International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 10 (2018) pp. 174–182

Adaptive Feature Selection and Classification Using Modified Whale Optimization Algorithm

Heba F. Eid[1] **and Ajith Abraham**[2]

[1]Faculty of Science,Al-Azhar University,
Cairo, Egypt
eid.heba@gmail.com,heba.fathy@azhar.edu.eg

[2]Machine Intelligence Research Labs, MIR Labs,
USA
abraham.ajith@gmail.com

Abstract: **Meta-heuristic algorithms gain researchers interest due to their simplicity and adaptability. Whale optimization algorithm (WOA) is a recent meta-heuristic algorithm, which showes competition performance to other swarm algorithms. However, success and challenges concerning the WOA algorithm are based on its control parameter tuning and search space diversity. In this paper, two modified WOA algorithms are proposed to set a proper exploration and exploitation exchange and increase the search space diversity. Furthermore, a novel feature selection algorithm is proposed with integrated information gain to enhance its initialization phase. Experimental results based on twenty mathematical benchmark test functions demonstrate the effectiveness and stability of the modified WOA when compared with the basic WOA and some well-known algorithms. In addition, experimental results on nine UCI datasets shows the ability of the novel feature selection algorithm for selecting the most informative features for classification tasks, due to its fast convergence and fewer chances to get stuck at local minima.**

Keywords: Meta-heuristic algorithm , Whale Optimization , Feature selection , Information Gain , Classification.

I. Introduction

Feature selection plays a pivotal role in data mining and pattern recognition. For high-dimensional datasets, huge number of features may contain lot of redundancy. This significantly degrade the learning speed of the classification models as well as their accuracy [19]. Therefore, a good dimensionality reduction method is a critical procedure in pattern recognition, which contributes towards boosting the performance of a classification model.

Global optimization concerns about finding the optimal values of the solutions variables in order to meet a certain criteria, has captured the research interest over the years [8] [7]. However, classical optimization algorithms require enormous computational efforts, which tend to fail as the problem search space increases. This motivates for employing meta-heuristics algorithms which show higher computational efficiency in avoiding local minima [15] [6] [21]. Meta-heuristic algorithms solve many kind of optimization problems by imitating biological or physical phenomenas. They can be divided into three main categories: evolutionary , trajectory, and swarm methods [1] [2].

Swarm-based algorithms imitate the social behavior of natural creatures such as ants [3], bees [16], fishes [12], particle swarms [4] and bats [20]. The intelligence derived with swarm based algorithms is self-organizing, distributed and decentralized control. Due to their inherent advantages, such algorithms can be applied to various applications including Feature selection problems.

Whale optimization algorithm (WOA) is a relatively new meta-heuristic optimization technique proposed by Mirjalili and Lewis [13], which mimics the hunting behavior of the humpback whales. However, WOA is easily trapped into local optimum and sometimes provide poor convergence, as the dimension of the search space expansion. Consequently, a number of variants are proposed to improve the performance of the basic WOA.

Ling et al. developed an improved version of WOA based on a *Lévy* flight trajectory, and called the *Lévy* flight trajectory-based whale optimization algorithm (LWOA). The *Lévy* flight trajectory help to increase the diversity of the population and enhancing its capability of avoiding the local optimal optima [10].

Hu et al. introduce different inertia weights into whale optimization algorithm (IWOA). Results illustrates a very competitive performance of IWOAs for prediction compared with

PSO and basic WOA [9].

Mafarja and Mirjalili proposed two hybridized feature selection models based on WOA. For which, simulated annealing (SA) algorithm is embedded to WOA algorithm in the first model, while it is used to enhance the best solution found so far by the WOA algorithm in the second model. The experimental results confirm the efficiency of the proposed SA-WOA models in improving the classification accuracy [11].

This paper aims to introduce two modified algorithms based WOA. The first is Cosine adapted WOA algorithm (CaWOA) which employs the cosine function to tune the control parameter of the WOA for varying exploration and exploitation combinations over the course of iterations. While the second is based on Cosine adapted mutation / crossover WOA (CaXWOA), which enhances the local search capability of the CaWOA algorithm by increasing the search space diversity. Moreover, a novel Information gain CaXWOA algorithm (ICaXWOA) is proposed for solving feature selection problems. The proposed CaWOA and CaXWOA are tested with twenty benchmark functions, while ICaXWOA is tested on nine UCI datasets. Experimental results reveals the efficiency of the proposed algorithms in most cases. The rest of this paper is structured as follows: Section II briefly overviews the whale optimization algorithm while Section III presents the details of the proposed CaWOA and CaXWOA algorithms. Section IV, discusses the proposed ICaXWOA based feature selection method. Experimentation design, results and comparative analysis occupy the remainder of the paper in Section V. Finally, Section VIII summarizes the main findings of this study.

II. Whale optimization Algorithm

Whale optimization algorithm (WOA) is a recently proposed bio-inspired optimization algorithm [13]. It simulates the Humpback whales social hunting behavior in finding and attacking preys. WOA simulates the upward-spirals and double-loops bubble-net hunting strategy; for which, whales dive down and start creating bubbles in a spiral shape around the prey and then swim up toward the surface; as shown in figure 1.

To find the global optimum for a given optimization problem using WOA; the search process starts with assuming a set of random solutions (candidate solutions). Then, a population of search agents will update their positions towards the best search agent until the termination criteria is met.

