

# An Improved Bat Algorithm based on Whale Optimization Algorithm for Data Clustering

Neda Damya<sup>1</sup>, Farhad Soleimanian Gharehchopogh<sup>2</sup>

1- Department of Computer Engineering, Urmia Branch, Islamic Azad University, Urmia, IRAN.

2- Department of Computer Engineering, Urmia Branch, Islamic Azad University, Urmia, IRAN. (bonab.farhad@gmail.com)

Received (2020-08-17)

Accepted (2020-10-16)

**Abstract:** Clustering is a method of data analysis and one of the important methods in data mining that has been considered by researchers in many fields as well as in many disciplines. In this paper, we propose combining WOA with BA for data clustering. To assess the efficiency of the proposed method, it has been applied in data clustering. In the proposed method, first, by examining BA thoroughly, the weaknesses of this algorithm in exploitation and exploration are identified. The proposed method focuses on improving BA exploitation. Therefore, in the proposed method, instead of the random selection step, one solution is selected from the best solutions, and some of the dimensions of the position vector in BA are replaced. We change some of the best solutions with the step of reducing the encircled mechanism and updating the WOA spiral, and finally, after selecting the best exploitation between the two stages of WOA exploitation and BA exploitation, the desired changes are applied on solutions. We evaluate the performance of the proposed method in comparison with other meta-heuristic algorithms in the data clustering discussion using six datasets. The results of these experiments show that the proposed method is statistically much better than the standard BA and also the proposed method is better than the WOA. Overall, the proposed method was more robust and better than the Harmony Search Algorithm (HAS), Artificial Bee Colony (ABC), WOA and BA.

**Keywords:** Bat Algorithm, Whale Optimization Algorithm, Data Clustering, Optimization.

**How to cite this article:**

Neda Damya, Farhad Soleimanian Gharehchopogh. An Improved Bat Algorithm based on Whale Optimization Algorithm for Data Clustering. J. ADV COMP ENG TECHNOL, 6(4) Autumn 2020 : 227-23

## I. INTRODUCTION

Data clustering is one of the most important and practical issues in data processing research and clustering is one of the basic steps in the data mining process [1-3]. In machine learning, data reduction, statistics, image processing, pattern recognition, network and other areas of clustering is used to analyze the data [4, 5]. In the data clustering process, a set of similar objects are grouped into a cluster. For each of these clusters, a header is considered. As a result, in clustering, the data of each cluster includes objects that are

similar and they are different and less similar to objects from other groups [6].

Clustering is an unsupervised learning method. In contrast, classification is a supervised learning method and input samples are labeled, but in clustering, input samples do not have an initial tag, and in fact, using clustering methods, similar data are clustered and implicitly labeled [7]. Before the data grouping operation, clustering should be done on the samples and then the centers of the resulting clusters should be calculated and a label should be assigned to the cluster centers and then the grouping operation should be performed for the new input samples [8]. As



This work is licensed under the Creative Commons Attribution 4.0 International Licence.

To view a copy of this licence, visit <https://creativecommons.org/licenses/by/4.0/>

mentioned, the k-means algorithm is the most popular clustering algorithm, and although the k-means algorithm is simple and fast, in the K-means algorithm everything depends on the initial value [9]. In addition, in the process of minimizing the objective function, it is possible to converge it with local minima. To solve this problem, different clustering algorithms have been proposed so far and each of these algorithms has its own strengths and weaknesses; the purpose of this paper is to cluster the data with the proposed new method based on the combination of BA [10] and WOA [11]. The weaknesses of BA are more in exploitation and exploration in proposed method focuses more on improving BA exploitation. To solve this problem, different clustering algorithms have been proposed so far and each of these algorithms has its own strengths and weaknesses; the purpose of this paper is to cluster the data with the proposed new method based on the hybrid of BA and WOA. The weaknesses of BA are more in exploitation and exploration in which the proposed method focuses more on improving BA exploitation. In fact, WOA is a population-based optimization algorithm inspired by the hunting behavior of humpback whales. WOA includes three operators to simulate whale behavior, including hunting search, prey siege, and bubble behavior. An extensive study has been performed on 29 mathematical benchmark functions for WOA search and convergence analysis and escape from local optimization, and finally, based on the results, WOA is sufficiently competitive with other existing meta-heuristic methods.

All meta-heuristic algorithms have strengths and weaknesses, the greatest weakness meta-heuristic algorithms are early convergence and trapping in the local minimum[12-16]. To eliminate these points in meta-heuristic algorithms, the weakness of the algorithm is first identified and then covered by another meta-heuristic algorithm that does not have this problem or weakness; therefore, in this case, both hybrid algorithms compensate each other's weaknesses and improve performance [17]. In this paper, based on the combination of WOA and BA, it is proposed a robust approach to clustering.

In this paper, a novel hybrid algorithm that combines WOA with BA is proposed for data clustering. Proposed method will be used to exploitation with the best objective values.

