I.J. Modern Education and Computer Science, **2018, 7, 37-49**
Published Online July 2018 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2018.07.04

Dimensionality Reduction Using an Improved Whale Optimization Algorithm for Data Classification

Ah. E. Hegazy
Faculty of Computers & Informatics, Suez Canal University, Ismailia, 41511, Egypt.
Email: ahmedhegazy281@yahoo.com

M. A. Makhlouf and Gh. S. El-Tawel
Faculty of Computers & Informatics, Suez Canal University, Ismailia, 41511, Egypt.
Email: {m.abdallah, ghada}@{ci.suez.edu.eg}

Received: 01 May 2018; Accepted: 18 June 2018; Published: 08 July 2018

***Abstract*—**Whale optimization algorithm is a newly proposed bio-inspired optimization technique introduced in 2016 which imitates the hunting demeanor of hump-back whales. In this paper, to enhance solution accuracy, reliability and convergence speed, we have introduced some modifications on the basic WOA structure. First, a new control parameter, inertia weight, is proposed to tune the impact on the present best solution, and an improved whale optimization algorithm (IWOA) is obtained. Second, we assess IWOA with various transfer functions to convert continuous solutions to binary ones. The pro-posed algorithm incorporated with the K-nearest neighbor classifier as a feature selection method for identifying feature subset that enhancing the classification accuracy and limiting the size of selected features. The proposed algorithm was compared with binary versions of the basic whale optimization algorithm, particle swarm optimization, genetic algorithm, antlion optimizer and grey wolf optimizer on 27 common UCI datasets. Optimization results demonstrate that the proposed IWOA not only significantly enhances the basic whale optimization algorithm but also performs much superior to the other algorithms.

***Index Terms*—**Feature Selection, Whale Optimization Algorithm, Bio-inspired Optimization, Classification.

I. INTRODUCTION

In numerous data mining problems, datasets hold an extensive number of unessential or repetitive features which may reduce the classification accuracy and increase the dimensionality of datasets [1]. Feature selection aims to discover the most illustrative set of features by eliminating pointless/repetitive features for the classification procedure, diminishing the classification error ratio and furthermore the training time for some datasets. Feature selection can be used to enhance the classifier accuracy and acquire equivalent or even best classification performance than utilizing the entire list of features [2, 3]. The feature selection can be considered as combinatorial optimization problem, where the best data fitting depends on the selected subset of features [4]. In real-world applications, feature selection is compulsory as the datasets hold noisy, insignificant or misdirecting features which negatively affect the classification accuracy during the learning procedure [5, 6]. A regular feature selection procedure comprises of two main steps:

- Subset generation: the search strategy used to produce feature subsets for assessment.
- Subset assessment: the measure used to weigh the goodness of feature subsets.

Feature selection strategies extensively fall into two classes: filter-based strategies and wrapper-based strategies. The filter-based type (uses statistical measures) [7] depends on general qualities of the data for assessing and identifying feature subsets without including any mining technique. The wrapper-based type (apply data mining techniques) requires the mining technique and utilizes the classifier performance as the assessment measure (uses the classifier as a black box for evaluating the sets of features in light of their classification accuracy). Wrappers produce better accuracy than filter strategies because of the cooperation between the classifier and the chosen subset of features throughout the selection task [8, 9].

Generally, the feature selection task has two goals: limit the number of the identified feature subset and enhance the classification performance. The size of the search space exponentially increments regarding the number of attributes in the dataset [1]. If the dataset contains M number of features, there are 2M possible subsets of features, thus a complete search to discover ideal solution is unrealistic unless M is little. An extensive variety of search procedures can be utilized, such as sequential backward selection [10] and sequential forward selection [11]. However, these methods still suffer from trapping in local optima and costly computational time [12]. In the previous few decades, many global optimization techniques have been produced that are dependent on the nature-inspired analogy [13]. Nature-inspired meta-heuristic techniques simulate physical or biological behavior. As shown in Figure 1, they can be assembled in three basic classes: evolution-based (mimic the natural evolution laws), physics-based (mimic the physical guidelines in the world) and swarm-based techniques (imitate the social demeanor of animals). Genetic Algorithm (GA) is the most prevalent evolution-inspired algorithm. Meta-heuristic techniques can be utilized in the feature selection domain to choose the optimal subset; such techniques include particle swarm optimization, ant colony optimization [14] and genetic algorithms [15].

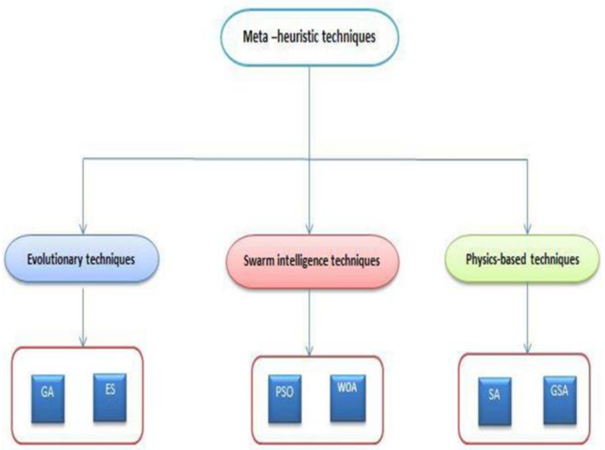

Fig.1. Classification of meta-heuristic techniques.

Exhaustive search is too time-consuming and impractical for solving most real-world problems. Therefore, meta-heuristic algorithms were developed that tried to solve these problems approximately in a reasonable amount of time. PSO [16] and GAs [17] are the most well-known population-based techniques. When utilizing any bio-inspired technique for feature selection using the binary encoding, the representation of every solution is an n-bit binary string. Every agent position is Boolean, where 1 indicates that this feature will be chosen and 0 otherwise.

Lately, [18] suggested an optimization nature-inspired technique (namely, Whale Optimization Algorithm (WOA)), mimicking the hunting demeanor of humpback whales. In this work, an improved version of WOA for feature selection called IWOA is introduced. The purpose of IWOA is to pick a small number of features and acquiring similar or even best classification performance from utilizing all features. In IWOA every search agent is related to a series of binary values that signify whether a feature will be selected or not. The objective function here is the classifier accuracy. The proposed approach has been contrasted against five recent wrapper feature selection methods over 27 datasets. Results showed that IWOA outperforms the other approaches. The fundamental concentration of this work is to utilize the binary whale algorithm for feature selection for choosing good features and enhancing the classification performance of selected features. We are especially keen on applying our model to data with an extensive number of features and have a little number of samples, which causes the feature selection task more complex.

The organization of the rest of this paper is as follows. Section II presents the related work. Section III presents the proposed methods, while the optimization results with discussions are stated in Section IV. Section V concludes the paper.

II. RELATED WORK

In the last years, numerous techniques for feature selection have been introduced. Comparing to the prior research where mostly feature selection depends on the filter-type, late advances trend towards utilizing wrapper strategies where the classification accuracy used to control the feature selection. Different nature-inspired techniques such as GA, ACO and PSO are utilized to produce the best solution. A variety of swarm algorithms have been utilized to solve feature selection problems. [19] Presents a comprehensive survey of the state-of-the-art work on nature-inspired algorithms for feature selection. GA utilizes the classifier accuracy as a fitness function and removes or adds features based on the ranking measure. A fuzzy set used as a fitness function in GA for feature selection has been introduced in [20]. The same fitness function with PSO produces better performance than GA in [21]. Genetic programming for feature selection has been introduced in [22].

