Expert Systems With Applications 183 (2021) 115312

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Hybrid filter-wrapper feature selection using whale optimization algorithm: A multi-objective approach

Adel Got [a,1], Abdelouahab Moussaoui [a,2], Djaafar Zouache [b,*,3]

[a] Department of Computer Science, University of Setif-1, Setif, Algeria
[b] Department of Computer Science, University of Mohamed El Bachir El Ibrahimi, Bordj Bou Arreridj, Algeria

ARTICLE INFO

Keywords:
Feature selection
Filter and wrapper approaches
Multi-objective optimization
Whale optimization algorithm (WOA)

ABSTRACT

Feature selection aims at finding the minimum number of features that result in high classification accuracy. Accordingly, the feature selection is considered as a multi-objective problem. However, most existing approaches treat feature selection as single-objective problem, and they are often divided into two main categories: filter and wrapper methods. Filters are known as fast methods but less accurate, while wrappers are computationally expensive but with high classification performance. This paper proposes a novel hybrid filter-wrapper feature selection approach using whale optimization algorithm (WOA). The proposed method is a multi-objective algorithm in which a filter and wrapper fitness functions are optimized simultaneously. Our algorithm's efficiency is demonstrated through an extensive comparison with seven well-known algorithms on twelve benchmark datasets. Experimental results show the ability of the proposed algorithm to obtain several subsets that include smaller number of features with excellent classification accuracy.

1. Introduction

Feature selection plays a crucial role in machine learning as it is the best weapon against the so-called "curse of dimensionality" problem. The feature space usually contains oversized features that can be relevant, irrelevant or redundant. Indeed, the presence of irrelevant and redundant features decreases the accuracy of the final classifier. Therefore, feature selection is invoked to identify the optimal subset of features to decrease feature space's dimensionality, reduce the running time, and improve the classification performance (Dash & Liu, 1997; Unler & Murat, 2010). Logically speaking, the optimal feature subset is the one containing a minimum number of features and maximize the classification accuracy. The conflicting design between number/accuracy makes such a situation a multi-objective problem.

Feature selection algorithms are generally divided into two main families: filter and wrapper approaches (Liu & Motoda, 1998). In filter model, the candidate subset is evaluated by using statistical measures such as mutual information (Chandrashekar & Sahin, 2014) and rough set theory (Swiniarski & Skowron, 2003). Wrapper methods use a learning algorithm (classifier) to assess the goodness of features. Usually, wrapper methods outperform filter methods in terms of accuracy, while filter methods run faster. In contrast, some recent and prominent survey papers (Jović, Brkić, & Bogunović, 2015; Li et al., 2017; Brezočnik, Fister, & Podgorelec, 2018) consider the methods that embed feature selection within the learning algorithm itself as a third category (embedded methods). However, according to (Brezočnik et al., 2018), the main drawback of these methods is that they require more effort as they need the modification or even the development of a new learning algorithm.

Recently, Evolutionary Computation (EC) algorithms (Xue, Zhang, Browne, & Yao, 2015) have become a powerful tool for solving feature selection problems as they are proper for global search and thanks to their ability to escape from local optima subset. However, most existing EC algorithms follow either filter or wrapper model, i.e. few studies try to combine filter and wrapper models using EC approaches. Additionally, most of them handle feature selection as a single-objective problem.

Whale Optimization Algorithm (WOA) (Mirjalili & Lewis, 2016) is a recent evolutionary computation approach that has been widely applied in feature selection (Mafarja & Mirjalili, 2018; Sharawi, Zawbaa, & Emary, 2017; Hussien, Houssein, & Hassanien, 2017; El Aziz, Ewees, &

* Corresponding author.
E-mail address: djaafarzouache@yahoo.fr (D. Zouache).
[1] ORCID ID: https://orcid.org/0000-0002-6304-9341
[2] ORCID ID: https://orcid.org/0000-0003-3669-1264
[3] ORCID ID: https://orcid.org/0000-0002-0337-6105

https://doi.org/10.1016/j.eswa.2021.115312
Received 10 November 2020; Received in revised form 29 April 2021; Accepted 30 May 2021
Available online 8 June 2021

Hassanien, 2018). However, compared with other EC techniques, very few number of feature selection approaches WOA-based have been developed. Furthermore, the existing feature selection WOA-based are either single-objective, filter, or wrapper approaches. To the best of our knowledge, combining filter and wrapper models in multi-objective WOA has not been investigated yet.

In this paper, we develop a new multi-objective hybrid filter-wrapper approach for tackling the feature selection problem. The proposed algorithm takes on one hand; the searching ability of WOA to gets promising regions in the feature space. On the other hand, it merges the merits of filter and wrapper models into a single system in order to achieve better performance. Two objective functions are taken into account during the optimization process. The first objective estimate the relevance and redundancy between features by using mutual information (MI) as filter fitness function to identify the non-dominated subsets that result in minimum redundancy and maximum relevance of features to the target class, while the second objective estimate the classification accuracy by using a learning classifier as wrapper fitness function. The structure of paper is given as follows. Section 2 provides background information. Section 3 presents the proposed approach and Section 4 gives the experimental results. Finally, Section 5 concludes the paper.

2. Background

2.1. Whale optimization algorithm (WOA)

Whale optimization algorithm (WOA) proposed by (Mirjalili & Lewis, 2016) imitates the foraging behavior of humpback whales. The mathematical formulas of WOA algorithm are described as follows:

- **Searching and encircling prey**
  The whales attempt to hunt prey by encircling them. Hence, they update their positions with respect to the best hunting agent:

$$D = |C.X^{*}(t) - X(t)| \tag{1}$$

$$X(t+1) = X^{*}(t) - A.D \tag{2}$$

where, $X^{*}(t)$ is the best solution obtained so far. A and C are two vectors given by:

$$\begin{cases} A = 2a.r - a \\ C = 2.r \end{cases} \tag{3}$$

$r \in [0,1]$ and a linearly decrease from 2 to 0. If $|A| \geqslant 1$, the exploration phase is favored. Thereby, whales update their positions as follows:

$$\begin{cases} D = |C.X_{rand} - X(t)| \\ X(t+1) = X_{rand} - A.D \end{cases} \tag{4}$$

X_{rand} is a point in the search space chosen randomly.

- **Spiral model**
  To simulate the helical-shape updating of whales, a spiral equation is used as follows:

$$X(t+1) = X^{*}(t) + D.e^{bk}.\cos(2\pi.k) \tag{5}$$

$$D = |X^{*}(t) - X(t)| \tag{6}$$

k is a random number between -1 and 1. b is a constant to define the spiral shape.

Whales swim within a shrinking circle along a spiral way; this behavior can be formulated mathematically by:

$$X(t+1) = \begin{cases} X^{*}(t) - A.D & \text{if } (p < 0.5) \\ X^{*}(t) + D.e^{bk}.\cos(2\pi.k) & \text{if } (p \geqslant 0.5) \end{cases} \tag{7}$$

where, p is a random number between 0 and 1, which indicates that there is probability of 50% to apply the encircling mechanism or the spiral behavior.