Proposed method uses the exploitation to keep a record of the best candidate solutions. These solutions are used to guide the bats members to avoid getting trapped into local optima. Proposed method improves the exploitation of the bats members by employing the encircled mechanism and the spiral inspired by WOA on bats members. This increases the ability of proposed method to achieve faster convergence rates. Also, proposed method ensures obtaining solutions with high quality by applying encircled mechanism among solutions in bats members. These modifications enable the proposed method to converge in a smaller number of iterations than other similar algorithms. Proposed method provides the balance between global and local search by proposing a modified exploitation step, which focuses on searching neighbors of bats members instead of neighbors of the best solution.

The main contributions of the paper are portrayed below.

- The proposed method involves the blending of WOA and BA.
- The WOA cover the solution space efficiently in BA based on exploitation.
- In this paper, we propose combining WOA with BA for data clustering.
- Proposed method was tested on six datasets, which their sizes vary from small to large-sized datasets.

The rest of the paper is organized as follows. Section 2 presents the previously proposed methods to solve data clustering. Section 3, describes both WOA and BA used in the proposed method. Section 4, reports a performance analysis for proposed method. Section 5, summarizes the main points of the paper and introduces the future work.

## II. RELATED WORKS

Gandomi and Yang [18] have used chaos mapping to replace BA dependent parameters. For this purpose, four algorithms have been proposed and in each of them, the parameters of the algorithm have been replaced with 13 turbulence maps. The results showed that different turbulence maps can lead to different BA behaviors and thus improve it, and of course some turbulence maps increased the performance

of BA and some of them were not at all suitable for the bat algorithm. Lin et al. [19] proposed an irregularly improved bat algorithm for estimating the parameters of dynamic biological systems. The simulation results of this system with the turbulence bat algorithm showed better accuracy than the original bat algorithm. As a result, turbulence and irregularity functions can be used to initialize bats.

Saba et al. [20] used the chaos sequence to initialize the BA parameter. They studied the effect of different sequences on the convergence behavior of the BA. The result showed that the turbulent BA is better than BA. Another way to improve the BA is to combine it with other meta-heuristic algorithms and add adaptive operators. Wang et al. [21] have proposed an improved version of the algorithm called the Improved BA. They combined BA with differential evolution to select the best solution in the bat population. They used this algorithm in 3D programming problems and concluded that the proposed method could work better than BA. On the other hand, Zhou et al. [22] successfully developed a comparative random search method and connected it to a standard bat algorithm. They used this method in routing issues and it has been very effective.

To solve multi-objective numerical problems, Kai and colleagues [23] proposed an improved version of the BA. They improved the local search algorithm of the BA using the optimal search strategy. They also introduced a random strategy to increase global search capabilities in a multi-objective environment. Zhou et al. [24] presented the mean quantum position of the bat algorithm to improve the convergence speed of the bat algorithm. In the early stages of this algorithm, the position of each bat is determined by the best current solution, and in the next stage, the position of the bat depends on the average of the best position.

In [25], the bats recovery algorithm with global optimization and inspired by the Tabu Search algorithm is investigated. In this proposed method, a global optimization is used and then all bats focus around it and search. One of the methods presented in this paper is to improve BA exploration by maintaining a set of best solutions, and each bat randomly focuses around one of the best. Using different methods to avoid being in the local optimization and trying to get out of the

local optimization, increases the execution time of the algorithm and increases the computational volume. Limiting the number of times that a solution is attempted to be out of local optimality is another idea used in this paper.

Nakamura et al. [26] proposed a discrete BA to solve classification and feature selection problems. Compared to binary particle swarm optimization (BPSO), HSA and gravitational search, the discrete BA showed better results for the feature selection problem on breast cancer cases. Pilmiz et al. [27] proposed three algorithms to improve the search and extraction properties of the BA. In the first two algorithms, improved models of BPSO with BA were used and in the third algorithm, invasive weed optimization algorithm was used to replace local search in bat algorithm.

To improve the BA search feature, Liangliang and Y.Zhou [28] changed the formulas for updating the emitted pulse rate and volume. The proposed method achieves better results on 92% of the benchmark mathematical functions than the main bat algorithm. Li and Zhou also focused on this issue in their paper [29] and used complex value encryption to improve the BA; in this way, each parameter contains a real part and an imaginary part, and each part is updated separately. The proposed BA increased the population diversity and increased the efficiency of the algorithm.

Wang et al. [30] used BA to better select and improve the K-means algorithm for data clustering. Experimental results from this improvement show that the improved algorithm has better clustering performance than the basic K-mean algorithm. In another study [31], Sood et al. proposed the K-Medoids clustering method using the BA. They proposed a new method based on bat behavior for initial quantification and overcoming the K-Medoids problem in data clustering. This paper introduces a combination of the K-Medoids clustering algorithm and the BA algorithm. Finally, the obtained results show the superiority of K-Medoid clustering technique with Bat and K-medoid algorithms.

Nguyen et al. [32] presented a hybridization of BA and ABC with a communication strategy. Additionally, Murugan et al. [33] presented an algorithm hybridizing BA and ABC with a chaotic based self-adaptive search strategy.

An efficient and robust bat algorithm with fusion of opposition-based learning and whale optimization algorithm has been proposed [34]. Numerous experiments showed that hybrid model had been remarkable advantages in accuracy and stability for many high dimensions, unimodal and multimodal problems.