Several researches utilized binary bio-inspired techniques for feature selection. Based on particle swarm optimization [23] suggested a wrapper method for feature selection named CBPSO in-tended to enhance the classification accuracy assessed by the k-nearest neighbor. Experiments indicated that the tent map in CBPSO got higher accuracy over a logistic map. [24] Used the advanced ACO for the feature selection domain. Features are dealt with as nodes to build a graph model. The experimental comparison checked that ABACO has good classification performance utilizing a little set of features than the original ACO approach. In [25] rough-sets used as a fitness function consolidated with the bat algorithm for feature selection. The utilized fitness function guarantees enhanced classification accuracy and obtains a minimal feature size.

[26] Used the Multi-Objective GA for feature selection. The results affirmed that the proposed algorithm can determine diverse ideal feature subsets and achieve good classification accuracy. In [27] the accuracy of Optimum-Path Forest Classifier is utilized as fitness function with the cuckoo search technique for feature selection. The experiments demonstrated the proposed model outperforms three other approaches BBA, BFA and BPSO. The results affirmed that cuckoo algorithm has great abilities to discover the best set on two datasets.

A binary grey wolf algorithm (BGWO) is utilized for feature selection task in [28]. Results demonstrated the ability of the proposed bGWO to find ideal feature sets. A feature selection based on a binary version of the krill herd algorithm (BKH) has been introduced in [29]. The experiments demonstrated the proposed BKH beats three other approaches BFA, BHS and BPSO over six datasets. In [30] a binary bat algorithm (BBA) with the Optimum-Path Forest classifier is utilized for feature selection area. Tests directed at five datasets have exhibited that the BBA can beat other algorithms such as PSO, HS, FFA and GSA. The system suggested for [31] utilized a binary ant lion algorithm (BALO) for feature selection. Results demonstrated the capacity of the proposed algorithm to scan the search space for ideal feature subsets.

III. Methods

A. Whale Optimization Algorithm (WOA)

Whale Optimization Algorithm (WOA) is a newly introduced swarm-based algorithm that was proposed by Seyedali Mirjalili and Andrew Lewis [18], which imitates the hunting procedure of humpback whales.

The mathematical model for WOA is given as follows:

1. Encircling Prey: Whales can chase a prey by encircling it. This demeanor is mathematically represented by the following Equations (1), (2), (3) and (4).

$$\vec{Y}(i+1) = \vec{Y}^*(i) - \vec{A}.\vec{Z} \tag{1}$$

$$\vec{Z} = \left|\vec{C}.\vec{Y}^*(i) - \vec{Y}(i)\right| \tag{2}$$

Where i is the present iteration, $\vec{C}$ and $\vec{A}$ are coefficient vectors, Y^* is the position vector of the optimal solution gained so far, $\vec{Y}$ is the position vector, | | is the absolute value and **.** is an element-by-element multiplication. The position vector Y^* is redesigned in every iteration if there exists a superior better solution.

Coefficient vectors and are calculated as follows:

$$\vec{C} = 2.\vec{v} \tag{3}$$

$$\vec{A} = 2\vec{a}.\vec{v} - \vec{a} \tag{4}$$

Where $\vec{a}$ is linearly reduced from 2 to 0 through iterations and $\vec{v}$ is random vector in [0, 1].

2. Bubble-net attacking procedure (exploitation stage): In this procedure, two strategies are demonstrated as follows:

- Shrinking encircling strategy: This demeanor is accomplished by reducing the value of $\vec{a}$ from 2 to 0. Random value for vector $\vec{A}$ in [-1, 1].

- Spiral updating position: The mathematical equation of this demeanor given as follows:

$$\vec{Y}(i+1) = \vec{Z}^{'}.e^{bl}.cos(2\pi l) + \vec{Y}^*(i) \tag{5}$$

Where $\vec{Z}^{'} = |\vec{Y}^*(i) - \vec{Y}(i)|$ and demonstrates the distance of the i^{th} whale to the prey (best solution gained so far), b is constant determines the logarithmic shape, l is a random number in [-1, 1].

Note: To change whale's position, there is 50-50% probability that whale either apply shrinking encircling or spiral strategy. Mathematically we demonstrated as follows:

$$\vec{Y}(i+1) = \begin{cases} \vec{Y}^*(i) - \vec{A}.\vec{Z}, & \text{if} \quad r < 0.5 \\ \vec{Z}^{'}.e^{bl}.cos(2\pi l) + \vec{Y}^*(i), & \text{if} \quad r \geq 0.5 \end{cases} \tag{6}$$

Where r is random value in [0, 1].

3. Search for prey (exploration stage): The vector $\vec{A}$ can be used for exploration to scan for prey; $\vec{A}$ takes the values greater than 1 or less than -1.

$$\vec{Z} = \left|\vec{C}.\vec{Y}_{rand} - \vec{Y}\right| \tag{7}$$

$$\vec{Y}(i+1) = \vec{Y}_{rand} - \vec{A}\vec{Z} \tag{8}$$

Where $\vec{Y}_{rand}$ is random position vector picked from the present population.

Algorithm 1 presents the whale optimization algorithm (WOA).

Input: whales *W*, iterations *T*.
Output: *Y** Optimal whale binary position, *f(Y)* Best fitness value.
1. Initialize a population of *W* whale's positions randomly.
2. Compute the fitness of every whale.
3. Locate the fittest search agent as *Y**.
4. while Stopping criteria not reached do
For each whale do
  Update a, C, A, l, and r
  If1 $r < 0.5$
    If2 $|A| < 1$
    Update the present whale position by Equation (1)
    Else If2 $|A| >= 1$
    Select a random search agent (Y^{rand})
    Update the present whale position by Equation (8)
    End If2
  Else If1 $r >= 0.5$
    Update the present whale position by Equation (5)
  End If1
End For
Compute the fitness of every whale.
Update *Y** if there is a superior solution.
End while
Return the best solution *Y** and its fitness value f(Y^*).

B. Proposed Improved Whale Optimization Algorithm (IWOA)

In WOA, whales continuously update their positions to whatever point in the search space. For the feature selection task, the solutions are confined to binary {0, 1} values which encourage applying a binary version of the WOA. As indicated by [18], each whale position is updated by Equations 1, 5, 8. In WOA, the updated solution is generally depended on the present best solution. Like PSO algorithm, an inertia weight ω∈ [0, 1] is introduced into WOA to get the improved binary whale optimization algorithm (IWOA). In this paper, we used a binary version of the improved whale optimization algorithm (IWOA) for the feature selection tasks. The problem here is to choose or not a given feature, each solution is a binary vector, where 1 indicates that a feature will be selected and 0 otherwise. The improved algorithm is stated by the following equations:

$$\vec{Y}(i+1) = \omega\vec{Y}^*(i) - \vec{A}.\vec{Z} \tag{9}$$

$$\vec{Z} = \left|\vec{C}.\omega\vec{Y}^*(i) - \vec{Y}(i)\right| \tag{10}$$

$$\vec{Y}(i+1) = \vec{Z}^{'}.e^{bl}.cos(2\pi l) + \omega\vec{Y}^*(i) \tag{11}$$

$$\vec{Y}(i+1) = \begin{cases} \omega\vec{Y}^*(i) - \vec{A}.\vec{Z}, & \text{if} \quad r < 0.5 \\ \vec{Z}^{'}.e^{bl}.cos(2\pi l) + \omega\vec{Y}^*(i), & \text{if} \quad r \geq 0.5 \end{cases}$$

We also evaluate how the IWOA work with various transfer functions for feature selection task using a binary, sigmoid and hyperbolic tangent functions to map the continuous values to binary ones. In this work, there is an M solution (agents) each solution is a one-dimensional vector that holds N elements, where N is the total number of features in the original dataset. Each cell in the vector has a value of ”1” (selected) or ”0” (not selected).