2.2. Multi-objective optimization

Multi-objective optimization problem (MOP) involves handling several objective functions in conflict with each other. Mathematically, an MOP can be formulated by:

$$\begin{cases} \textit{minimize} \quad F(x) = [f_1, f_2, \ldots, f_m] \\ g_i(x) \leq 0, \; i = 1, 2, \ldots, k \\ h_i(x) = 0, \; i = 1, 2, \ldots, I \end{cases} \tag{8}$$

$x \in X$ is the vector of decision variables, X is the search space. m is the number of objective functions, g_i and h_i are the constraint functions.

In MOP, the optimal solution is represented by a set of trade-off solutions called *Pareto optimal solutions*. Since there are multiple conflicting objective functions, a particular relationship should be defined for comparing between solutions. Let u and v two candidate solutions, u is said to dominate v (denoted by $u \prec v$) if and only if:

$$\forall i \in \{1, 2, \ldots, m\} : f_i(u) \leqslant f_i(v) \; \textit{and} \; \exists j \in \{1, 2, \ldots, m\} : f_j(u) < f_j(v) \tag{9}$$

where m is the number of objective functions.

The solution u is called Pareto-optimal solution if it is not dominated by any other feasible solution. The image of Pareto solutions vector in the objective space is called *Pareto front*.

2.3. Entropy and mutual information

Information theory, including entropy and mutual information, offers a means for measuring the information of random variables (Shannon, 1948). The entropy is a metric of uncertainty of random variable X. It is given as:

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x) \tag{10}$$

where, x a discrete value and $p(x) = Pr(X = x)$ stands for the probability density function of X. For two discrete random variables X and Y with their probability density function $p(x, y)$, the joint entropy $H(X, Y)$ is defined as:

$$H(X, Y) = -\sum_{x \in X, y \in Y} p(x, y) \log_2 p(x, y) \tag{11}$$

In the case when a certain variable Y is identified and others are unidentified, the remaining uncertainty is estimated by the conditional entropy $H(X|Y)$:

$$H(X|Y) = -\sum_{x \in X, y \in Y} p(x, y) \log_2 p(x|y) \tag{12}$$

The common information quantity of two random variables X and Y

is known as the mutual information:

$$I(X;Y) = -\sum_{x\in X, y\in Y} p(x,y)\log_2\frac{p(x,y)}{p(x)p(y)} \tag{13}$$

If the two variables X and Y are independant, that is $p(x,y) = p(x)p(y)$, then $I(X;Y)$ will be zero. Otherwise, $I(X;Y)$ will be large.

2.4. Related work on feature selection

Based on Genetic algorithms (GAs), (Hamdani, Won, Alimi, & Karray, 2007) proposed a wrapper approach using NSGA2 and K-Nearest Neighbours (KNN) algorithm to minimize both the classification error and the number of features. In (Xue, Cervante, Shang, Browne, & Zhang, 2013), NSGA2 has investigated again to develop two filter approaches, namely NSGAIIMI and NSGAIIE, by applying mutual information and entropy, respectively, as the evaluation criterion. Experiments showed that the proposed algorithms could better handle the feature selection problem compared to classical wrapper methods. Recently, (Labani, Moradi, & Jalili, 2020) proposed a filter-based multi-objective algorithm for text feature selection. The Relative Discriminative Criterion (RDC) is used as the first objective to compute the relevance of text features, and the correlation metric as the second objective to compute the redundancy between features. Experiment results showed the ability of the proposed algorithm to select a subset of features with maximum relevance and minimum redundancy. However, only three databases have been used in the experimental phase. In (Rathee & Ratnoo, 2020), wrapper multi-objective non-traditional genetic algorithm (CHC) to optimize features reduction rate and KNN performance. Despite the notable performance of the proposed algorithm, it has been compared only against two multi-objective approaches.

Based on Particle Swarm Optimization (PSO), (Xue, Zhang, & Browne, 2012) proposes two multi-objective algorithms called NSPSOFS and CMDPSOFS by introducing the non-dominated sorting technique in the first algorithm, and applying the concepts of mutation, dominance and crowding mechanism in the second algorithm. The results showed that CMDPSOFS could evolve a Pareto front which includes fewer features and achieve better classification accuracy. (Cervante, Xue, Shang, & Zhang, 2013) used multi-objective binary PSO (MOBPSO) to perform filter feature selection by using rough set theory. Also, (Xue, Cervante, Shang, Browne, & Zhang, 2012) proposed two multi-objective filter feature selection based on BPSO and two modified mutual information and entropy, which allowed to achieve better prediction accuracy than that of all features. In (Yong, Dun-wei, & Wan-qiu, 2016), hybrid mutation operator is used to design an enhanced MOPSO in which, the reliability level and the prediction accuracy were the two criteria to be optimized. The algorithm has showed its ability to deal with unreliable datasets. Recently, (Amoozegar & Minaei-Bidgoli, 2018) proposed a new multi-objective PSO based on a ranking method to store the set of features into an external archive according to their frequencies. Objective comparison was performed through nine datasets and three quantitative metrics, and the algorithm gave very satisfactory results.

Ant Bee Colony (ABC) and Grey Wolf Optimizer (GWO) are other EC algorithms which were also investigated for solving feature selection problem. For example, (Hancer, Xue, Zhang, Karaboga, & Akay, 2015) developed three multi-objective ABC approaches (MOABC) by integrating three filter fitness based on information theory. (Hancer, Xue, Zhang, Karaboga, & Akay, 2018) inspired the idea of non-dominated sorting to develop two wrapper approaches ABC-based named Bin-MOABC (for binary version) and Num-MOABC (for continuous version). (Zhang, Cheng, Shi, Gong, & Zhao, 2019) addressed the problem of cost-sensitive features at the aim of minimizing the features cost and maximizing the classification performance by using multi-objective ABC with two-archive (external repository and leader archive). The external archive was attributed to maintain the diversity in the objective space, while the leader archive was attributed to the solutions' quality in the variable space. Based on GWO algorithm, (Sahoo & Chandra, 2017) proposed non-dominated sorting GWO (NSGWO) to perform feature selection in order to improve the classification of cervix lesions by minimizing the number of textural features and maximizing the classification accuracy of cervix lesions.

More recently, some studies have demonstrated that combining filter model with wrapper model can achieve excellent results such as in (Hammami, Bechikh, Hung, & Said, 2019) where a multi-objective GA is proposed for handling two filter objectives and one wrapper objective by using an indicator-based local search strategy. In (Emary, Yamany, Hassanien, & Snasel, 2015) a new multi-objective GWO for feature selection is proposed by using mutual information as filter fitness to search the set of solutions with minor redundancy, and wrapper fitness to enhance the obtained solutions towards better classification performance. The developed method showed a good ability to avoid local optima. However, despite the authors considered their algorithm as multi-objective approach, which means logically a simultanous optimization process, but the feature selection itself was treated as single-objective problem because they optimize, in sequential manner, a single objective function in the first stage (filter-stage) and a single objective function in the second stage (wrapper-stage).

