection methods, due to the use of the learning algorithm in the feature selection process, will be able to effectively select a feature subset of relevant and non-redundant features. Therefore, these wrapper-based methods are usually highly accurate. On the other, these methods have high computational complexity and will have a high execution time in high-dimensional datasets. Also, filter-based methods are much more efficient than wrapper-based methods in terms of computational complexity due to the lack of learning algorithms in the feature selection process. But most filter-based methods converge to local optimal, and the quality of the selected subset in this approach is usually less than the wrapper-based methods.

6. Conclusions

With the advancement of data collection technologies and the increasing capacity of data storage over the last decades, high-dimensional datasets have grown significantly. Usually, many features of these datasets are irrelevant or redundant, which reduces the performance of the prediction model. Feature selection plays an essential role in machine learning and, more specifically, in the high-dimensional dataset. Reducing the size of the medical dataset, on the one hand, reduces the computational complexity and, on the other hand, decrees the parameters of the classification algorithm. As a result, the accuracy of the prediction model will be increased. In the past decades, the rapid growth of computer and database technologies has led to the rapid growth of large-scale datasets. On the other hand, data mining applications with high dimensional datasets that require high speed and accuracy are rapidly increasing. An important issue with these applications is the curse of dimensionality, where the number of features is much greater than the number of patterns. One of the dimensionality reduction approaches is feature selection that can increase the accuracy of the data mining task and reduce its computational complexity. The feature selection method aims at selecting a subset of features with the lowest inner similarity and highest relevancy to the target class. It reduces the dimensionality of the data by eliminating irrelevant, redundant, or noisy data.

In this paper, comparative analysis and categorization of different feature selection methods are presented. Moreover, in this paper, wrapper and filter SI-based methods (i.e., PSO, ACO, ABC. DE, GSA, FA, BA, COA, GWO, WOA, and SSA) and its application in feature selection are studied. Furthermore, the strengths and weaknesses of the different studied SI-based feature selection methods are evaluated, and the factors that can lead to the superiority of these methods are analyzed.

CRediT authorship contribution statement

Mehrdad Rostami: Conceptualization, Writing - original draft, Validation, Software. **Kamal Berahmand:** Conceptualization, Software. **Elahe Nasiri:** Validation, Formal analysis. **Saman Forouzandeh:** Writing - review & editing, Formal analysis.

Declaration of competing interest

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

References

Abdel-Basset, M., et al., 2020. A new fusion of grey wolf optimizer algorithm with a two-phase mutation for feature selection. Expert Syst. Appl. 139, 112824.

Al-Ani, A., Alsukker, A., Khushaba, R.N., 2013. Feature subset selection using differential evolution and a wheel based search strategy. Swarm Evol. Comput. 9, 15–26.

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

Al-Zoubi, A.M., et al., 2020. Salp chain-based optimization of support vector machines and feature weighting for medical diagnostic information systems. In: Mirjalili, S., Faris, H., Aljarah, I. (Eds.), Evolutionary Machine Learning Techniques: Algorithms and Applications. Springer Singapore, Singapore, pp. 11–34.

Alshamlan, H.M., Badr, G.H., Alohali, Y.A., 2015. Genetic Bee Colony (GBC) algorithm: A new gene selection method for microarray cancer classification. Comput. Biol. Chem. 56, 49–60.

Arslan, S., Ozturk, C., 2019. Multi hive artificial bee colony programming for high dimensional symbolic regression with feature selection. Appl. Soft Comput. 78, 515–527.

Asuncion, A., Newman, D., 2007. UCI repository of machine learning datasets. Available from: http://archive.ics.uci.edu/ml/datasets.html.

Bandyopadhyay, S., et al., 2014. Integration of dense subgraph finding with feature clustering for unsupervised feature selection. Pattern Recognit. Lett. 40, 104–112.