- Binary

$$Y_{MN} = \begin{cases} 1 & \text{if} Y_{MN} > 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

Where Y_{MN} is the dimension value for search agent M at dimension N.

- Sigmoid

$$S(Y_{MN}) = \frac{1}{1+e^{-Y_{MN}}} \tag{14}$$

$$Y_{MN} = \begin{cases} 1 & \text{if} S(Y_{MN}) > \sigma \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

Where $S(Y_{MN})$ is a sigmoid function to normalize the value of Y_{MN} into [0, 1] and σ is a random value in [0, 1].

- Hyperbolic Tangent

$$Y_{MN} = \left|\tanh(Y_{MN})\right| \tag{16}$$

Note that all optimizers (IWOA, WOA, PSO, GA, ALO, and GWO) used the same solution representation method (binary encoding) by utilizing the sigmoid function to map continuous values into binary ones.

C. The Improved Whale Optimization Algorithm for Feature Selection

In this section, the improved binary whale optimizer is used in feature selection for classification purposes. For a feature set sized N, the diverse feature subsets would be 2N which is an enormous space of features to be searched exhaustively. So, IWOA is utilized to scan adaptively the search space for optimal feature subset. The ideal feature set is the one with least classification error and least number of identified features. The fitness function is utilized in IWOA to assess individual agents is defined in Equation 17.

$$f = \alpha * (1-C) + (1-\alpha) * \frac{F}{T} \tag{17}$$

Where C is the classifier accuracy of the identified subset, α is a constant to control the classification accuracy and the feature reduction, F is the size of identified feature subset, T is the total number of features and in [0, 1]. In this work α = 0.9.

We used a wrapper feature selection mode by applying the KNN classifier as a fitness function. Note the proposed algorithm can be utilized with any other classifier. The K-nearest neighbor (KNN) is a simple and very common classifier. In this work, the KNN is utilized as a classification algorithm to assess the quality of the selected subset of features. The proposed model comprises of two fundamental stages; features selection and classification. First, the model begins from taking the dataset as input, then IWOA bio-inspired algorithm is received to choose good features, then the resulted features are utilized to feed KNN classifier. Finally, the results are assessed.

- **Features Selection Stage**: In this paper, IWOA algorithm is utilized as feature selection method in a wrapper mode. At the end of every run, the best solution is the subset which gives the minimum fitness value.
- **Classification Stage**: In this paper, KNN is the used classifier and cross-validation is the used method to test the robustness of the proposed algorithm.

The procedure for describing proposed IWOA-KNN is as follows:

1. Initialize the parameters of IWOA such as number of whales (W), Dataset Z, number of iterations (T), optimal whale position (Y^*) and best fitness value ($f(Y^*)$).
2. Divide Dataset Z into a training set Z_1 and Test set Z_2.
3. Initialize a population of W whale's positions randomly.
4. Compute the fitness of every whale on the Training set Z_1 using Equation 17.
5. Set number of iteration (t) to 0.
6. Update the position of each whale through Equations 9 to 12.
7. Calculate the classification accuracy of the selected feature subset on the Test set Z_2.
8. Increment t to 1.
9. Repeat Steps 6 to 8 until t=T is satisfied.
10. Return the best solution Y^* (optimal subset of features) and its fitness value $f(Y^*)$.

The proposed IWOA algorithm is schematically presented in Figure 2.

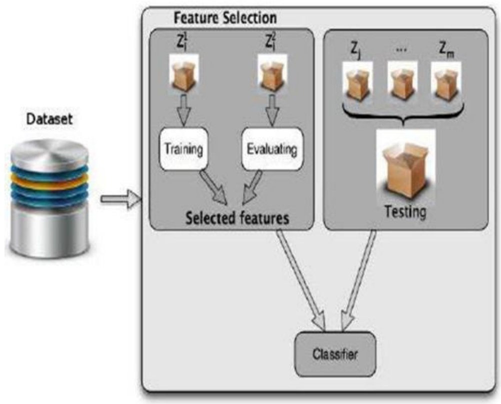

Fig.2. Pipeline of the proposed algorithm [29]

IV. Experimental Results and Discussions

A. Data Description

The improved binary whale optimization algorithm (IWOA) was utilized to choose the ideal feature set to enhance classification accuracy and limit the length of identified features. In this work, 27 datasets from the UCI archive [32] are used to weigh the proposed algorithm. We choose these datasets based on the following terms:

- These data sets are UCI standard real data sets (it is on the machine learning website) is that almost all the result of recent studies in the field of data mining in the world are reported using these data sets.
- These datasets hold various numbers of attributes, extending from 6 to 856. Most of them have a vast number of attributes, so they are suitable for the feature selection process.
- These datasets are likewise various as far as the number of classes (from 2 to 16) and instances (from 32 to 5000).

Table 1. Datasets Description

Category	DS No.	Dataset	# Features	# Samples	# Classes
Small < 20	1	Wine	13	178	3
	2	Zoo	16	101	7
	3	Hepatitis	19	155	2
	4	Fertility	9	100	2
	5	Vehicle	18	846	4
	6	Heart	13	270	2
	7	Ecoli	7	336	8
	8	Liver	6	345	2
	9	Diabetes	8	768	2
	10	Breastcancer	9	699	2
Medium [20-100]	11	Ionosphere	34	351	2
	12	Lung Cancer	56	32	3
	13	Dermatology	34	366	6
	14	Sonar	60	208	2
	15	BreastEW	30	569	2
	16	Soybean Small	35	47	4
	17	Movementlibras	90	360	15
	18	Parkinsons	22	195	2
	19	Spambase	57	4601	2
	20	Waveform	40	5000	3
	21	Hillvalley	100	606	2
Large >100	22	Arrhythmia	279	452	16
	23	Multiple Features	649	2000	10
	24	Semeion	256	1593	10
	25	Clean	166	476	2
	26	CNAE	856	1080	9
	27	DNA	180	2000	3

Details of the used datasets are presented in Table 1. In addition, we chose a set of high dimensional datasets to guarantee the performance of algorithms in large feature spaces. The experiments were implemented in MATLAB-R2015a on a pc with windows 8, AMD A10 CPU 2.30 GHz and 4GB memory.