In summary, several multi-objective evolutionary techniques have been applied in the feature selection field. However, most of them are filter-based or wrapper-based. There is no sufficient number of studies in the literature that handle feature selection as multi-objective problem by incorporating filter and wrapper models in a single system. This motivated our attempts to develop yet a new multi-objective hybrid filter-wrapper approach by using WOA algorithm.

3. Proposed approach

The proposed algorithm, which we shall call FW-GPAWOA (for Filter-Wrapper Guided Population Archive WOA), investigates our recent work (Got, Moussaoui, & Zouache, 2020) in which a new multi-objective optimization approach WOA-based (GPAWOA) has been developed. An external archive update strategy is used for ensuring the elitism, and an effective leader selection strategy based on the crowding distance is applied to guide the population towards a well-spread Pareto front. GPAWOA algorithm has demonstrated its applicability in practical continuous problems, and the results were very satisfactory. However, the feature selection occurs in a discrete search space - i.e., the solutions are restricted to 0 or 1 values, for this reason, it is preferable to develop firstly a binary variant of the GPAWOA.

A simple way to convert the continuous version of GPAWOA to the binary version is to use a transfer function. In our case, the hyperbolic tangent function is used as follows:

$$T(x^{i+1}) = \frac{e^{-\lambda(x^{i+1})} - 1}{e^{-\lambda(x^{i+1})} + 1} \tag{14}$$

where, λ is assigned to 1 and x^{i+1} is the real-value of the current search agent, thus, the binary update is carried out as:

Table 1
Summary of utilized datasets.

Datasets	No. of samples	No. of features	No. of classes
Breast cancer	699	9	2
Lymphography (Lymph)	148	18	4
Musk	476	166	2
Semeion	1593	265	2
Cervical cancer (Sobar)	72	19	2
Spect	267	22	2
Spectf	160	44	2
Sports articles	1000	59	2
Vehicle	846	18	4
Wholesale customers	440	7	2
Optical digits (Optical)	5620	64	10
Letter recognition (Letter)	20000	16	26

Table 2
Parameter settings.

Algorithm	Parameters
BDE	Crossover rate: $CR = 0.9$
BPSO	$c_1 = 2, c_2 = 2, V_{max} = 6, w_{max} = 0.9, w_{min} = 0.4$
BGWO	-
WOA	$b = 1$
jDE-FS	$F_0 = 0.5$, $CR_0 = 0.9$
MOPSO	$\phi_1 = \phi_2 = 2.05, \phi = 4.1$, Inertia weight: $w = 0.73$
	Personal coefficient: $c_1 = 1.4962$, Social coefficient: $c_2 = 1.4962$, $nGrid = 10$
MOGWO	Grid inflation: $\alpha = 0.1$, Archive member selection pressure: $\gamma = 2$, $nGrid = 10$
	Leader selection pressure parameter: $\beta = 4$
GPAWOA	$b = 1$, Percentage of the top part of the archive: $a = 0.1, T_{max} = 30$

$$x_{bin} = \begin{cases} 1 & if\ \mu < T(x^{i+1}) \\ 0 & Otherwise \end{cases} \tag{15}$$

with, μ is a random number with uniform distribution in (0, 1).

In FW-GPAWOA, each search agent in the population represents a subset of features. For a n-dimensional feature search space, each whale is encoded by a n-bit binary string. The i^{th} whale's position is a n-dimensional vector $X_i = [x_1, x_2, \ldots, x_n]$ with $x_j = \{0, 1\}$. $x_j = 1$ indicates that the corresponding j^{th} feature is selected, whereas $x_j = 0$ means that the corresponding feature is not selected.

One of important elements to successfully address the feature selection problem is choosing an appropriate objective function. As mentioned above, the proposed algorithm combines filter and wrapper measures into a single model. Here, it should be noted that FW-GPAWOA considers both filter and wrapper measures simultaneously during the optimization process. In other words, all iterations are attributed to evaluate and optimize the two objectives in the same time, which make the difference between our algorithm and other existing hybrid filter-wrapper feature selection algorithms which attribute a certain number of iterations to filter evaluation and attribute the rest of iterations to wrapper evaluation (two-stage), which may lead to eliminate some relevant features in the filter stage without considering to their contribution in the wrapper stage.

In this paper, mutual information (filter function) and classification accuracy (wrapper function) are used to formulate our objective function. Mutual information (MI) is a filter metric that can be used to estimate the relevance of any feature to the target class and estimate the redundancy among a set of features. Therefore, we propose to use a mixed filter fitness function MI-based, which attempts to maximize the average relevance of the selected features subset and minimize the average redundancy within the selected subset:

$$Fit_1 = Avg(I(x;c)) - Avg(I(x_i;x_j)) \tag{16}$$

where, $x, x_i, x_j \in S, c$ is the target class and S is the selected feature subset. Fit_1 is a maximization function that attempts to maximize $Avg(I(x;c))$ which indicates the average relevance of the selected subset to the target class, and reduces the number of features by minimizing $Avg(I(x_i;x_j))$ which indicates the average redundancy among the

Table 3
Experimental results of FW-PGAWOA, BDE, BPSO, BGWO, WOA and jDE-FS.

Datasets	Breast cancer (9, 98.35%)						Lymph (18, 76.92%)					
Method	FW	BDE	BPSO	BGWO	WOA	jDE	FW	BDE	BPSO	BGWO	WOA	jDE
Size	**2.06**	6.1	5.7	7.3	6.5	4.2	**3.3**	11.6	7.3	10.8	10	10.8
Acc %	**97.14**	96.70	96.85	95.05	96.75	94.94	**85.89**	80.76	60.25	82.56	82.58	77.43
Datasets	Musk (166, 84.67%)						Semeion (265, 97.35%)					
Method	FW	BDE	BPSO	BGWO	WOA	jDE	FW	BDE	BPSO	BGWO	WOA	jDE
Size	**36**	139.9	82.8	110.7	93.8	95.7	**120**	218.6	133.6	197.8	210	135.4
Acc %	**85.16**	79.67	84.03	84.75	83.87	83.87	97.37	97.40	96.78	96.58	**97.59**	97.40
Datasets	Sobar (19, 89.47%)						Spect (22, 82.85%)					
Method	FW	BDE	BPSO	BGWO	WOA	jDE	FW	BDE	BPSO	BGWO	WOA	jDE
Size	**3.8**	12.4	9.4	11.2	11.9	10.4	**3.4**	15.9	8.7	15.6	13	11.2
Acc %	**93.68**	86.31	90	88.42	80.52	85.26	85.71	83.57	70	82.14	**86.57**	81.14
Datasets	Spectf (44, 81.42%)						Sports (59, 81.60%)					
Method	FW	BDE	BPSO	BGWO	WOA	jDE	FW	BDE	BPSO	BGWO	WOA	jDE
Size	**4.82**	34.9	21.5	29.9	22.5	14	**9.1**	45.2	30.3	42.7	31.8	29.4
Acc %	**84.28**	76.28	73.85	74.57	78.57	75.42	**83.18**	82.26	78.35	78.65	81.91	82.37
Datasets	Vehicle (18, 66.06%)						Wholesale (7, 86.08%)					
Method	FW	BDE	BPSO	BGWO	WOA	jDE	FW	BDE	BPSO	BGWO	WOA	jDE
Size	**9.05**	13.1	9.3	12.1	9.1	9.8	**2.49**	2.9	5,1	4	2.9	4.4
Acc %	71.49	70.22	69.55	73.48	73.12	**73.81**	**92.17**	86.95	90.78	90.60	77.91	87.30
Datasets	Optical (64, 98.26%)						Letter (16, 94.55%)					
Method	FW	BDE	BPSO	BGWO	WOA	jDE	FW	BDE	BPSO	BGWO	WOA	jDE
Size	**36.95**	58.2	40.5	54.9	59.9	42.3	**11.1**	13.3	12	13.9	14.8	12.8
Acc %	97.96	98.13	97.51	98.09	**98.23**	98.10	**93.65**	93.52	92.40	93.19	93.41	93.60