The WOA mathematical model is given by equation 1; where,a probability of 0.5 is assumed to choose between updating either the shrinking encircling or the spiral mechanism during optimization:

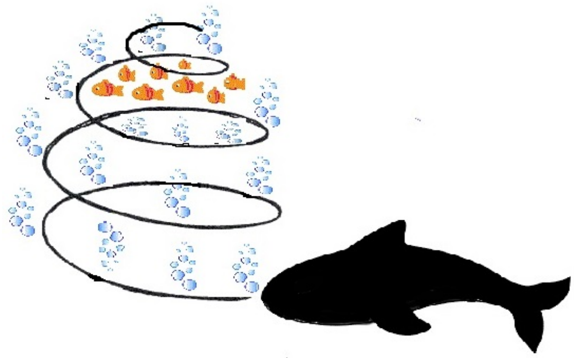


Figure. 1: Humpback Whales bubble-net hunting strategy

$$\overrightarrow{X}(t+1) = \begin{cases} \overrightarrow{\acute{X}}(t) - \overrightarrow{A}.\overrightarrow{D}, & \text{if } p < 0.5. \\ \overrightarrow{\acute{D}}.e^{bl}.cos(2\pi l) + \overrightarrow{\acute{X}}(t), & \text{if p} \geq 0.5. \end{cases} \quad (1)$$

For which, p random number $\in [0,1]$, t the current iteration, $\acute{X}$ the best solution obtained so far, X the position vector, b is a constant defining the spiral shape and l random number $\in [-1,1]$; $\overrightarrow{D}$ is given by:

$$\overrightarrow{D} = |\overrightarrow{C}.\overrightarrow{\acute{X}}(t) - \overrightarrow{X}(t)| \quad (2)$$

While, $\overrightarrow{A}$ and $\overrightarrow{C}$ are coefficient vectors, calculated by:

$$\overrightarrow{A} = 2\overrightarrow{a}.\overrightarrow{r} - \overrightarrow{a} \quad (3)$$

$$\overrightarrow{C} = 2.\overrightarrow{r} \quad (4)$$

where $\overrightarrow{a}$ decreased linearly from 2 to 0 over the course of iterations and $\overrightarrow{r}$ is a random vector $\in [0,1]$.

The distance of the i th whale to the prey is indicated by:

$$\overrightarrow{\acute{D}} = |\overrightarrow{\acute{X}}(t) - \overrightarrow{X}(t)| \quad (5)$$

In order to have a global optimizer, vector $\overrightarrow{A}$;$1 < \overrightarrow{A} < -1$; is used for exploration. Whereby; the search agent position is update according to a randomly selected search agent $\overrightarrow{X_{rand}}(t)$:

$$\overrightarrow{D} = |C.\overrightarrow{X_{rand}}(t) - \overrightarrow{X}(t)| \quad (6)$$

$$\overrightarrow{X}(t+1) = \overrightarrow{X_{rand}}(t) - \overrightarrow{A}.\overrightarrow{D} \quad (7)$$

III. Modified Whale optimization Algorithm

In WOA algorithm the desirable way to converge towards the global minimum can be divided into two conflicting phases: exploration versus exploitation. In the exploration phase whales scatter throughout the entire search space instead of clustering around the local minima. While, in the exploitation phase whales try to converge to the global minimum by searching locally around the obtained solutions .

The WOA algorithm transits between both exploration and exploitation phase by linearly decreasing the distance control parameter a from 2 to 0 using equation 8. Wherein, half iterations are devoted to exploration; when $|A| \geq 1$; best solution is the pivot to update the search agents positions . While, the other half are dedicated to exploitation; $|A| < 1$; the best solution obtained so far plays the role of the pivot point.

$$a = 2(1 - \frac{t}{t_{max}}) \tag{8}$$

Where t and t_{max} indicate the current iteration and the maximum number of iterations respectively.

Generally, higher exploration is similar to much randomness; while higher exploitation is related to too little randomness and will probably give low quality optimization results. The proposed Cosine adapted WOA algorithm (CaWOA) aims to set a right balance between exploration and exploitation phase; to guarantee an accurate approximation of the global optimum.

CaWOA employs a cosine function instead of the linear function for the decay of the control parameter a over the course of iterations; as given in equation 9.

$$a = 1 + 0.5Cosin(\Pi \frac{t}{t_{max}}) \tag{9}$$

Moreover, to enhance the exploitation capability (local search) of the CaWOA algorithm, Cosine adapted mutation / crossover WOA (CaXWOA) is proposed. In the CaXWOA algorithm, mutation operator attempts to change the solution around the best solution obtained so far $\grave{X}$; or around a randomly selected solution X_{rand}. Furthermore, Crossover operator is employed to obtain an intermediate solution between the resultant solution from the mutation operation X_{mut} and the solution X_t.

IV. Adaptive Modified WOA for Feature Selection Problem

For solving feature selection problem, a novel Information gain CaXWOA algorithm (ICaXWOA) is proposed. ICaXWOA algorithm aims to deal with the binary optimization problems. Therefore, the whale position is represented by a binary vector; either 1 indicating that the corresponding feature is selected or 0 for non selected features. The length of the vector is based on the number of features of the original dataset. ICaXWOA adapted information gain (IG) for population initialization; for which, features with corresponding entropy is represented by 1; otherwise the value is set to 0. The IG initialization methods of ICaXWOA are used to guarantee a large initialization in order to improve the local search capability; as the agents positions are commonly closest to the optimal solution.