In each dataset, the samples are randomly partitioned into three diverse equivalent segments namely training, validation and testing parts via cross-validation approach. The training part is utilized to train the KNN while the validation part is utilized to measure the KNN accuracy and is used inside the fitness function. The test part is kept covered up for both the optimization algorithm and the classifier to test the finally selected features given the KNN classifier and is let for last evaluation. The best choice of the value of K in KNN (k= 5) is based on trial and error rule as the best performing on all datasets. The running time increments when changing to another classifier, such as random forest or support vector machine. The proposed method is compared with WOA, PSO, GA, ALO and GWO for evaluation. All the used parameters are presented in Table 2 and Table 3. For a fair comparison, the best choice of the value of ω in IWOA is 0.7298 as in PSO. All the parameters are set based on domain specific-knowledge as the parameter in the fitness function, or by trial and error such as the remainder of parameters. PSO parameters values are chosen based on the common settings in [19].

Table 2. Global Parameters Setting

Parameter	Value	Meaning
α	0.9999	Fitness function constant
Niters	100	Max number of iterations
NAgents	30	Number of search agents used in the optimization
NRuns	20	The number of runs
Problem Dimension		Number of features in the dataset
Search Domain		Binary vector [0 1]
K	5	K-value in KNN

Table 3. Individual Optimizer Parameters Setting

Parameter	Value	Meaning
PSO parameters		
ω	0.72980	Inertia factor
c1 , c2	1.49618	Individual-best acceleration factor
GA parameters		
Cross_Val	0.9	Crossover Fraction
Mut_Val	0.1	Mutation Fraction

B. Performance Metrics

Each optimizer has been executed 20 times with random values for the whale's positions to guarantee steadiness and statistical significance of the results and to test the convergence ability. The measures (calculated in every run) used to compare the optimizers are as follows:

- Classification average accuracy: is the classifier accuracy on the resulted set of features.
- Statistical best fitness: is the minimum value of fitness given by an optimizer
- Statistical worst fitness: is the maximum value of fitness given by an optimizer
- Statistical mean fitness: is the average value of fitness given by an optimizer
- STD: is utilized as a marker of the optimizer stability and robustness.
- Average selection size: is the average number of chosen features to the aggregate number of features.
- Average execution time: is the run time for a given optimizer in seconds.

Algorithms used for comparison: our comparisons incorporate the following algorithms:

- WOA: standard whale optimization algorithm
- PSO: particle swarm optimization
- GA: genetic algorithm
- ALO: antlion optimization algorithm
- GWO: grey wolf optimization algorithm

C. Numerical Results and Discussion

Tables 4, 5 and 6 present the performance of all optimizers utilizing fitness function defined in Equation 17 in a minimization type. These tables present the statistical best, mean and worst fitness values acquired over all runs. For each dataset, the best values are appeared in boldface. We can comment that the best performance is accomplished by the proposed IWOA in the acquired fitness values, which demonstrates the ability of the IWOA for scanning the feature space adaptively superior to the other methods. We can remark that the enhanced parameter helps WOA to enhance its obtained solutions by providing a good balance between exploration and exploitation capabilities.

Table 7 summarizes the results for the number of the selected features. We can observe that IWOA, while outperforming all other methods in classification accuracy, IWOA has selected minimal features than all other techniques considering the entire collection of datasets. So IWOA can be considered as a contender for selecting least number of features with better performance. Concerning Hillvalley dataset, for instance, IWOA selected 187.40% less features than WOA, which has been the second most exact method in this dataset. For DNA dataset, IWOA selected 3.75% less features than WOA, which has been the second most exact method.

The results for the classification accuracy exhibited in Table 8 demonstrate that IWOA gets the best results for 10 of the datasets, showing the ability of IWOA to discover ideal feature sets guaranteeing good test accuracy on the test data, and subsequently it can be utilized as a contender for feature selection. Table 9 outlines the average execution time of all optimizers. All optimizers utilizing the same number of iterations. In Table 9, the GA has the best computational time in contrast with the other optimizers. We can observe GA has been the fastest technique in 9 datasets, trailed by WOA in 7 datasets and IWOA in 6 datasets. Thus, if we consider the best trade-off among classification accuracy, feature reduction and computational time, the best decision depends on IWOA.

Table 4. Best (Min) Fitness Values for all Algorithms after 20 Runs

DS.	IWOA	WOA	PSO	GA	ALO	GWO
Wine	**0.0116**	0.0253	0.0610	0.0388	0.0310	0.0410
Zoo	0.0951	**0.0445**	0.1002	0.0617	0.1772	0.1006
Hepatitis	**0.0125**	0.0787	0.1158	0.1555	0.1290	0.1954
Fertility	**0.0825**	**0.0825**	0.1312	**0.0825**	0.1199	0.1397
Ecoli	**0.1118**	0.1323	0.1145	0.1441	0.1427	0.1604
Vehicle	0.2591	0.2864	0.2630	0.2542	0.2923	**0.2535**
Heart	**0.1219**	0.1235	0.1513	0.1717	0.1513	0.1293
Liver	**0.1934**	0.3076	0.3083	0.3369	0.2952	0.2545
Diabetes	0.2266	**0.2014**	0.2525	0.2448	0.2434	0.2473
Breastcancer	0.0188	**0.0186**	0.0299	0.0321	0.0399	0.0395
Ionosphere	**0.0721**	0.0808	0.1083	0.0968	0.1166	0.1412
Lung Cancer	0.1140	0.1275	0.1277	**0.064**	0.1887	0.2530
Dermatology	**0.0084**	**0.0084**	0.0209	0.0089	0.0185	0.0221
Sonar	0.1149	**0.0701**	0.1471	0.1851	0.1652	0.2132
BreastEW	**0.0411**	0.0415	0.0538	0.0496	0.0548	0.0652
Soybean Small	0.0057	**0.0006**	0.0011	0.0006	0.0017	0.1943
Movementlibras	0.3408	0.2558	0.3448	**0.2344**	0.3125	0.2700
Parkinsons	0.0625	**0.0624**	0.0946	0.1028	0.0932	0.1033
Spambase	0.0712	**0.0683**	0.0827	0.0904	0.0814	0.0794
Waveform	**0.1495**	0.1519	0.1772	0.1572	0.1589	0.1596
Hillvalley	**0.4395**	0.3762	0.4057	0.4321	0.3987	0.3891
Arrhythmia	**0.2132**	0.2709	0.3067	0.3493	0.3634	0.3526
Multiple Features	**0.0248**	0.0273	0.0404	0.0419	0.0432	0.0441
Semeion	**0.0305**	0.0675	0.0893	0.0867	0.0954	0.0824
Clean	**0.0177**	0.0540	0.1497	0.1034	0.1435	0.1696
CNAE	**0.0014**	0.0061	0.1142	0.0091	0.1119	0.0041
DNA	0.1114	**0.1062**	0.1345	0.1747	0.1125	0.1421
Average	**0.1093**	0.1139	0.1454	0.1374	0.1512	0.1572