In [35], an artificial neural network (ANN) is employed as a classifier to detect the gender of an email author and the WOA is used to find optimal weights and biases for improving the accuracy of the ANN classification.

Data Clustering Method Based on Improved Bat Algorithm with Six Convergence Factors and Local Search Operators has been proposed [36]. BA has the disadvantages of being easily trapped into local minima and not being highly accurate. WOA position updating strategy are adopted to improve the local optimization ability of the BA.

### III. PROPOSED METHOD

BA has attracted the attention of researchers due to its simplicity and good performance, and this algorithm had some downsides from the beginning, and after the initial version was presented by Young, variety of improved models of the BA have been introduced so far. The main problem with this algorithm is that it does not have good exploration and exploitation. As a result, to improve the efficiency and convergence speed of the BA, a combined approach with the WOA is proposed to integrate the data clustering problem. WOA is one of the newest meta-heuristic algorithms that has been considered by many students and researchers in order to optimize various problems. In fact, WOA is a population-based optimization algorithm inspired by the hunting behavior of humpback whales.

One of the main advantages of BA is that it can provide very quick convergence at an early stage by switching from exploration to exploitation. In the proposed method, the exploitation stage is amplified by WOA. WOA has two main stages of exploration and exploitation. In the exploitation stage, WOA uses two methods based on reducing the encircled mechanism and the spiral position according to Eq. (1) [11].

$$\vec{X}(t+1) = \begin{cases} \vec{D}' \cdot e^{bl} \cdot \cos 2 \cdot \Pi \cdot l - \vec{X}^*(t) & \text{if } P < 0.5 \\ \vec{X}^*(t) - \vec{A} \cdot \vec{D} & \text{if } P \geq 0.5 \end{cases} \quad (1)$$

Where  $\vec{D}' = |\vec{X}^*(t) - \vec{X}(t)|$  and indicates the distance of the  $i^{\text{th}}$  whale to the prey (best solution obtained so far),  $b$  is a constant for defining the shape of the logarithmic spiral,  $l$  is a random number in  $[-1, 1]$ , and  $\cdot$  is an element-by-element multiplication. where  $t$  indicates the current iteration,  $A$  is coefficient vector,  $X^*$  is the position vector of the best solution obtained so far,  $\vec{X}$  is the position vector.

Eq. (1) must be replaced in BA. In BA, the exploitation stage is concentrated according to Eq. (2) and by changing the  $X_i$  solution, it focuses around one of these best [11]. 0.001 is obtained based on the test.

$$X_i = Best + 0.001 \times rand(1, d) \quad (2)$$

The proposed method uses two Eq. (1) and Eq. (2) for exploitation. In this way, the exploitation steps of the BA and the exploitation of the WOA are applied around the best, and then the solution whose exploitation is better than the other solutions is selected as the  $X_t$  solution. Thus, the exploitation of the proposed method can be defined as Eq. (3).

$$\vec{X}(t+1) = \text{Min\_Fitness} \begin{cases} \vec{D}' \cdot e^{bl} \cdot \cos 2 \Pi \cdot l - \vec{X}1^*(t) \\ \vec{X}2^*(t) - \vec{A} \cdot \vec{D} \\ \vec{X}3^* + 0.001 \times rand(1, d) \end{cases} \quad (3)$$

Therefore, in order to improve the proposed method, Eq. (3) is used. In Eq. (3), each of the parameters that have better exploitation is selected as an alternative solution, and also in the proposed method, we use a set of best answers

that include  $\overline{X1^*}$ ,  $\overline{X2^*}$  and  $\overline{X3^*}$  better

exploitation, instead of using only the best solution ( $X^*$ ). As a result, the efficiency of the BA is improved by using a set of better solutions as well as the powerful exploitation stage of the WOA. Therefore, according to the definition of Eq. (1) and Eq. (2) and the proposed of a new exploitation method for the BA, according to Eq. (3), the pseudocode of the proposed method is shown in Figure (1).