Feature selection has two main objectives; maximizing the classification accuracy and minimizing the number of features. ICaXWOA is used to adaptively search for the best feature combination, which considers these two objectives. The fitness function adopted to evaluate each individual whale positions is given by:

$$Fitness = \alpha E_R + (1 - \alpha)\frac{|S^*|}{|S|} \tag{10}$$

where E_R is the classification error rate, S^* is the number of selected features and S is the total number of features. α and $(1 - \alpha)$ present the relative importance of the classification accuracy and the selected features number; where, $\alpha \in (0.5, 1]$.

The pseudocode of ICaXWOA is given in Algorithm 1:

V. Experiments and Discussion

VI. Results and Analysis of CaWOA and CaXWOA

The efficiency of the CaWO and CaXWO algorithm proposed in this study was tested using 20 optimization functions. The benchmark functions are divided into three categories: unimodal, multimodal and fixed-dimension multimodal; as given in table 1-3. Figure 2 shows the cost function for F_2, F_{10}, F_{14} *and* F_{19} test problem considered in this study.

For each benchmark function, the CaWOA and CaXWOA algorithms was run 30 independent times and statistical results; average cost function (av) and standard deviation (std) are recorded. CaWOA and CaXWOA were compared against each other and with Particle Swarm Optimization (PSO) [4], Differential Evolution (DE)[18] and Gravitational Search Algorithm (GSA) [17]; as reported in Table 5. Most comparative algorithms results are taken from [14].

Algorithm 1 Pseudocode of ICaXWO Algorithm

Input:
Number of whales n
Number of optimization iterations Max_Iter

Output:
Optimal whale binary position X^*

```
Calculate the entropy of each feature f ∈ dataset.
Initialize the n whales population positions ∈ entropy(f) > 1.
Initialize a, A and C.
t=1
while t ≤ Max_Iter do
    Calculate the fitness of every search agent.
    X* = the best search agent.
    for each search agent do
        Update a by equation 9
        Update A, C and l
        Generate randomly p ∈ [0,1]
        if p < 0.5 then
            if |A| < 1 then
                perform X_mut = mutation(X*)
                update X_{t+1} = Crossover(X_mut, X_t)
            else if |A| ≥ 1 then
                choose a random search agent X_rand
                perform X_mut = mutation(X_rand)
                update X_{t+1} = Crossover(X_mut, X_t)
            end if
        else if p > 0.5 then
            Update position X_{t+1} by equation 1(b)
        end if
        Calculate the fitness of every search agent
        Update X* if a better solution exist
    end for
    t=t+1
end while
return X*
```

Unimodal functions have only one global optimum; thus, they allow to evaluate the exploitation capability of the algorithms. According to Table 5, CaXWOA delivers better results than WOA, CaWOA,PSO, GSA and DE. In particular, CaXWOA is the most efficient optimizer for functions F_1, F_2 *and* F_7 and the second best for functions $F_3 - F_6$. Also, the test remarks a large difference in performance of CaWOA versus CaXWOA which is directly related to applying the mutation/crossover operators. Hence, the CaXWOA algorithm can provide very good exploitation.

Multimodal functions present a good optimization challenge as they possess many local minima; whose number increases exponentially as the expansion of the problem dimensions. As a result, multimodal functions allow to asses the exploration capability. Fixed-dimension multimodal functions provide a different search space compared to multimodal functions.

Table 5, results indicate that CaXWOA shows the best performance in case of functions $F_8, F_{10}, F_{16} - F_{18}$ *and* F_{20}. Produces a similar results to WOA and CaWOA for function F_9, and similar to DE for function F_{11}. While given the second best performance for function F_{12}, F_{14} *and* F_{19}. This is due to adapting cosine function for a better exchange between exploration and exploitation, which leads CaXWOA algorithm towards the global optimum.

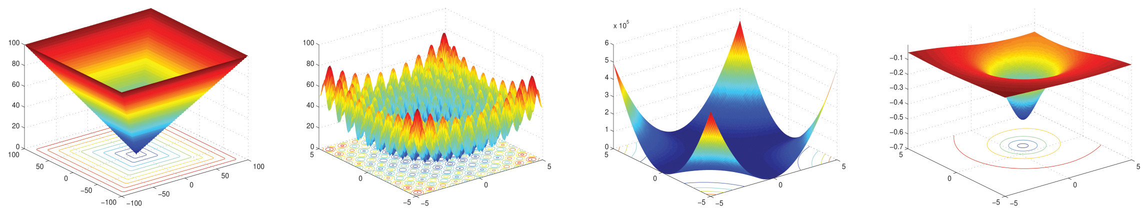
The convergence curves of the WOA, CaWOA , and CaXWOA over the different runs are provided in Figure 3 and 4. As illustrated, the CaXWOA algorithm shows a rapid convergence behavior from the initial steps of iterations when optimizing the test functions. This behavior shows that the CaXWOA algorithm benefits from a good balance of exploration and exploitation; and the crossover diversity; which consequently assists the CaXWOA algorithm to avoid being trapped into local optimal solutions.

VII. Results and Analysis of ICaXWOA

To estimate the performance of the proposed ICaXWO algorithm; experiments are performed on Nine datasets from the UCI machine learning repository [5], as given in Table4. The 9 datasets were chosen to have various numbers of features, classes and instances.

