IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 26, 2019, accepted March 15, 2019, date of publication March 21, 2019, date of current version April 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2906757

Binary Optimization Using Hybrid Grey Wolf Optimization for Feature Selection

QASEM AL-TASHI[1,2], SAID JADID ABDUL KADIR[1,3], HELMI MD RAIS[1], SEYEDALI MIRJALILI[4], AND HITHAM ALHUSSIAN[1]

[1]Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Seri Iskandar 32160, Malaysia
[2]Faculty of Administrative and Computer Sciences, University of Albaydha, Rada'a CV46+6X, Yemen
[3]Centre for Research in Data Science, Universiti Teknologi PETRONAS, Seri Iskandar 32160, Malaysia
[4]Institute for Integrated and Intelligent Systems, Griffith University, Brisbane, QLD 4111, Australia

Corresponding author: Qasem Al-Tashi (qasemacc22@gmail.com)

This Research was funded by Centre for Research in Data Science (CERDAS), Universiti Teknologi PETRONAS.

ABSTRACT A binary version of the hybrid grey wolf optimization (GWO) and particle swarm optimization (PSO) is proposed to solve feature selection problems in this paper. The original PSOGWO is a new hybrid optimization algorithm that benefits from the strengths of both GWO and PSO. Despite the superior performance, the original hybrid approach is appropriate for problems with a continuous search space. Feature selection, however, is a binary problem. Therefore, a binary version of hybrid PSOGWO called BGWOPSO is proposed to find the best feature subset. To find the best solutions, the wrapper-based method K-nearest neighbors classifier with Euclidean separation matric is utilized. For performance evaluation of the proposed binary algorithm, 18 standard benchmark datasets from UCI repository are employed. The results show that BGWOPSO significantly outperformed the binary GWO (BGWO), the binary PSO, the binary genetic algorithm, and the whale optimization algorithm with simulated annealing when using several performance measures including accuracy, selecting the best optimal features, and the computational time.

INDEX TERMS Feature selection, hybrid binary optimization, grey wolf optimization, particle swarm optimization, classification.

I. INTRODUCTION

Data mining is regarded as the fastest growing subfield of information technology, which is due to the massive data collected daily and the necessity of converting this data into useful information [1]. Data mining involves a several preprocessing (integration, filtering, transformation, reduction, etc.), knowledge presentation and also pattern evaluation [2]. One of the main preprocessing steps is called feature selection that aims to remove irrelevant and redundant attributes of specific dataset. Generally speaking, algorithms of feature selection are classified into two classes: filters or wrapper approaches [3], [4]. The former class include methods independent from classifiers and work directly on date. Such methods normally find correlations between variables. On the other hand, wrapper feature selection methods involve classifiers and find interaction between variables. As the literature shows, wrapper-based methods are better than filter-based techniques for classification algorithms [5], [6]. Commonly, three key elements must be specified when utilizing a wrapper-based methods including: classifiers (e.g., Support Vector Machine (SVM), KNN, etc.), evaluation criteria of feature subset and the search algorithm to find a subset including the optimal features [7].

Finding an optimal set of features is challenging and computationally expensive task. Recently, metaheuristics seem to be effective and reliable tools for solving several optimization problems (e.g., machine learning, data mining problems, engineering design, and feature selection) [8].

In contrast to the exact search mechanisms, metaheuristics show an outstanding performance, as they do not have to search the entire search space. In fact, they are not complete search algorithm. Exact methods, however, are complete and guarantee finding the best solution for a problem subject to having enough time and memory. They are not efficient for problems with high computational complexity. In the problem of feature selection, for instance, a dataset with n features includes the total number of 2^n solutions. Therefore, the

The associate editor coordinating the review of this manuscript and approving it for publication was Aysegül Ucar.

problem of feature selection has an exponential computational growth [4]. Another search method for selecting best feature subsets is a random search which randomly searches the next set [9]. Metaheuristics can be considered as "directed" random algorithms [8], [10]; they find an acceptable solution but does not guarantee determining the optimal solution in each run [15]. Metaheuristics have been largely employed to solve feature selection problems including: GWO [11], [12], Genetic Algorithm (GA) [13], Ant Colony Optimization (ACO) [14], PSO [15], Differential Evolution (DE) [16], Dragon algorithm (DA) [17], to name a few.

Exploring the space of the search and exploiting the optimal solutions found are two contradictory principles to be considered when using or modeling a metaheuristic [8]. Balancing exploration and exploitation in a good manner will led to the improvement of the search algorithm's performance. In order to achieve a good balance, one option is to utilize a hybrid approach where two algorithms or more are combined to improve each algorithm's performance and the resulted hybrid approach is named a memetic method [18]. This motivated our attempt to proposed a binary version of the hybrid PSOGWO [19], which has been utilized for continues search space problem and develop a binary version of it to enhance the feature selection and classification tasks.

According to [19] and based on Talbi [8], two algorithms belong to the class of co-evolutionary techniques can be hybridized at a high or low level. In this paper, the hybrid PSO and GWO in [19] uses low evolutionary mixed hybrid. The hybrid proposed in [19] can be considered low because the functionality of the two algorithms is combined. It is coevolutionary because both algorithms are not used one after the other which means the run made in parallel. It is mixed, due to two separate algorithms implicated in the final solution of the problems. The main intention of Singh [19] has been to use exploration of PSO and exploitation of GWO to solve optimization problems.

PSO was invented in [20], which inspired by the flocking and schooling behaviours of birds and fish [21]. This algorithm is easy to implement, it monitors three global variables namely: target value, global best (gBest), and stopping value. Besides, every particle in PSO contains a position vector and a vector to solve the personal best (pBest), and a variable to store the objective value of pBest.

GWO [22] is a recently-proposed swarm intelligence technique that has gain great reception in the optimization community. This algorithm mimics the hunting and dominancy behaviour of grey wolves in nature [23]. This algorithm has been largely applied to a wide range of problems in the literature.

This study proposes a binary version of the hybrid PSOGWO in [19] and uses it as a wrapper feature selection method. The main contribution is to use suitable operators to solve binary problems using PSOGWO. The rest of this paper is organized as follows: Section II presents the state-of-art approaches. Section III introduces the methods. The proposed binary approach is clearly described in Section IV. In Section V, the results are given, and required analyses are provided. Lastly, in Section VI, conclusions of the study and future work are explained.

II. RELATED WORK

Recently, the area of optimization has gained much attention from researchers especially in hybrid metaheuristics field [8]. For instance, the first proposed feature selection method using hybrid metaheuristic was in 2004 [24] using local search methods and the GA algorithm.

In the literature, PSO has been hybridized with other metaheuristics for continuous search space problems. In [25], for instance, a hybrid PSO with GA (PSOGA) was proposed. Other similar works are: a PSO with DE (PSODE) [26], hybrid PSO and Gravitational Search Algorithm (GSA) (PSOGSA) [27]. Moreover, PSO was hybridized with Bacterial Foraging Optimization algorithm for power system stability enhancement in [28]. These hybrid approaches are aimed to share the strength of each other to expand the capability of exploitation and reducing the chances of dropping in local optimum.

Similarly, GWO has gained much attention in the hybrid metaheuristics field. For instance, in [29] and [30], the authors have hybridized GWO with DE for test scheduling and continuous optimization. Tawhid and Ali [31] have hybridized GWO with GA for minimizing potential energy functions. Gaidhane and Nigam [32] proposed a hybridized GWO and Artificial Bee Colony (ABC) to improve the complex systems performance. Another hybrid method is GWOSCA proposed in [33] using GWO and Sine Cosine Algorithm (SCA). These studies have shown that the hybrid methods performed much better compared to other global or local search methods.