**Proposed Algorithm**

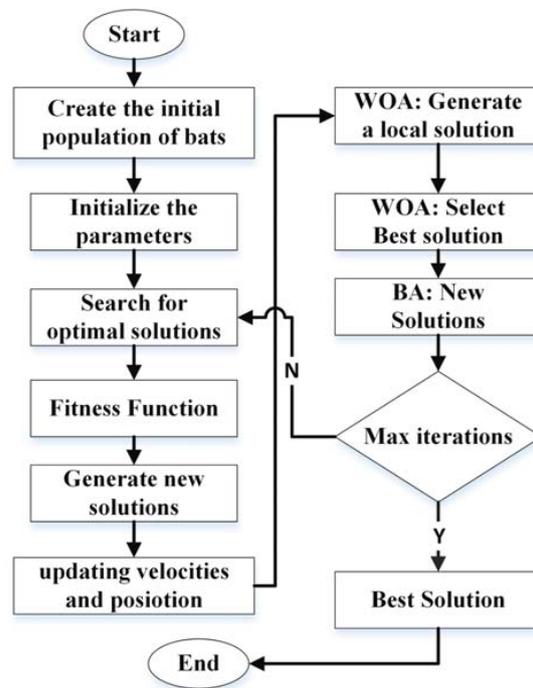
```

Objective function f(x), x = (x1 ... XD)
Initialize the bat population Xi (i = 1, 2... n)
Define pulse frequency fi
Initialize pulse rates and the loudness A
While (t < Max number of iterations)
    Generate new solutions by adjusting frequency,
    And updating velocities and position/solutions [Equations
    (2) to (4)]
    If (rand > r)
        %... New Exploitation .....
        Select  $\overline{X1^*}$ ,  $\overline{X2^*}$ ,  $\overline{X3^*}$  among the best solutions
        Calculation Whale Optimization Algorithm Parameter:
        C=2*rand
        a=2-t*(2)/Max_iter;
        a2=-1+t*(-1)/Max_iter;
        A=2*a*rand-a;
        b=1;
        %Generate a local solution around the Select best solution
         $X_{inew} = \overline{D^i} \cdot e^{bl} \cdot \cos(2 \cdot \Pi \cdot l - \overline{X1^*}(t))$ ;
         $X_{jnew} = \overline{X2^*}(t) - A \cdot \overline{D}$ 
         $X_{knew} = \overline{X3^*} + 0.001 \times rand(1, d)$ ;
        Select Best solution among {Xinew, Xjnew, and Xknew}
    End if
    Generate a new solution by
    flying randomly
    If (rand < A & f(xi) < f(x*))
        Accept the
        new
        solutions
    End if
    Rank the bats and find the
    current best x*
End while
Post process results and visualization
    
```

**Fig. 1. Pseudo-code of the proposed method**

As it can be seen in Figure (1), a new method of WOA and BA is presented. In fact, instead of the random selection step, we change one of the best solutions ( $G_{best}$ ) from the best solutions and replace some of the location vector dimensions with dimensions in the BA, with some of the best solutions with the WOA exploitation stage and the BA exploitation stage.

Before performing the exploitation stage of the WOA, a series of dynamic parameters of this algorithm such as C, A, a, a2 are calculated and the values of each are obtained, and then the two stages of exploitation of the siege mechanism and spiral upgrade are performed. Of course, in addition to these two stages, the efficiency stage of the BA is also maintained. Finally, after selecting the best efficiency, the desired changes are applied to the solution ( $X_i$ ). This somehow increases the efficiency and convergence of the BA. In the following, the steps of the proposed algorithm are explained step by step. the flowchart of the proposed method is shown in Figure (2).



**Fig. 2. Flowchart of the proposed method**

**3.1. Initialization**

In these steps, like other meta-heuristic algorithms, the initial population value and parameters are performed. In this step, the parameters of the BA such as determining the initial population locations with random initial velocity and determining the pulse frequency  $Fi$ , assigning initial values of pulse rate and volume  $Ai$ , and so on are determined. Problem parameters such as the dimension of the problem and the maximum value of each dimension and the minimum value of the dimension, and so

on are also determined in this step. Problem parameters such as the dimension of the problem and the maximum value of each dimension and the minimum value of the dimension, etc. are also determined in this step. In this step, the information about the clustering data set and the dimensions of the problem are also determined.

### 3.2. Objective Function

Proposed method combines WOA with BA to take advantage of efficient search of the solution space in data clustering. In proposed method, each bat in the population represents a complete solution (a set of centers of the clusters). These centers are selected based on the best value of the objective function (4).

Since the goal is to improve the BA with the operators of the WOA to solve the data clustering problem, so the objective function is defined according to the clustering problem as Eq. (4). The objective function of clustering is a standard objective function that is defined in the same way in all scientific papers. In this function, it tries to divide the data into clusters or groups that maximize similarity between data within each cluster and minimize similarity between data of different clusters. Also, in most sources, the Euclidean distance criterion is used to determine the similarity between the patterns. The Euclidean distance given in Eq. (4) is the distance between the members to the desired headers.

$$D_{ij} = \sqrt{\sum_{d=1}^D (X_{jd} - X_{id})^2} \quad (4)$$

In Eq. (4),  $i$  represents the index of objects,  $j$  represents the center of the cluster,  $D$  represents the dimension, and  $D_{ij}$  represents the Euclidean distance between objects and the center of the cluster. After determining the distance between the members to the clusters, each object is first assigned to a cluster so that the sum of the squares of the Euclidean distance between each object and the center of the cluster to which the object is assigned is minimal. In fact, the goal of clustering is to find the centers of the clusters in a way that minimizes the objective function presented in Eq. (5) [37, 38]. Where  $k$  is the number of clusters and

$z_j$  is the center of the  $j^{\text{th}}$  cluster and  $N$  is iteration numbers.  $W_{ij}=1$  if the object  $i (x_i)$  is in the cluster  $i$  and 0 otherwise.

$$f(w, z) = \sum_{i=1}^N \sum_{j=1}^k w_{ij} \|x_i - z_j\|^2 \quad (5)$$

In Eq. (5),  $K$  represents the number of clusters,  $N$  the number of objects,  $X_i \in \{x_1, x_2, \dots, x_n\}$

represents the center of the  $j^{\text{th}}$  of the cluster.