For each dataset, the instances are divided randomly into three sets: training, validation and test sets. In order to ensure the statistical significance and the stability of the obtained results; the partitioning of the data instances are repeated for 30 independent runs. For each run, the average accuracy (Av_Acc), best accuracy ($Best\_Acc$) and the standard deviation (Std); are recorded on the unseen test sets.

Table 6, illustrates the overall performance of the proposed ICaXWOA feature selection algorithm, to asses the effect of hybridizing IG with CaXWOA algorithm. Likewise, ICaXWOA is compared with state of the art feature selection methods such as particle swarm optimization (PSO), genetic algorithm (GA) and ant colony optimization (ACO). From Table 6, it is evident that the ICaXWOA outperforms GA, PSO and ACO feature selection algorithm in term of the average accuracy on all datasets, except for the Diabetic dataset. Meanwhile, in all datasets, ICaXWOA shows a better performance in term of standard deviation values, which indicates the stability of ICaXWOA feature selection algorithm against GA, PSO and ACO feature selection algorithm. To examine the feature selection capability of the ICaXWOA, it is tested on different well known classifiers SVM, NB, J48 and KNN; as shown in table 7. ICaXWOA shows a significant superiority for reducing the number of feature, hence increasing the classification accuracy. The superior performance of the ICaXWOA is justifiable since it adopts IG to guarantee large initialization to enhance the local searching capability.



(a) F_2 (b) F_{10} (c) F_{14} (d) F_{19}

Figure. 2: Graphical presentations of the benchmark functions

Table 1: Unimodal optimization functions.

Function	Dim	Range	F_{min}
$F_1(x)=\sum_{i=1}^{n} x_i^2$	30	[-100,100]	0
$F_2(x)=\sum_{i=1}^{n} \lvert x_i\rvert + \prod_{i=1}^{n} \lvert x_i\rvert$	30	[-10,10]	0
$F_3(x)=\sum_{i=1}^{n}(\sum_{j-1}^{i} x_j)$	30	[-100,100]	0
$F_4(x)=max_i\{\lvert x_i\rvert, 1\leq i\leq n\}$	30	[-100,100]	0
$F_5(x)=\sum_{i=1}^{n-1}[100(x_{i+1}-x_i^2)^2+(x_i-1)^2]$	30	[-30,30]	0
$F_6(x)=\sum_{i=1}^{n}([x_i+0.5])^2$	30	[-100,100]	0
$F_7(x)=\sum_{i=1}^{n} ix_i^4+random[0,1)$	30	[-1.28,1.28]	0

Table 2: Multimodal optimization functions

Function	Dim	Range	F_{min}
$F_8(x)=\sum_{i=1}^{n} -x_i sin(\sqrt{\lvert x_i\rvert})$	30	[-500,500]	-418.98295
$F_9(x)=\sum_{i=1}^{n}[x_i^2-10cos(2\pi x_i)+10]$	30	[-5.12,5.12]	0
$F_{10}(x)=-20exp(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}-exp(\frac{1}{n}\sum_{i=1}^{n} cos(2\pi x_i))+20+e$	30	[-32,32]	0
$F_{11}(x)=\frac{1}{4000}\sum_{i=1}^{n} x_i^2-\prod_{i=1}^{n} cos(\frac{x_i}{\sqrt{i}})+1$	30	[-600,600]	0
$F_{12}(x)=\frac{\pi}{n}10sin(\pi y_1)+\sum_{i=1}^{n-1}(y_i-1)^2[1+10sin^2(\pi y_{i+1})]+(y_n-1)^2+\sum_{i=1}^{n} u(x_i,10,100,4)$	30	[-50,50]	0
$y_i=1+\frac{x_i+1}{4}u(x_i,a,k,m)=\begin{cases}k(x_i-a)^m & x_i>a\\ 0-a<x_i<a & \\ k(-x_i-a)^m & x_i<-a\end{cases}$			
$F_{13}(x)=0.1\{sin^2(3\pi x_1)+\sum_{i=1}^{n}(x_i-1)^2[1+sin^2(3\pi x_i+1)]+(x_n-1)^2[1+sin^2(2\pi)x_n]\}+\sum_{i=1}^{n} u(x_i,5,100,4)$	30	[-50,50]	0

Table 3: Fixed-dimension multimodal optimization functions.

Function	Dim	Range	F_{min}
$F_{14}(x)=\sum^{1} 1_{i=1}[a_i-\frac{x_1(b_i^2+b_iX_2)}{b_i^2+b_iX_3+x_4}]^2$	4	[-5,5]	0.00030
$F_{15}(x)=(x_2-\frac{5.1}{4\pi^2}x_1^2+\frac{5}{\pi}x_1-6)^2+10(1-\frac{1}{8\pi})cosx_1+10$	2	[-5,5]	0.398
$F_{16}(x)=[1+(x_1+x_2+1)^2(19-14x_1+3x_1^2-14x_2+6x_1x_2+3x_2^2)]\times[30+(2x_1-3x_2)^2\times(18-32x_1=12x_1^2+48x_2-36x_1x_2+27x_2^2)]$	2	[-2,2]	3
$F_{17}(x)=\sum_{i=1}^{4} c_i exp(-\sum_{j=1}^{3} a_{ij}(x_j-p_{ij})^2)$	3	[1,3]	-3.86
$F_{18}(x)=\sum_{i=1}^{4} c_i exp(-\sum_{j=1}^{6} a_{ij}(x_j-p_{ij})^2)$	6	[0,1]	-3.32
$F_{19}(x)=\sum_{i=1}^{7}[(X-a_i)(X-a_i)^T+c_i]^{-1}$	4	[0,10]	-10.4028
$F_{20}(x)=\sum^{1} 0_{i=1}[(X-a_i)(X-a_i)^T+c_i]^{-1}$	4	[0,10]	10.5363