Metaheuristics have been popular in the field of feature selection as well. For instance, a hybrid filter feature selection approach has been proposed in [34] using SA with GA to improve the search ability of GA, the performance was evaluated on eight datasets collected from UCI and obtained a good outcomes considering the selected number of attributes. Another study hybridized GA with SA and evaluated on the Farsi characters hand-printed [35].

Moreover, a hybrid PSO with novel local search strategy based on information correlation was proposed in [36]. A hybrid GA with PSO named GPSO for wrapper feature selection using SVM classifier for classifying microarray data [37]. In the same filed unreliable data, the authors proposed a hybrid mutation operator for an improved multi-objective PSO [38]. For Digital Mammogram datasets, a hybrid GA with PSO to enhance the feature set was proposed in [39]. In [40] and [41], two hybrids were proposed using ACO and GA to perform feature selection. Another similar method can be found in [42]. In [16], a hybrid of DE and ABC was used as a feature selector. For the same purpose, Nekkaa and Boughaci [43], proposed a hybrid harmony search algorithm with a local stochastic search. Recently, in [18] a hybrid WOA and SA was proposed for wrapper feature selection. Besides, a hybrid between GWO and antlion optimization (ALO) for feature selection was proposed in [44].

Consistent with (No Free Lunch) theorem, there has been, is, and will be no optimization algorithm to solve all optimization problems. While an algorithm shows a good performance on specific datasets, its performance might degrades on similar or other types of datasets [45]. In spite of the good performance of above-mentioned methods we can state that none of them is capable to solve all problems related to feature selection. As such, improvements can be made to the existing methods to enhance the solutions of feature selection problems. In the next section, the methodology is discussed, and the proposed binary hyper metaheuristic is clearly explained.

III. METHODS

A. GREY WOLF OPTIMIZATION ALGORITHM

GWO proposed in [22] has been inspired from the social intelligence of grey wolves that prefer living in a group of 5-12 individuals. In order to simulate the leadership hierarchy of GWO four levels are considered in this algorithm: alpha, beta, delta, and omega. Alpha known as male and female and the leaders of a pack, making decisions (e.g. hunting, sleep place and wake-up time) are the main responsibility of alpha. Beta known to assist alpha in making decisions and the main responsibility of beta is the feedback suggestions. Delta performs as scouts, sentinels, caretakers, elders and hunters. Delta controls omega wolves by obeying alpha and beta wolves. The omega wolves must obey every other wolves.

In the GWO, α, β, and δ, guides the hunting process and ω wolves follows them. The encircling behavior of GWO can be calculated as follows:

$$\vec{\mathrm{X}}\,(\mathrm{t}+1) = \vec{\mathrm{X}}_p\,(t) + \vec{\mathrm{A}} \cdot \vec{\mathrm{D}} \quad (1)$$

where $\vec{A}$, $\vec{C}$ are coefficient vectors, $\vec{\mathrm{X}}_{\mathrm{p}}$ is the prey's positions vector, X mimics the position of wolves in a d-dimensional space where d is the number of variables, (t) is the iterations number, and $\vec{D}$ is denoted as follows:

$$\vec{\mathrm{D}} = \left|\vec{\mathrm{C}} \cdot \vec{\mathrm{X}}_p(t) - \vec{\mathrm{X}}\,(t)\right| \quad (2)$$

where $\vec{\mathrm{A}}$, $\vec{C}$ are donated as following:

$$\vec{\mathrm{A}} = 2\vec{a} \cdot \vec{r_1} - \vec{a} \quad (3)$$
$$\vec{\mathrm{C}} = 2 \cdot \vec{r_2} \quad (4)$$

where $\vec{r_1}$, $\vec{r_2}$ are vectors randomly in [0, 1]. $\vec{a}$ a set vector linearly decreases from 2 to 0 over iterations.

In the hunting process of grey wolves, alpha is considered the optimal applicant for the solution, beta and delta expected to be knowledgeable about the prey's possible position. Thus, three best solutions that have been found until a certain iteration are kept and forces others (e.g. omega) to modify their positions in the decision space consistent with the best place. The updating positions mechanism can be calculated as follows:

$$\vec{\mathrm{X}}\,(t+1) = \frac{\overrightarrow{x_1 + x_2 + x_3}}{3} \quad (5)$$

where $x_1, x_2; x_3$ are defined and calculated as following:

$$\begin{aligned} \vec{x_1} &= \vec{X_\alpha} - A_1 \cdot \left(\vec{D_\alpha}\right), \\ \vec{x_2} &= \vec{X_\beta} - A_2 \cdot \left(\vec{D_\beta}\right), \\ \vec{x_3} &= \vec{X_\delta} - A_3 \cdot \left(\vec{D_\delta}\right) \end{aligned} \quad (6)$$

where $\vec{x_1}$, $\vec{x_2}$ and $\vec{x_3}$ are the three best wolves (solutions) in the swarm at a given iteration t. Where, A_1, A_2 and A_3 are calculated as in Eq (3). $\vec{D_\alpha}, \vec{D_\beta}, \vec{D_\delta}$ are calculated as in Eq (7).

$$\begin{aligned} \vec{D_\alpha} &= \left|\vec{\mathrm{C}}_1 \cdot \vec{\mathrm{X}}_\alpha - \vec{\mathrm{X}}\right|, \\ \vec{D_\beta} &= \left|\vec{\mathrm{C}}_2 \cdot \vec{\mathrm{X}}_\beta - \vec{\mathrm{X}}\right|, \\ \vec{D_\delta} &= \left|\vec{\mathrm{C}}_3 \cdot \vec{\mathrm{X}}_\delta - \vec{\mathrm{X}}\right| \end{aligned} \quad (7)$$

where $\vec{\mathrm{C}}_1, \vec{\mathrm{C}}_2, \vec{\mathrm{C}}_3$ are calculated based on Eq (4).

In GWO one of the main components to tune exploration and exploitation is the vector $\vec{a}$. In the main paper of this algorithm, it is suggested to decrease the vector for each of dimension linearly proportional to the number of iterations from 2 to 0. The equation to update it is as follows:

$$\vec{a} = 2 - t \cdot \frac{2}{max\,ter_i} \quad (8)$$

where t is the iteration number, ter is the optimization total iterations number.

B. PARTICLE SWARM OPTIMIZATION ALGORITHM

PSO was introduced in [20]. As a swarm intelligence technique, it mimics the intelligence of bird swarms of fish schools in nature. In PSO, each particle is represented with a position vector and velocity vector. Every particle has individual intelligence and search a search space around the best solution that it has found so far. Particles also know the best position that all particles (as a swarm) has found so far. The position and velocity vectors are updated using the following equations.

$$v_i^{k+1} = v_i^k + c_1 r_1 \left(Pbest_i^k - x_i^k\right) + c_2 r_2 \left(g_{best} - x_i^k\right) \quad (9)$$
$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad (10)$$

C. THE HYBRID PSOGWO (CONTINUOUS VERSION)

The hybrid PSOGWO algorithm was proposed in [19]. The PSOGWO's basic idea is to increase the algorithm's capability to exploit PSO with the ability to explore GWO to achieve both optimizer strength. In HPSOGWO, first three agents' position is updated in the search space, instead of using usual mathematical equations, the exploitation and exploration of the grey wolf were controlled by inertia constant. This was mathematically modeled as follows:

$$\overrightarrow{D_\alpha} = \left|\overrightarrow{C_1} \cdot \overrightarrow{X}_\alpha - w * \overrightarrow{X}\right|,$$
$$\overrightarrow{D_\beta} = \left|\overrightarrow{C_2} \cdot \overrightarrow{X}_\beta - w * \overrightarrow{X}\right|,$$
$$\overrightarrow{D_\delta} = \left|\overrightarrow{C_3} \cdot \overrightarrow{X}_\delta - w * \overrightarrow{X}\right| \quad (11)$$