Banka, H., Dara, S., 2015. A Hamming distance based binary particle swarm optimization (HDBPSO) algorithm for high dimensional feature selection, classification and validation. Pattern Recognit. Lett. 52, 94–100.

Barak, S., Dahooie, J.H., Tichý, T., 2015. Wrapper ANFIS-ICA method to do stock market timing and feature selection on the basis of Japanese Candlestick. Expert Syst. Appl. 42 (23), 9221–9235.

Basir, M., Ahmad, F., 2014. Comparison on swarm algorithms for feature selections reductions. Int. J. Sci. Eng. Res. 5, 479–486.

Battiti, R., 1994. Using mutual information for selecting features in supervised neural net learning. IEEE Trans. Neural Netw. 5 (4), 537–550.

Berahmand, K., et al., 2020. A new attributed graph clustering by using label propagation in complex networks. J. King Saud Univ., Comput. Inf. Sci..

Bi, N., et al., 2019. A multi-feature selection approach for gender identification of handwriting based on kernel mutual information. Pattern Recognit. Lett. 121, 123–132.

Brezočnik, L., Fister, I., Podgorelec, V., 2018. Swarm intelligence algorithms for feature selection: A review. Appl. Sci. 8 (9), 1521.

Cadenas, J.M., Garrido, M.C., Martínez, R., 2013. Feature subset selection Filter–Wrapper based on low quality data. Expert Syst. Appl. 40 (16), 6241–6252.

Chandrashekar, G., Sahin, F., 2014. A survey on feature selection methods. Comput. Electr. Eng. 40 (1), 16–28.

Chen, B., Chen, L., Chen, Y., 2013. Efficient ant colony optimization for image feature selection. Signal Process. 93 (6), 1566–1576.

Chen, H., et al., 2019. Feature selection for imbalanced data based on neighborhood rough sets. Inform. Sci. 483, 1–20.