Table 4: Datasets Description

Dataset	#Features	#Instances	#Classes
Australian	14	690	2
German Credit	24	1000	2
Sonar	60	208	2
Zoo	17	101	7
NSL-KDD	41	5960	4
Diabetic	19	1151	2
Heart Disease	13	270	2
Segment	19	2310	7
Liver Disorders	6	345	2

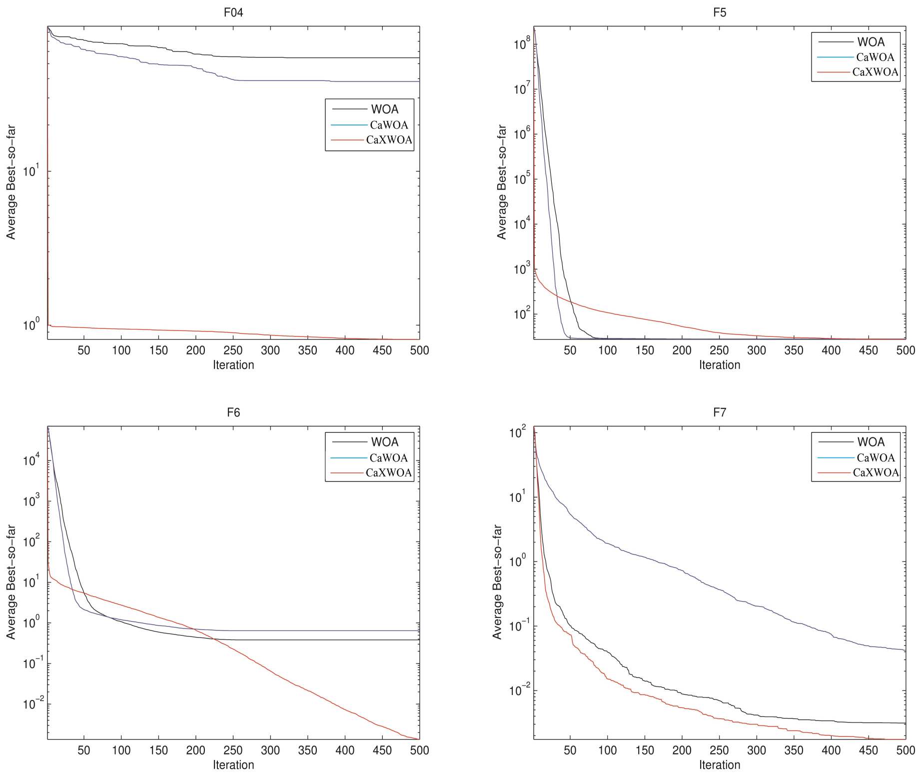


Figure. 3: Best fitness convergence curves of WOA, CaWOA and CaXWOA

Table 5: Comparison results obtained for WOA, CaWOA and ICaXWOA for different optimization functions.

Function	WOA		CaWOA		CaXWOA		PSO		GSA		DE	
	av	std	av	std	av	std	av	std	av	std	av	std
F1	1.41e-30	4.91e-30	2.963e-94	1.5669e-93	**0**	**0**	0.000136	0.000202	2.53e-16	9.67e-17	8.2e-14	5.9e-14
F2	1.06e-21	2.39e-21	2.3851e-60	7.9734e-60	**0**	**0**	0.042144	0.045421	0.055655	0.194074	1.5e-09	9.9e-10
F3	5.3901e-07	2.9310e-06	4.8321e-8	1.530e-3	9.5739e-10	1.918e-08	70.12562	22.11924	896.5347	318.9559	**6.8e-11**	**7.4e-11**
F4	0.072581	0.39747	0.020328	0.14324	8.0795e-4	0.026033	1.086481	0.317039	7.35487	1.741452	**0**	**0**
F5	27.86558	0.763626	27.481	10.399	26.93	0.64762	96.71832	60.11559	67.54309	62.22534	**0**	**0**
F6	3.116266	0.532429	0.64294	0.36901	0.0013559	0.00076554	0.000102	8.28e-05	2.5e-16	1.74e-16	**0**	**0**
F7	0.001425	0.001149	0.017488	0.0024131	**0.00042081**	**0.0010888**	0.122854	0.044957	0.089441	0.04339	70.00463	0.0012
F8	-5080.76	695.7968	-62569	232.7	**-12565**	**173.4**	-4841.29	1152.814	-2821.07	493.0375	-11080.1	574.7
F9	**0**	**0**	**0**	**0**	**0**	**0**	46.70423	11.62938	25.96841	7.470068	69.2	38.8
F10	7.4043	9.897572	**8.8818e-16**	2.234e-15	**8.8818e-16**	**0**	0.276015	0.50901	0.062087	0.23628	9.7e-08	4.2e-08
F11	0.000289	0.00158	0.0000604	0.033101	**0**	**0**	0.009215	0.007724	27.70154	5.040343	**0**	**0**
F12	0.339676	0.214864	0.031696	0.01724	3.8796e-05	0.00010826	0.006917	0.026301	1.799617	0.95114	**7.9e-15**	**8e-15**
F13	1.889015	0.266088	0.71698	0.29682	**1.3498e-32**	**0**	0.006675	0.008907	8.899084	7.126241	5.1e-14	4.8e-14
F14	0.000572	0.000324	0.00066615	0.00024207	0.00034207	**0.00013148**	0.000577	0.000222	0.003673	0.001647	**4.5e-14**	0.00033
F15	0.397914	2.7e-05	**0.39789**	7.7385e-06	**0.39789**	1.3038e-06	**0.39789**	**0**	**0.39789**	**0**	**0.39789**	9.9e-09
F16	**3**	4.22e-15	**3**	0.00082618	**3**	**1.13e-15**	**3**	1.33e-15	**3**	4.17e-15	**3**	2e-15
F17	-3.85616	0.002706	**-3.8628**	0.021346	**-3.8628**	**0**	**-3.8628**	2.58e-15	**-3.8628**	2.29e-15	N/A	N/A
F18	-3.2202	0.098696	-3.321	0.198	**-3.322**	**0.017434**	-3.26634	0.060516	-3.31778	0.023081	N/A	N/A
F19	-8.18178	3.829202	-8.8001	3.0643	**-10.403**	1.3485	-8.45653	3.087094	-9.68447	2.014088	**-10.403**	**3.9e-07**
F20	-9.34238	2.414737	**-10.536**	3.1414	**-10.536**	**1.8067e-15**	-9.95291	1.782786	**-10.536**	2.6e-15	**-10.536**	1.9e-07