selected feature subset.

In Fit_1, the relevance and the redundancy have the same weight. However, the classification accuracy is more important than the number of features. Therefore, Fit_1 is weighted as follows:

$$Fit_1 = \alpha.Avg(I(x;c)) - (1-\alpha).Avg(I(x_i;x_j)) \tag{17}$$

where, α is a constant number in the interval]0.5, 1[.

The second objective function Fit_2 maintains the wrapper evaluation which aimed to maximize the classification accuracy (Acc) given by the following equation:

$$Fit_2 = Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{18}$$

where, TP, TN refer to true positives and true negatives, whereas FP, FN stand for false positives and false negatives, respectively. Fit_2 is evaluated by using a preselected classifier; K-nearest neighbors (KNN) in the current case.

In summary and according to Eqs. 17,18, our feature selection problem is formulated by the following multi-objective optimization problem:

$$\begin{cases} Maximize\ Fit_1 = \alpha.Avg(I(x;c)) - (1-\alpha).Avg(I(x_i;x_j)) \\ Maximize\ Fit_2 = \dfrac{TP+TN}{TP+TN+FP+FN} \\ Subject\ to\ x_i \in \{0,1\} \end{cases} \tag{19}$$

Algorithm 1. Pseudocode of FW-GPAWOA

```
Divide Dataset into a training set and a test set
Initialize the whales population X_i =( 1,2,...,n)
Evaluate each individual in X_i (Fit_1 and Fit_2 in Equation 19) on the training set
Initialize Archive Ar (with maximum size equal to T_max)
while t < iter_max do
    Calculate the crowding distance (CD) of each solution in Ar
    Sort Ar in descending order according to the CD-values
    for each individual do
        update a, A, C, k and p
        randomly select a leader from the top part of the sorted Ar
        if (p<0.5) then
            if (|A| < 1) then
                calculate D using Equation (1)
                update the position of current individual (Equation 2)
            else
                select a random search agent X_rand
                calculate D in Equation (4)
                update the position of current individual (Equation 4)
            end
        else
            calculate D using Equation (6)
            update the position of current individual (Equation 5)
        end
    end
    Generate the binary representation of X_i by using Equations 14-15
    Evaluate each individual in {X_i ∪ Ar} (Fit_1 and Fit_2 in Equation 19) on the
      training set
    Update the Archive Ar
    if (the archive Ar is full) then
        calculate the crowding distance (CD) of each solution in Ar.
        sort Ar in descending order according to the CD-values.
        remove the last solutions in the sorted archive Ar
    end
    t = t+1.
end
Return the solutions in Ar, the selected feature subsets by each solution and their
  testing classification performance.
```

Algorithm 1 outlines the pseudocode of FW-GPAWOA. All the non-dominated solutions (feature subsets) are stored in an external-archive in descending order by considering their crowding distance values. For each individual, a leader is selected from the less crowded part of the sorted archive to guide the optimization process to well-distributed Pareto front and update the position of the current population by using Eqs. 2, 4 and 5, and then binarized it based on Eqs. 14,15. After that, the current population and the current archive are combined into a temporary repository from which the newly non-dominated solutions are identified to prepare the archive to the next iteration (updating phase). Whenever the archive attains its allowable size, the most crowded $|Archive| - T_{max}$ solutions are removed from the sorted archive, where T_{max} is the maximum size of the external-archive (i.e. the number of feature subsets allowed to be stored in the archive). The final archive contains the obtained Pareto front (i.e. the optimal feature subsets).

4. Experimental design and result analysis

This section presents the employed benchmarking datasets and their characteristics, the selected state-of-the-art algorithms and their parameters setting, and outlines the comparison between the obtained results.

4.1. Datasets

In order to assess the performance of the proposed algorithm, 12 benchmark datasets chosen from UCI machine learning repository (Frank, 2010) are used in the experiments. 70% of each dataset represents the training set on which the algorithm will be executed. In contrast, the remain 30% corresponding to the test set is used to evaluate the performance of the selected features. In this work, K-Nearest Neighbour (KNN) with K = 5 is employed to calculate the classification accuracy (Acc) of the selected features on the test set.

4.2. Competitor algorithms and parameter settings

To validate the performance of the proposed FW-GPAWOA algorithm, seven algorithms including five algorithms are single-objective, BDE (Too, Abdullah, & Mohd Saad, 2019), BPSO (Too, Abdullah, Mohd Saad, & Tee, 2019), BGWO (Too, Abdullah, Mohd Saad, Mohd Ali, & Tee, 2018), WOA (Mirjalili & Lewis, 2016), jDE-FS (Fister, Fister, Jagrič, & Brest, 2018), and two algorithms are multi-objective, MOPSO (Coello, Pulido, & Lechuga, 2004), MOGWO (Mirjalili, Saremi, Mirjalili, & Coelho, 2016), are used as benchmark techniques in the comparative study. BDE (for Binary Differential Evolution) is a wrapper feature selection method. Similar to the basic DE, BDE uses crossover and mutation operators, and generates a trial vector based on the crossover operator in order to compare it with the current vector (individual) to choose, based on a greedy selection, the best one according to the classification error rate obtained by KNN classifier. BGWO introduces a binary variant of the standard Grey Wolf Optimizer through the use of a specific transfer function in order to deal with the discrete design of feature selection problem. Identical to BDE, BGWO adopts the classification error rate obtained by the classifier as wrapper fitness function. BPSO is a wrapper-based binary Particle Swarm Optimization which introduce a new personal best guide strategy to enhance the search ability of PSO, and utilizes a modified sigmoid function to corvert the continuous representation of particles into a binary representation in order to solve feature selection problem. jDE-FS is a self-adaptive differential evolution method which adopt the AUC-indicator as fitness function and introduces a dynamic parameter as threshold of selecting and rejecting the candidate feature. MOPSO algorithm is the most popular multi-objective PSO-based algorithm in which a secondary repository is incorporated into PSO to store the non-dominated solutions, and an archive controller with an adaptive grid mechanism are applied to improve the convergence and diversity of the swarm. MOGWO is a relatively recent multi-objective GWO which apply the same techniques of MOPSO (i.e. archive controller and adaptive grid), and utilizes a sigmoid function to make wolves able to move in the binary feature space. Both MOPSO and MOGWO adopt the number of features as first objective function, and classification error rate as the second objective function.