To combine PSO and GWO variants, the velocity and positions have been updated as follows:

$$v_i^{k+1} = w * (v_i^k + c_1 r_1 \left(x_1 - x_i^k\right) + c_2 r_2 \left(x_2 - x_i^k\right) + c_3 r_3 \left(x_3 - x_i^k\right)) \quad (12)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad (13)$$

IV. THE PROPOSED BINARY APPROACH (BGWOPSO)

Feature selection is a binary problem by nature. Therefore, the algorithm presented in Section III-C cannot be used to solve such problems without modifications. A binary version of the hybrid PSOGWO should be developed to be suitable for the problem of feature selection. Agents can move around the search space continuously in the original PSOGWO [19] since they have position vectors with a continuous real domain. According to [11] the updating mechanism of wolves is a function of three vectors position namely x_1,x_2;x_3 which promotes every wolf to the first three best solutions. For the agents to work in a binary space, the position updating (5) can be modified into the following equation [11]:

$$x_d^{t+1} = \begin{cases} 1 \ if \ sigmoid\left(\frac{x_1 + x_2 + x_3}{3}\right) \geq rand \\ 0 \qquad otherwise \end{cases} \quad (14)$$

where x_d^{t+1} is the binary updated position at iteration t in dimension d, *rand* is a random number drawn from uniform distribution ∈ [1,0], and sigmoid(a) is denoted as following [11]:

$$sigmoid\,(a) = \frac{1}{1 + e^{-10(x-0.5)}} \quad (15)$$

x_1, x_2, x_3 in (6) are updated and calculated using the following equations [11]:

$$x_1^d = \begin{cases} 1 \ if \ \left(x_\alpha^d + bstep_\alpha^d\right) \geq 1 \\ 0 \qquad otherwise \end{cases}$$
$$x_2^d = \begin{cases} 1 \ if \ \left(x_\beta^d + bstep_\beta^d\right) \geq 1 \\ 0 \qquad otherwise \end{cases}$$
$$x_3^d = \begin{cases} 1 \ if \ \left(x_\delta^d + bstep_\delta^d\right) \geq 1 \\ 0 \qquad otherwise \end{cases} \quad (16)$$

where $x_{\alpha,\beta,\delta}^d$ the position's vector of the alpha, beta, delta wolves in d dimension, and $bstep_{\alpha,\beta,\delta}^d$ is a binary step in d dimension, which can be formulated as follow [11]:

$$bstep_{\alpha,\beta,\delta}^d = \begin{cases} 1 \ if \, cstep_{\alpha,\beta,\delta}^d \geq rand \\ 0 \qquad otherwise \end{cases} \quad (17)$$

where *rand* a random value derived from uniform distribution ∈ [1,0], d indicates dimension, and $cstep_{\alpha,\beta,\delta}^d$ is d's continuous value. This component is calculated using the following equation [11]:

$$cstep_{\alpha,\beta,\delta}^d = \frac{1}{1 + e^{-10\left(A_1^d D_{\alpha,\beta,\delta}^d - 0.5\right)}} \quad (18)$$

In BGWOPSO, and based on the best three solutions positions updated in (16), the exploration and exploitation are controlled by an inertia constant weight mathematically modeled as follows [19]:

$$\overrightarrow{D_\alpha} = \left|\overrightarrow{C_1} \cdot \overrightarrow{X}_\alpha - w * \overrightarrow{X}\right|,$$
$$\overrightarrow{D_\beta} = \left|\overrightarrow{C_2} \cdot \overrightarrow{X}_\beta - w * \overrightarrow{X}\right|,$$
$$\overrightarrow{D_\delta} = \left|\overrightarrow{C_3} \cdot \overrightarrow{X}_\delta - w * \overrightarrow{X}\right| \quad (19)$$

Accordingly, the velocity and positions have been updated as follows [19]:

$$v_i^{k+1} = w * (v_i^k + c_1 r_1 \left(x_1 - x_i^k\right) + c_2 r_2 \left(x_2 - x_i^k\right) + c_3 r_3 \left(x_3 - x_i^k\right) \quad (20)$$

Note that in (20) the best three solutions x_1,x_2, x_3 are updated according to (16).

$$x_i^{k+1} = x_d^{t+1} + v_i^{k+1} \quad (21)$$

where x_d^{t+1} and v_i^{k+1} are calculated based on Eq (14) and Eq (20) respectively.

Pseudocode 1 Pseudocode of the Proposed Binary (BGWOPSO)

Initialization
Initialize *A, a, C* and *w*
Randomly Initialize an agent of *n* wolves positions ∈ [1,0].
Based on the fitness function attain the α; β; δ solutions.
Evaluate the fitness of agents by using Eq (19)
While (t < Max_iter)
  For each population
    Update the velocity using Eq (20)
    Update the position of agents into a binary position based on Eq (21)
  end
  Update *A, a, C and w*
  Evaluate all particles using the objective function
  Update the positions of the three best agents α, β, δ t= t+ 1
end while

The solution in this study is illustrated in a one-dimensional vector. The length of this vector is equal to the number of features. In this binary vector, 0 and 1 have the following meaning:

- 0: feature is not selected
- 1: feature is selected

TABLE 1. Benchmark datasets used.

	Dataset	Instances	No. Features
1	Breastcancer	699	9
2	BreastEW	569	30
3	CongressEW	435	16
4	Exactly	1000	13
5	Exactly2	1000	13
6	HeartEW	270	13
7	IonosphereEW	351	34
8	KrvskpEW	3196	36
9	Lymphography	148	18
10	M-of-n	1000	13
11	PenglungEW	73	325
12	SonarEW	208	60
13	SpectEW	267	22
14	Tic-tac-toe	958	9
15	Vote	300	16
16	WaveformEW	5000	40
17	WineEW	178	13
18	Zoo	101	16

Feature selection problem is bi-objective by nature. One objective is to find the minimum number of features, and the other is to maximization the classification accuracy. To consider both, the following equations used as a fitness function (the classifier is KMN [11], [46]:

$$fitness = \alpha \rho_R (D) + \beta \frac{|S|}{|T|} \tag{22}$$

where $\alpha = [0,1]$ and $\beta = (1 - \alpha)$ they are a parameters adapted form [11], $\rho_R (D)$ indicates the error rate of the KNN classifier. Moreover, $|S|$ is the selected subset of features and $|T|$ is the whole features in the dataset.

It should be noted that the mathematical equations used in this section are obtained from [11] and [19]. In fact, this work proposed the use of ideas/equations in [11] in conjunction with the hybrid PSOGSA [19] to solve binary problems.

V. EXPERMINTAL RESULTS

A. DATASETS

To validate the proposed binary algorithm, the BGWOPSO is tested against 18 benchmark datasets (see Table 1) collected from the UC Irvine Machine Learning Repository [47].

B. PARAMETER SETTINGS

To produce the best solutions, the wrapper-based method KNN classifier with the Euclidean separation matric is utilized (where K = 5).

