A novel feature selection method for data mining tasks using hybrid Sine Cosine Algorithm and Genetic Algorithm

Laith Abualigah[1] · Akram Jamal Dulaimi[1]

Abstract

Feature selection (FS) is a real-world problem that can be solved using optimization techniques. These techniques proposed solutions to make a predictive model, which minimizes the classifier's prediction errors by selecting informative or important features by discarding redundant, noisy, and irrelevant attributes in the original dataset. A new hybrid feature selection method is proposed using the Sine Cosine Algorithm (SCA) and Genetic Algorithm (GA), called SCAGA. Typically, optimization methods have two main search strategies; exploration of the search space and exploitation to determine the optimal solution. The proposed SCAGA resulted in better performance when balancing between exploitation and exploration strategies of the search space. The proposed SCAGA has also been evaluated using the following evaluation criteria: classification accuracy, worst fitness, mean fitness, best fitness, the average number of features, and standard deviation. Moreover, the maximum accuracy of a classification and the minimal features were obtained in the results. The results were also compared with a basic Sine Cosine Algorithm (SCA) and other related approaches published in literature such as Ant Lion Optimization and Particle Swarm Optimization. The comparison showed that the obtained results from the SCAGA method were the best overall the tested datasets from the UCI machine learning repository.

Keywords Optimization problems · Feature selection · Sine Cosine algorithm · Genetic algorithm · Hybridization

1 Introduction

When addressing the data mining field, one must expand briefly and mention the broader concept of Knowledge Discovery in Databases (KDD), which categorizes datasets to find useful information from a large-scale dataset. The Discovery of knowledge has four main processes: data warehouse, preprocessing, data mining, and evaluation process [1, 2]. This process is essential in information acquisition, Machine Learning, pattern discovery, data visualization, databases, statistics, and artificial intelligence.

Data mining goes through multiple steps such as reversion, categorizing, grouping, deviation-change discovery, dependence modeling, and summarization [3]. For the data mining process to obtain the best results, a vital step must be conducted: reduction, transformation, normalization, discretization, integration, feature extraction, data cleaning, and feature selection of the data. Feature selection (variable or subsets selection) is the highlight of the study; it points to the process of choosing relevant and essential features (features are also identified as attributes, properties, characteristics, and dimensions). It discards any irrelevant, redundant, and noisy features that may decrease the accuracy of the classification, therefore lowering the algorithm's performance and the accuracy of the output. With that being said, problems are created by a faulty feature selection result in a more complicated learning process and high computational expenses [4–6].

Feature selection can encounter many problems. The most significant and commonly encountered problem is the curse of dimensionality, which happens when the features' attributes or numbers exceed the samples' number resulting in problems that lower accuracy and affect learning speed. To address this problem, datasets must be summarized in order to find smaller or narrower attributes and samples that are more relevant to the original matrix while

✉ Laith Abualigah
 Aligah.2020@gmail.com

[1] Faculty of Computer Sciences and Informatics, Amman Arab University, Amman 11953, Jordan

eliminating noise and redundancy, resulting in an enhanced discrimination power and a better classification performance; this process is named dimensionality reduction [2, 7]. Feature selection techniques are applied in several applications such as image processing [8], signal processing [9], pattern recognition [10], text clustering [11–13], and machine learning [14, 15].

Approaches of feature selection are usually categorized into three wide groups: wrapper model, filter model, and embedded model; such approaches are selected based on how the selection algorithm and the process of building the model are combined. The wrapping-based model employs learning algorithms to find the optimal subset and evaluate it, which results in finding the predictions that have better performance. Filter model approaches are rankers; they rank attributes based on their relevance to the output variable while having an independent evaluator from any learning algorithm. A filter method has adequate generalization capacity and low computational cost; this method can also handle high-scale dataset. The embedded approach is the method that combines the benefits of the wrapper and filter methods while trying to eliminate their disadvantages. This method starts just like the filter method by independently trying to find the optimal subset. It then employs linear classifiers such as Support Vector Machine (SVM) to enhance this subset by finding correlated features locally to have better local discrimination, resulting in a final optimal subset but at a lower computational cost than that of the wrapper model [16].

Feature Selection (FS) problems are considered real-world problems that affect the classification accuracy and learning speed. Metaheuristic algorithms are high procedure algorithms designed to search for a good enough solution to the search space. Such algorithms use two conflicting criteria in determining the best solution: exploration and exploitation of the search space. Exploring the search space is trying to find a good enough solution in the whole search area. In contrast, exploitation is a method that tries to determine the optimal solutions. In the native sine–cosine Algorithm, the weakness in its exploration strategy is noted, which leads to weakness in its performance during the search space. However, enhancement or modification by creating a hybrid technique can be done by introducing a new version of metaheuristic algorithms to improve performance by balancing the search space's exploitation and exploration. This stimulus underlies all of our attempts to create a predictive model based on hybridization approaches for solving feature selection problems by reducing the number of features, weakly relevant and irrelevant features. Practically, an optimal subset is likely to contain only powerfully relevant features.

The objectives of the proposed feature selection approach are reducing dimensionality and eliminate noise from data. This leads to an increase in learning speed, ease of rules, easy visualization of the data, and predictive accuracy. So, this study aims to achieve maximal accuracy of classification with a minimal number of features. This study assesses the ability of hybridization between metaheuristic algorithms to create a new feature selection approach to solve a feature selection problem by enhancing search space performance. To evaluate the new feature selection method's performance, classification accuracy, best, worst, and mean fitness, Standard Deviation (Std). An average number of features are used as evaluation criteria. This study is significant as it tries to solve Feature Selection (FS) problems by building a new hybrid feature selection approach (SCAGA), which discards redundant, irrelevant, noisy, and weak features from the original dataset. This leads to increased learning speed, ease of rules, dataset visualization, and predictive accuracy for a classification task.

The paper organization is as follows: Sect. 2 explains the literature survey that talks about previous studies and related works. Section 3 introduces the procedures and methodology and discusses the proposed schemes. Section 4 shows the evaluation criteria. Section 5 portrays the results and the discussion of the results, and Sect. 6 discusses the conclusions and future work.

2 Related works

Feature Selection (FS) technique is one technique employed to enhance the prediction accuracy of the searching space problems [17–19]. Approaches of searching may be summarized as follows: thorough searching, probabilistic searching, heuristic searching, and involuntary hybrid exploration algorithms [20]. Metaheuristic algorithms plan to decrease the time consumption and only search for a particular path to obtain the optimal solution [2]. Metaheuristic search is typically used on real-world problems and to exact varied computer science series [21]. Heuristics are also suitable to treat other parts of massive data, such as diversity and speed [22].

Different metaheuristic methods are applied to treat feature selection problems [23]. Genetic algorithm (GA) is the furthermost inspected metaheuristic algorithm. Population- and single-based metaheuristic algorithms are proposed [24]. Metaheuristic algorithms that are single-based such as hill climbing and simulated annealing, have been used. Scatter, random, harmony, and hill-climbing searches have main disadvantages; they are very tricky for opening solutions. They often drop in local optima [25]. The subset features have been selected by using the spider monkey optimization approach. The primary population algorithms have been given for the dataset. The assessment of the

fitness calculation was done using the SVM for classification accuracy. In order to continue or stop the process, a stopping criterion is tested. The best final subset of attributes with high classification accuracy is defined as the best optimal results [26].

A hybrid binary technique between coral reefs optimization and simulated annealing for attribute selection (BCROSAT) can discover the maximal accuracy and select minimal features' number for most datasets utilized [27]. Instance selection is a method that can reduce the size of the original training data. A combination of instance selection and feature extraction reduces the large volume of computation time of training the classifier [28]. For the global optimizer and FS algorithm, a novel chaotic slap swarm technique is efficient for two problems: FS problems and global optimization problems [29]. FS is the procedure to statistically identify the utmost relevant features to improve the predictive capability of the classifiers [30]. A method for FS to enhance clustering of documents is by using particle swarm optimization; this approach focuses on enhancing the current implementation of Bayesian calibration to building energy simulation [31].

The feature selection technique of Water Wave Optimization (WWO) builds the text FS technique based on Water Wave Optimization (WWOTFS) [32]. A hybrid approach based on binary chemical reaction optimization and a Tabu search optimization algorithm for FS has been developed. Once the four essential reactions are performed, in the iteration step, the best solution is checked. Then Tabu search is utilized to search neighbors, which is a local search process. An enhanced FS based on the Ant Colony Optimization (FACO) algorithm and the classifier SVM has been used to solve FS problems [33]. With the growing volume of data in networks and the number of feature sets, the security of the network is threatened by extra network attacks, such as APT and DDoS attacks. To speedily detect anomalies in networks, a classification technique is extensively used in the anomaly field of data discovery. However, there are massive irrelevant and redundant features in the dataset, which are considered difficulties that prevent the classification algorithm from creating efficient anomaly detection classifiers. To enhance the performance of classification for classifiers, the ant colony optimization method searches for the optimal features subset. It selects the relevant features independently from the classifier, which can efficiently decrease the algorithms' complexity to classify and improve the classification accuracy of classification.