For all algorithms and during the training optimization process, K-NN classifier (with K = 5) is used based on 10-fold cross-validation procedure to assess the classification performance of the selected features on the training set. To provide an equitable comparison, the population size and the number of iterations in all algorithms are set to 30 and 50, respectively. All the algorithms have been conducted for 10 independent runs for each dataset, and they are implemented using Matlab2016b under 32-bit Windows 8.1, except jDE-FS which is executed on NiaAML[4] environment. The experiments have been performed on the same machine: Intel Core i3, 2.39 GHz and 4 GB RAM.

4.3. Results and discussion

This subsection discusses the obtained results of the proposed FW-GPAWOA versus the selected state-of-the-art approaches. First, we compare FW-GPAWOA with the five single-objective algorithms (BDE, BPSO, BGWO, WOA and jDE-FS). Second, we compare the performance of FW-GPAWOA with that of the two multi-objective algorithms (MOGWO and MOPSO). Besides, a statistical comparison using the Wilcoxon test is performed on the Hyper-Volume metric (HV) to check if there is a significant difference between the performance of algorithms. Finally, a computational time analysis must be done to verify the speed of the algorithms (see Tables 1 and 2).

4.3.1. FW-GPAWOA vs Single-objective algorithms

Table 3 shows the results obtained by FW-GPAWOA, BDE, BPSO, BGWO, WOA and jDE-FS where, "Size" means the average number of the selected features and "Acc" means the average of the classification accuracies achieved by the algorithms. The values in parentheses show the classification accuracies resulted using all features, while the best results on each dataset are marked in **boldface**. In addition, Table 4 presents

Table 4
Some selected feature subsets.

Dataset	Selected subsets	Accuracy (%)	Dataset	Selected subsets	Accuracy (%)
Breast	2 - 6	97.25	Spectf	11 - 20 - 30	85.71
	3 - 6	96.70		12 - 18 - 28 - 30 - 36 - 37 - 42	91.42
	3 - 6 - 9	96.70		6 - 27 - 30	84.28
Lymph	7 - 8 - 13	89.74	Sports	2 - 17 - 22 - 50 - 53	85.05
	13 - 18	87.17		22 - 34 - 42 - 46	82.37
	2 - 13 - 15 - 18	84.61		6 - 9 - 16 - 19 - 22 - 25 - 37 - 44 - 53	84.29
Sobar	10 - 11 - 18	100	Vehicle	2 - 3 - 5 - 6 - 10 - 11 - 13 - 14	73.30
	1 - 13	94.73		1 - 5 - 9 - 10 - 11 - 15 - 17 - 18	72.85
	4 - 11 - 9	100		1 - 3 - 5 - 6 - 7 - 9 - 11 - 16 - 17 - 18	74.20
Spect	4 - 7 - 16	85.71	Wholesale	3 - 6	92.17
	7 - 13	85.71		5 - 6	92.17
	2 - 3 - 18 - 22	85.71		1 - 6 - 7	91.30

[4] Platform for feature selection based on nature-inspired algorithms, and it is available at https://github.com/lukapecnik/NiaAML

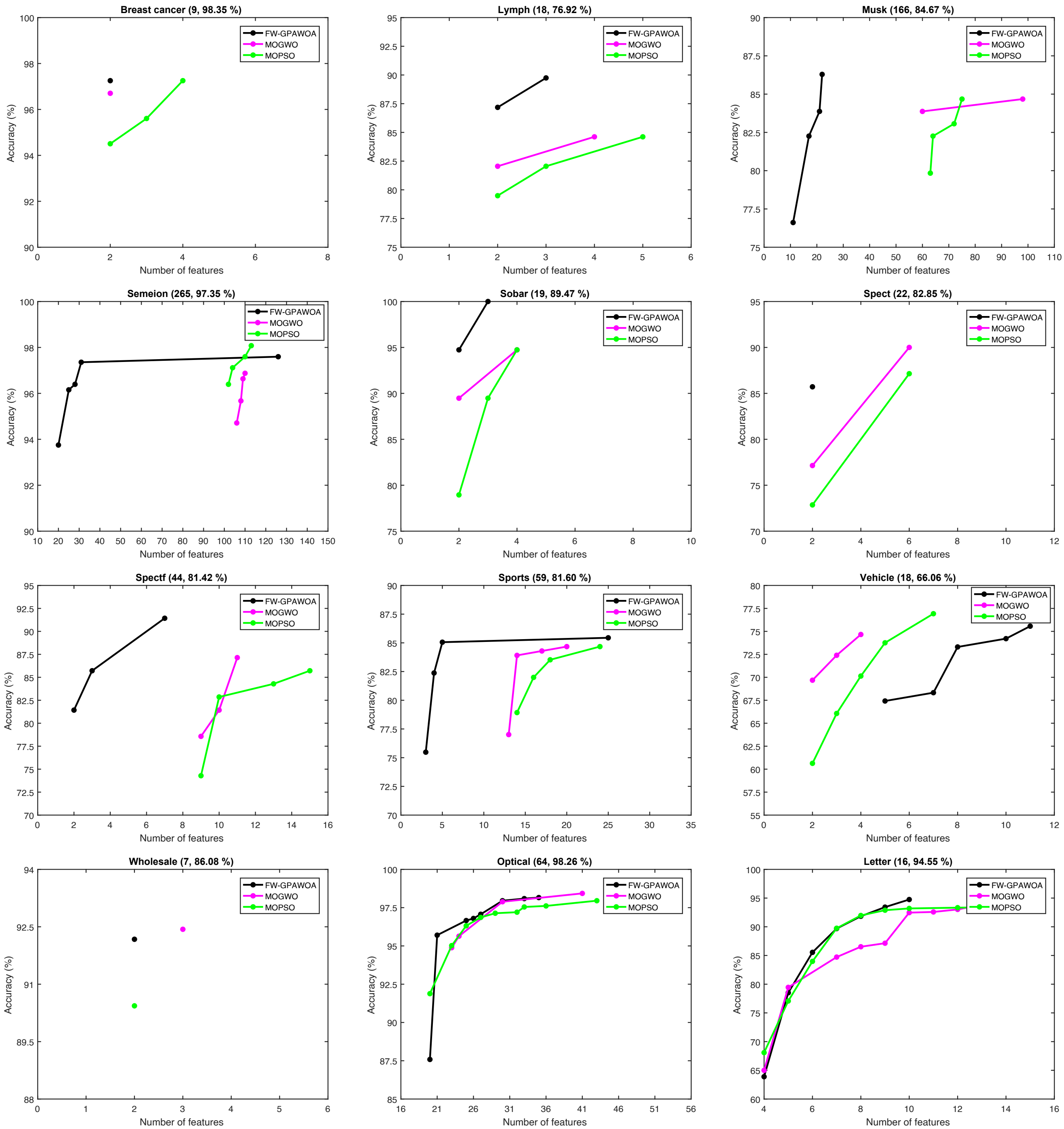

Fig. 1. Comparison between FW-GPAWOA, MOGWO and MOPSO.