Table 5. Mean (Average) Fitness Values for all Algorithms after 20 runs

DS.	IWOA	WOA	PSO	GA	ALO	GWO
Wine	**0.0165**	0.0325	0.0654	0.0654	0.0372	0.0430
Zoo	0.117	**0.0622**	0.1008	0.1138	0.1875	0.1027
Hepatitis	**0.0144**	0.0998	0.1158	0.1680	0.1293	0.2195
Fertility	0.0848	0.0855	0.1397	**0.0838**	0.1199	0.1408
Ecoli	**0.1118**	0.1337	0.1152	0.1641	0.1427	0.1633
Vehicle	0.2594	0.2988	0.2659	**0.2542**	0.3019	0.2945
Heart	**0.1270**	0.1334	0.1513	0.1762	0.2026	0.1301
Liver	**0.2716**	0.3202	0.3282	0.3381	0.2952	0.2716
Diabetes	0.2382	**0.2302**	0.2525	0.2547	0.2434	0.2473
Breastcancer	0.0238	**0.0231**	0.0302	0.0327	0.0425	0.0415
Ionosphere	**0.0894**	0.0976	0.1122	0.1165	0.1199	0.1415
Lung Cancer	0.1650	0.1948	0.1895	**0.1047**	0.1902	0.2844
Dermatology	**0.0106**	**0.0106**	0.0213	0.0128	0.0214	0.0226
Sonar	0.1558	**0.1057**	0.1515	0.2154	0.1769	0.2256
BreastEW	**0.0500**	0.0501	0.0539	0.0508	0.0564	0.0654
Soybean Small	0.0258	0.0040	0.0016	**0.0006**	0.0019	0.2913
Movementlibras	0.3539	0.2672	0.3534	**0.2511**	0.3889	0.2784
Parkinsons	0.1160	**0.0666**	0.1035	0.1120	0.0936	0.1242
Spambase	0.0959	**0.0729**	0.0862	0.1016	0.0904	0.084
Waveform	**0.1515**	0.1558	0.1854	0.1634	0.1681	0.1648
Hillvalley	0.4426	**0.3831**	0.4136	0.4451	0.4052	0.4088
Arrhythmia	**0.2451**	0.2888	0.3200	0.3634	0.3744	0.3527
Multiple Features	**0.0269**	0.0304	0.0407	0.0448	0.0499	0.0448
Semeion	**0.0351**	0.0746	0.0908	0.0944	0.1012	0.0885
Clean	**0.0181**	0.0763	0.1600	0.1179	0.1611	0.1843
CNAE	**0.0016**	0.0063	0.1311	0.0135	0.1662	0.100
DNA	**0.0112**	0.1189	0.1481	0.1953	0.1371	0.1452
Average	**0.1207**	0.1267	0.1528	0.1501	0.1631	0.1726

Table 6.Worst (Max) Fitness Values for all Algorithms after 20 Runs

DS.	IWOA	WOA	PSO	GA	ALO	GWO
Wine	**0.0217**	0.0499	0.0698	0.1293	0.0487	0.0533
Zoo	0.1238	**0.0652**	0.1014	0.1453	0.1985	0.1587
Hepatitis	**0.0155**	0.1152	0.1158	0.1829	0.1358	0.3954
Fertility	0.0957	**0.0925**	0.1521	0.0948	0.1298	0.1597
Ecoli	0.1205	0.1526	**0.1118**	0.1841	0.1591	0.1736
Vehicle	0.2604	0.3249	0.2688	**0.2547**	0.3542	0.3125
Heart	**0.1305**	0.1435	0.1513	0.2076	0.2125	0.1459
Liver	**0.2957**	0.3345	0.3598	0.3419	0.3251	0.3874
Diabetes	**0.2451**	0.2471	0.2525	0.2834	0.2987	0.2874
Breastcancer	**0.0241**	0.0292	0.0352	0.0338	0.0435	0.0456
Ionosphere	**0.0957**	0.1172	0.1354	0.1338	0.1359	0.1547
Lung Cancer	0.1883	0.2508	0.2513	**0.1879**	0.2215	0.2954
Dermatology	**0.0112**	0.0146	0.0218	0.0205	0.0255	0.0298
Sonar	0.1621	**0.1265**	0.1558	0.2423	0.1894	0.2365
BreastEW	0.0532	**0.0509**	0.0541	0.0742	0.0623	0.0755
Soybean Small	0.0295	0.0048	0.0020	**0.0008**	0.0022	0.2956
Movementlibras	0.3613	0.2853	0.3621	**0.2628**	0.4125	0.3215
Parkinsons	0.1180	**0.1120**	0.1125	0.1334	0.1123	0.1354
Spambase	0.0984	**0.0791**	0.0898	0.1091	0.1025	0.0954
Waveform	**0.1598**	0.1723	0.1957	0.1691	0.1987	0.1789
Hillvalley	0.4568	**0.3935**	0.4987	0.4548	0.4987	0.4879
Arrhythmia	**0.2645**	0.3056	0.3375	0.3728	0.3987	0.3974
Multiple Features	**0.0298**	0.0330	0.0410	0.0498	0.0521	0.0542
Semeion	**0.0388**	0.0825	0.0923	0.1004	0.1235	0.1264
Clean	**0.0214**	0.0906	0.1711	0.1351	0.1954	0.1987
CNAE	**0.0021**	0.0075	0.1398	0.0850	0.1987	0.1542
DNA	**0.0122**	0.1321	0.1504	0.2046	0.1478	0.1524
Average	**0.1272**	0.1412	0.1640	0.1701	0.1845	0.2040

Table 7. Average Number of Selected Features after 20 Runs

DS.	IWOA	WOA	PSO	GA	ALO	GWO
Wine	**4.23**	4.70	5.50	4.75	5.32	7.24
Zoo	**5.00**	6.45	6.45	11.25	5.00	9.00
Hepatitis	**3.00**	4.00	6.00	4.10	4.50	7.00
Fertility	2.00	3.00	**1.00**	3.25	4.47	5.24
Ecoli	**4.00**	**4.00**	5.50	6.25	5.00	5.00
Vehicle	8.40	**8.20**	11.00	12.65	8.90	12.00
Heart	**3.90**	4.10	6.00	6.35	4.40	7.28
Liver	**2.47**	3.00	4.48	3.75	3.15	5.24
Diabetes	**4.00**	**4.00**	**4.00**	**4.00**	5.14	4.48
Breastcancer	**3.38**	4.00	5.65	5.65	5.50	5.78
Ionosphere	**8.25**	8.65	17.38	9.10	13.56	22.79
Lung Cancer	**13.09**	14.45	21.75	31.75	25.06	33.05
Dermatology	**11.33**	13.20	17.35	27.30	21.05	21.44
Sonar	18.08	**17.75**	23.45	30.20	21.08	40.23
BreastEW	4.04	**3.05**	5.50	15.30	10.02	8.05
Soybean Small	**2.01**	2.20	5.05	14.20	6.45	9.18
Movementlibras	**35.14**	37.25	40.47	38.20	55.05	35.47
Parkinsons	3.07	**3.05**	5.55	10.60	6.11	6.27
Spambase	**21.48**	23.49	28.71	27.20	27.49	45.73
Waveform	21.63	**20.90**	35.47	33.00	28.38	26.79
Hillvalley	**33.35**	39.60	50.79	41.20	33.94	69.48
Arrhythmia	**88.14**	98.07	127.23	108.05	113.27	168.15
Multiple Features	**255.41**	273.00	284.84	439.20	499.05	296.07
Semeion	**105.94**	108.26	130.45	133.95	199.05	141.08
Clean	**50.08**	69.55	73.14	115.65	124.47	113.79
CNAE	**210.17**	216.23	440.35	227.00	573.78	512.49
DNA	**51.08**	53.00	98.55	78.25	140.49	112.07
Average	**36.02**	38.78	54.13	53.40	72.21	64.09