A new technique for the subset of feature selection in ML, FSS-MGSA (Feature Subset Selection by Modified Gravitational Search Algorithm), is presented. FSS-MGSA is a sophisticated chaotic search algorithm based on the gravity law and interaction of mass. It can be performed

when knowledge of the domain is not accessible [34]. The binary of bare-bones particle swarm optimization (BPSO): the stimulus for this method is to design a global search technique using a small number of parameters, which has a better performance when solving feature selection problems also is easy to implement as well. Also, FS methods have been used to find feature subsets that have maximal classification ability [35]. In 2014, Moradi et al. presented a hybrid approach for selecting features in two phases: In the first phase, they seek a decrease in the original set's feature set by using the filtering model. Then, in the second phase, the wrapper model is applied for selecting the best features subset from the reduced feature set [36].

In the FS technique of binary PSO and GA, when determining coronary artery disease using a support vector machine (BPSO-FST), every particle is created from 23 binary cells, which refer to all features in the dataset. The cells' value shows whether the feature would be selected or not, where a cell value of 1 means that the feature is selected. With a value of a cell of 0, the feature is unselected into the dataset [37]. Feature selection is used to identify a powerfully predictive of fields inside a database and decreases the field number presented to the computational process. Feature selection affects some pattern classification aspects containing the learning classification procedure's accuracy, such as a support vector machine [38]. FS technique based on improved binary-coded Ant Colony Optimization technique (MPACO) is established to increase the accuracy of classification while reducing redundant features [39]. A novel FS technique using PSO has been used for cancer microarray datasets. This method is used for classifying high-dimensional cancer microarray datasets after solving feature selection problems. Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Probabilistic Neural Network (PNN) are used as classifiers and evaluators [40].

The experiment results showed that a heuristic FS method for text categorization using Chaos Optimization and Genetic technique (CGFSO) found feature subsets that resulted in the maximum accuracy of classification, while it found compact feature subsets. The performance of this approach is quicker than other traditional approaches [41]. Genetic algorithms and particle swarm optimization method can be combined in various ways. In the hybrid of GA and PSO (HGAPSO) FS method, hybridization is done by integrating the updated PSO rules and the standard velocity with selection, mutation, and crossover from the genetic algorithm [42]. FS based on the antlion optimization algorithm is working on the wrapper-based model. The wrapper methodology's main characteristic is using the classifier as a leader of the feature selection technique. Wrapper-based feature selection can classify depending on

the next three chief items: classification method, feature-evaluation criteria, and search method.

The first hybrid approach is proposed; Oh and Lee presented aggregation between algorithms and local search methods inserted inside a GA to improve search space by searching the utmost hopeful area discovered by the genetic algorithm procedure [43]. Lately, hybrid meta-heuristic techniques have displayed high performance for solving a hard combinatorial optimization problem. For example, combination methods, such as the hybrid between PSO and GA [44] and ant colony optimization with a genetic algorithm, have also been suggested. Simulated Annealing (SA) with a genetic algorithm are some local search approaches inside operative algorithms that balance between exploration and exploitation [45]. A new wrapper-based method for hybrid simulated annealing with a crossover operator was proposed [46]. Moreover, Talbi, Jourdan, Garcia-Nieto, and Alba, proposed GA was hybridized with the PSO technique, which used the SVM as a classifier. Table 1 shows an overview of different FS approaches with their details.

3 Procedures and methodology

This section supplies a detailed depiction of the research objectives methodology. Generally, wrapper methodology approaches use the method of learning as a classifier to assess the usefulness of features' subset and subsequently obtain the best performance of a predictive. Wrapper-based approaches gain better results of the quality measurement than filter-based methods. It obtained a subset of features that are enhanced for learning algorithm utilization. The technical details and a good understanding of the existing algorithms are necessary to make the appropriate choice unpractical in most statuses. Inside the field of this area, there is no metaheuristic-based method capable of solving all FS problems [20]. For newly unknown datasets, it will be more intricate to select a suitable method. Nevertheless, enhancements can be made to current algorithms to improve performance during balancing of the search space. This incentive mostly motivates us to make a prediction wrapper model depending on the hybridization technique for solving a feature selection problem.

3.1 Binary version of Sine Cosine Algorithm (SCA)

Recently, algorithms of metaheuristic proved to obtain high performance for solving real-world problems. Feature selection problems are considered binary search problems. In the proposed method, a Binary Sine Cosine Algorithm (SCA) adjusts the continuous SCA to handle feature

selection problems in the binary domain. The SCA begins with random positions whereas the search agent $(X_i) = 5$. Sine Cosine Algorithm works according to mathematical Eq. (3.1) [47].

$$x_{ij}^{t+1} \begin{cases} x_{ij}^t + r_1 * sin(r2) * \left| r_3(Xb_j^t) - x_{ij}^t \right| \text{if } R_1 < 0.5 \\ x_{ij}^t + r_1 * cos(r2) * \left| r_3(Xb_j^t) - x_{ij}^t \right| \text{if } R_1 \geq 0.5, \end{cases}$$
(3.1)

$$r_1 = a - t\frac{a}{Tmax}$$
(3.2)

The fitness function of the SCA will increase if the performance of classification over the testing dataset is increased when validating and achieving the minimal number of selected feature selections together.

$$f\theta = \omega * E + (1 - \omega)\frac{\sum_i \theta i}{n},$$
(3.3)

where the fitness function $f\theta$ gives vector θ sized n with 0/1 elements representing unselected/selected feature; n is the features number in the dataset. E is the classifier error rate and ω is a constant value (fixed 0.05) to control the classification accuracy performance to the features number that is selected. The variables used are the same features number in the given dataset. The variables are limited in the range [0, 1], where the variable value approaching 1 means that its corresponding features are candidate to be selected in the classification [28, 29, 44]. In individual fitness calculations, the variable is the threshold to decide the exact features to be assessed as in Eq. (3.4):

$$f_{ij} = 1 \text{ if } X_{ij} > 0.5, \text{ otherwise } 0,$$
(3.4)

where X_{ij} is the value of dimension for search agent i at dimension j. While updating each position of search agent at some dimensions, the updated value can violate the limiting constraints: [0, 1]; hence, we use simple truncation rule to ensure variable limit. Each candidate feature is represented as a binary vector with one dimension. The vector is used for features mapped to be in [0, 1] interval based on threshold value that is set to 0.5, indicating to the upper bound (ub = 1) and the lower bound (lb = 0).

3.2 Initial population

The Sine Cosine Algorithm (SCA) begins first randomly with positions to converge the global optima. It then calculates the value of fitness for every individual. It allocates the utmost remarkable location to FS as candidate features. Every solution is presented as a binary vector in one dimension. The number of the vectors is equal to the number of features in the original dataset. All cells within the vectors are labeled with 0 or 1. One value is indicating

Table 1 Overview of the classification and optimization methods used for feature selection

FS methods	Dataset	Classifier	Year	Researchers	Advantage
Ant Colony Optimization for FS (ACOFS)	Colon cancer dataset	NN	2012	Kabir, Shahjahan & Murase	Solving feature selection problems. The ACOFS has a powerful ability in solving the given problem
Binary gray wolf optimization approaches for FS	Benchmark dataset from UCI Repository of Machine Learning	KNN	2015	Emary, Aboul Ella Hassanien & Hossam M. Zawbaa	Solving feature selection problems by optimizing the selected features
Whale Optimization Algorithm (WOA)	Arabic dataset	KNN	2018	Mohammad, Mohammad, Abushariah, Idris & Aljarah	Discarding irrelevant and noisy features
Salp swarm technique with particle swarm optimization (SSAPSO)	Breastcancer, Clean1, Waveform, Lymphography, & WBDC	KNN	2018	Rehab Ali Ibrahim Ewees, Oliva, Abd Elaziz & Songfeng Lu	Accuracy of classification with minimal no. of features. A powerful hybrid algorithm is used
Chaotic Salp Swarm Technique (CSSA)	Datasets of benchmark from UCI repository	KNN	2018	Sayed, Horiba & Haggag	Accuracy of classification with minimal no. of features
Tunable Particle Swarm Size Optimization (TPSO)	Benchmark dataset from UCI Machine Learning Repository	ADT	2018	Naresh Mallenahalli & Hitendra Sarma	Solving feature selection problems by optimizing the selected features
Water Wave Optimization Text FS (WWOTFS)	History, energy, art, education, philosophy, compute, & literature dataset	KNN,SVM & Bayesian classifier	2018	Chen, Hou, Luo, Hu, & Yan	Eliminating redundant and irrelevant text
Binary Chemical Reaction Optimization and Tabu search (BCRO-TC)	Benchmark dataset	KNN	2018	Yan, Ma, Luo, & Wang	Accuracy of classification with minimal no. of features
Improved FS Based on Ant Colony Optimization (FACO)	KDD cpu99 datasets	SVM	2018	PENG, YING, TAN, BING, HU & SUN	Accuracy of classification with minimal no. of features. This method worked to find the maximum accuracy with minimal number of features
Spider Monkey Optimization (SMO)	Dataset of cancer	SVM	2018	Rania & Ramyachitraa	This method worked to find the maximum accuracy with minimal number of features

that the feature is selected, otherwise indicating that the feature is ignored.