some selected feature subsets with their accuracies on eight datasets. According to Table 3, it can be seen that the proposed approach could effectively select fewer number of features with higher classification accuracy in almost all datasets. For example, in Breast cancer, FW-GPAWOA selected on average of 2.06 from 9 features, which means around 23% of the initial feature set, selected 2.49 from 7 features in Wholesale dataset, i.e, 36% of the original feature set, and it was able to select approximately 11% of the available features (on average of 4.82 from 44 features) in Spectf dataset.

Regarding the classification performance, it can be seen clearly that FW-GPAWOA could achieve higher accuracy than using all features in most datasets. Indeed, except on Breast cancer, Optical and Letter datasets where the classification accuracy was slightly decreased from 98.35% to 97.14%, 98.26% to 97.96% and 94.55% to 93.65%, respectively, FW-GPAWOA was able to increase the prediction accuracy in all remaining cases. Generally speaking, the results suggest that the proposed algorithm can obtain feature subsets that included a smaller number of features with superior classification performance than using all features. Hence, FW-GPAWOA can be successfully used in feature selection to reduce the dimensionality and improve the classification performance.

Compared with the other algorithms, the results of Table 3 indicate that FW-GPAWOA performs better in terms of classification accuracy and the number of selected features in almost all datasets. For example,

Table 5
The results of HV metric obtained by FW-GPAWOA, MOGWO and MOPSO.

Algorithms	FW-GPAWOA			MOGWO			MOPSO		
	Average	Best	Worst	Average	Best	Worst	Average	Best	Worst
Breast cancer	**0.8237**	0.8498	0.8061	0.7988	0.8165	0.7017	0.8108	0.8144	0.8084
Lymph	0.6082	0.7453	0.4169	**0.7564**	0.8894	0.6931	0.7228	0.8339	0.6949
Musk	**1.0240**	1.0540	0.9777	0.7344	0.7964	0.7119	0.7094	0.7537	0.6832
Semeion	**0.5094**	0.5258	0.5026	0.3327	0.3450	0.3239	0.3505	0.3599	0.3366
Sobar	**0.8598**	0.9302	0.7558	0.8139	0.8550	0.7182	0.6720	0.8276	0.3693
Spect	**0.9328**	0.9990	0.8192	0.7253	0.7967	0.6802	0.8233	0.8833	0.7484
Spectf	**1.2404**	1.3120	1.1358	1.0126	1.0633	0.9410	0.8233	0.8833	0.7484
Sports	**1.7042**	1.7648	1.6217	1.3923	1.5150	1.2931	1.3997	1.4851	1.3461
Vehicle	0.4819	0.6168	0.4264	**0.6739**	0.7649	0.6055	0.6365	0.6781	0.5797
Wholesale	**0.7035**	0.7079	0.6879	0.5565	0.6492	0.4715	0.6823	0.6946	0.6679
Optical	**1.7435**	1.7621	1.7171	1.5271	1.7166	1.2921	1.5471	1.6388	1.4891
Letter	**0.6419**	0.7115	0.5234	0.3925	0.5849	0.2132	0.6072	0.7290	0.5156

compared against BPSO, in all datasets, FW-GPAWOA accomplished higher classification accuracy and selected feature subsets with smaller size. Further, on nine from the twelve used datasets (Breast cancer, Lymph, Musk, Sobar, Spect, Spectf, Sports, Wholesale and Letter), FW-GPAWOA achieved better results than BDE and BGWO in terms of both the number of features and classification performance. However, it obtains a lower classification accuracy than BDE, BGWO, WOA and jDE-FS in four datasets (i.e. Semeion, Vehicle,Spect and Optical), but with smaller feature subsets size. Accordingly, the proposed algorithm showed a higher ability to better explore the features space than the selected single-objective approaches. From the above discussion, we can say that FW-GPAWOA effectively reduces the dimensionality of features space by removing the unnecessary (irrelevant/redundant) features, while improving the classification performance in almost all cases.

4.3.2. FW-GPAWOA vs Multi-objective algorithms

For each dataset, FW-GPAWOA, MOGWO and MOPSO generate different subsets of non-dominated features in each of the 10 independent runs. For comparing these results, the features subsets provided by each algorithm are merged into one set. From this set, the non-dominated solutions are identified as the *best* Pareto front and they are used in the comparison. Fig. 1 illustrates the non-dominated results of FW-GPAWOA, MOGWO and MOPSO. The horizontal axis of each chart indicates the number of selected features, while the vertical axis shows the classification accuracy.

According to Fig. 1, in seven datasets, we can observe that the best Pareto front produced by FW-GPAWOA, contains at least two solutions that selected less than half of the available features and increased the classification accuracy regarding to that provided using all the features. For example, in Lymphography dataset, FW-GPAWOA selected 3 from 18 features and increased the classification accuracy from 76.92% to approximately 90%, and selected around 3 features from the 19 features of Sobar dataset, while it was able to increase the classification accuracy until 100%. However, in Musk dataset, our approach selected only one subset which includes fewer features and obtained higher classification performance than using all features. Furthermore, in Optical dataset, the Pareto front of FW-GPAWOA contains no solution which improve the accuracy of features, and it provides a unique solution which increase the accuracy of features in Letter dataset, but with more than half of available features (i.e. > 8 features). Despite that the results were very satisfactory, however, the performance of the algorithm slightly decreased in datasets which contain more classes (i.e. number of classes > 10).

Comparing FW-GPAWOA against MOGWO and MOPSO, in seven datasets (Breast, Musk, Lymph, Sobar, Spect, Spectf and Wholesale), it can be seen that the Pareto fronts resulted by FW-GPAWOA significantly dominate those of MOGWO and MOPSO. Also, the graphical results of Optical and Letter datasets show that the Pareto fronts of FW-GPAWOA are slighly better with respect to the other algorithms. In other words, the proposed algorithm outperforms its competitors in terms of both the number of features and classification performance. However, in Vehicle dataset, the Pareto front of FW-GPAWOA is dominated by those of MOGWO and MOPSO, especially for the number of feature criterion. In the other datasets, our method can always get better results than MOGWO and MOPSO in terms of feature subsets size or classification accuracy. For instance, compared with MOGWO on Semeion dataset, FW-GPAWOA obtained some solutions that include smaller features size, but with lower classification performance, and inversely, compared with MOPSO on Sports dataset, the proposed approach can achieve better classification performance, but with a larger feature subset.