Table 6: Performance Results of ICaXWOA, GA, PSO and ACO Feature Selection algorithm on different Datasets

Dataset		ICaXWOA	WOA	GA	PSO	ACO
Australian	Av_Acc	**0.88464**	0.8256	0.8289	0.8246	0.8390
	Std	**0.0047**	0.0202	0.0228	0.0731	0.0240
	Best_Acc	**0.8898**	0.8656	0.8553	0.8744	0.8530
German Credit	Av_Acc	**0.7436**	0.7140	0.7133	0.6889	0.7081
	Std	**0.0110**	0.0367	0.0200	0.0207	0.0168
	Best_Acc	**0.7540**	0.7490	0.7451	0.7333	0.7240
Sonar	Av_Acc	**0.9519**	0.8543	0.7540	0.7857	0.8130
	Std	**0.01065**	0.0341	0.0691	0.0346	0.0255
	Best_Acc	**0.9519**	0.9188	0.8720	0.8571	0.8751
Zoo	Av_Acc	**0.9818**	0.9569	0.8550	0.9512	0.9406
	Std	**0.0080**	0.0278	0.0690	0.0646	0.0324
	Best_Acc	**0.9960**	0.9647	0.9601	0.9714	0.9730
NSL-KDD	Av_Acc	**0.9577**	0.9318	0.9051	0.9241	0.9260
	Std	**0.0015**	0.0214	0.0349	0.0251	0.0351
	Best_Acc	0.9567	0.9408	0.9252	**0.9581**	0.9411
Diabetic	Av_Acc	0.6872	0.6031	**0.7504**	0.6931	0.6451
	Std	**0.00161**	0.0393	0.0169	0.0347	0.0394
	Best_Acc	0.6944	0.6231	**0.7748**	0.6897	0.6681
Heart Disease	Av_Acc	**0.8356**	0.7633	0.7801	0.7700	0.8260
	Std	**0.0155**	0.0209	0.0210	0.0360	0.0240
	Best_Acc	**0.9518**	0.7801	0.9102	0.9059	0.8871
Segment	Av_Acc	**0.9652**	0.9515	0.9150	0.9431	0.9152
	Std	**0.0038**	0.0043	0.0177	0.0147	0.0167
	Best_Acc	**0.9679**	0.9605	0.9515	0.9521	0.9462
Liver Disorders	Av_Acc	**0.7120**	0.7004	0.6780	0.7030	0.6120
	Std	**0.0168**	0.1185	0.0524	0.1263	0.0460
	Best_Acc	**0.7589**	0.7354	0.7373	0.7573	0.6551

Table 7: Comparison Results of ICaXWOA feature selection Algorithm on different Datasets