Chen, R.-C., et al., 2020. Selecting critical features for data classification based on machine learning methods. J. Big Data 7 (1), 52.
Dadaneh, B.Z., Markid, H.Y., Zakerolhosseini, A., 2016. Unsupervised probabilistic feature selection using ant colony optimization. Expert Syst. Appl. 53, 27–42.
Ding, D., et al., 2020. Unsupervised feature selection via adaptive hypergraph regularized latent representation learning. Neurocomputing 378, 79–97.
Dorigo, M., Caro, G.D., 1999. Ant colony optimization: A new meta-heuristic. In: Proceeding of the Congress on Evolutionary Computing.
Elyasigomari, V., et al., 2017. Development of a two-stage gene selection method that incorporates a novel hybrid approach using the cuckoo optimization algorithm and harmony search for cancer classification. J. Biomed. Inform. 67, 11–20.
Emary, E., Zawbaa, H.M., Hassanien, A.E., 2016. Binary grey wolf optimization approaches for feature selection. Neurocomputing 172, 371–381.
Emary, E., et al., 2015. Multi-objective gray-wolf optimization for attribute reduction. Procedia Comput. Sci. 65, 623–632.
Estévez, P.A., et al., 2009. Normalized mutual information feature selection. IEEE Trans. Neural Netw. 20 (2), 189–201.
Farahat, Ahmed K., Ghodsi, Ali, Kamel, M.S., 2013. Kamel efficient greedy feature selection for unsupervised learning. Knowl. Inf. Syst. 35 (2), 285–310.
Faris, H., et al., 2018. An efficient binary salp swarm algorithm with crossover scheme for feature selection problems. Knowl.-Based Syst. 154, 43–67.
Ferreira, A.J., Figueiredo, M.A., 2012. An unsupervised approach to feature discretization and selection. Pattern Recognit. 45 (9), 3048–3060.
Figueiredo, E., et al., 2019. Swarm intelligence for clustering — A systematic review with new perspectives on data mining. Eng. Appl. Artif. Intell. 82, 313–329.
Forouzandeh, S., Berahmand, K., Rostami, M., 2020. Presentation of a recommender system with ensemble learning and graph embedding: A case on MovieLens. Multimedia Tools Appl..
Forsati, et al., 2014. Enriched ant colony optimization and its application in feature selection. Neurocomputing 142, 354–371.
Friedman, M., 1940. A comparison of alternative tests of significance for the problem of m rankings. Ann. Math. Stat. 11 (1), 86–92.
Ghaemi, M., Feizi-Derakhshi, M.-R., 2016. Feature selection using forest optimization algorithm. Pattern Recognit. 60, 121–129.
Ghimatgar, H., et al., 2018. An improved feature selection algorithm based on graph clustering and ant colony optimization. Knowl.-Based Syst. 159, 270–285.
Gokalp, O., Tasci, E., Ugur, A., 2020. A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification. Expert Syst. Appl. 146, 113176.
Gong, D., et al., 2020. A similarity-based cooperative co-evolutionary algorithm for dynamic interval multiobjective optimization problems. IEEE Trans. Evol. Comput. 24 (1), 142–156.
González, J., et al., 2019. A new multi-objective wrapper method for feature selection – Accuracy and stability analysis for BCI. Neurocomputing 333, 407–418.
Gu, Quanquan, Li, Zhenhui, Han, J., 2011. Generalized Fisher score for feature selection. In: Proceedings of the International Conference on Uncertainty in Artificial Intelligence.
Gunasundari, S., Janakiraman, S., Meenambal, S., 2018. Multiswarm heterogeneous binary PSO using win-win approach for improved feature selection in liver and kidney disease diagnosis. Comput. Med. Imaging Graph. 70, 135–154.
Hall, M., et al., The WEKA data mining software. Available from: http://www.cs.waikato.ac.nz/ml/weka.
Han, X., et al., 2014. Feature subset selection by gravitational search algorithm optimization. Inform. Sci. 281, 128–146.
Hancer, E., 2020. A new multi-objective differential evolution approach for simultaneous clustering and feature selection. Eng. Appl. Artif. Intell. 87, 103307.
Hancer, E., Xue, B., Zhang, M., 2018a. Differential evolution for filter feature selection based on information theory and feature ranking. Knowl.-Based Syst. 140, 103–119.
Hancer, E., et al., 2015a. A multi-objective artificial bee colony approach to feature selection using fuzzy mutual information. In: 2015 IEEE Congress on Evolutionary Computation, CEC. IEEE.
Hancer, E., et al., 2015b. A binary ABC algorithm based on advanced similarity scheme for feature selection. Appl. Soft Comput. 36, 334–348.
Hancer, E., et al., 2018b. Pareto front feature selection based on artificial bee colony optimization. Inform. Sci. 422, 462–479.
Hashemi, A., Dowlatshahi, M.B., Nezamabadi-pour, H., 2020. MGFS: A multi-label graph-based feature selection algorithm via PageRank centrality. Expert Syst. Appl. 142, 113024.
He, Xiaofei, Cai, Deng, Niyogi1, P., 2005. Laplacian score for feature selection. Adv. Neural Inf. Process. Syst. 18, 507–514.
Hegazy, A.E., Makhlouf, M.A., El-Tawel, G.S., 2020. Improved salp swarm algorithm for feature selection. J. King Saud Univ., Comput. Inf. Sci. 32 (3), 335–344.
Henni, K., Mezghani, N., Gouin-Vallerand, C., 2018. Unsupervised graph-based feature selection via subspace and pagerank centrality. Expert Syst. Appl. 114, 46–53.
Hu, L., et al., 2018. Feature selection considering two types of feature relevancy and feature interdependency. Expert Syst. Appl. 93, 423–434.
Hu, Y., et al., 2020a. A dynamic multi-objective evolutionary algorithm based on intensity of environmental change. Inform. Sci. 523, 49–62.