### 3.3. Bat Algorithm Steps

In the proposed method, at first, the initial steps of BA are performed. In this step, the frequency values are updated and the speeds are updated and the value of the conversion function is calculated using the bat location update. At this stage we have no new changes and the standard BA will be implemented except for the exploitation stage of the BA, in fact all the stages of the standard algorithm will be implemented until the condition of the exploitation stage is met. If the exploitation condition is met, then the new exploitation part will be implemented. Otherwise all BA steps will be executed without any changes.

### 3.4. Whale Optimization Algorithm for Exploitation

In the proposed method, instead of simple exploitation operations in the BA, two exploitation stages in WOA are used, namely the siege mechanism stage and the spiral upgrade stage. Therefore, BA enters this stage when its exploitation condition is met. After entering this stage, a series of dynamic WOA parameters such as  $C$ ,  $A$ ,  $a$ ,  $a_2$  are calculated and the values of each are obtained, and then two exploitation steps, namely the siege mechanism and the spiral update, are performed. Also, the efficiency stage of the BA is implemented along with the WOA exploitation, and finally the solution that has better exploitation is selected as the final solution. Of course, for better exploitation at this stage, instead of using a better solution, a set of better solutions has been used together. The set of best

solutions is randomly selected from the best possible solutions.

### 3.5. Final Steps of Bat Algorithm

After the exploitation steps have been applied to the best solutions selected and the solution chosen, the BA will proceed to the next steps again. The next steps of the BA include generating a new answer by random flight and accepting new answers to increase the and decrease the , and then these steps rank the bats and find the current by the standard BA. The last step of the proposed method is to check the end condition of the algorithm. The condition for the termination of all metaheuristic algorithms is twofold: one is that the algorithm be repeated at a predetermined number, for example 20 times, since some algorithms may change in a repetition of the whole population and some only a percentage of their population. This method is almost unfair. For this reason, the second method, that is the number of tests or calls of the objective function, is mostly used. In the objective function number estimation method, all algorithms have the same conditions and each can change or improve its population to the same extent.

## IV. EVALUATION AND RESULTS

The proposed method is implemented in MATLAB software environment and the evaluation is performed using six datasets derived from UCI. The number of initial populations in all algorithms is equal to 50. Also,  $\gamma, \alpha, r, f_{max}, f_{min}$  are 0.9, 0.9, [0,1], 0,1 respectively. The parameters' values were set based on different run in WOA and BA. These six datasets are [39]: Iris (150 samples, 3 clusters and 4 features) Wine dataset (178 samples, 3 clusters and 13 features), CMC dataset (1473 samples, 3 clusters and 9 features), Wisconsin (683 samples, 2 clusters and 9 features), Glass (214 samples, 6 clusters and 9 features) and Vowel (871 samples, 6 clusters and 3 features).

To evaluate the proposed method, statistical criteria such as population mean, population standard deviation, best solution in the population, worst solution in the population have been used and each factor has been calculated and displayed the optimal clustering answer for

each data set separately. Each of the evaluation criteria is shown in Table 1.

Criterion	Equations
Best	Best= Min(All Fitness Population)
Worst	Worst= Max (All Fitness Population)
Avg	$Mean = \frac{\sum_{i=1}^{Npop} Fitness(i)}{Npop}$
Standard deviation	$STD = \sqrt{\frac{1}{N} \sum_{i=1}^{Npop} (Fitness_i - mean(All fitness))}$

The results of implementing the proposed algorithm on different datasets are shown in Table (2). Table (2) shows the results of statistical calculations such as population average, population standard deviation, best solution in the population, and worst solution in the population.

To evaluate the proposed method, it is necessary to compare the proposed method with other clustering algorithms. Therefore, in this section, we compare the proposed method with Harmony Search Algorithm (HAS), Artificial Bee Colony (ABC), and Standard WOA. Of course, for these experiments, all the parameters of the algorithms in terms of population and number of iterations are equal, as well as the number of iterations of the objective function 1000 to show which algorithm works better than other algorithms. Table (3) shows a comparison of the proposed method with other clustering algorithms on six clustering datasets.

The results of Table (3) show that the proposed algorithm has been able to show good results in all experiments. The results show that the proposed algorithm performs much better than the standard algorithm in terms of statistical criteria. Table (3) presents a comparison among proposed method, ABC, BA, HSA, and WOA, based on the quality of fitness function. Table (3) describes that proposed method improves clustering solutions in comparison with other algorithms over all the datasets except CMC, and glass. On two datasets, the HSA fitness function is better than the proposed method. HSA is a powerful meta-heuristic algorithm with excellent exploitation capabilities but suffers a very serious limitation of premature convergence. Thus,

TABLE II  
RESULTS OF STATISTICAL CALCULATIONS OF THE PROPOSED METHOD ON  
DIFFERENT DATASETS

Criterion	Iris	Wine	CMC	Wisconsin	Glass	Vowel
Best	110.606	16667.907	6202.690	4441.182	454.016	186654.637
Worst	384.576	36178.164	17604.466	9324.929	739.912	554245.197
Avg	136.295	18234.060	6652.258	5086.031	471.429	214089.112
Standard deviation	54.600	4307.421	1670.403	1143.355	39.888	52571.377