Dataset	Method	#Features	F-measure			
			SVM	NB	J48	KNN
Australian	All	14	0.5565	0.7710	**0.8565**	0.8434
	WOA	8	0.6985	0.7637	0.8362	**0.8608**
	ICaXWOA	3	0.8579	0.8681	0.8710	**0.8846**
German Credit	All	24	72.400	**75.500**	72.200	0.7140
	WOA	12	**0.7450**	0.7330	0.7240	0.7380
	ICaXWOA	9	0.7660	0.8070	**0.8340**	0.7436
Sonar	All	60	0.6346	0.6682	0.7115	**0.8365**
	WOA	38	0.6682	0.6923	0.7115	**0.8466**
	ICaXWOA	27	0.6635	0.7019	**0.9855**	0.9519
Zoo	All	17	0.9108	**0.96039**	0.9207	0.9405
	WOA	12	0.9307	**0.9505**	0.9209	0.9099
	ICaXWOA	10	0.9703	**0.9901**	0.9819	0.9818
NSL-KDD	All	41	0.7698	0.6355	**0.9582**	0.9337
	WOA	28	0.8602	0.6012	**0.9798**	0.9517
	ICaXWOA	14	0.9545	0.6629	**0.9827**	0.9577
Diabetic	All	19	0.5690	0.5638	**0.6359**	0.6159
	WOA	15	**0.6342**	0.5656	0.6299	0.6325
	ICaXWOA	7	**0.8279**	0.5943	0.6725	0.6872
Heart Disease	All	13	0.5592	**0.8518**	0.7778	0.7888
	WOA	9	**0.8333**	0.8259	0.8296	0.8111
	ICaXWOA	7	**0.9878**	0.8518	0.9148	0.8356
Segment	All	19	0.6450	0.8038	**0.9645**	0.9536
	WOA	13	0.8082	0.7969	**0.9636**	0.9580
	ICaXWOA	8	0.9808	0.8269	**0.9892**	0.9652
Liver Disorders	All	6	0.5942	0.5536	**0.6869**	0.5623
	WOA	4	0.6010	0.4986	**0.6289**	0.6226
	ICaXWOA	3	**0.9797**	0.5797	0.7841	0.7120

VIII. Conclusion

This paper proposed two variants of meta heuristic algorithms named CaWOA and CaXWOA based on WOA algorithm. In the proposed CaWOA, cosine decay function is used to balance the exploration and exploitation of the search space over the course of iterations. While, CaXWOA algorithm integrate mutation and crossover operators to insure the search space diversity. Twenty benchmark test functions were laboring to verify the performance of the proposed CaWOA and CaXWOA algorithm. Experimental results reveal that the proposed improved algorithms with nonlinearly distance control strategies and search space diversity can provide highly competitive results, due to fast convergence and fewer chances to get stuck at local minima.

This paper also consider the feature selection problem in which the ICaXWOA algorithm is proposed. For which, information gain (IG) is used to guarantee a large initialization for the ICaXWOA algorithm. Results on nine UCI datasets concluded that the proposed ICaXWOA is able to out perform the current well-known feature selection algorithms in the literature.

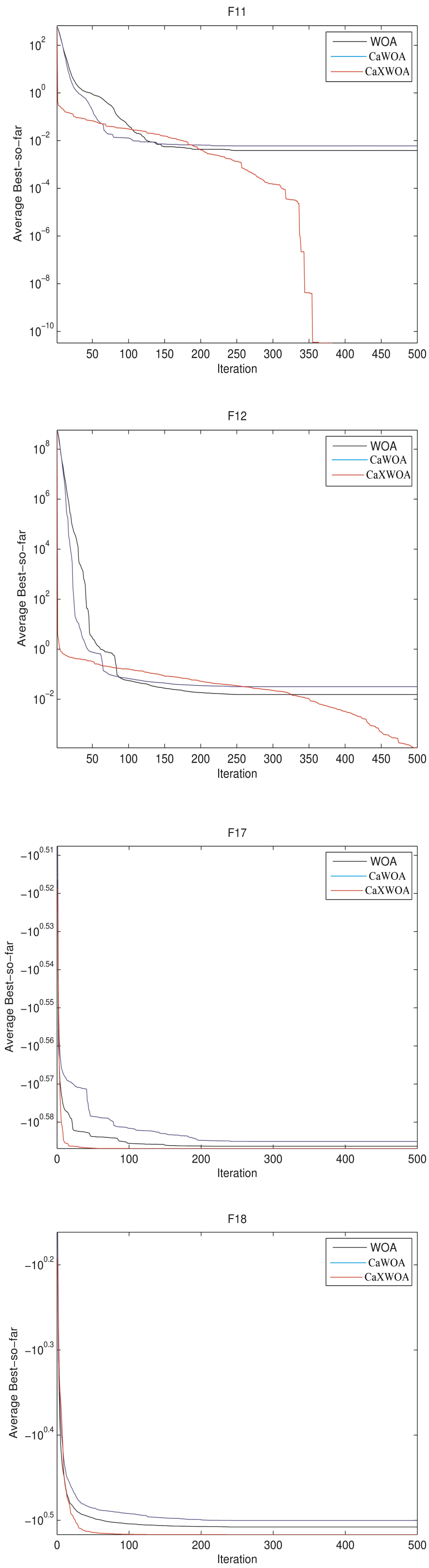


Figure. 4: Best fitness convergence curves of WOA, CaWOA and CaXWOA

References

[1] Blum, C., Roli, A.: A survey on optimization metaheuristics. Metaheuristics in combinatorial optimization: overview and conceptual comparison **35**, pp.268-308 (2003)

[2] BoussaD, I., Lepagnot, J., Siarry, P.: A survey on optimization metaheuristics. Inf.Sci. **237**, pp.82-117 (2013)

[3] Dorigo, M., Birattari, M.: Ant colony optimization. Encyclopedia of Machine Learning, Springer (2010)

[4] Eberhart, R., Kennedy, J.: A new optimizer using particle swarm theory. In: Proc. of the Sixth International Symposium on Micro Machine and Human Science, pp. 39-43. Nagoya, Japan (1995)

[5] Frank, A., Asuncion, A.: Uci machine learning repository (2010)

[6] Gang, X.: An adaptive parameter tuning of particle swarm optimization algorithm. Appl Math Comput **219**, pp.4560-4569 (2013)