3.3 Classifier

K-Nearest Neighbor (KNN) classifier, The KNN classifier is a predictor of variables weight at a distance based on trial-and-error process. K-Nearest Neighbor is utilized as a part of the fitness function in all the experiments due to its excellent performance in classifying.

3.4 Fitness function

In this study, the fitness function is applied to assess each feature subset in the search space of Sine Cosine Algorithm based on K-Nearest Neighbor (KNN) as a classifier, where $K = 5$. The proposed fitness function is calculated by using Eq. (3.5):

$$f\theta = \omega * E + (1 - \omega)\frac{\sum_i \theta i}{n}, \tag{3.5}$$

where the fitness function $f\theta$ is given vector θ sized n with 0/1 element representing select/unselect feature; n is the

features' amount of dataset. E is the error rate of classifier and ω is a constant to control the classification accuracy performance to the features number selected.

3.5 The crossover operator

Crossover is the leading exploration operator in the genetic algorithm. It searches the area for possible solutions depending on present solutions [34]. The binary crossover operators are to exchange bits between two parents selected to reproduce two new individuals. Both will be different from their parents yet hold some parent features. The type of crossover operator depends on the encoding method. Therefore, several types of crossover techniques, such as one-point, two-point, and uniform crossover. Uniform crossover is more exploratory, better for small populations, while the two-point crossover is suitable for large populations. In general, recombining parts of the right individuals gives a better opportunity to produce better individuals.

In this study, the crossover operator is designed by using three critical formulas, which are as follows: single, double, and uniform, enhancing the worst solution that has been selected by the Sine Cosine Algorithm through recombining the worst solution with the best one obtained from the previous iteration. The appropriate formula is chosen in each iteration depending on the roulette wheel selection function to take advantage of the features for each type of species mentioned. The crossover is utilized as an internal agent within Sine Cosine Algorithm, as shown in Fig. 1; Algorithm 3.1 shows the crossover operator's main steps.

3.6 The mutation operator

The mutation operator is utilized to produce new individuals with various features not present in their predecessors. Mutations can be applied for integer, binary, or real representations and categorized into several types. In general, the mutations are generated by randomly selecting one or more bits and then flipping their value with a certain probability ($pm = 0.02$). The mutation operator is utilized to act as an internal function employed within Sine Cosine Algorithm (SCA) to generate a new solution and improve the exploration ability after applying the crossover operator. The following Eq. (3.6) shows the work of the mutation operator:

$$X_i^{t+1} = Mutation(X_i^t). \tag{3.6}$$

At first, the metaheuristic algorithms' process displays two contradictory criteria: strategies of exploration of the search space and exploitation to detect the optimal solution. In the native Sine Cosine Algorithm, we note weakness in its exploration strategy, leading to weakness in its performance during the search space. The main steps of the proposed framework are shown in Fig. 1. In the proposed hybrid feature selection approach to improve the exploration strategy of the Sine Cosine Algorithm and its performance during the search space for solving feature selection problems, we used the Genetic Algorithm as an internal function within the SCA as a hybrid feature selection method named SCAGA.

Algorithm 3.1. Crossover Operator

INPUT: Best Solution (Sol), Worst Solution (Sol*).

OUTPUT: New Solution (Sol*).

    1. s=0.1,d=0.2, u=S-D;

    2. T= Roulette Wheel Selection ([s, d, u]);

    3. Switch (T)

       Case 1: Sol*=Single (Best _Sol, Worst_ Sol*);

       Case 2: Sol*=Double (Best _Sol, Worst_ Sol*);

       Case 3: Sol*=Uniform (Best _Sol, Worst_ Sol*);

    4. End Switch

Algorithm 3. 2. Main Steps of Hybrid (SCA-GA) Approach

INPUT: Search agent number, dataset, dimension, max_ iteration, upper bound (*ub*), lower bound (*lb*), *pm*.

OUTPUT: vector with 20 best solutions

---

1. Initialize the SCA population *yi*, (*i*=1, 2 …n)
2. FS=the best search agent by equation (3.5)
3. *l*=1
4. While (*l*<Max_iter+1)
5. *For i=1:N*
6. *Calculate r₁,r₂,r₃ and R₁*
7. *identify best solution Xb*
8. If (R₁< 0.5)
9. Update the position of the SCA by Sin equation (3.1);
10. Else
11. Update the position of the SCA by cos equation (3.2);
12. End if
13. Improvement the current position by proposed genetic algorithm.
14. End For
15. Amend the SCA Agent based on the upper and lower bounds of variables;
16. FS=the best search agent by equation (3.5);
17. End For
18. *l = l +* 1;
19. End While
20. End For

The wrapper model applies the classifier method as evidence in the FS method depending on some optimization techniques. The SCA is used as an FS method for balancing the accuracy of classification (maximal value) and the feature number selected (minimal value) in all solutions. In the beginning, population solutions with the SCA algorithm depend on updating its roles on the functions of sine and cosine according to Eq. (3.1).

Proposed Binary version of Sine Cosine algorithm then generates feature subsets. The fitness function is applied to evaluate the feature subsets. Update the iteration. *R1* is switched between sine and cosine functions based on the random switch parameter, where it is applied according to Eq. (3.1). The proposed genetic algorithm improves the current position or solution. The crossover operator (embedded within SCA) is applied to generate new offspring based on the best and candidate solutions and then enhanced by the mutation operator. The mutation operator is acting as an internal function within the SCA to prevent the algorithm from falling in the local optima problem and getting an optimal solution or near-optimal solution.

The probability of mutation controls the mutation operator. The optimal rate of mutation is a common problem in this arena. It has to be set at a low rate where the probability of mutation (*pm = 0.02)* is the best value during tuning parameters. On the other hand, if the rate is set at a max value, then the search will deflect into a random search and avoid the technique of converging to an optimum solution. All features are within this range [0, 1], so the position must be amended based on ub = 1 & lb = 0. Then fitness function of the new solution that is generated by applying the genetic algorithm is computed. After terminating the SCAGA search, we get the best solution, apply the evaluation measurements, and finally get results. An example of this rule can be viewed in Fig. 2. In this Figure, ten features are given, and number 1 presents the selected feature, and number 0 presents the unselected feature.