In order to further evaluate the performance of the multi-objective algorithms, the hyper-Volume (HV) metric (Auger, Bader, Brockhoff, & Zitzler, 2009) is applied to compute the region of the objective space bounded by the obtained Pareto front and a reference point r. This metric has been frequently used for comparing multi-objective optimization algorithms as it can evaluate both convergence and diversity of solutions.

$$HV = volume\left(\bigcup_{s\in P} v(s,r)\right) \tag{20}$$

For each algorithm, 10 Hyper-volumes over the 10 independent runs are calculated based on the obtained testing Pareto fronts, and then Wilcoxon test (with significance level α equal to 5%) is applied to check whether there exists a significant difference between the algorithms by considering the hypotheses H_0 and H_1 below, knowing that μ_1 and μ_2 represent the results of FW-GPAWOA and the algorithm in comparison, respectively.

- H_0 $(p\_value > \alpha) : \mu_1 = \mu_2$, indicates that the results of FW-GPAWOA are similar "=" to the results of the algorithm in comparison.
- H_1 $(p\_value < \alpha) : \mu_1 \neq \mu_2$, indicates that the results of FW-GPAWOA are significantly better "+" (or worse "-") than those of the compared algorithm.

The results of HV-metric, as shown in Table 5, indicate that FW-GPAWOA achieves larger HV values for ten out of the twelve datasets. Comparing by MOGWO, the proposed method outperforms its competitor on ten cases, and compared against MOPSO, the results show that FW-GPAWOA has better HV-metric on all datasets. Therefore, we can say that our algorithm presents a set of non-dominated feature subsets near to the true Pareto front as well as preserving the diversity within the whale's population. Additionally, p-values of Wilcoxon test, as listed in Table 6, indicate that the HV-metrics obtained by FW-GPAWOA were significantly better than those obtained by MOGWO for nine datasets, similar for one dataset, and significantly worse for two datasets, while the superior performance of the proposed algorithm is more obvious with respect to MOPSO algorithm. Broadly speaking and for the 24 p-values (2 algorithms $\times$ 12 datasets), our algorithm obtains similar or

Table 6
The *p* values of Wilcoxon test for HV metric.

FW-GPAWOA vs	MOGWO	Sig. diff	MOPSO	Sig. diff
Breast	**4.860E−3**	+	**1.697E−2**	+
Lymph	**1.213E−2**	−	0.060	=
Musk	**7.936E−3**	+	**7.936E−3**	+
Semeion	**1.082E−5**	+	**1.082E−5**	+
Sobar	0.089	=	**8.324E−3**	+
Spect	**1.082E−5**	+	**5.033E−3**	+
Spectf	**1.082E−5**	+	**1.082E−5**	+
Sports	**1.082E−5**	+	**1.082E−5**	+
Vehicle	**7.577E−5**	−	**4.330E−3**	−
Wholesale	**1.418E−2**	+	**1.082E−5**	+
Optical	**4.570E−5**	+	**1.082E−5**	+
Letter	**7.577E−5**	+	0.2474	=

Table 7
The average running time consumed by the algorithms (unit: minutes).

	FW-GPAWOA	MOGWO	MOPSO
Breast	**2.62**	**3.73**	4.25
Lymph	4.42	**4.15**	4.35
Musk	**3.54**	3.66	4.02
Semeion	6.97	**6.55**	7.63
Sobar	3.87	**3.49**	3.57
Spect	4.18	**3.81**	4.04
Spectf	4.23	**3.88**	3.98
Sports	4.30	**3.76**	4.12
Vehicle	4.85	4.99	**4.33**
Wholesale	**2.29**	3.64	4.21
Optical	10.89	**10.12**	10.52
Letter	12.49	**11.07**	11.43

significantly better results for 21 cases, and it gets significantly worse results than its counterparts only for 3 cases.

4.3.3. Computational time analysis

Table 7 shows the average running time (in minutes) of FW-GPAWOA, MOGWO and MOPSO over the 10 independent runs. Here, it should be remembered that all the algorithms have the same population size, the same number of iterations, and they have been executed on the same computer.

From Table 7, it can be seen that our approach consumes a longer running time than the other approaches for most datasets. Although the proposed algorithm selected smaller feature subsets, which be supposed to leads to less wrapper evaluation, hence, low computational time should be required. So, the reason why the proposed approach spent more time can be explained, mainly, because it uses the Crowding distance (that requires high computational cost) in the archiving strategy and in the deletion process whenever the external archive is full. The other reason may be that MOGWO and MOPSO use directly the ratio between the selected features and the original features set as one of the objective functions, which requires obviously less computational effort than using mutual information. However, the growth in the running time consumed by FW-GPAWOA was acceptable for effectively reducing the dimensionality and keeping a highest prediction accuracy. Indeed, the FW-GPAWOA tends to increase the computational time in the hope of improving its performance.

As a summary, all the above results showed that our algorithm gets a superior performance and outperforms the benchmarking algorithms in almost all cases. The experiments demonstrated that the proposed algorithm could effectively explore the Pareto front to accomplish the main challenges of the feature selection problem which are the curse of dimensionality of the features space and the improvement of the classification performance.

5. Conclusion and future work

This paper presented a new multi-objective optimizer for solving feature selection problem. The proposed algorithm investigates our recent GPAWOA algorithm and combines a filter and wrapper models into a single scheme in the hope of benefits from each model's merits. Therefore, mutual information and KNN classifier are used as filter and wrapper evaluation criteria during the training process. In addition, the hyperbolic tangent function is employed to make GPAWOA able to deal with discrete problems.

A comparative study with seven well-regarded algorithms has been performed on twelve benchmark datasets. Experimental results show that the proposed algorithm generally outperforms the selected approaches in terms of both number of features and classification performance. However, we have observed that the performance of the proposed algorithm is not the same when the number of classes increases. Moreover, we have observed that FW-GPAWOA usually requires more running time due to the crowding distance computation and the used filter function. Therefore, as part of future work, we plan to investigate other objective functions with the aim of maintaining better performance without increasing the running time. The ration between the selected features and the original set may be a future direction. Also, we are interested in applying other evolutionary computation techniques with different learning algorithms such as decision tree (DT) and support vector machine (SVM).

CRediT authorship contribution statement

Adel Got: Supervision, Conceptualization, Methodology, Software, Investigation, Validation, Writing - original draft. **Abdelouahab Moussaoui:** Writing - original draft, Visualization, Investigation. **Djaafar Zouache:** Conceptualization, Methodology, Writing - original draft, Visualization, Investigation.