Table 8. Average Classification Accuracy (%) for all Optimizers after 20 Runs

DS.	IWOA	WOA	PSO	GA	ALO	GWO
Wine	94.78	94.47	93.49	93.88	**95.16**	93.41
Zoo	87.04	86.52	82.54	83.49	**87.64**	84.35
Hepatitis	**90.09**	89.23	86.05	87.50	88.25	84.79
Fertility	91.18	91.18	**92.07**	91.18	88.13	90.47
Ecoli	81.09	**82.14**	80.99	82.14	81.17	81.04
Vehicle	**72.07**	70.09	66.95	71.63	68.14	61.72
Heart	**88.19**	87.71	82.67	82.39	86.62	82.05
Liver	**67.48**	65.14	64.87	64.70	65.49	64.16
Diabetes	77.31	**77.34**	73.69	72.27	70.05	72.05
Breastcancer	**97.06**	96.48	96.47	96.34	95.14	95.17
Ionosphere	**89.27**	88.07	89.25	85.32	85.22	82.38
Lung Cancer	**60.56**	48.20	56.27	48.2	50.56	50.14
Dermatology	**97.17**	96.56	90.71	96.45	93.22	94.88
Sonar	**75.04**	74.21	74.03	72.07	70.96	68.26
BreastEW	97.05	**97.16**	88.91	96.37	84.91	88.42
Soybean Small	**98.01**	97.81	86.48	94.38	90.98	92.05
Movementlibras	68.01	68.97	64.66	**69.02**	65.97	68.66
Parkinsons	87.19	**88.46**	86.53	84.92	83.67	83.67
Spambase	**89.79**	88.96	87.35	82.29	88.04	88.39
Waveform	81.03	80.50	78.89	77.61	83.01	**83.96**
Hillvalley	**61.44**	60.41	55.07	56.27	57.09	55.44
Arrhythmia	**65.79**	62.78	57.07	58.02	54.62	56.41
Multiple Features	**95.09**	94.68	93.59	92.86	92.55	91.06
Semeion	**98.78**	96.95	93.14	95.78	98.07	97.32
Clean	79.55	**79.65**	77.84	76.48	79.41	79.53
CNAE	**86.79**	84.58	81.47	82.46	79.62	84.07
DNA	**85.05**	83.56	79.79	79.54	76.66	84.25
Average	**83.77**	82.66	80.03	80.50	80.01	79.93

Table 9. Average execution time (sec) for all optimizers after 20 runs

DS.	IWOA	WOA	PSO	GA	ALO	GWO
Wine	**19.23**	21.20	28.13	19.60	26.27	32.45
Zoo	18.89	**18.70**	21.79	19.20	24.65	21.44
Hepatitis	25.74	**25.30**	29.36	27.20	29.35	28.15
Fertility	21.48	**20.50**	27.71	20.70	24.23	21.25
Ecoli	**21.80**	**21.80**	23.12	22.60	**21.80**	27.45
Vehicle	48.78	43.60	42.32	**40.10**	52.03	54.74
Heart	**27.38**	29.50	31.74	30.30	35.23	38.12
Liver	21.79	23.50	22.79	20.40	**19.24**	21.45
Diabetes	22.64	**21.90**	24.23	22.80	29.15	27.35
Breastcancer	22.53	21.90	23.40	**21.70**	27.14	23.78
Ionosphere	37.48	36.90	**30.27**	36.20	32.74	37.46
Lung Cancer	32.74	**32.50**	33.48	33.70	36.48	39.73
Dermatology	55.20	56.70	61.24	**55.20**	57.81	64.48
Sonar	63.43	62.50	66.74	65.20	**58.48**	68.16
BreastEW	47.32	46.50	50.75	44.20	**41.48**	49.12
Soybean Small	45.76	44.80	45.48	**40.20**	47.23	45.19
Movementlibras	65.11	65.00	69.48	64.80	68.92	**64.78**
Parkinsons	57.05	58.90	61.49	55.30	**54.27**	62.48
Spambase	225.24	215.70	210.27	**198.2**	236.69	253.74
Waveform	291.67	290.60	290.37	**288.50**	301.07	305.41
Hillvalley	85.59	95.40	97.58	**80.30**	115.47	105.38
Arrhythmia	152.78	145.10	125.75	**108.60**	242.78	233.18
Multiple Features	**480.45**	510.20	498.38	489.60	530.78	515.37
Semeion	**279.06**	280.70	298.64	281.20	315.45	305.47
Clean	**240.48**	241.50	249.48	250.20	287.35	277.73
CNAE	564.79	**543.50**	591.74	546.50	588.66	566.91
DNA	489.25	488.40	470.91	**461.30**	505.45	513.48
Average	128.28	128.25	130.61	**123.84**	141.11	140.89

Tables 10 and 11 demonstrate specific feature reduction selection examples for the Heart dataset with 13 features, and for the Wine dataset, with 13 features. We can observe from the Heart dataset that IWOA proposes that only six of the features are sufficient for the classification. For the Wine dataset, our method recommends that only seven of the attributes will guarantee the same accuracy in performing the classification as if we consider all the set. Over-all, while comparing IWOA with other optimizers, We observe that IWOA almost always gets better classification accuracy with a minimal number of selected features. In most of the tests performed, around 80% of the features identified by IWOA are just the same as features identified by other optimizers, but in many cases, the subset of features identified by IWOA is incorporated into the subset of features identified by other optimizers.

Table 10. An Example of the Features Selected for all Optimizers using the Heart Dataset

Algorithm	All features	No. of selected	Features Indices	% Selected
IWOA	13.00	6.00	2,7,9,10,12,13	53.85
WOA	13.00	3.00	9,12,13	23.93
PSO	13.00	6.00	3,7,9,10,12,13	53.85
GA	13.00	4.00	2,9,10,12	69.24
ALO	13.00	6.00	2,3,7,10,12,13	53.85
GWO	13.00	8.00	2,3,6,9,10,11,12,13	38.47

The fitness values standard deviation is computed on the 20 runs and showed in Figure 3. We can observe that IWOA has a small value of standard deviation which demonstrates the repeatability, stability and capacity to achieve ideal solutions regardless of the stochastic process. From this figure, we can see that the IWOA is still performing superior to other optimizers which affirms the searching ability of IWOA.

Figure 4 illustrates the classification accuracy, the average number of selected features and the computational time averaged over all the datasets using all optimizers. We can highlight from the figure that the classification performance of IWOA is greatly improved than other optimizers and obtains the best feature reduction rates. We can observe that GA has been the fastest algorithm followed by WOA and then IWOA. Thus, if we consider the best trade-off among classification accuracy, feature reduction rate and computational time, the best choice depends on IWOA.