TABLE III COMPARISON OF THE PROPOSED METHOD WITH DIFFERENT DATA CLUSTERING ALGORITHM

Datasets	Criterion	ABC	BA	HSA	WOA	Proposed Mode
Iris	Best	124.770	143.9	137.2300	135.75	<b>119.78</b>
	Worst	281.910	152.91	163.5900	1169.71	402.06
	Avg	191.510	146.02	154.4100	327.37	158.61
	Standard deviation	34.6600	2.130	7.11000	293.58	74.950
Wine	Best	17219.84	18494.0	16865.90	16873.02	<b>16553.82</b>
	Worst	22187.91	18637.3	18693.42	151714.29	107457.98
	Avg	19470.62	18554.3	18026.25	40083.12	20984.97
	Standard deviation	1287.43	54.77	484.6300	35252.18	15050.97
CMC	Best	6703.19	7145.2	<b>6243.880</b>	6514.67	6851.12
	Worst	11811.28	7357.45	7159.400	44015.37	7221.31
	Avg	8145.71	7167.45	6904.260	13007.21	6921.34
	Standard deviation	846.240	33.10	212.6500	10293.95	121.95
Cancer	Best	4398.84	4886.78	4245.470	3743.6	<b>3469.53</b>
	Worst	7713.82	5309.18	5318.020	8915.37	9602.91
	Avg	5510.22	4910.22	5000.160	4420.67	4539.59
	Standard deviation	828.270	58.66	256.5100	1072.99	1455.01
glass	Best	425.780	593.88	<b>403.8100</b>	516.57	445.36
	Worst	809.980	609.23	534.700	14493.57	1149.45
	Avg	593.500	597.74	495.4000	3049.2	514.07
	Standard deviation	93.3100	3.980	34.77000	3460.12	156.53
vowel	Best	208491.30	215304.65	197298.1	196839.7	<b>192876.62</b>
	Worst	328605.82	215318.73	228070.41	3729568.75	469053.83
	Avg	255866.8	215308.20	219376.94	732966.42	244265.36
	Standard deviation	23928.83	3542.170	7610.51	738333.36	73749.41