In this research, every dataset is partitioned using cross validation similar to that in[18] for assessment. In K - overlap cross-approval, K − 1 folds are utilized for training and validation and the rest of the overlay is utilized for testing. The proposed approach is repetitive for M times. Hereafter, the proposed approach is assessed k * M times for every dataset. The information for training and validation are similarly estimated. The parameters of the proposed approach are set as following:

- Number of wolves: 10,
- maximum number of iterations is 100,
- $c1 = c2 = 0.5$, $c3 = 0.5$,
- $w = 0.5 + \text{rand}()/2$, and $l \in [1,0]$

all these parameter settings are applied to test the quality of the proposed hybrid approach. Furthermore, the algorithm is run 20 times with random seed on an Intel(R) Core$^{\text{TM}}$i7-6700 machine, 3.4GHz CPU and 16GB of RAM. The parameters applied in bGWO2[11], BPSO, BGA and WOASAT-2 [18] are identical to their own parameters setting used.

C. EVALUATION MEASURES

The datasets are randomly partitioned into three diverse equivalent portions (e.g., validation, training, and testing datasets). The dividing of the data is repeated for multiple times to guarantee strength and measurable noteworthiness of the outcomes. The following statistical measures are tested from the validation data in each run:

1) THE AVERAGE OF CLASSIFICATION ACCURACY

It is an indicator depicts how precise is the classifier given the chosen set of features when algorithm run N times, and it is calculated as follows:

$$AvgAcc = \frac{1}{N} \sum_{k=1}^{N} AvgAcc^k \tag{23}$$

where $AvgAcc^k$ is the value of accuracy gained at run k.

2) THE AVERAGE OF SELECTED FEATURE

It is an indicator to the average selected features to the overall features when algorithm run N times, and it is calculated as follows:

$$AvgSelection = \frac{1}{N} \sum_{k=1}^{N} \frac{AvgSelection^k}{M} \tag{24}$$

where $AvgSelection^k$ is the selected features at run k, and M shows the dataset's total number of features.

3) THE MEAN FITNESS FUNCTION

Is an indicator to the average value of the fitness function gained when algorithm run N times, and it is calculated as follows:

$$mean = \frac{1}{N} \sum_{k=1}^{N} g_k^* \tag{25}$$

where g_k^* the mean fitness value gained at run k.

TABLE 2. Results obtained from the proposed approach BGWOPSO.

	Dataset	Average Accuracy	Mean Fitness	Feature Selected	Computational time(s)
1	Breastcancer	0.98	0.03	4.4	5.69
2	BreastEW	0.97	0.04	13.6	5.38
3	CongressEW	0.98	0.03	4.4	5.35
4	Exactly	1.0	0.0046	6	5.61
5	Exactly2	0.76	0.24	1.6	5.41
6	HeartEW	0.85	0.15	5.8	5.25
7	IonosphereEW	0.95	0.05	13	5.38
8	KrvskpEW	0.98	0.02	15.8	19.61
9	Lymphography	0.92	0.08	9.2	5.33
10	M-of-n	1.0	0.004	6	5.60
11	PenglungEW	0.96	0.05	130.8	4.58
12	SonarEW	0.96	0.05	31.2	5.44
13	SpectEW	0.88	0.12	8.4	5.71
14	Tic-tac-toe	0.81	0.19	5.2	5.30
15	Vote	0.97	0.03	3.4	5.72
16	WaveformEW	0.80	0.21	14.2	33.53
17	WineEW	1.00	0.009	6	5.62
18	Zoo	1.00	0.008	6.8	5.22
	Average	**0.93**	**0.073**	**15.88**	**7.76**

4) THE BEST FITNESS FUNCTION

Is an indicator to the minimum value of fitness function gained when algorithm run N times, and it is calculated as follows:

$$Best = \min_{k} g_k^* \tag{26}$$

where g_k^* the best fitness value gained at run k.

5) THE WORST FITNESS FUNCTION

Is an indicator to the maximum value of the fitness function gained when algorithm run N times, and it is calculated as follows:

$$Worst = \max_{k} g_K^* \tag{27}$$

where g_k^* the worst fitness (maximum) value gained at run k.

6) AVERAGE COMPUTIONAL TIME

Is an indicator to the average of computational time in seconds gained when algorithm run N times, and it is calculated as follows:

$$AvgCT = \frac{1}{N}\sum_{k=1}^{N} AvgCT^k \tag{28}$$

where $AvgCT^k$ is the value of computational time gained at run k .

D. RESULTS AND DISCUSSION

This section presents and discussed the results of the proposed binary hybrid algorithm.

1) THE COMPLETE RESULTS OF BGWOPSO

Table 2 presents the results of the proposed hybrid BGWOPSO for feature selection after running each algorithm 20 times. Using the proposed method four datasets named Exactly, M-of-N, WineEW, and Zoo had the highest average accuracy rate with 100%. Followed by Breastcancer, CongressEW, and KrvskpEW with average accuracy of 98%. Besides, BreastEW, vote, SonarEW, PenglungEW, and IonosphereEW with average accuracy of 97%, 97%, 96%, 96% and 95% respectively. Moreover, the most reduction of features were in the following datasets: Exactly2, vote, CongressEW, and Breastcancer with number of selected features of 1.6, 3.4, 4.4 and 4.4 respectively. Besides, the less computational time (in second) spent by the proposed approach for feature selection were in the following datasets: PenglungEW, zoo, HeartEW and Tic-tac-toe with 4.58, 5.22, 5.25, 5.30 seconds respectively. The average of classification accuracy, fitness function, reduced feature, and computational time for all 18 datasets using the proposed method are achieved as following: 93% accuracy, 0.073, 15.88 attributes, 7.76 seconds.

Clearly speaking, the good achievements of the proposed BGWOPSO method which combines the strengths of both

TABLE 3. Classification accuracy comparison between the proposed BGWOPSO and related work methods.

Dataset	BGWOPSO	bGWO2	BPSO	BGA	WOASAT-2
Breastcancer	**0.98**	0.97	0.95	0.96	0.97
BreastEW	0.97	0.93	0.94	0.94	**0.98**
CongressEW	**0.98**	0.93	0.94	0.94	**0.98**
Exactly	**1.0**	0.77	0.68	0.67	**1.00**
Exactly2	**0.76**	0.75	0.75	0.76	0.75
HeartEW	**0.85**	0.77	0.78	0.82	**0.85**
IonosphereEW	0.95	0.83	0.84	0.83	**0.96**
KrvskpEW	**0.98**	0.95	0.94	0.92	**0.98**
Lymphography	**0.92**	0.70	0.69	0.71	0.89
M-of-n	**1.0**	0.96	0.86	0.93	**1.00**
PenglungEW	**0.96**	0.58	0.72	0.70	0.94
SonarEW	0.96	0.72	0.74	0.73	**0.97**
SpectEW	**0.88**	0.82	0.77	0.78	0.88
Tic-tac-toe	**0.81**	0.72	0.73	0.71	0.79
Vote	**0.97**	0.92	0.89	0.89	**0.97**
WaveformEW	**0.80**	0.78	0.76	0.77	0.76
WineEW	**1.00**	0.92	0.95	0.93	0.99
Zoo	**1.00**	0.87	0.83	0.88	0.97
Average	**0.93**	0.83	0.82	0.83	0.92