Hua, Z., et al., 2020. Strong approximate Markov blanket and its application on filter-based feature selection. Appl. Soft Comput. 87, 105957.
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.
Huang, Cheng-Lung, Tsai, C.-Y., 2009. A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting. Expert Syst. Appl. 36 (2), 1529–1539.
Ibrahim, R.A., et al., 2019. Improved salp swarm algorithm based on particle swarm optimization for feature selection. J. Ambient Intell. Humaniz. Comput. 10 (8), 3155–3169.
Inbarani, H.H., Azar, A.T., Jothi, G., 2014. Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis. Comput. Methods Programs Biomed. 113 (1), 175–185.
Jain, Anil K., Duin, Robert P.W., Mao, J., 2000. Statistical pattern recognition: A review. IEEE Trans. Pattern Anal. Mach. Intell. 22 (1), 4–37.
Jain, I., Jain, V.K., Jain, R., 2018. Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification. Appl. Soft Comput. 62, 203–215.
Jayaraman, V., Sultana, H.P., 2019. Artificial gravitational cuckoo search algorithm along with particle bee optimized associative memory neural network for feature selection in heart disease classification. J. Ambient Intell. Humaniz. Comput..
Ji, X., et al., 2019. Visual exploration of neural document embedding in information retrieval: Semantics and feature selection. IEEE Trans. Vis. Comput. Graphics 25 (6), 2181–2192.
Kabir, Monirul, Shahjahan, B., M., K., 2012. A new hybrid ant colony optimization algorithm for feature selection. Expert Syst. Appl. 39 (3), 3747–3763.
Ke, L., Feng, Z., Ren, Z., 2008. An efficient ant colony optimization approach to attribute reduction in rough set theory. Pattern Recognit. Lett. 29 (9), 1351–1357.
Ke, L., et al., 2010. A multiobjective ACO algorithm for rough feature selection. In: 2010 Second Pacific-Asia Conference on Circuits, Communications and System. IEEE.
Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. In: The Proceedings of the 1995 IEEE International Conference on Neural Network. pp. 1942–1948.
Koide, K., Miura, J., Menegatti, E., 2020. Monocular person tracking and identification with on-line deep feature selection for person following robots. Robot. Auton. Syst. 124, 103348.
Kou, G., et al., 2020. Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods. Appl. Soft Comput. 86, 105836.
Kwak, N., Choi, C.-H., 2002. Input feature selection for classification problems. IEEE Trans. Neural Netw. 13 (1), 143–159.
Labani, M., et al., 2018. A novel multivariate filter method for feature selection in text classification problems. Eng. Appl. Artif. Intell. 70, 25–37.
Lai, Carmen, Reinders, Marcel J.T., Wessels, L., 2006. Random subspace method for multivariate feature selection. Pattern Recognit. Lett. 27 (10), 067–1076.
Larabi Marie-Sainte, S., Alalyani, N., 2020. Firefly algorithm based feature selection for Arabic text classification. J. King Saud Univ., Comput. Inf. Sci. 32 (3), 320–328.
Li, A.-D., Xue, B., Zhang, M., 2020a. Multi-objective feature selection using hybridization of a genetic algorithm and direct multisearch for key quality characteristic selection. Inform. Sci. 523, 245–265.
Li, Ying, et al., 2013. An ant colony optimization based dimension reduction method for high-dimensional datasets. J. Bionic Eng. 10 (2), 231–241.
Li, S., et al., 2019a. Dual graph regularized compact feature representation for unsupervised feature selection. Neurocomputing 331, 77–96.
Li, J., et al., 2019b. Exclusive feature selection and multi-view learning for Alzheimer's disease. J. Vis. Commun. Image Represent. 64, 102605.
Li, J., et al., 2020b. An enhanced-indicator based many-objective evolutionary algorithm with adaptive reference point. Swarm Evol. Comput. 55, 100669.
Liang, Y., Zhang, M., Browne, W.N., 2017. Image feature selection using genetic programming for figure-ground segmentation. Eng. Appl. Artif. Intell. 62, 96–108.
Lin, Chuen-Horng, Chen, Huan-Yu, Wua, Y.-S., 2014. Study of image retrieval and classification based on adaptive features using genetic algorithm feature selection. Expert Syst. Appl. 41 (15), 6611–6621.
Liu, Q., Gu, Q., Wu, Z., 2017. Feature selection method based on support vector machine and shape analysis for high-throughput medical data. Comput. Biol. Med. 91, 103–111.
Liu, F., Yan, X., Lu, Y., 2020. Feature selection for image steganalysis using binary bat algorithm. IEEE Access 8, 4244–4249.
Liu, H., Yu, L., 2005. Toward integrating feature selection algorithms for classification and clustering. IEEE Trans. Knowl. Data Eng. 17 (4), 491–502.
Liu, Y., Zheng, Y.F., 2006. FS_SFS: A novel feature selection method for support vector machines. Pattern Recognit. 39 (7), 1333–1345.
Liu, Y., et al., 2019a. A classification method based on feature selection for imbalanced data. IEEE Access 7, 81794–81807.
Liu, Y., et al., 2019b. Flexible unsupervised feature extraction for image classification. Neural Netw. 115, 65–71.
Lyu, H., et al., 2017. A filter feature selection method based on the maximal information coefficient and Gram–Schmidt orthogonalization for biomedical data mining. Comput. Biol. Med. 89, 264–274.
Mafarja, M.M., Mirjalili, S., 2017. Hybrid whale optimization algorithm with simulated annealing for feature selection. Neurocomputing 260, 302–312.
Mafarja, M., Mirjalili, S., 2018. Whale optimization approaches for wrapper feature selection. Appl. Soft Comput. 62, 441–453.