4 Evaluation criteria

The proposed hybrid feature selection method SCAGA is run 20 times. Iteration = 80 at each time because we have come to stability at run = 20 to test both the FS approach's stability and the statistical significance. The informative feature subset is evaluated using the evaluation criteria (accuracy of classification, mean, best and worst fitness, Standard Deviation (*std*), and average selected size) so the best feature subsets are obtained. The proposed FS method results were at maximal accuracy of a classification and a
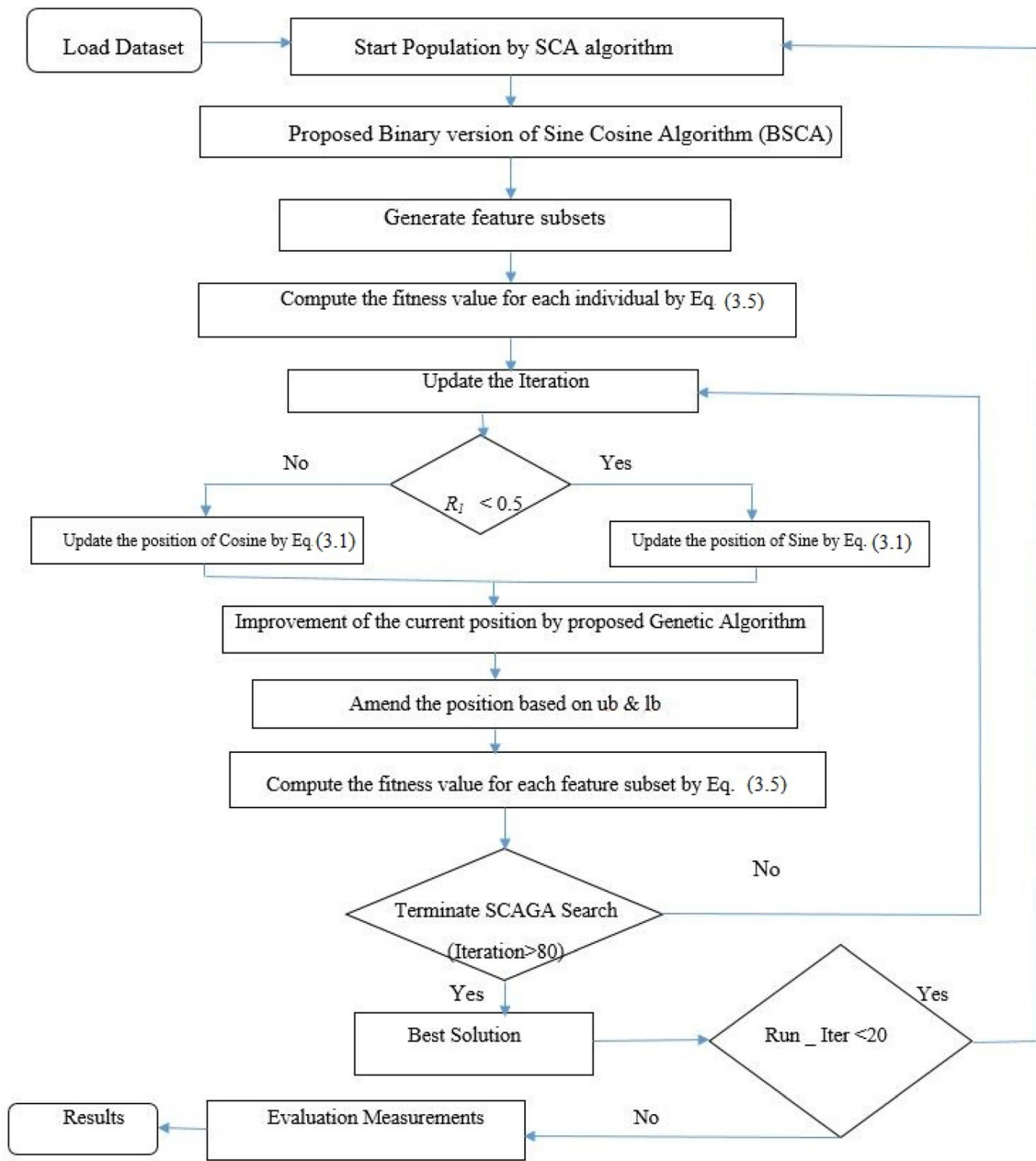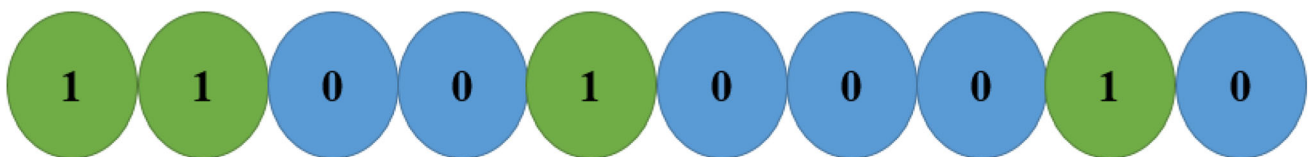
Fig. 1 The proposed SCAGA method

Fig. 2 The solution representation of the feature selection problem

minimal number of features. The evaluation criteria are explained as follows:

- *Classification accuracy* it is used to evaluate the performance of the feature selection method on the

Table 2 Datasets description

Datasets	Features number	Instances number	Subject of datasets
Breast_EW	30	699	Life
Breast cancer	9	699	Life
Congress	16	535	Social
Exactly	13	1000	Life
Heart_Ew	13	270	Life
Exactly-2	13	1000	Life
Ionosphere	34	351	Physical
Kr$v$skp_EW	36	3196	Agricultural
L$y$mpho$g$raphy	18	148	Life
Sonar_EW	60	208	Physical
m_of_n	13	1000	Rules
tic_tac_toe	9	958	Game
Spect_Ew	22	267	Life
Waveform_EW	40	5000	Physical
Zoo	16	101	Life
Wine_EW	13	178	Physical

Table 3 Experimental parameters setting of proposed method

Parameters	Values
Number of search agents	5
Number of iterations	80
Runs repetition	20
Dimension	No. of features in dataset
Search domain	Binary [0 1]
Probability of mutation (pm)	0.02
α parameter	0.01
β parameter	$1 - \alpha$

dataset that the classifier has been given. The classification accuracy can be calculated by Eq. (4.1) [34]:

$$Test = \frac{1}{N}\sum_{j=1}^{N}\sqrt{\sum_{i=1}^{K}(A_i - E_i)^2} \qquad (4.1)$$

where K is the number of test sample points and A_i, E_i are actual and expected class labels for data point i.

- *Best fitness* represents the smallest fitness function value for each optimization algorithm at the dissimilar M operations of an optimization algorithm and can be formulated as in the following Eq. (4.2) [48]:

$$Best = Min_{i=1 g_*^i}^{M}. \qquad (4.2)$$

- *Worst fitness* represents the maximum solution among the best solutions found for running each optimization algorithm for M times as in Eq. (4.3) [48]:

$$Worst = Max_{i=1 g_*^i}^{M}. \qquad (4.3)$$

- *Mean fitness* it represents the average of solutions acquired from running an optimization algorithm for diverse M running as in Eq. (4.4) [48]:

$$Mean = \frac{1}{M}\sum_{i=1}^{M} g_*^i. \qquad (4.4)$$

- *Standard Deviation (std)* represents the variance of the best solutions found for running each optimization algorithm for M diverse as runs in Eq. (4.5) [48]:

$$Std = \sqrt{\frac{1}{M-1}\sum (g^i - Mean)^2} \qquad (4.5)$$

- *Average selection size (average number of features selected)* represents the number of features selected to the entire number of features and may be formulated as Eq. (4.6) [20]:

$$Average\ selection\ size\ = \frac{1}{M}\sum_{i=1}^{M}\frac{size(g_*^i)}{D}, \qquad (4.6)$$

where $size(x)$ is the number of values for the vector x, D is the number of features in the original dataset, and g_*^i is the optimal solution resulting from run number i.

5 The results and discussion

In this section, the proposed method's performance is evaluated and compared using other similar methods using several feature selection datasets. All experiments are conducted using the same conditions. The maximal

Table 4 The Comparison between the SCAGA method with other optimizations methods in terms of accuracy of classification

No	Dataset	PSO	ALO	SCA	SCAGA
1.	Breast cancer	0.959714	0.961286	0.960571	0.970143
2.	Exactly	0.706800	0.717200	0.716000	0.984600
3.	Exactly2	0.672600	0.697200	0.693000	0.736700
4.	BreastEW	0.932105	0.936491	0.939298	0.962105
5.	CongressEW	0.905505	0.929128	0.926606	0.962385
6.	HeartEW	0.764815	0.773333	0.784444	0.827407
7.	IonosphereEW	0.865057	0.882102	0.881250	0.926705
8.	KrvskpEW	0.899186	0.903066	0.914111	0.979693
9.	Lymphography	0.761486	0.785006	0.778878	0.870816
10.	M-of-n	0.839800	0.836800	0.843800	1.000000
11.	SonarEW	0.837981	0.846154	0.854808	0.928365
12.	SpectEW	0.755597	0.786940	0.798881	0.849254
13.	Tic-tac-toe	0.727035	0.761587	0.760334	0.795929
14.	WaveformEW	0.701920	0.708380	0.713500	0.759600
15.	WineEW	0.934270	0.954494	0.957865	0.979213
16.	Zoo	0.937582	0.952584	0.945640	0.967461