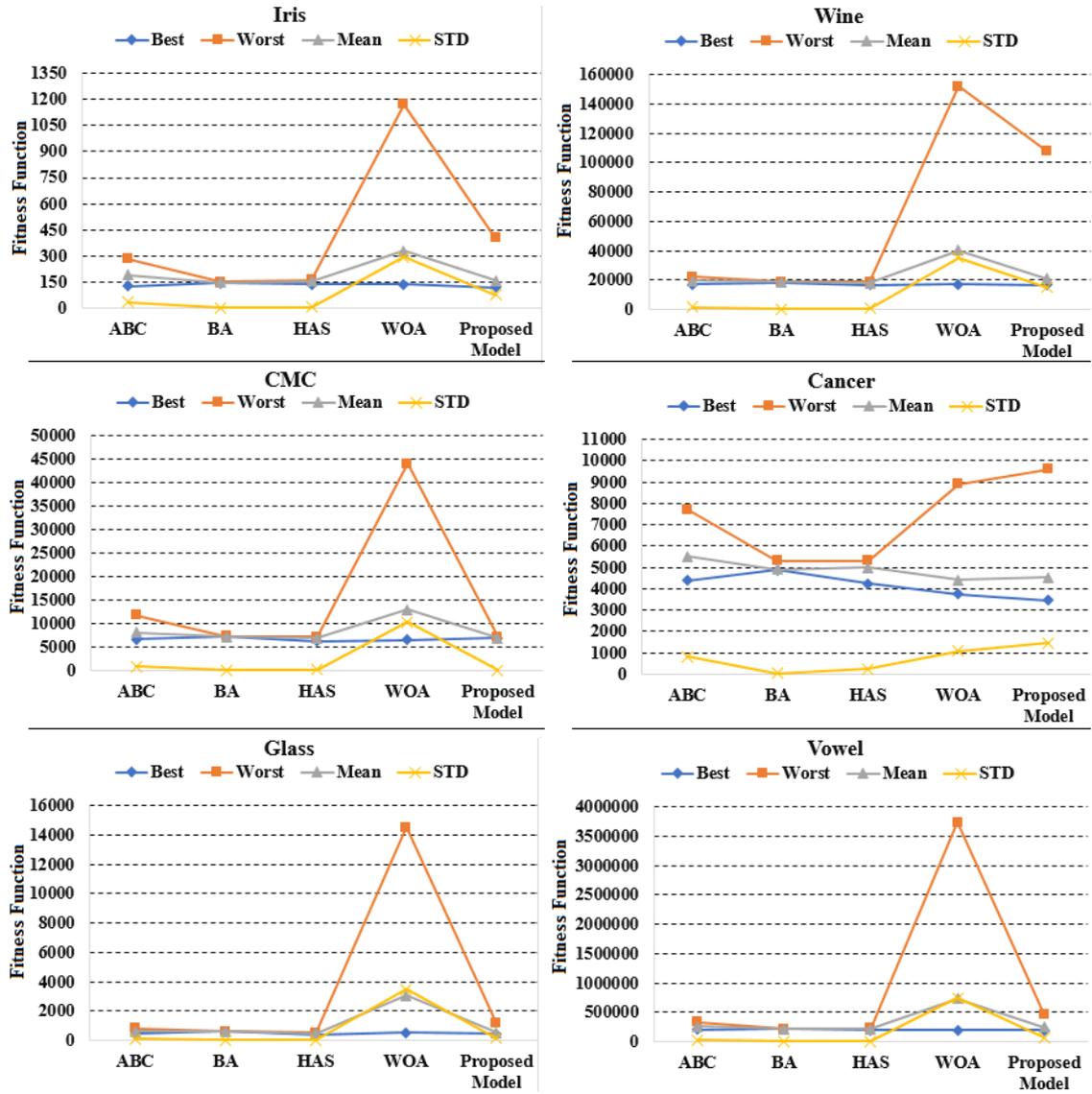


Fig. 3. Comparison diagram of the proposed method with different data clustering algorithms

proposed method is stronger for data clustering.

A careful review of the results of the best solution shows that the proposed method performs better than the standard algorithm in all six datasets, hence the efficiency improvement in the BA was done well, and on the other hand the proposed method compared to other comparison algorithms in four datasets better performed. It can also be seen that the proposed method has shown its superiority in 5 datasets compared to WOA. Examining the standard deviation of the proposed method, we find that the proposed method improves the whole population in most datasets and in general, compared to the meta-heuristic algorithm, the HSA, the ABC, and the WOA and BA, it has been able to perform stronger and better. In general, the results show the superiority of the proposed method compared to other algorithms.

## V. CONCLUSION

In the proposed method, after a thorough review of BA, the weaknesses of this algorithm in exploitation and exploration were identified, and the proposed method focused more on improving BA exploitation. Therefore, in the proposed method, instead of randomly selecting one of the best answers and replacing some of the location vector dimensions with the dimensions in BA, we changed some of the best solutions by reducing the siege mechanism and updating the WOA helix and the BA exploitation stage. Finally, after selecting the best exploitation from the two stages of WOA exploitation and BA exploitation, the desired changes were applied to the optimal solutions. The performance of the proposed method was evaluated on six UCI datasets. The proposed method had a better on five datasets compared to WOA. Examining the standard deviation of the proposed method, we found that the proposed method in most of the data sets has improved the whole population well and in general, compared to the harmonic search algorithm and the artificial bee algorithm and WOA and BA, it has been able to perform stronger and better. For future work, we plan to use a hybrid of WOA and the Gray Wolf Algorithm to cluster the data.