Majbouri Yazdi, K., et al., 2020. Prediction optimization of diffusion paths in social networks using integration of ant colony and densest subgraph algorithms. J. High Speed Netw. 26, 141–153.

Martín-Smith, P., et al., 2017. A supervised filter method for multi-objective feature selection in EEG classification based on multi-resolution analysis for BCI. Neurocomputing 250, 45–56.

Mirjalili, S., Lewis, A., 2016. The whale optimization algorithm. Adv. Eng. Softw. 95, 51–67.

Mirjalili, S., Mirjalili, S.M., Lewis, A., 2014. Grey wolf optimizer. Adv. Eng. Softw. 69, 46–61.

Mirjalili, S., et al., 2017. Salp swarm algorithm: A bio-inspired optimizer for engineering design problems. Adv. Eng. Softw. 114, 163–191.

Mitchell, T.M., 1997. Machine Learning. McGraw-Hill, NewYork.

Mlakar, U., et al., 2017. Multi-objective differential evolution for feature selection in facial expression recognition systems. Expert Syst. Appl. 89, 129–137.

Moradi, P., Gholampour, M., 2016. A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy. Appl. Soft Comput. 43, 117–130.

Moradi, P., Rostami, M., 2015a. A graph theoretic approach for unsupervised feature selection. Eng. Appl. Artif. Intell. 44, 33–45.

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

Morita, M., et al., 2003. Unsupervised Feature Selection Using Multi-Objective Genetic Algorithms for Handwritten Word Recognition. IEEE.

Nayar, N., Ahuja, S., Jain, S., 2019. Swarm intelligence for feature selection: A review of literature and reflection on future challenges. In: Advances in Data and Information Sciences. Springer Singapore, Singapore.

Neggaz, N., et al., 2020. Boosting salp swarm algorithm by sine cosine algorithm and disrupt operator for feature selection. Expert Syst. Appl. 145, 113103.

Nematzadeh, H., et al., 2019. Frequency based feature selection method using whale algorithm. Genomics.

Nguyen, B.H., Xue, B., Zhang, M., 2020. A survey on swarm intelligence approaches to feature selection in data mining. Swarm Evol. Comput. 54, 100663.