Table 5 The percentage of the selected features for the comparative methods

No	Dataset	PSO	ALO	SCA	SCAGA
1.	Breast cancer	0.861111	0.794444	0.650000	0.588889
2.	Exactly	0.838462	0.961538	0.780769	0.442308
3.	Exactly2	0.703846	0.676923	0.688462	0.546154
4.	BreastEW	0.795000	0.770000	0.638333	0.470000
5.	CongressEW	0.765625	0.581250	0.578125	0.384375
6.	HeartEW	0.826923	0.826923	0.711538	0.588462
7.	IonosphereEW	0.730882	0.666176	0.600000	0.427941
8.	KrvskpEW	0.862500	0.918056	0.775000	0.501389
9.	.Lymphography	0.819444	0.780556	0.638889	0.533333
10	M-of-n	0.850000	0.911538	0.884615	0.461538
11.	SonarEW	0.749167	0.781667	0.702500	0.482500
12.	.SpectEW	0.752273	0.622727	0.515909	0.354545
13.	.Tic-tac-toe	0.861111	0.900000	0.783333	0.688889
14.	.WaveformEW	0.873750	0.987500	0.843750	0.601250
15.	.WineEW	0.792308	0.873077	0.757692	0.542308
16.	.Zoo	0.778125	0.781250	0.612500	0.390625

Table 6 Results of best fitness

No	Dataset	PSO	ALO	SCA	SCAGA
1.	Breast cancer	0.033235	0.037276	0.035559	0.022527
2.	Exactly	0.081394	0.103174	0.163343	0.004615
3.	Exactly2	0.219338	0.239138	0.286431	0.218569
4.	BreastEW	0.042070	0.040070	0.043877	0.021702
5.	CongressEW	0.042747	0.050413	0.054329	0.024581
6.	HeartEW	0.174821	0.178667	0.173282	0.135077
7.	IonosphereEW	0.092647	0.088456	0.089669	0.036985
8.	KrvskpEW	0.045205	0.070158	0.049500	0.017113
9.	Lymphography	0.170541	0.166096	0.125961	0.073559
10.	M-of-n	0.079414	0.033874	0.072263	0.004615
11.	SonarEW	0.062615	0.089006	0.102192	0.041910
12.	SpectEW	0.155034	0.165265	0.153670	0.107524
13.	Tic-tac-toe	0.216681	0.216681	0.207147	0.179167
14.	WaveformEW	0.281136	0.287200	0.276300	0.226614
15.	WineEW	0.017277	0.017277	0.027632	0.004615
16.	Zoo	0.005625	0.005625	0.006250	0.002500

Table 7 Results of worst fitness

No	Dataset	PSO	ALO	SCA	SCAGA
1.	Breast cancer	**0.061521**	0.055257	0.057581	0.043438
2.	Exactly	**0.362112**	0.289587	0.360902	0.305689
3.	Exactly2	0.332526	0.332854	**0.333623**	0.303595
4.	BreastEW	0.085561	**0.096842**	0.085088	0.054298
5.	CongressEW	0.127448	0.109283	**0.150946**	0.054954
6.	HeartEW	0.243128	**0.280564**	0.268615	0.225385
7.	IonosphereEW	0.151250	**0.167500**	0.158346	0.106250
8.	KrvskpEW	**0.207137**	0.125851	0.178083	0.030678
9.	Lymphography	**0.305023**	0.264189	0.286502	0.192297
10.	M-of-n	**0.252443**	0.209980	0.230991	0.004615
11.	SonarEW	0.205571	**0.228776**	0.206571	0.119564
12.	SpectEW	0.263128	**0.278358**	0.241872	0.218799
13.	Tic-tac-toe	0.335133	0.268351	**0.341489**	0.231949
14.	WaveformEW	**0.318110**	0.316108	0.311898	0.256730
15.	WineEW	**0.105497**	0.087096	0.073665	0.051417
16.	Zoo	**0.122721**	0.122721	0.122096	0.120221

iteration number is 80, and the number of search agents is 5. All results were calculated with an average of 20 runs using the framework of Matlab on an Intel Core i5 computer, 2.50G CPU, and 4.00 G of RAM with the 64-bit operating system.

5.1 Datasets and parameters

Sixteen datasets with two high dimensions were collected from the University of California Irvine (UCI) Machine Learning Repository, which is available at https://archives. ics.uci.edu/ml/datasets.html [49]. All details of the datasets are represented in Table 2. The proposed new hybrid method is a wrapper-based procedure. Every solution in the

Table 8 Results of mean fitness

No	Dataset	PSO	ALO	SCA	SCAGA
1.	Breast cancer	0.043928	0.046272	0.045534	0.035447
2.	Exactly	0.278815	0.289587	0.288968	0.019669
3.	Exactly2	0.295603	0.306541	0.310815	0.266129
4.	BreastEW	0.067961	0.070574	0.066478	0.042216
5.	CongressEW	0.077048	0.075975	0.078442	0.041082
6.	HeartEW	0.218200	0.232669	0.220515	0.176751
7.	IonosphereEW	0.120099	0.123381	0.123563	0.076842
8.	KrvskpEW	0.099386	0.105145	0.092780	0.025117
9.	Lymphography	0.228492	0.220650	0.225299	0.133225
10.	M-of-n	0.154459	0.170683	0.163484	0.004615
11.	SonarEW	0.144596	0.160124	0.150765	0.075743
12.	SpectEW	0.212781	0.217156	0.204267	0.152784
13.	Tic-tac-toe	0.257536	0.245029	0.245103	0.208919
14.	WaveformEW	0.297716	0.298579	0.292073	0.244008
15.	WineEW	0.058532	0.053781	0.049290	0.026002
16.	.Zoo	0.054526	0.054754	0.059942	0.036120

Table 9 The most remarkable solutions in terms of standard deviation of all optimizations

No	Dataset	PSO	ALO	SCA	SCAGA
1.	Breast cancer	0.006657	0.005596	0.006214	**0.005224**
2.	Exactly	0.062513	**0.046208**	0.049940	0.067322
3.	Exactly2	0.031287	0.020623	**0.011287**	0.022324
4.	BreastEW	0.010154	0.012369	0.013421	**0.007668**
5.	CongressEW	0.023501	0.016556	0.021639	**0.007879**
6.	HeartEW	**0.020174**	0.026402	0.022313	0.024696
7.	IonosphereEW	0.018169	0.020568	0.018584	**0.018159**
8.	KrvskpEW	0.037420	0.014467	0.027533	**0.003511**
9.	Lymphography	0.033453	**0.030470**	0.039642	0.037254
10.	M-of-n	0.039209	0.036393	0.034826	**0.000000**
11.	SonarEW	0.033012	0.036837	0.032373	**0.021587**
12.	SpectEW	0.029539	0.031767	**0.022189**	0.030058
13.	Tic-tac-toe	0.030516	0.015733	0.029619	**0.015026**
14.	WaveformEW	0.010249	0.007797	0.008591	**0.006564**
15.	WineEW	0.019416	0.017030	**0.012172**	0.013518
16.	Zoo	0.034248	0.033974	0.032794	**0.032444**

population is represented as an index binary vector for the features in the dataset. We take only the optimal solution and its fitness, which attained maximal accuracy of classification with minimal features. Besides, the parameter settings are summarized in Table 3.

The relevant features are known previously. In this case, we can validate the features selected by prior knowledge.

In contrast, in furthermost real-world problems, the relevant features are unknown previously. So, we have to utilize classification performance for testing datasets to indicate the quality and provide an unbiased evaluation of a final method. As FS is naturally multiobjective, results have been compared in both features' count instances and obtained classification accuracy.

To conduct a comparison between the diverse FS methods with our proposed hybrid method, the various next indicators are used: First, classification accuracy is an indicator for describing how accurate the classifier is given the selected feature set. Classification accuracy is formulated in Eq. (4.1). Secondly, the best fitness represents the most optimistic solution gained. The criteria of best fitness are formulated in Eq. (4.2). Third, the worst fitness represents the worst solution among all possible solutions that can be obtained for running optimization. The criteria of the worst fitness are formulated in Eq. (4.3). Fourth, mean fitness is the average performance, indicating the average of solutions obtained from running an optimizer with diverse 20 runs; the criteria of mean fitness are formulated in Eq. (4.4). Fifth, Standard Deviation (*std*) refers to the variation of the acquired optimum solutions from running a stochastic optimizer with 20 diverse runs and formulated in Eq. (4.5). Sixth and finally, the average selection size represents the average number of features selected to the whole number of features defined by Eq. (4.6).

5.2 Experimental results and discussions

In the proposed method, the SCAGA embedded a Genetic Algorithm inside Sine Cosine Algorithm to act as an internal function to improve the exploration ability of the SCA algorithm. The performance of the SCAGA was compared with native SCA and other approaches published in the literature survey as follows: PSO and ALO methods based on two critical goals, namely, the accuracy of the classification and average selection size. The SCAGA was also compared with other FS methods based on the next evaluation criteria: worst fitness, mean fitness, best fitness, and Std. An average of 20 runs-based frameworks calculates all the results of the evaluation criterion.