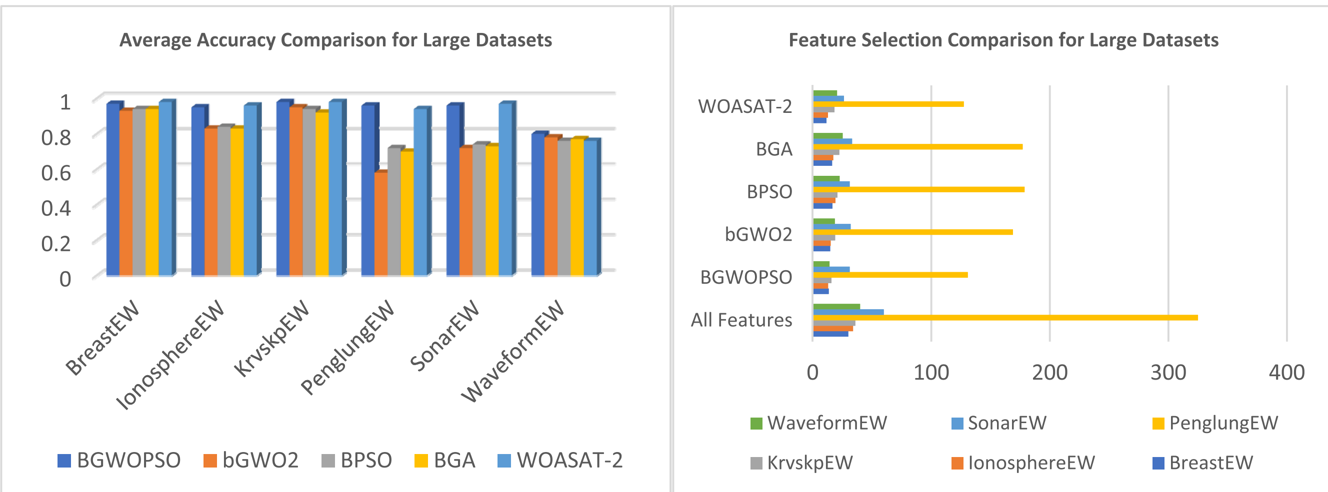

FIGURE 1. Large data comparison in term of average accuracy and number of selected features.

GWO and PSO indicates its ability to control the trade-off between the exploitation and exploration during optimization iterations.

2) COMPARISON OF THE PRPOSED BGWOPSO WITH THE STATE-OF-ART

In this subsection, the results of the proposed method are verified against some of the methods in the literate of feature selection. Table 3 introduced the results of BGWOPSO, bGWO2, BPSO, BGA and WOASAT-2. This table shows that the accuracy of the proposed BGWOPSO is performed better that all other methods on all datasets, excluding three datasets where WOASAT-2 achieves better than other methods with a minor difference from BGWOPSO, and BGWOPSO arises in the second position.

TABLE 4. Average selected features comparison between the proposed BGWOPSO and related work methods.

Dataset	BGWOPSO	bGWO2	BPSO	BGA	WOASAT-2
Breastcancer	4.4	5.2	5.7	5.09	**4.2**
BreastEW	13.6	14.8	16.6	16.35	**11.6**
CongressEW	**4.4**	5.2	6.8	6.62	6.4
Exactly	**6.0**	7.4	9.8	10.82	**6.0**
Exactly2	**1.6**	5.3	6.2	6.18	2.8
HeartEW	5.8	6.8	7.9	9.49	**5.4**
IonosphereEW	13	15.2	19.2	17.31	**12.8**
KrvskpEW	**15.8**	18.9	20.8	22.43	18.4
Lymphography	9.2	8.3	9.0	11.05	**7.2**
M-of-n	**6.0**	7.2	9.1	6.83	**6.0**
PenglungEW	130.8	168.9	178.8	177.13	**127.4**
SonarEW	31.2	32.0	31.2	33.30	**26.4**
SpectEW	**8.4**	11.8	12.5	11.75	9.40
Tic-tac-toe	**5.2**	5.9	6.6	6.85	6.00
Vote	**3.4**	7.4	8.8	6.62	5.20
WaveformEW	**14.2**	18.7	22.7	25.28	20.60
WineEW	**6.0**	7.2	8.4	8.63	6.40
Zoo	6.8	8.3	9.7	10.11	**5.60**
Average	**15.88**	19.69	21.66	21.77	15.99

TABLE 5. Main fitness function comparison between the proposed BGWOPSO and related work methods.

Dataset	BGWOPSO	bGWO2	BPSO	BGA	WOASAT-2
Breastcancer	**0.03**	0.03	0.03	0.03	0.04
BreastEW	0.04	**0.03**	**0.03**	0.04	**0.03**
CongressEW	**0.03**	0.04	0.04	0.04	**0.03**
Exactly	**0.00**	0.22	0.28	0.28	0.01
Exactly2	**0.24**	0.25	0.25	0.25	0.25
HeartEW	0.15	**0.13**	0.15	0.14	0.16
IonosphereEW	0.05	0.08	0.14	0.13	**0.04**
KrvskpEW	**0.02**	0.04	0.05	0.07	**0.02**
Lymphography	**0.08**	0.15	0.19	0.17	0.11
M-of-n	**0.00**	0.04	0.11	0.08	0.01
PenglungEW	**0.05**	0.23	0.22	0.22	0.06
SonarEW	0.05	0.10	0.13	0.13	**0.03**
SpectEW	**0.12**	0.15	0.13	0.14	0.13
Tic-tac-toe	**0.19**	0.23	0.24	0.24	0.21
Vote	**0.03**	**0.03**	0.05	0.05	0.04
WaveformEW	0.21	0.20	0.22	**0.20**	0.25
WineEW	**0.00**	0.01	0.02	0.01	0.01
Zoo	**0.00**	0.11	0.10	0.08	0.04
Total	**1.29**	2.07	2.38	2.3	1.47

TABLE 6. Best fitness function comparison between the proposed BGWOPSO and related work methods.

Dataset	BGWOPSO	bGWO2	BPSO	BGA	WOASAT-2
Breastcancer	**0.02**	0.02	0.03	0.02	0.03
BreastEW	0.03	0.02	0.02	0.02	**0.02**
CongressEW	**0.02**	0.02	0.03	0.03	**0.02**
Exactly	**0.00**	0.07	0.21	0.27	0.01
Exactly2	**0.23**	0.23	**0.22**	**0.22**	0.23
HeartEW	**0.12**	0.10	0.13	0.12	0.13
IonosphereEW	**0.04**	0.07	0.12	0.09	**0.03**
KrvskpEW	**0.02**	**0.02**	0.03	0.03	**0.02**
Lymphography	**0.06**	0.10	0.14	0.12	0.09
M-of-n	**0.00**	**0.00**	0.06	0.02	0.01
PenglungEW	**0.03**	0.17	0.13	0.13	**0.03**
SonarEW	**0.02**	0.07	0.07	0.07	**0.01**
SpectEW	**0.11**	0.12	0.10	0.12	0.11
Tic-tac-toe	**0.17**	0.21	0.21	0.21	0.20
Vote	0.03	**0.00**	0.03	0.03	0.02
WaveformEW	**0.20**	**0.19**	0.21	**0.19**	0.23
WineEW	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Zoo	**0.00**	**0.00**	0.03	**0.00**	**0.00**
Total	**1.1**	1.41	1.77	1.69	1.19