Pashaei, E., Pashaei, E., Aydin, N., 2019. Gene selection using hybrid binary black hole algorithm and modified binary particle swarm optimization. Genomics 111 (4), 669–686.

Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. 27 (8), 1226–1238.

Prabukumar, M., Agilandeeswari, L., Ganesan, K., 2019. An intelligent lung cancer diagnosis system using cuckoo search optimization and support vector machine classifier. J. Ambient Intell. Humaniz. Comput. 10 (1), 267–293.

Prasad, Y., Biswas, K.K., Hanmandlu, M., 2018. A recursive PSO scheme for gene selection in microarray data. Appl. Soft Comput. 71, 213–225.

Qasim, O.S., Algamal, Z.Y., 2018. Feature selection using particle swarm optimization-based logistic regression model. Chemometr. Intell. Lab. Syst. 182, 41–46.

Raileanu, L.E., Stoffel, K., 2004. Theoretical comparison between the Gini index and information gain criteria. Ann. Math. Artif. Intell. 41, 77–93.

Rajabioun, R., 2011. Cuckoo optimization algorithm. Appl. Soft Comput. 11 (8), 5508–5518.

Rashedi, E., Nezamabadi-pour, H., Saryazdi, S., 2009. GSA: A gravitational search algorithm. Inform. Sci. 179 (13), 2232–2248.

Rehman, A., Javed, K., Babri, H.A., 2017. Feature selection based on a normalized difference measure for text classification. Inf. Process. Manage. 53 (2), 473–489.

Rehman, A., et al., 2015. Relative discrimination criterion–A novel feature ranking method for text data. Expert Syst. Appl. 42 (7), 3670–3681.

Renuka Devi, D., Sasikala, S., 2019. Online Feature Selection (OFS) with Accelerated Bat Algorithm (ABA) and Ensemble Incremental Deep Multiple Layer Perceptron (EIDMLP) for big data streams. J. Big Data 6 (1), 103.

Rostami, M., Berahmand, K., Forouzandeh, S., 2020a. A novel method of constrained feature selection by the measurement of pairwise constraints uncertainty. J. Big Data 7 (1), 83.

Rostami, M., Moradi, P., 2014. A clustering based genetic algorithm for feature selection. Inf. Knowl. Technol. 112–116.

Rostami, M., et al., 2020b. Integration of multi-objective PSO based feature selection and node centrality for medical datasets. Genomics 112 (6), 4370–4384.

Saeys, Y., Inza, I., Larranaga, P., 2007. A review of feature selection techniques in bioinformatics. Bioinformatics 23 (19), 2507–2517.

Santucci, V., Baioletti, M., Milani, A., 2020. An algebraic framework for swarm and evolutionary algorithms in combinatorial optimization. Swarm Evol. Comput. 55, 100673.

Sayed, S., et al., 2019. A nested genetic algorithm for feature selection in high-dimensional cancer Microarray datasets. Expert Syst. Appl. 121, 233–243.

Schiezaro, Mauricio, Pedrini, H., 2013. Data feature selection based on artificial bee colony algorithm. EURASIP J. Image Video Process. 2013, 47.

Sebastiani, F., 2002. Machine learning in automated text categorization. ACM Comput. Surv. 34 (1), 1–47.

Selvakumar, B., Muneeswaran, K., 2019. Firefly algorithm based feature selection for network intrusion detection. Comput. Secur. 81, 148–155.

Senawi, A., Wei, H.-L., Billings, S.A., 2017. A new maximum relevance-minimum multicollinearity (MRmMC) method for feature selection and ranking. Pattern Recognit. 67, 47–61.

Shu, W., Qian, W., Xie, Y., 2020. Incremental feature selection for dynamic hybrid data using neighborhood rough set. Knowl.-Based Syst. 105516.

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.