As shown in Table 4, the proposed hybrid method (SCAGA) is considerably better than the SCA, ALO, and PSO methods in terms of both goals: the number of features selected and the accuracy of classification. The comparison of SCAGA with other FS methods showed that SCAGA performance is better than SCA, PSO, and ALO through all datasets regarding the accuracy of the classification. As noted, the optimal solution is obtained in the M-of-n dataset, so the value of the classification accuracy is one, as it is clearly stated in Table 4. This means that the
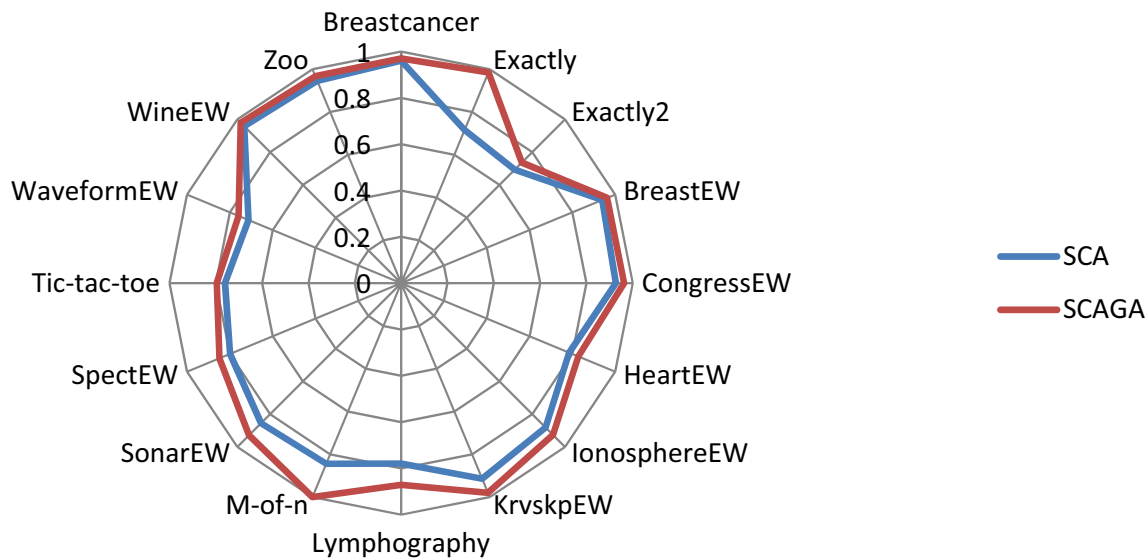
Fig. 3 Explanation of the performance of the proposed method for preventing local optima problem

proposed method can be considered a suitable FS method for small-scale datasets and high-dimensional datasets.

In Table 4, the proposed method SCAGA achieved maximal accuracy of classification in all used datasets. So, SCAGA is better than SCA, PSO, and ALO. Furthermore, the average number of selected features in Table 5 signifies that the performance of SCAGA is better than that of other methods over all datasets.

As shown in Table 6, the obtained results in best fitness criteria are the best when used the proposed SCAGA compared to other methods.

Table 7 shows that the proposed method (SCAGA) has never taken the worst fitness value, which is straightforward compared with other methods. The bold font refers to the worst value in Table 7. Also, the best results are obtained in mean fitness, as shown in Table 8.

In Table 9, the results are gained from (*Std*) evaluation referring to the variation of the acquired optimal solutions from running stochastic optimization with 20 diverse runs. The bold font refers to the best value in Table 9. SCAGA method outperforms other native SCA and related methods in the literature over ten datasets. Here the compared results of the proposed hybrid feature selection method (SCAGA) are given, the results of the native Sine Cosine Algorithm (SCA), and few other approaches related to feature selection picked from the literature survey such as Antlion Optimization (ALO) and Particle Swarm Optimization (PSO). Our results show that the performance of SCAGA is significantly better than SCA, ALO, and PSO that is common in wrapper-based feature selection. In Fig. 3, the proposed method's performance (SCAGA) for preventing local optima problem is the best over the native sine cosine algorithm.

Figure 4a and b summarizes empirical results obtained from proposed methods (SCA and SCAGA). It is observed that the SCAGA method gave a high performance as a multiobjective optimization method where it achieves two conflicting goals, maximum accuracy of classification with the least number of selected attributes on all datasets. All evaluation results fall between [0, 1] when the accuracy of classification is at the maximum values, and average selection size is at the minimum values.

A high-dimensional dataset means data with significant features number that leads to the dimensionality curse; with the high-dimensional dataset, the number of features can exceed the number of observations. Therefore, the calculations become very difficult. FS method has become a critical stage of analyzing high-dimensional datasets.

The SCAGA method is designed for high-dimensional data. It has been shown that it is useful in discarding redundant features and irrelevant features (see Fig. 5). In this study, to evaluate the proposed FS method's performance on high-dimensional datasets, we use two massive datasets, namely, Krvskp.EW (3196 objects with 36 attributes) & Waveform.EW (5000 objects with 40 attributes). As shown in Fig. 5, the proposed SCAGA method got better results than the other methods (SCA, PSO, and ALO) with low-dimensional datasets. It achieves two inconsistent targets, maximum accuracy of classification with the least number of features on all datasets. This means the proposed method can be considered as a suitable FS method for high-dimensional datasets.
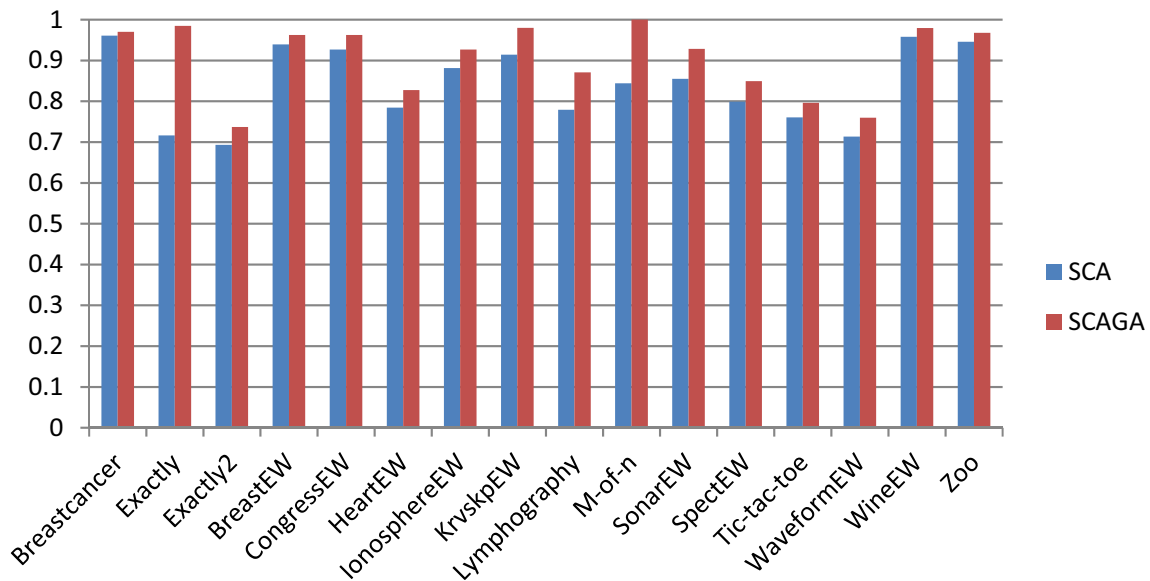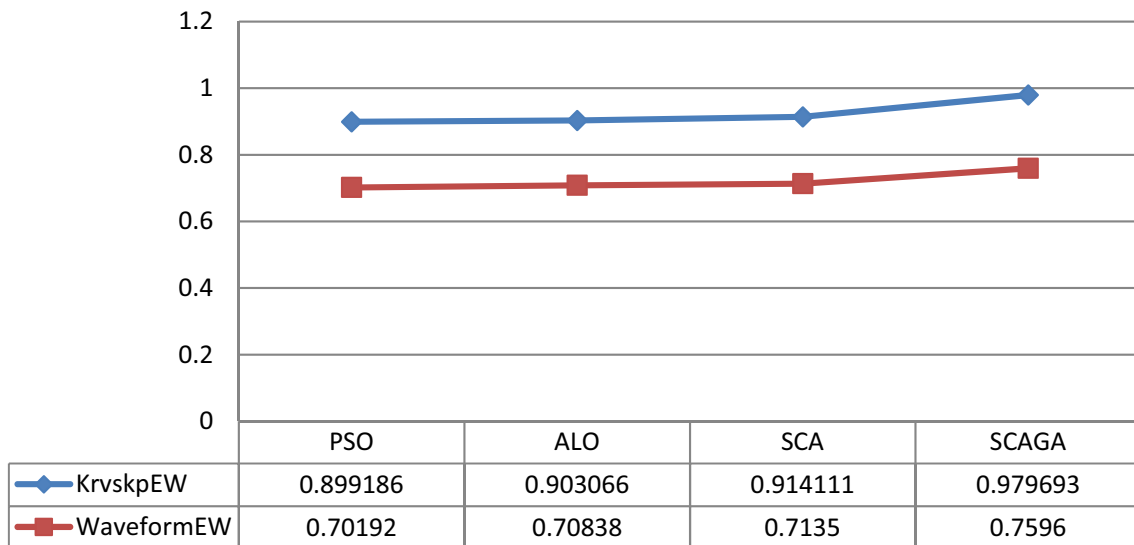
(a) Classification Accuracy
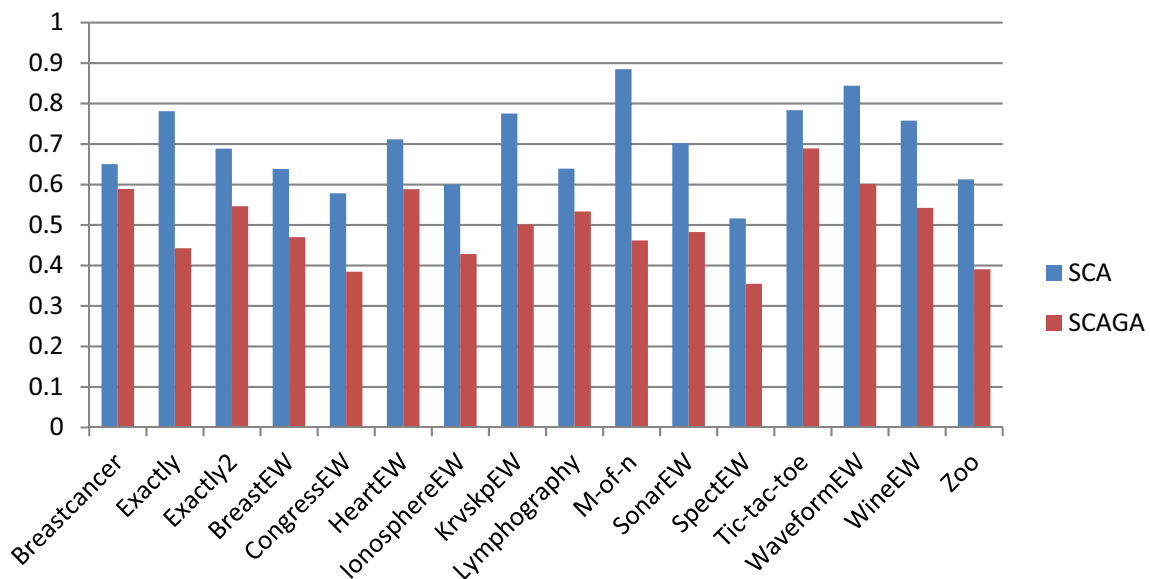
(b) Average Selected Features