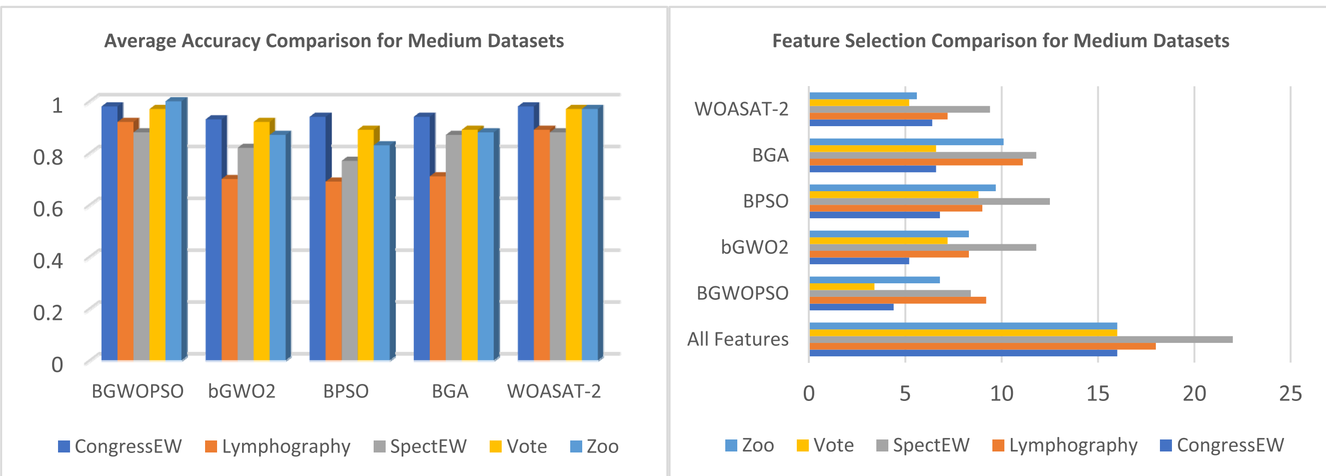

FIGURE 2. Medium data comparison in term of average accuracy and number of selected features.

As per the results in Table 4, the average of selected features using BGWOPSO and other methods are given. The proposed BGWOPSO confirms significantly better results than other methods on the majority of the datasets where BGWOPSO provides average number of selected features of 15.88, while the number of selected features of the WOASAT-2 is 15.99 which is the closest method to our proposed method. Besides, the number of selected features obtained by bGWO2, BPSO, and BGA are 19.69, 21.66, and 21.77 respectively. Inspecting the results in Tables 3 and 4, a considerable disparity may be found in the classification accuracy and the selected features when contrasting BGWOPSO and other methods. We can clearly see that BGWOPSO's performance is superior in selecting fewer features while maintaining its classification in good performance. This demonstrates the BGWOPSO's ability to search

TABLE 7. Worst fitness function comparison between the proposed BGWOPSO and related work methods.

Dataset	BGWOPSO	bGWO2	BPSO	BGA	WOASAT-2
Breastcancer	0.03	0.03	0.03	**0.04**	**0.04**
BreastEW	0.04	0.03	**0.05**	**0.05**	0.04
CongressEW	0.03	0.04	0.04	**0.06**	0.05
Exactly	0.00	0.22	**0.32**	0.31	0.01
Exactly2	0.25	0.25	**0.31**	**0.30**	0.27
HeartEW	0.17	0.13	**0.18**	0.14	**0.18**
IonosphereEW	0.07	0.08	**0.17**	**0.16**	0.05
KrvskpEW	0.03	0.04	0.07	**0.13**	0.02
Lymphography	0.10	0.15	**0.27**	**0.27**	0.14
M-of-n	0.00	0.04	**0.16**	0.15	0.01
PenglungEW	0.08	0.23	**0.29**	**0.29**	0.11
SonarEW	0.07	0.10	0.22	0.23	0.05
SpectEW	0.13	0.15	**0.16**	0.15	**0.15**
Tic-tac-toe	0.20	0.23	**0.27**	0.26	0.23
Vote	0.05	0.03	**0.08**	**0.08**	0.04
WaveformEW	0.22	0.20	0.23	0.21	**0.26**
WineEW	0.01	0.01	**0.03**	**0.03**	**0.03**
Zoo	0.02	0.11	**0.21**	0.18	0.10
Total	**1.5**	2.07	3.09	3.04	1.78

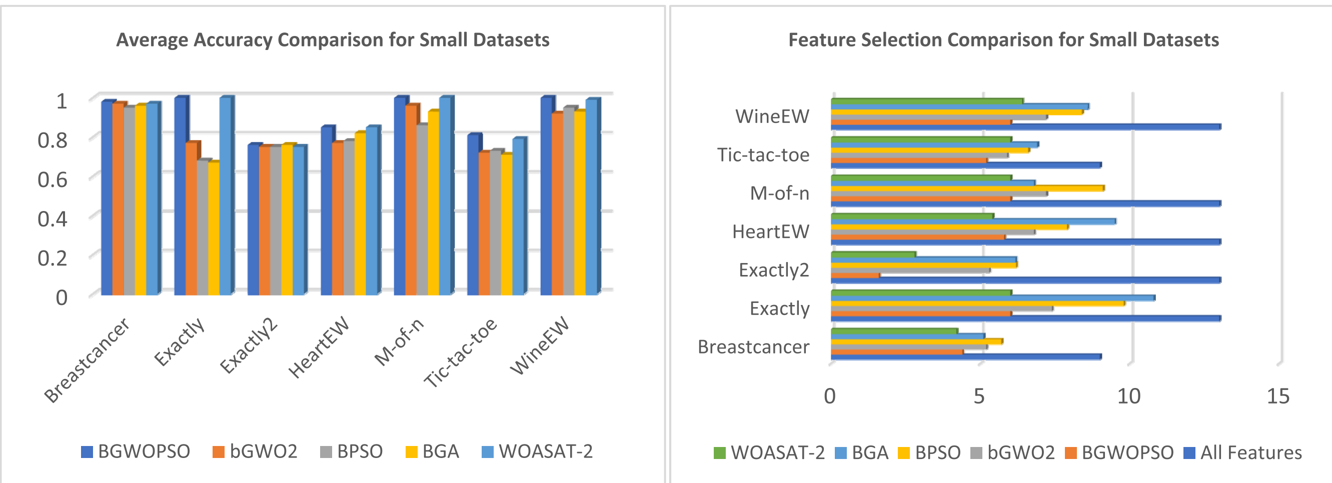

FIGURE 3. Small data comparison in term of average accuracy and number of selected features.

for both optimization objectives and can be regarded as a candidate for the selection of features with reduced number and high classification accuracy.

The three tables labeled as 5, 6, 7 summarized the attained statistical measures for all the data sets based on different runs of the different optimizers. Here, the proposed BGWOPSO method is used in comparison with other methods. As can be seen in these tables, GWOPSO outperforms bGWO2, BPSO, BGA and WOASAT-2 in terms of mean fitness function on thirteen datasets, while in the best fitness function BGWOPSO outperforms all other methods on sixteen datasets. Moreover, the proposed BGWOPSO is not worse than any other methods on all eighteen datasets. The good achievements of the proposed BGWOPSO method indicates its ability to control the trade-off between the exploitation and exploration during optimization iterations.

In addition, as shown in Table 8 the computational time spent on running the proposed approach per second shows

TABLE 8. Comparison between the proposed BGWOPSO with WOASAT-2 in term of computational time (in seconds).

Dataset	BGWOPSO	WOASAT-2
Breastcancer	**5.69**	41.74
BreastEW	**5.38**	44.30
CongressEW	**5.35**	35.67
Exactly	**5.61**	51.79
Exactly2	**5.41**	54.88
HeartEW	**5.25**	29.79
IonosphereEW	**5.38**	30.84
KrvskpEW	**19.61**	589.56
Lymphography	**5.33**	26.17
M-of-n	**5.60**	51.54
PenglungEW	**4.58**	30.49
SonarEW	**5.44**	27.76
SpectEW	**5.71**	31.38
Tic-tac-toe	**5.30**	56.89
Vote	**5.72**	30.79
WaveformEW	**33.53**	1633.27
WineEW	**5.62**	26.33
Zoo	**5.22**	27.02
Total	**139.73**	2820.21

an outstanding performance when compared to the hybrid WOASAT-2. Where the total time spent in running all the datasets using the proposed approach is 139.73 seconds, while the hybrid WOASAT-2 is 2820.21 seconds which shows that BGWOPSO is more reliable in giving superior quality of solutions with reasonable computational iteration. This is due to the fewer parameters of the GWO, the quality of the PSO's velocity and the combination of both strengths.