Singh, U., Singh, S.N., 2019. A new optimal feature selection scheme for classification of power quality disturbances based on ant colony framework. Appl. Soft Comput. 74, 216–225.

Song, Q., Ni, J., Wang, G., 2013. A fast clustering-based feature subset selection algorithm for high-dimensional data. IEEE Trans. Knowl. Data Eng. 25 (1), 1–14.

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

Tabakhi, S., Moradi, P., Akhlaghian, F., 2014. An unsupervised feature selection algorithm based on ant colony optimization. Eng. Appl. Artif. Intell. 32, 112–123.

Tadist, K., et al., 2019. Feature selection methods and genomic big data: A systematic review. J. Big Data 6 (1), 79.

Tang, X., Dai, Y., Xiang, Y., 2019. Feature selection based on feature interactions with application to text categorization. Expert Syst. Appl. 120, 207–216.

Tang, C., et al., 2018. Robust unsupervised feature selection via dual self-representation and manifold regularization. Knowl.-Based Syst. 145, 109–120.

Taradeh, M., et al., 2019. An evolutionary gravitational search-based feature selection. Inform. Sci. 497, 219–239.

Tawhid, M.A., Dsouza, K.B., 2018. Hybrid binary bat enhanced particle swarm optimization algorithm for solving feature selection problems. Appl. Comput. Inform..

Theodoridis, S., Koutroumbas, C., 2009. Pattern Recognition, fourth ed. Elsevier Inc.

Tu, Q., Chen, X., Liu, X., 2019. Multi-strategy ensemble grey wolf optimizer and its application to feature selection. Appl. Soft Comput. 76, 16–30.

Tubishat, M., et al., 2020. Improved salp swarm algorithm based on opposition based learning and novel local search algorithm for feature selection. Expert Syst. Appl. 145, 113122.

Unler, A., Murat, A., Chinnam, R.B., 2011. mr2PSO: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification. Inform. Sci. 181 (20), 4625–4641.

Uysal, A.K., Gunal, S., 2012. A novel probabilistic feature selection method for text classification. Knowl.-Based Syst. 36, 226–235.

Vignolo, L.D., Milone, D.H., Scharcanski, J., 2013. Feature selection for face recognition based on multi-objective evolutionary wrappers. Expert Syst. Appl. 40 (13), 5077–5084.

Wang, H., Hong, M., 2019. Supervised hebb rule based feature selection for text classification. Inf. Process. Manage. 56 (1), 167–191.

Wang, Z., Li, M., Li, J., 2015. A multi-objective evolutionary algorithm for feature selection based on mutual information with a new redundancy measure. Inform. Sci. 307, 73–88.

Wang, T., Lu, J., Zhang, G., 2018. Two-stage fuzzy multiple kernel learning based on Hilbert–Schmidt independence criterion. IEEE Trans. Fuzzy Syst. 26 (6), 3703–3714.

Wang, C., Pan, H., Su, Y., 2020a. A many-objective evolutionary algorithm with diversity-first based environmental selection. Swarm Evol. Comput. 53, 100641.

Wang, H., et al., 2019. A factor graph model for unsupervised feature selection. Inform. Sci. 480, 144–159.

Wang, X.-h., et al., 2020b. Multi-objective feature selection based on artificial bee colony: An acceleration approach with variable sample size. Appl. Soft Comput. 88, 106041.

Welikala, R.A., et al., 2015. Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy. Comput. Med. Imaging Graph. 43, 64–77.

Wu, H., Liu, H., Duan, Z., 2020. PM2.5 concentrations forecasting using a new multi-objective feature selection and ensemble framework. Atmos. Pollut. Res..

Wu, J., et al., 2019. Hyperparameter optimization for machine learning models based on Bayesian optimizationb. J. Electron. Sci. Technol. 17 (1), 26–40.

Xiang, J., et al., 2015. A novel hybrid system for feature selection based on an improved gravitational search algorithm and k-NN method. Appl. Soft Comput. 31, 293–307.