Table 11. An Example of the Features Selected for all Optimizers using the WINE Dataset

Algorithm	All features	No. of selected	Features Indices	% Selected
IWOA	13.00	7.00	1,2,3,7,8,9,10	46.16
WOA	13.00	5.00	1,2,7,9,11	61.36
PSO	13.00	7.00	1,3,4,7,8,9,12	46.16
GA	13.00	7.00	1,2,4,7,9,10,12	46.16
ALO	13.00	5.00	2,6,7,10,11	61.36
GWO	13.00	7.00	1,2,3,4,6,10,12	46.16

We also evaluate how the IWOA work with various transfer functions for feature selection task using a binary, sigmoid and hyperbolic tangent functions to convert the continuous values to binary ones. Figure 5 displays the accuracy performance of IWOA. We can highlight from the figure that the sigmoid function worked well and provided good classification accuracy in all datasets. Figure 6 displays the average number of selected features over all datasets. We can observe from the figure that the sigmoid function worked well and provided good feature reductions in all datasets. In regard to transfer functions, we can observe the sigmoid function works well with all datasets to convert the continuous values to binary ones. With this function, the proposed algorithm can reduce the number of features. Therefore, we used the sigmoid function as a transfer function with IWOA and other optimizers to build binary solutions with 1 and 0 values.

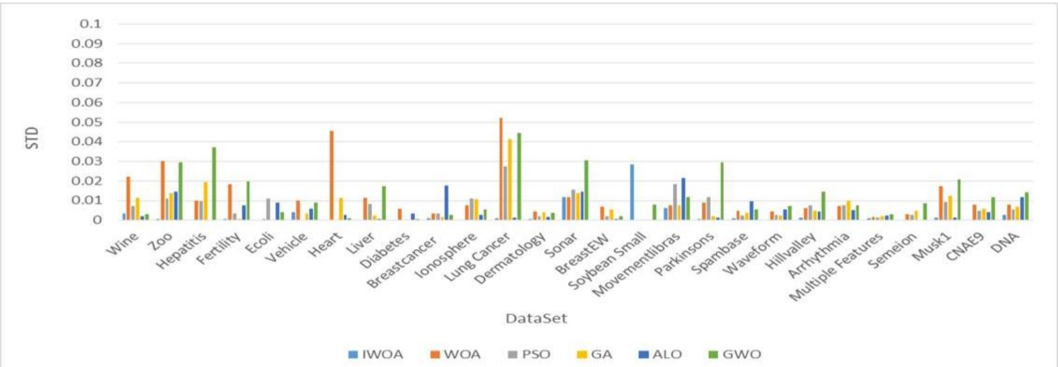

Fig.3. STD fitness values acquired of all optimizers averaged over all the datasets

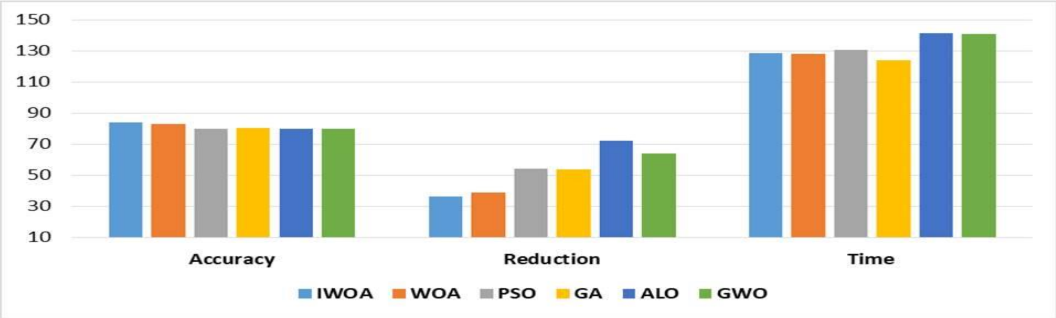

Fig.4. Average performance of all optimizers averaged over all the datasets

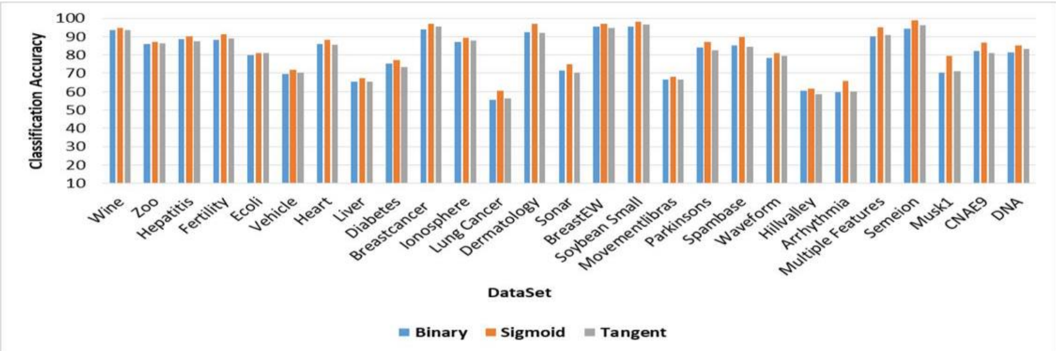

Fig.5. Average classification accuracy of IWOA using different transfer functions

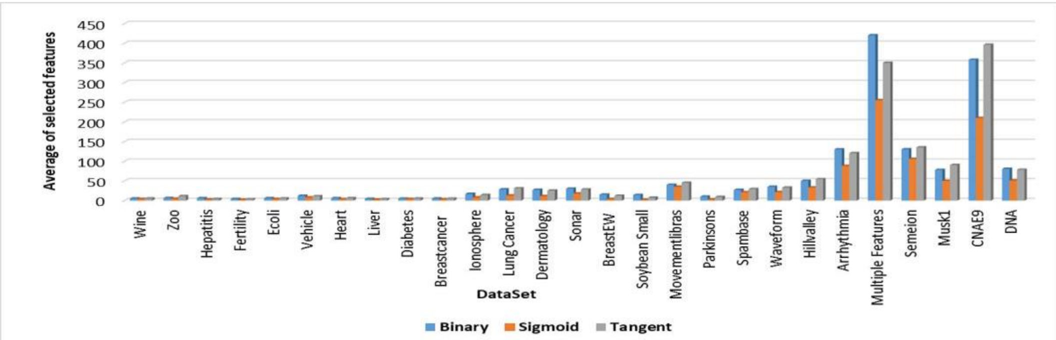

Fig.6. Average feature reduction ratio of IWOA using different transfer functions

According to these results, the improved whale optimization algorithm is the appropriate optimizer. Such an improvement of the results came from embedding inertia weight parameter in the searching mechanism of WOA. This helps the algorithm to improve the solution accuracy and the ability for finding the feature subset in the feature space better than the other optimizers. Optimization results proved that IWOA is a powerful search algorithm since it is simple in concept and effective to explore global solutions. IWOA has only a few control parameters and has wide applications in real-world engineering optimization problems.

V. Conclusion

In this work, we have introduced an improved binary version of the standard whale optimization algorithm for feature selection tasks in wrapper approach, which was inferred so as to position the whale agents to only binary values, which represents a series of bits that means whether a feature will be chosen or not. We led tests against five optimizers to test the proposed algorithm robustness, and also its good generalization capacity. We have utilized 27 datasets to achieve this task, in which IWOA has been compared against WOA, PSO, GA, ALO and GWO. The proposed optimizer has outperformed the other optimizers in classification accuracy, being the second fastest optimizer and the one that has selected the minimal number of features. In regard to transfer functions, we can observe the sigmoid function works well with all datasets to convert the continuous values to binary ones.