3) CATEGORIZING THE DATASETS

This subsection categorized the eighteen datasets considered in this work into three categories: Large dataset, where the number of features should be in range of 30 - 350 features. Medium datasets where the number of features should be in range of 16 - 25 features. Small datasets, where the number of features should be in range of 0 - 15 features. Figure 1 shows the high performance of BGWOPSO on large datasets considering both number features selected and accuracy of classification. In addition, Figure 2 illustrates the superiority of the proposed BGWOPSO in medium datasets on both performance of the classification and less number of attributes of features.

Figure 3 shows the superior performance of BGWOPSO in small datasets on both performance measures. According to the results gained, the superiority of the proposed BGWOPSO is verified on the large, medium and small size datasets. We can also see that in terms of the best and worst solution gained, the BGWOPSO performs better than bGWO2, BPSO, BGA and WOASAT-2.

The fewer parameters of the GWO, the quality of the PSO's velocity and the combination of both strengths has demonstrated a good performance of the proposed BGWOPSO algorithm compared to the sate-of- art techniques.

VI. CONCLUSION AND FUTURE WORK

A binary version of an existing algorithm called BGWOPSO was proposed and used to solve the problem of feature selection in this work. To confirm the effectiveness and the efficiency of the proposed method, 18 standard UCI benchmark datasets were employed. A set of evaluation measures were used to assess the proposed method. The proposed hybrid was compared with a number of feature selection algorithms called bGWO2, BPSO, BGA, and the hybrid WOASAT-2.

The results demonstrated the superiority of the proposed method compared to a wide range of algorithms on the majority of datasets in terms of accuracy and number of features selected. Besides, computational time analysis was conducted between the proposed binary hybrid approach and hybrid WOASAT-2 and the results showed that the propose method benefits from a higher execution time. Moreover, a comparison in term of statistical measures Mean, Best, Worst Fitness was conducted, and the results proved much better performance of the proposed approach against the other state-of-art methods. The superior results of the proposed BGWOPSO method indicates its ability to control the trade-off between the exploratory and exploitative behaviors during optimization iterations.

As future works, we recommend employing the proposed hybrid algorithm to solve another real-world problem such as engineering optimization problems, scheduling problems and/or molecular potential energy function. Moreover, the proposed method can be experimented with two other popular classifiers such as SVM and Artificial Neural Network (ANN) which are strong competitors of KNN and see whether performance is stable or varies. Another possible future work is to hybridize the GWO with recent optimization algorithm such as Dragon algorithm (DA).

REFERENCES

[1] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. Amsterdam, The Netherlands: Elsevier, 2011.

[2] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, vol. 454. Springer, 2012.

[3] M. Dash and H. Liu, "Feature selection for classification," *Intell. Data Anal.*, vol. 1, nos. 1–4, pp. 131–156, 1997.

[4] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Jan. 2003.

[5] H. Liu and Z. Zhao, "Manipulating data and dimension reduction methods: Feature selection," in *Encyclopedia Complexity Systems Science*. New York, NY, USA: Springer, 2009, pp. 5348–5359.

[6] H. Liu, H. Motoda, R. Setiono, and Z. Zhao, "Feature selection: An ever evolving frontier in data mining," in *Proc. 4th Workshop Feature Selection Data Mining*, May 2010, pp. 4–13.

[7] A. Zarshenas and K. Suzuki, "Binary coordinate ascent: An efficient optimization technique for feature subset selection for machine learning," *Knowl.-Based Syst.*, vol. 110, pp. 191–201, Oct. 2016.

[8] E.-G. Talbi, *Metaheuristics: From Design to Implementation*, vol. 74, Hoboken, NJ, USA: Wiley, 2009.

[9] C. Lai, M. J. T. Reinders, and L. Wessels, "Random subspace method for multivariate feature selection," *Pattern Recognit. Lett.*, vol. 27, no. 10, pp. 1067–1076, 2006.

[10] H. Liu and H. Motoda, *Feature Extraction, Construction and Selection: A Data Mining Perspective*, vol. 453. Springer, 1998.

[11] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing*, vol. 172, pp. 371–381, Jan. 2016.

[12] Q. Al-Tashi, H. Rais, and S. Jadid, "Feature selection method based on grey wolf optimization for coronary artery disease classification," in *Proc. Int. Conf. Reliable Inf. Commun. Technol.*, 2018, pp. 257–266.

[13] M. M. Kabir, M. Shahjahan, and K. Murase, "A new local search based hybrid genetic algorithm for feature selection," *Neurocomputing*, vol. 74, no. 17, pp. 2914–2928, Oct. 2011.

[14] S. Kashef and H. Nezamabadi-pour, "An advanced ACO algorithm for feature subset selection," *Neurocomputing*, vol. 147, pp. 271–279, Jan. 2015.

[15] R. Bello, Y. Gomez, A. Nowe, and M. M. Garcia, "Two-step particle swarm optimization to solve the feature selection problem," in *Proc. 7th Int. Conf. Intell. Syst. Des. Appl. (ISDA)*, Oct. 2007, pp. 691–696.

[16] E. ZorarpacI and S. A. Özel, "A hybrid approach of differential evolution and artificial bee colony for feature selection," *Expert Syst. Appl.*, vol. 62, pp. 91–103, Nov. 2016.

[17] M. M. Mafarja, D. Eleyan, I. Jaber, A. Hammouri, and S. Mirjalili, "Binary dragonfly algorithm for feature selection," in *Proc. Int. Conf. New Trends Comput. Sci. (ICTCS)*, Oct. 2017, pp. 12–17.

[18] M. M. Mafarja and S. Mirjalili, "Hybrid whale optimization algorithm with simulated annealing for feature selection," *Neurocomputing*, vol. 260, pp. 302–312, Oct. 2017.

[19] N. Singh and S. B. Singh, "Hybrid algorithm of particle swarm optimization and Grey Wolf optimizer for improving convergence performance," *J. Appl. Math.*, vol. 2017, 2017, Art. no. 2030489.

[20] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. ICNN Int. Conf. Neural Netw.*, Dec. 1995, pp. 1942–1948.

[21] J. Kennedy, "The particle swarm: Social adaptation of knowledge," in *Proc. IEEE Int. Conf. Evol. Comput. (ICEC)*, Apr. 1997, pp. 303–308.

[22] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014.

[23] H. Faris, I. Aljarah, M. A. Al-Betar, and S. Mirjalili, "Grey wolf optimizer: A review of recent variants and applications," *Neural Comput. Appl.*, vol. 30, no. 2, pp. 413–435, 2018.

[24] I.-S. Oh, J.-S. Lee, and B.-R. Moon, "Hybrid genetic algorithms for feature selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1424–1437, Nov. 2004.

[25] X. Lai and M. Zhang, "An efficient ensemble of GA and PSO for real function optimization," in *Proc. 2nd IEEE Int. Conf. Comput. Sci. Inf. Technol.*, Aug. 2009, pp. 651–655.

[26] B. Niu and L. Li, "A novel PSO-DE-based hybrid algorithm for global optimization," in *Proc. Int. Conf. Intell. Comput.*, 2008, pp. 156–163.