[7] Horst, R., Pardalos, P.: Handbook of Global Optimization. Springer Science & Business Media (2013)

[8] Horst, R., Tuy, H.: Global Optimization: Deterministic Approaches. Springer Science & Business Media (2013)

[9] Hu, H., Bai, Y., Xu, T.: Improved whale optimization algorithms based on inertia weights and theirs applications. International Journal of Circuits, Systems and Signal Processing **11**, pp.12-26 (2017)

[10] Ling, Y., Zhou, Y., Luo, Q.: Levy flight trajectory-based whale optimization algorithm for global optimization. IEEE Access pp. 6168 - 6186 (2017)

[11] Mafarja, M.M., Mirjalili, S.: Hybrid whale optimization algorithm with simulated annealing for feature selection. Neurocomputing **260**,pp. 302-312 (2017)

[12] Miller, P.: The Smart Swarm: How Understanding Flocks, Schools, and Colonies Can Make Us Better at Communicating, Decision Making, and Getting Things done. Avery Publishing Group, Inc. (2010)

[13] Mirjalili, S., Lewis, A.: The whale optimization algorithm. Adv Eng Softw **95**, pp.51-67 (2016)

[14] Mirjalili, S., Mirjalili, S., Lewis, A.: Grey wolf optimizer. Advances in Engineering Software **69**, pp.46-61 (2014)

[15] Nesmachnow, S.: An overview of metaheuristics: accurate and efficient methods for optimisation. International Journal of Metaheuristics **3**, pp.320-347 (2014)

[16] Pham, D., Ghanbarzadeh, A., Koc, E., Otri, S., Rahim, S., Zaidi, M.: The bees algorithm. Manufacturing Engineering Centre, Cardiff University pp. 1-57 (2005)

[17] Rashedi, E., Nezamabadi-Pour, H., Saryazdi, S.: Gsa: a gravitational search algorithm. Information Science, pp. 2232-2248 (2009)

[18] Storn, R., Price, K.: Differential evolutiona simple and efficient heuristic for global optimization over continuous spaces. J. Glob Optim, pp. 341-359 (1997)

[19] Xue, B., Zhang, M., Browne, W., Yao, X.: A survey on evolutionary computation approaches to feature selection. IEEE Transactions on Evolutionary Computation **20**, pp.606-626 (2016)

[20] Yang, X.S.: Bat algorithm for multi-objective optimisation. Int. J. Bio-Inspired Comput. **3**, pp.267-274 (2011)

[21] Zhang, Y., Wang, S., Ji, G.: A comprehensive survey on particle swarm optimization algorithm and its applications. Math Probl Eng p. doi:10.1155/2015/931256 (2015)

Author Biographies

Heba F. Eid is an Assistant Professor at Mathematics and Computer Department, Faculty of Science, Al-Azhar University, Egypt. She received her M.S. degree in Distributed database systems and Ph.D. degree in Network Intrusion Detection, both from Faculty of Science, Al-Azhar University, Egypt. Her research interests include multi-disciplinary environment involving machine learning, computational intelligence, pattern recognition, computer vision, metaheuristics and cyber security. Dr.Heba has served as the program committee member of various international conferences and reviewer for various international journals.

Ajith Abraham is the Director of Machine Intelligence Research Labs (MIR Labs), a Not-for-Profit Scientific Network for Innovation and Research Excellence connecting Industry and Academia. The Network with Head quarters in Seattle, USA has currently more than 1,000 scientific members from over 100 countries. He also works as a Research Professor in VSB-Technical University of Ostrava, Czech Republic. As an Investigator / Co-Investigator, he has won research grants worth over 100+ Million US from Australia, USA, EU, Italy, Czech Republic, 100+ France, Malaysia and China. Dr. Abraham works in a multi-disciplinary environment involving machine intelligence, cyber-physical systems, Internet of things, network security, sensor networks, Web intelligence, Web services, data mining and applied to various real world problems. In these areas he has authored / coauthored more than 1,300+ research publications out of which there are 100+ books covering various aspects of Computer Science. One of his books was translated to Japanese and few other articles were translated to Russian and Chinese. About 1000+ publications are indexed by Scopus and over 700 are indexed by Thomson ISI Web of Science. Some of the articles are available in the ScienceDirect Top 25 hottest articles. He has 700+ co-authors originating from 40+ countries. Dr. Abraham has more than 30,000+ academic citations (h-index of 81 as per google scholar). He has given more than 100 plenary lectures and conference tutorials (in 20+ countries). For his research, he has won seven best paper awards at prestigious International conferences held in Belgium, Canada, Bahrain, Czech Republic, China and India.

Since 2008, Dr. Abraham is the Chair of IEEE Systems Man and Cybernetics Society Technical Committee on Soft Computing (which has over 200+ members) and served as a Distinguished Lecturer of IEEE Computer Society representing Europe (2011-2013).

Currently Dr. Abraham is the editor-in-chief of Engineering Applications of Artificial Intelligence (EAAI) and serves/served the editorial board of over 15 International Journals indexed by Thomson ISI. He is actively involved in the organization of several academic conferences, and some of them are now annual events. Dr. Abraham received Ph.D. degree in Computer Science from Monash University, Melbourne, Australia (2001) and a Master of Science Degree from Nanyang Technological University, Singapore (1998). More information at: http://www.softcomputing.net/