Fig. 4 Comparison between the different proposed methods in terms of **a** Accuracy of classification and **b** Average selected size
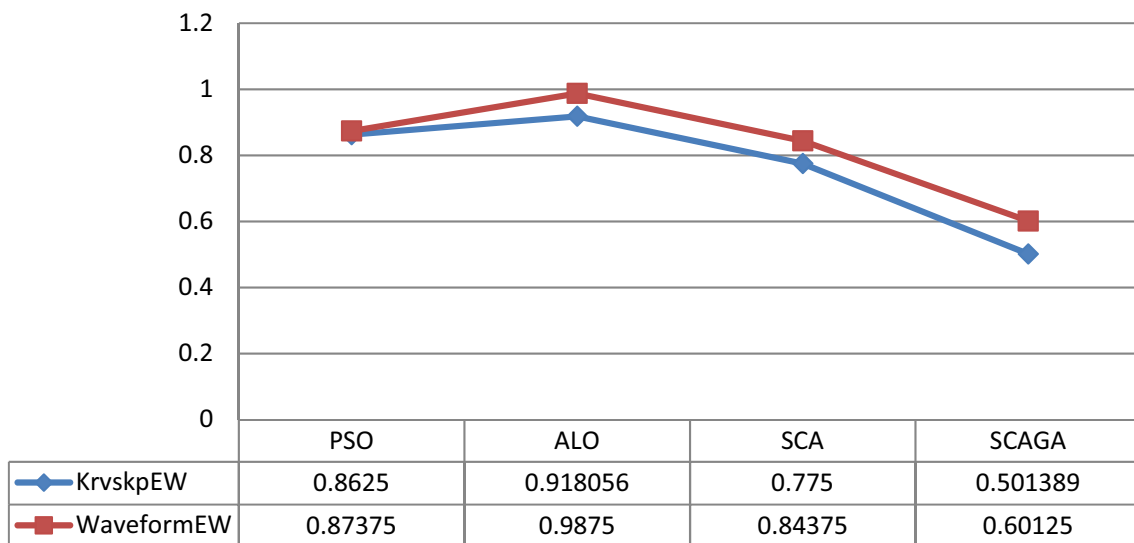
6 Conclusion and future work

In this paper, an enhanced version of the Sine Cosine Algorithm (SCA) is proposed with a wrapper model to solve the feature selection problems, called SCAGA. The proposed SCAGA is based on utilizing the genetic algorithm's crossover operator to generate the new solution and then apply the genetic algorithm of the genetic algorithm to enhance the solution generated by applying it to improve the exploration of the Sine Cosine Algorithm (SCA). It allows a more extensive search to prevent falling in local optima and then find the enhanced best solutions. The proposed hybrid SCAGA has significantly improved the performance of the native Sine Cosine Algorithm (SCA) to solve the FS problem from the reported results. It became more robust through the results showing the qualities of the

(a) Classification Accuracy

	PSO	ALO	SCA	SCAGA
KrvskpEW	0.899186	0.903066	0.914111	0.979693
WaveformEW	0.70192	0.70838	0.7135	0.7596

(b) Average Selected Features

	PSO	ALO	SCA	SCAGA
KrvskpEW	0.8625	0.918056	0.775	0.501389
WaveformEW	0.87375	0.9875	0.84375	0.60125

Fig. 5 Comparison between the different proposed FS methods in high-dimensional dataset: **a** Classification accuracy and **b** Average selected size

method proposed for solving real-world problems with unknown and challenging search spaces. Therefore, the proposed method (SCAGA) got better results than other methods (SCA, PSO, and ALO) with all datasets. It achieved two contradictory goals, maximal accuracy of classification with a minimal size of features on all datasets, either with small datasets or high-dimensional datasets. For future work, the proposed approach can be applied to other different datasets to generalize the approach for different domains, such as fault diagnosis in wind turbine test rig datasets. Other new optimizers can solve the feature selection problem, such as Arithmetic Optimization Algorithm (AOA).

References

1. Liu, H., Motoda, H.: Computational Methods of Feature Selection. Chapman and Hall/CRC Press, Boca Raton (2007)

2. Liu, H., Motoda, H.: Feature Selection for Knowledge Discovery and Data Mining, vol. 454. Springer, New York (2012)

3. Abdullah, S., Shaker, K., Shaker, H.: Investigating a round robin strategy over multi algorithms in optimizing the quality of university course timetables. Int. J. Phys. Sci. **6**(6), 1452–1462 (2011)

4. Holland. Genetic Algorithm for Solving Optimization Problems (1975)

5. Abualigah, L.M., Khader, A.T., Al-Betar, M.A., Alomari, O.A.: Text feature selection with a robust weight scheme and dynamic dimension reduction to text document clustering. Expert Syst. Appl. **84**, 24–36 (2017)

6. Abualigah, L., Alsalibi, B., Shehab, M., Alshinwan, M., Khasawneh, A.M., Alabool, H.: A parallel hybrid krill herd algorithm for feature selection. Int. J. Mach. Learn. Cybern. 1–24 (2020)

7. Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A.: Feature Selection for High-Dimensional Data. Springer , Cham (2015)

8. Nakamura, R.Y., Pereira, L.A., Costa, K.A., Rodrigues, D., Papa, J.P., Yang, X.S.: BBA: a binary bat algorithm for feature selection. In: 2012 25th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), (pp. 291–297). IEEE (2012)

9. Choi, S.I., Oh, J., Choi, C.H., Kim, C.: Input variable selection for feature extraction in classification problems. Signal Process. **92**(3), 636–648 (2012)

10. Fu, K.S., Min, P.J., Li, T.J.: Feature selection in pattern recognition. IEEE Trans. Syst. Sci. Cybern. **6**(1), 33–39 (1970)