[27] S. Mirjalili and S. Z. M. Hashim, "A new hybrid PSOGSA algorithm for function optimization," in *Proc. Int. Conf. Comput. Inf. Appl.*, Dec. 2010, pp. 374–377.

[28] S. M. Abd-Elazim and E. S. Ali, "A hybrid particle swarm optimization and bacterial foraging for power system stability enhancement," *Complexity*, vol. 21, no. 2, pp. 245–255, 2015.

[29] A. Zhu, C. Xu, Z. Li, J. Wu, and Z. Liu, "Hybridizing grey wolf optimization with differential evolution for global optimization and test scheduling for 3D stacked SoC," *J. Syst. Eng. Electron.*, vol. 26, no. 2, pp. 317–328, Apr. 2015.

[30] D. Jitkongchuen, "A hybrid differential evolution with grey wolf optimizer for continuous global optimization," in *Proc. 7th Int. Conf. Inf. Technol. Elect. Eng. (ICITEE)*, Oct. 2015, pp. 51–54.

[31] M. A. Tawhid and A. F. Ali, "A Hybrid grey wolf optimizer and genetic algorithm for minimizing potential energy function," *Memetic Comput.*, vol. 9, no. 4, pp. 347–359, 2017.

[32] P. J. Gaidhane and M. J. Nigam, "A hybrid grey wolf optimizer and artificial bee colony algorithm for enhancing the performance of complex systems," *J. Comput. Sci.*, vol. 27, pp. 284–302, Jul. 2018.

[33] N. Singh and S. B. Singh, "A novel hybrid GWO-SCA approach for optimization problems," *Eng. Sci. Technol. Int. J.*, vol. 20, no. 6, pp. 1586–1601, Dec. 2017.

[34] M. Mafarja and S. Abdullah, "Investigating memetic algorithm in solving rough set attribute reduction," *Int. J. Comput. Appl. Technol.*, vol. 48, no. 3, pp. 195–202, Oct. 2013.

[35] R. Azmi, B. Pishgoo, N. Norozi, M. Koohzadi, and F. Baesi, "A hybrid GA and SA algorithms for feature selection in recognition of hand-printed Farsi characters," in *Proc. IEEE Int. Conf. Intell. Comput. Intell. Syst.*, Oct. 2010, pp. 384–387.

[36] P. Moradi and M. Gholampour, "A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy," *Appl. Soft Comput.*, vol. 43, pp. 117–130, Jul. 2016.

[37] E.-G. Talbi, L. Jourdan, J. Garcia-Nieto, and E. Alba, "Comparison of population based metaheuristics for feature selection: Application to microarray data classification," in *Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl.*, Mar./Apr. 2008, pp. 45–52.

[38] Z. Yong, G. Dun-wei, and Z. Wan-qiu, "Feature selection of unreliable data using an improved multi-objective PSO algorithm," *Neurocomputing*, vol. 171, pp. 1281–1290, Jan. 2016.

[39] J. Jona and N. Nagaveni, "A hybrid swarm optimization approach for feature set reduction in digital mammograms," *WSEAS Trans. Inf. Sci. Appl.*, vol. 9, pp. 340–349, 2012.

[40] M. E. Basiri and S. Nemati, "A novel hybrid ACO-GA algorithm for text feature selection," in *Proc. IEEE Congr. Evol. Comput.*, May 2009, pp. 2561–2568.

[41] R. S. Babatunde, S. O. Olabiyisi, and E. O. Omidiora, "Feature dimensionality reduction using a dual level metaheuristic algorithm," *Optimization*, vol. 7, no. 1, pp. 49–52, 2014.

[42] J. B. Jona and N. Nagaveni, "Ant-cuckoo colony optimization for feature selection in digital mammogram," *Pakistan J. Biol. Sci.*, vol. 17, no. 2, pp. 266–271, 2014.

[43] M. Nekkaa and D. Boughaci, "Hybrid harmony search combined with stochastic local search for feature selection," *Neural Process. Lett.*, vol. 44, no. 1, pp. 199–220, 2016.

[44] H. M. Zawbaa, E. Emary, C. Grosan, and V. Snasel, "Large-dimensionality small-instance set feature selection: A hybrid bio-inspired heuristic approach," *Swarm Evol. Comput.*, vol. 42, pp. 29–42, Oct. 2018.

[45] Q. Al-Tashi, H. Rais, and S. J. Abdulkadir, "Hybrid swarm intelligence algorithms with ensemble machine learning for medical diagnosis," in *Proc. 4th Int. Conf. Comput. Inf. Sci. (ICCOINS)*, Aug. 2018, pp. 1–6.

[46] N. S. Altman, "An introduction to kernel and nearest-neighbor non-parametric regression," *Amer. Statist.*, vol. 46, no. 3, pp. 175–185, 1992.

[47] C. L. Blake, "CJ Merz UCI repository of machine learning databases," Ph.D. dissertations, Dept. Inform. Comput. Sci., Univ. California, Irvine, CA, USA, 1998.

QASEM AL-TASHI received the B.Sc. degree in software engineering from Universiti Teknologi Malaysia, in 2012, and the M.Sc. degree in software engineering from Universiti Kebangsaan Malaysia, in 2017. He is currently pursuing the Ph.D. degree with Universiti Teknologi PETRONAS. He is currently an Academic Staff with Albaydaa University, Yemen. His research interests include machine learning, multi-objective optimization, and feature selection and classification.

SAID JADID ABDUL KADIR received the B.Sc. degree in computer science from Moi University, the M.Sc. degree in computer science from Universiti Teknologi Malaysia, in 2012, and the Ph.D. degree in information technology from Universiti Teknologi PETRONAS, in 2016, where he is currently a Senior Lecturer with the Department of Computer and Information Sciences. His current research interests include machine learning, data clustering, and predictive analytics. He is currently serving as a Journal Reviewer for *Artificial Intelligence Review*, the IEEE ACCESS, and *Knowledge-Based Systems*.

HELMI MD RAIS received the B.Sc. degree of science in business administration from Drexel University, USA, the M.Sc. degree in information technology from Griffith University, Australia, and the Ph.D. degree in science and system management from Universiti Kebangsaan Malaysia. He is currently a Senior Lecturer with Universiti Teknologi PETRONAS. His research interests include ant colony algorithm, optimization, swarm intelligence, and database technologies.

SEYEDALI MIRJALILI is currently a Lecturer with Griffith University and internationally recognized for his advances in swarm intelligence (SI) and optimization, including the first set of SI techniques from a synthetic intelligence standpoint—a radical departure from how natural systems are typically understood—and a systematic design framework to reliably benchmark, evaluate, and propose computationally cheap robust optimisation algorithms. He has published more than 100 journal articles, many in high-impact journals with more than 8000 citations in total with an H-index of 33 and G-index of 90. From Google Scholar metrics, he is globally the third most cited researcher in Engineering Optimization and Robust Optimization. He is serving an Associate Editor of *Advances in Engineering Software* and the *Journal of Algorithms*.

HITHAM ALHUSSIAN received the B.Sc. and M.Sc. degrees in computer science from the School of Mathematical Sciences, University of Khartoum, Sudan, and the Ph.D. degree from Universiti Teknologi PETRONAS (UTP), Malaysia, where he was a Postdoctoral Researcher with the High-Performance Cloud Computing Center and he is currently a Senior Lecturer with the Department of Computer and Information Sciences. His current research interests include real-time parallel and distributed systems, big data and cloud computing, data mining, and real-time analytics.