Xu, Y., et al., 2007. A study on mutual information-based feature selection for text categorization. J. Comput. Inf. Syst. 3 (3), 1007–1012.

Xue, B., Zhang, M., Browne, W.N., 2013. Particle swarm optimization for feature selection in classification: A multi-objective approach. IEEE Trans. Cybern. 43 (6), 1656–1671.

Xue, B., Zhang, M., Browne, W.N., 2014. Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms. Appl. Soft Comput. 18, 261–276.

Xue, Y., et al., 2020. Self-adaptive parameter and strategy based particle swarm optimization for large-scale feature selection problems with multiple classifiers. Appl. Soft Comput. 88, 106031.

Yamada, M., et al., 2014. High-dimensional feature selection by feature-wise kernelized lasso. Neural Comput. 26 (1), 185–207.

Yan, K., et al., 2018. Cost-sensitive and sequential feature selection for chiller fault detection and diagnosis. Int. J. Refrig. 86, 401–409.

Yan, C., et al., 2019. A novel hybrid feature selection strategy in quantitative analysis of laser-induced breakdown spectroscopy. Anal. Chim. Acta 1080, 35–42.

Yang, X.-S., 2010a. Firefly Algorithm, Lévy Flights and Global Optimization. Springer London, London.

Yang, X.-S., 2010b. A new metaheuristic bat-inspired algorithm. In: González, J.R., et al. (Eds.), Nature Inspired Cooperative Strategies for Optimization, NICSO 2010. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 65–74.

Yang, Wenzhu, Li, Daoliang, Zhu, L., 2011. An improved genetic algorithm for optimal feature subset selection from multi-character feature set. Expert Syst. Appl. 38, 2733–2740.

Yazdi, K.M., et al., 2019. Improving recommender systems accuracy in social networks using popularity. In: 2019 20th International Conference on Parallel and Distributed Computing, Applications and Technologies, PDCAT.

Yong, Z., Dun-wei, G., Wan-qiu, Z., 2016. Feature selection of unreliable data using an improved multi-objective PSO algorithm. Neurocomputing 171, 1281–1290.

Yuan, X., et al., 2020. Integrated long-term stock selection models based on feature selection and machine learning algorithms for China Stock Market. IEEE Access 8, 22672–22685.

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

Zhang, Z., Hancock, E.R., 2012. Hypergraph based information-theoretic feature selection. Pattern Recognit. Lett. 33 (15), 1991–1999.

Zhang, Y., Song, X.-f., Gong, D.-w., 2017b. A return-cost-based binary firefly algorithm for feature selection. Inform. Sci. 418–419, 561–574.

Zhang, X., et al., 2015a. Embedded feature-selection support vector machine for driving pattern recognition. J. Franklin Inst. B 352 (2), 669–685.

Zhang, S., et al., 2015b. Swarm intelligence applied in green logistics: A literature review. Eng. Appl. Artif. Intell. 37, 154–169.

Zhang, T., et al., 2018a. A fast feature selection algorithm based on swarm intelligence in acoustic defect detection. IEEE Access 6, 28848–28858.

Zhang, L., et al., 2018b. Feature selection using firefly optimization for classification and regression models. Decis. Support Syst. 106, 64–85.

Zhang, Y., et al., 2019a. Cost-sensitive feature selection using two-archive multi-objective artificial bee colony algorithm. Expert Syst. Appl. 137, 46–58.

Zhang, H., et al., 2019b. An efficient framework for unsupervised feature selection. Neurocomputing 366, 194–207.

Zhang, Y., et al., 2020. Binary differential evolution with self-learning for multi-objective feature selection. Inform. Sci. 507, 67–85.

Zhou, X., et al., 2017. Eye tracking data guided feature selection for image classification. Pattern Recognit. 63, 56–70.

Zhu, P., et al., 2017. Subspace clustering guided unsupervised feature selection. Pattern Recognit. 66, 364–374.