11. Abualigah, L., Gandomi, A.H., Elaziz, M.A., Hussien, A.G., Khasawneh, A.M., Alshinwan, M., Houssein, E.H.: Nature-inspired optimization algorithms for text document clustering—a comprehensive analysis. Algorithms **13**(12), 345 (2020)

12. Abualigah, L.: Multi-verse optimizer algorithm: a comprehensive survey of its results, variants, and applications. Neural Comput. Appl. 1–21 (2020)

13. Abualigah, L.: Group search optimizer: a nature-inspired metaheuristic optimization algorithm with its results, variants, and applications. Neural Comput. Appl. 1–24 (2020)

14. Yan, M.: Hybrid Bainary Coral Reefs Optimazation Algorithm with Samulated Annealing for Feature Selection in High Dimentional Bieomedical Datasets, pp. 102–111. Elsevier, Amsterdam (2018)

15. Abualigah, L., Diabat, A., Mirjalili, S., AbdElaziz, M., Gandomi, A.H.: The arithmetic optimization algorithm. Comput. Methods Appl. Mech. Eng. **376**, 113609 (2021)

16. Kumar, V., Minz, S.: Feature selection: a literature review. Smart Comput. Rev. **4**(3), 211–229 (2014). https://doi.org/10.6029/smartcr.2014.03.007

17. Kang, S.H., Kim, K.J.: A feature selection approach to find optimal feature subsets for the network intrusion detection system. Clust. Comput. **19**(1), 325–333 (2016)

18. Manoj, R.J., Praveena, M.A., Vijayakumar, K.: An ACO–ANN based feature selection algorithm for big data. Clust. Comput. **22**(2), 3953–3960 (2019)

19. Gokulnath, C.B., Shantharajah, S.P.: An optimized feature selection based on genetic approach and support vector machine for heart disease. Clust. Comput. **22**(6), 14777–14787 (2019)

20. Khamees, A.A., Khalid, S.: Multi-objective Feature Selection: Hybrid of Salp Swarm and Simulated Annealing Approach, pp. 1–14. Springer, Switzerland (2018)

21. Du, K.L., Swamy, M.N.S.: Search and Optimization by Metaheuristics, p. 434. Springer, New York City (2016)

22. Dhaenens, C., Jourdan, L.: Metaheuristics for Big Data. Wiley, New York (2016)

23. Diao, R., Shen, Q.: Nature inspired feature selection metaheuristics. Artif. Intell. Rev. **44**(3), 311–340 (2015)

24. Mallenahalli, S.: A Tunable particle swarm size optimization algorithm for feature selection. In: 2018 IEEE Congress on Evolutionary Computation. IEEE (2018)

25. Diao, R., Shen, Q.: Feature selection with harmony search. IEEE Trans. Syst. Man Cybern. Part B **42**(6), 1509–1523 (2012)

26. Peng, Y.T., Hu, S.: An improved feature selection algorithm based on ant colony optimization. IEEE Access. **6**, 69203–69209 (2018)

27. Yan, M., Luo, W.: A hybrid algorithm based on binary chemical reaction optimization and tabu search for feature selection of high-dimensional biomedical data. Tsinghua Sci. Technol. **23**(6), 733–743 (2018)

28. Sayed, G.I., Khoriba, G.: A Novel Chaotic Salp Swarm Algorithm for Global Optimization and Feature Selection. Springer, New York (2018)

29. Sahu, B., Debahut, M.: A novel feature selection algorithm using particle swarm optimization for cancer microarray data. Procedia Eng. **38**, 27–31 (2012)

30. Abualigah, L.M.Q.: Feature Selection and Enhanced Krill Herd Algorithm for Text Document Clustering. Studies in Computational Intelligence. Springer, Berlin (2019)

31. Abualigah, L.M., Khader, A.T., Hanandeh, E.S.: A new feature selection method to improve the document clustering using particle swarm optimization algorithm. J. Comput. Sci. **25**, 456–466 (2018)

32. Chen, H., Hou, Y., Luo, Q., Hu, Z., Yan, L.: Text feature selection based on water wave optimization algorithm. In: International Conference on Advanced Computational Intelligence (ICACI). IEEE, pp. 546 551 (2018)

33. Padhy, N., Mishra, D., Panigrahi, R.: The survey of data mining applications and feature scope. *arXiv preprint* (2012).

34. Han, X.C., Quan, Y.X., Li, J., Zhang, L.: Feature subset selection by gravitational search algorithm optimization. Inf. Sci. **281**, 128–146 (2014)

35. Zanaty, E.A., Ghiduk, A.S.: A novel approach based on genetic algorithms and region growing for magnetic resonance image (MRI) segmentation. Comput. Sci. Inf. Syst. **10**(3), 1319–1342 (2013)

36. Mirjalili, S.: ALO: Antlion Optimization for solving feature selection problems. Adv. Eng. Softw. **83**, 80–98 (2015)

37. Linoff, G.S., Berry, M.J.: Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. Wiley, New York (2011)

38. Zhang, Z., Ning, Y.: Effective semi-supervised nonlinear dimensionality reduction for wood defects recognition. Comput. Sci. Inf. Syst. **7**(1), 127–138 (2010)

39. Wan, M.W., Ye, L.: A feature selection method based on modified binary coded ant colony optimization algorithm. Appl. Soft Comput. **49**, 248–258 (2016)

40. Zhao, Z.A., Liu, H.: Spectral Feature Selection for Data Mining. CRC Press, Boca raon (2011)

41. Chen, W.J., Li, L.: A heuristic feature selection approach for text categorization by using chaos optimization and genetic algorithm. In: Hindawi Publishing Corporation, Mathematical Problems in Engineering, pp. 1–6, (2013)

42. Ghamisi, P., Jon, A.B.: Feature selection based on hybridization of genetic algorithm and particle swarm optimization. IEEE Geosci. Remote Sens. Lett. **12**(2), 309–313 (2014)

43. Oh, I.S., Lee, J.S., Moon, B.R.: Hybrid genetic algorithms for feature selection. IEEE Trans. Pattern Anal. Mach. Intell. **26**(11), 1424–1437 (2004)

44. Atyabi, A., Luerssen, M., Fitzgibbon, S., Powers, D.M.: Evolutionary feature selection and electrode reduction for EEG classification. In: IEEE Congress on Evolutionary Computation (CEC), (pp. 1–8). IEEE (2012)

45. Vasant, P.: Hybrid simulated annealing and genetic algorithms for industrial production management problems. Int. J. Comput. Methods **7**(02), 279–297 (2010)

46. Wu, J., Lu, Z., Jin, L.: A novel hybrid genetic algorithm and simulated annealing for feature selection and kernel optimization in support vector regression. In: 2012 IEEE 13th International Conference on Information Reuse and Integration (IRI), (pp. 401–406). IEEE (2012)

47. Mirjalili, S.: SCA: a sine cosine algorithm for solving optimization problems. Knowl.-Based Syst. **96**, 120–133 (2016)

48. Emary, E., Zawbaa, H.M., AboulElla, H.: Binary Gray Wolf optimization approaches for feature selection. Neuro computing **2312**(15), 1–33 (2015)

49. Abualigah, L.M., Khader, A.T.: Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering. J Super-comput. **73**(11), 4773–4795 (2017)
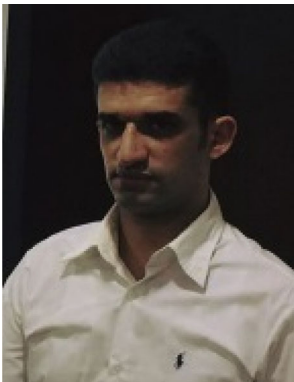
Laith Abualigah is an Assistant Professor at the Computer Science Department, Amman Arab University, Jordan. He received his first degree from Al-Albayt University, Computer Information System, Jordan, in 2011. He earned a Master's degree from Al-Albayt University, Computer Science, Jordan, in 2014. He received the Ph.D. degree from the School of Computer Science in Universiti Sains Malaysia (USM), Malaysia in 2018. According to the report published by Stanford University in 2020, Abualigah is one of the 2% influential scholars, which depicts the 100,000 top-scientists in the world. Abualigah has published more than 70 journal papers and books, which collectively have been cited more than 2500 times (H-index = 25). His main research interests focus on Bio-inspired Computing, Swarm Intelligence, Artificial Intelligence, Metaheuristic Modeling, and Optimization Algorithms, Evolutionary Computations, Information Retrieval, Text clustering, Feature Selection, Combinatorial Problems, Optimization, Advanced Machine Learning, Big data, and Natural Language Processing.

Akram Jamal Dulaimi has a Master degree in computer science and AI from Amman Arab University. His main interests are feature selection, optimization algorithms, and data mining.