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

Amoozegar, M., & Minaei-Bidgoli, B. (2018). Optimizing multi-objective pso based feature selection method using a feature elitism mechanism. *Expert Systems with Applications, 113*, 499–514.
Auger, A., Bader, J., Brockhoff, D., & Zitzler, E. (2009). Theory of the hypervolume indicator: optimal-distributions and the choice of the reference point. In *Proceedings of the tenth ACM SIGEVO workshop on Foundations of genetic algorithms* (pp. 87–102).
Brezočnik, L., Fister, I., & Podgorelec, V. (2018). Swarm intelligence algorithms for feature selection: a review. *Applied Sciences, 8*, 1521.
Cervante, L., Xue, B., Shang, L., & Zhang, M. (2013). A multi-objective feature selection approach based on binary pso and rough set theory. In *European conference on evolutionary computation in combinatorial optimization* (pp. 25–36). Springer.
Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering, 40*, 16–28.
Coello, C. A. C., Pulido, G. T., & Lechuga, M. S. (2004). Handling multiple objectives with particle swarm optimization. *IEEE Transactions on Evolutionary Computation, 8*, 256–279.
Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis, 1*, 131–156.
El Aziz, M. A., Ewees, A. A., & Hassanien, A. E. (2018). Multi-objective whale optimization algorithm for content-based image retrieval. *Multimedia Tools and Applications, 77*, 26135–26172.
Emary, E., Yamany, W., Hassanien, A. E., & Snasel, V. (2015). Multi-objective gray-wolf optimization for attribute reduction. *Procedia Computer Science, 65*, 623–632.
Fister, D., Fister, I., Jagrič, T., & Brest, J. (2018). A novel self-adaptive differential evolution for feature selection using threshold mechanism. In *2018 IEEE symposium series on computational intelligence (SSCI)* (pp. 17–24). IEEE.
Frank, A. (2010). *Uci machine learning repository*. http://archive. ics. uci. edu/ml.
Got, A., Moussaoui, A., & Zouache, D. (2020). A guided population archive whale optimization algorithm for solving multiobjective optimization problems. *Expert Systems with Applications, 141*, Article 112972.

Hamdani, T. M., Won, J.-M., Alimi, A. M., & Karray, F. (2007). Multi-objective feature selection with nsga ii. In *International conference on adaptive and natural computing algorithms* (pp. 240–247). Springer.

Hammami, M., Bechikh, S., Hung, C.-C., & Said, L. B. (2019). A multi-objective hybrid filter-wrapper evolutionary approach for feature selection. *Memetic Computing, 11*, 193–208.

Hancer, E., Xue, B., Zhang, M., Karaboga, D., & Akay, B. (2015). A multi-objective artificial bee colony approach to feature selection using fuzzy mutual information. In *2015 IEEE congress on evolutionary computation (CEC)* (pp. 2420–2427). IEEE.

Hancer, E., Xue, B., Zhang, M., Karaboga, D., & Akay, B. (2018). Pareto front feature selection based on artificial bee colony optimization. *Information Sciences, 422*, 462–479.

Hussien, A. G., Houssein, E. H., & Hassanien, A. E. (2017). A binary whale optimization algorithm with hyperbolic tangent fitness function for feature selection. In *2017 Eighth international conference on intelligent computing and information systems (ICICIS)* (pp. 166–172). IEEE.

Jović, A., Brkić, K., & Bogunović, N. (2015). A review of feature selection methods with applications. In *2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)* (pp. 1200–1205). IEEE.

Labani, M., Moradi, P., & Jalili, M. (2020). A multi-objective genetic algorithm for text feature selection using the relative discriminative criterion. *Expert Systems with Applications, 149*, Article 113276.

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM Computing Surveys (CSUR), 50*, 1–45.

Liu, H., & Motoda, H. (1998). *Feature extraction, construction and selection: A data mining perspective* (Vol. 453). Springer Science & Business Media.

Mafarja, M., & Mirjalili, S. (2018). Whale optimization approaches for wrapper feature selection. *Applied Soft Computing, 62*, 441–453.

Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in Engineering Software, 95*, 51–67.

Mirjalili, S., Saremi, S., Mirjalili, S. M., & Coelho, L.d. S. (2016). Multi-objective grey wolf optimizer: a novel algorithm for multi-criterion optimization. *Expert Systems with Applications, 47*, 106–119.

Rathee, S., & Ratnoo, S. (2020). Feature selection using multi-objective chc genetic algorithm. *Procedia Computer Science, 167*, 1656–1664.

Sahoo, A., & Chandra, S. (2017). Multi-objective grey wolf optimizer for improved cervix lesion classification. *Applied Soft Computing, 52*, 64–80.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379–423.

Sharawi, M., Zawbaa, H. M., & Emary, E. (2017). Feature selection approach based on whale optimization algorithm. In *2017 Ninth International Conference on Advanced Computational Intelligence (ICACI)* (pp. 163–168). IEEE.

Swiniarski, R. W., & Skowron, A. (2003). Rough set methods in feature selection and recognition. *Pattern Recognition Letters, 24*, 833–849.

Too, J., Abdullah, A. R., & Mohd Saad, N. (2019). Hybrid binary particle swarm optimization differential evolution-based feature selection for emg signals classification. *Axioms, 8*, 79.

Too, J., Abdullah, A. R., Mohd Saad, N., Mohd Ali, N., & Tee, W. (2018). A new competitive binary grey wolf optimizer to solve the feature selection problem in emg signals classification. *Computers, 7*, 58.

Too, J., Abdullah, A. R., Mohd Saad, N., & Tee, W. (2019). Emg feature selection and classification using a pbest-guide binary particle swarm optimization. *Computation, 7*, 12.

Unler, A., & Murat, A. (2010). A discrete particle swarm optimization method for feature selection in binary classification problems. *European Journal of Operational Research, 206*, 528–539.

Xue, B., Cervante, L., Shang, L., Browne, W. N., & Zhang, M. (2012). A multi-objective particle swarm optimisation for filter-based feature selection in classification problems. *Connection Science, 24*, 91–116.

Xue, B., Cervante, L., Shang, L., Browne, W. N., & Zhang, M. (2013). Multi-objective evolutionary algorithms for filter based feature selection in classification. *International Journal on Artificial Intelligence Tools, 22*, 1350024.

Xue, B., Zhang, M., & Browne, W. N. (2012). Particle swarm optimization for feature selection in classification: A multi-objective approach. *IEEE Transactions on Cybernetics, 43*, 1656–1671.

Xue, B., Zhang, M., Browne, W. N., & Yao, X. (2015). A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation, 20*, 606–626.

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.

Zhang, Y., Cheng, S., Shi, Y., Gong, D.-W., & Zhao, X. (2019). Cost-sensitive feature selection using two-archive multi-objective artificial bee colony algorithm. *Expert Systems with Applications, 137*, 46–58.