References

[1] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[2] D. Y. Harvey and M. D. Todd, "Automated Feature Design for Numeric Sequence Classification by Genetic

Programming," *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 4, pp. 474–489, 2015.
[3] S. Goswami and A. Chakrabarti, "Feature Selection: A Practitioner View," *I.J. Information Technology and Computer Science, MECS*, vol. 11, no. 11, pp. 66–77, 2014.
[4] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern classification," Wiley-Interscience, p. 680, 2000.
[5] Y. Chen, D. Miao, and R. Wang, "A rough set approach to feature selection based on ant colony optimization," Pattern Recognition Letters, vol. 31, no. 3, pp. 226–233, 2010.
[6] A. F. Alia and A. Taweel, "Feature Selection based on Hybrid Binary Cuckoo Search and Rough Set Theory in Classification for Nominal Datasets," *I.J. Information Technology and Computer Science, MECS*, vol. 4, no. 4, pp. 63–72, 2017.
[7] R. Kohavi and G. H. John, "Wrappers for feature subset selection," Artificial Intelligence, vol. 97, no. 1-2, pp. 273–324, 1997.
[8] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms," *Applied Soft Computing*, vol. 18, pp. 261–276, 2014.
[9] M. Toghraee, H. Parvin, and F. Rad, "The Impact of Feature Selection on Meta-Heuristic Algorithms to Data Mining Methods," *I.J. Modern Education and Computer Science, MECS*, vol. 10, no. 10, pp. 33–39, 2016.
[10] T. Marill and D. Green, "On the effectiveness of receptors in recognition systems," *IEEE Transactions on Information Theory*, vol. 9, no. 1, pp. 11–17, 1963.
[11] A. W. Whitney, "A Direct Method of Nonparametric Measurement Selection," *IEEE Transactions on Computers*, vol. C-20, no. 9, pp. 1100–1103, 1971.
[12] B. Xue, M. Zhang, and W. N. Browne, "Particle Swarm Optimization for Feature Selection in Classification: A Multi-Objective Approach," *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 1656–1671, 2013.
[13] R. Parimala and R. Nallaswamy, "Feature Selection using a Novel Particle Swarm Optimization and It's Variants," *I.J. Information Technology and Computer Science, MECS*, vol. 5, no. 5, pp. 16–24, 2012.
[14] C. L. Huang, "ACO-based hybrid classification system with feature subset selection and model parameters optimization," Neurocomputing, vol. 73, no. 1-3, pp. 438–448, 2009.
[15] A. E. Eiben, P. E. Raue, and Z. Ruttkay, "Genetic algorithms with multi-parent recombination," Parallel Problem Solving from NaturePPSN III, *International Conference on Evolutionary Computation. Springer*, pp. 78–87, 1997.
[16] J. Kennedy and R. Eberhart, "Particle swarm optimization," *IEEE International Conference on Neural Networks*, vol. 4, pp. 1942–1948, 1995.
[17] J. H. Holland, "Genetic Algorithms," Scientific American, vol. 267, no. 1, pp. 66–73, 1992.
[18] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016.
[19] B. Xue, M. Zhang, W. Browne, and X. Yao, "A Survey on Evolutionary Computation Approaches to Feature Selection," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 4, pp. 606–626, 2016.
[20] C. B, "Genetic algorithm with fuzzy fitness function for feature selection," *IEEE International Symposium on Industrial Electronics*, vol. 1, pp. 315–319, 2002.
[21] B. Chakraborty, "Feature subset selection by particle swarm optimization with fuzzy fitness function," 2008 3rd *International Conference on Intelligent System and Knowledge Engineering*, pp. 1038–1042, 2008.
[22] K. Neshatian and M. Zhang, "Genetic Programming for Feature Subset Ranking in Binary Classification Problems," *European conference on genetic programming*, pp. 121–132, 2009.
[23] C. Yang, L. Chuang, J. Li, and C. Yang, "Chaotic maps in binary particle swarm optimization for feature selection," 2008 *IEEE Conference on Soft Computing in Industrial Applications*, pp. 107–112, 2008.
[24] S. Kashef and H. Nezamabadi, "A new feature selection algorithm based on binary ant colony optimization," *The 5th Conference on Information and Knowledge Technology*, pp. 50–54, 2013.
[25] E. Emary, W. Yamany, and A. E. Hassanien, "New approach for feature selection based on rough set and bat algorithm," 2014 9th *International Conference on Computer Engineering & Systems (ICCES)*, pp. 346–353, 2014.
[26] Saroj and Jyoti, "Multi-objective genetic algorithm approach to feature subset optimization," *IEEE International Advance Computing Conference (IACC)*, pp. 544–548, 2014.
[27] L. A. M. Pereira, D. Rodrigues, T. N. S. Almeida, C. C. O. Ramos, A. N. Souza, X.-S. Yang, and J. P. Papa, "A Binary Cuckoo Search and Its Application for Feature Selection," *Springer International Publishing Switzerland*, vol. 516, pp. 141–154, 2014.
[28] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," Neurocomputing, vol. 172, pp. 371–381, 2016.
[29] D. Rodrigues, L. A. Pereira, J. P. Papa, and S. A. Weber, "A Binary Krill Herd Approach for Feature Selection," 22nd *International Conference on Pattern Recognition*, pp. 1407–1412, 2014.
[30] R. Y. M. Nakamura, L. A. M. Pereira, K. A. Costa, D. Rodrigues, J. P. Papa, and X. S. Yang, "BBA: A Binary Bat Algorithm for Feature Selection," 25th SIBGRAPI *Conference on Graphics, Patterns and Images*, pp. 291–297, 2012.
[31] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary ant lion approaches for feature selection," Neurocomputing, vol. 213, pp. 54–65, 2016.
[32] "UCI Machine Learning Repository." [Online]. Available: http://archive.ics.uci.edu/ml/index.php

Authors' Profiles

Ahmed Elsayed Hegazy was born in Ismailia, Egypt. He is currently a teaching Assistant in Faculty of Computer Science & informatics Suez Canal University. He received his First degree in Information system, Faculty of Computer Science & informatics Suez Canal University. He is currently a M.S. Student of the same institution. His research interests: Machine learning, data mining, intelligent Bioinformatics and metaheuristic optimization.

Mohamed Abd Allah makhlouf is currently a lecturer in Faculty of Computer Science & informatics Suez Canal University. He received his First degree in Computer Science and Operation Research, Faculty of Science, Master degree in Expert systems, Faculty of Science Cairo university. He received his PhD degree in computer science, Faculty of Science, Zagazig University. He got the Post-Doctoral studies in Computer science from Granada University Spain in 2016.His research interests: Machine learning, data mining, intelligent Bioinformatics, metaheuristic optimization, Decision support systems and predictive models.

Gh. El-Taweel She received B.S., M.S., in computer Sciences and Ph.D. Degree in Information technology from Cairo University in 1996, Helwan University in 2000, and Cairo University 2005 respectively. She is currently Professor in Suez Canal University, Ismailia, Egypt, since November 2016, Her research interests include image classification, DNA, image fusion, and image security

How to cite this paper: Ah. E. Hegazy, M. A. Makhlouf, Gh. S. El-Tawel, " Dimensionality Reduction Using an Improved Whale Optimization Algorithm for Data Classification", International Journal of Modern Education and Computer Science(IJMECS), Vol.10, No.7, pp. 37-49, 2018.DOI: 10.5815/ijmecs.2018.07.04