## REFERENCES

1. Soleimanian Gharehchopogh, F. and S. Haggi, An Optimization K-Modes Clustering Algorithm with Elephant Herding Optimization Algorithm for Crime Clustering. *Journal of Advances in Computer Engineering and Technology*, 2020. 6(2): p. 79-90.
2. Guha, R., et al., Introducing clustering based population in Binary Gravitational Search Algorithm for Feature Selection. *Applied Soft Computing*, 2020. 93: p. 106341.
3. Gharehchopogh, F.S., I. Maleki, and S.R. Khaze, A new optimization method for dynamic travelling salesman problem with hybrid ant colony optimization algorithm and particle swarm optimization. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 2013. 2(2): p. 352-358.
4. Gharehchopogh, F.S. and H. Gholizadeh, A comprehensive survey: Whale Optimization Algorithm and its applications. *Swarm and Evolutionary Computation*, 2019. 48: p. 1-24.
5. Xiao, Y., et al., Optimal mathematical programming and variable neighborhood search for k-modes categorical data clustering. *Pattern Recognition*, 2019. 90: p. 183-195.
6. Sun, L., et al., Combining density peaks clustering and gravitational search method to enhance data clustering. *Engineering Applications of Artificial Intelligence*, 2019. 85: p. 865-873.
7. Rabani, H. and F. Soleimanian Gharehchopogh, An Optimized Firefly Algorithm based on Cellular Learning Automata for Community Detection in Social Networks. *Journal of Advances in Computer Research*, 2019. 10(3): p. 13-30.
8. Rahnema, N. and F.S. Gharehchopogh, An improved artificial bee colony algorithm based on whale optimization algorithm for data clustering. *Multimedia Tools and Applications*, 2020: p. 1-26.
9. Huang, X., et al., DSKmeans: A new kmeans-type approach to discriminative subspace clustering. *Knowledge-Based Systems*, 2014. 70: p. 293-300.
10. Yang, X.-S., A New Metaheuristic Bat-Inspired Algorithm, in *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)*, J.R. González, et al., Editors. 2010, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 65-74.
11. Mirjalili, S. and A. Lewis, The Whale Optimization Algorithm. *Advances in Engineering Software*, 2016. 95: p. 51-67.
12. Osmani, A., J.B. Mohasefi, and F.S. Gharehchopogh, Sentiment Classification Using Two Effective Optimization Methods Derived From The Artificial Bee Colony Optimization And Imperialist Competitive Algorithm. *The Computer Journal*, 2020.
13. Shayanfar, H. and F.S. Gharehchopogh, Farmland fertility: A new metaheuristic algorithm for solving continuous optimization problems. *Applied Soft Computing*, 2018. 71: p. 728-746.
14. Gharehchopogh, F.S. and S.K. Mousavi, A New Feature Selection in Email Spam Detection by Particle Swarm Optimization and Fruit Fly Optimization Algorithms. *Journal of Computer and Knowledge Engineering*, 2019. 2(2).
15. Gharehchopogh, F.S., H. Shayanfar, and H. Gholizadeh, A comprehensive survey on symbiotic organisms search algorithms. *Artificial Intelligence Review*, 2019: p. 1-48.
16. Gharehchopogh, F.S., S.R. Khaze, and I. Maleki, A new approach in bloggers classification with hybrid of k-nearest neighbor and artificial neural network algorithms. 2015.
17. Abedi, M. and F.S. Gharehchopogh, An improved opposition based learning firefly algorithm with dragonfly algorithm for solving continuous optimization problems. *Intelligent Data Analysis*, 2020. 24: p. 309-338.
18. Gandomi, A.H. and X.-S. Yang, Chaotic bat algorithm. *Journal of Computational Science*, 2014. 5(2): p. 224-232.
19. Lin, J.-H., et al. A Chaotic Levy Flight Bat Algorithm for Parameter Estimation in Nonlinear Dynamic Biological Systems. in *CIT 2012*. 2012.
20. Sabba, S. and S. Chikhi, A discrete binary version of bat algorithm for multidimensional knapsack problem. *Int. J. Bio-Inspired Comput.*, 2014. 6(2): p. 140-152.
21. Zhou, Y., et al., A Hybrid Bat Algorithm with Path Relinking for the Capacitated Vehicle Routing Problem, in *Metaheuristics and Optimization in Civil Engineering*, X.-S. Yang, G. Bekdaş, and S.M. Nigdeli, Editors. 2016, Springer International Publishing: Cham. p. 255-276.
22. Cai, X., X.-z. Gao, and Y. Xue, Improved bat algorithm with optimal forage strategy and random disturbance strategy. *Int. J. Bio-Inspired Comput.*, 2016. 8(4): p. 205-214.
23. Zhu, B., et al., A Novel Quantum-Behaved Bat Algorithm with Mean Best Position Directed for Numerical Optimization. *Computational Intelligence and Neuroscience*, 2016. 2016: p. 6097484.
24. Yammani, C., S. Maheswarapu, and S.K. Matam, A Multi-objective Shuffled Bat algorithm for optimal placement and sizing of multi distributed generations with different load models. *International Journal of Electrical Power & Energy Systems*, 2016. 79: p. 120-131.
25. Nakamura, R.Y.M., et al. BBA: A Binary Bat Algorithm for Feature Selection. in *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images*. 2012.

26. Mirjalili, S., S.M. Mirjalili, and X.-S. Yang, Binary bat algorithm. *Neural Computing and Applications*, 2014. 25(3): p. 663-681.
27. Yilmaz, S., E.U. Kucuksille, and Y. Cengiz, Modified bat algorithm. *ELEKTRONIKA IR ELEKTROTECHNIKA*, 2014. 20(2): p. 71-78.
28. Li, L. and Y. Zhou, A novel complex-valued bat algorithm. *Neural Computing and Applications*, 2014. 25(6): p. 1369-1381.
29. Mallikarjuna, B., K. Reddy, and O. Hemakesaavulu, Economic load dispatch problem with valve-point effect using a binary bat algorithm. *ACEEE International Journal of Electrical and Power Engineering*, 2013. 4(3): p. 33-38.
30. Xiaodong, W., J. ZHANG, and H. XUE, K-Means Clustering Algorithm Based on Bat Algorithm. *Journal of Jilin University (Information Science Edition)*, 2016. 6.
31. Sood, M. and S. Bansal, K-Medoids Clustering Technique using Bat Algorithm. *International Journal of Applied Information Systems*, 2013. 5: p. 20-22.
32. Nguyen, T.-T., et al. Hybrid Bat Algorithm with Artificial Bee Colony. in *Intelligent Data analysis and its Applications*, Volume II. 2014. Cham: Springer International Publishing.
33. Murugan, R., et al., Hybridizing bat algorithm with artificial bee colony for combined heat and power economic dispatch. *Applied Soft Computing*, 2018. 72: p. 189-217.
34. Luo, J., F. He, and J. Yong, An efficient and robust bat algorithm with fusion of opposition-based learning and whale optimization algorithm. *Intelligent Data Analysis*, 2020. 24: p. 581-606.
35. Safara, F., et al., An Author Gender Detection Method Using Whale Optimization Algorithm and Artificial Neural Network. *IEEE Access*, 2020. 8: p. 48428-48437.
36. Zhu, L.F., et al., Data Clustering Method Based on Improved Bat Algorithm With Six Convergence Factors and Local Search Operators. *IEEE Access*, 2020. 8: p. 80536-80560.
37. Calixto, V. and G. Celani. A literature review for space planning optimization using an evolutionary algorithm approach: 1992-2014. 2015.
38. Lim, S.M. and K.Y. Leong, A Brief Survey on Intelligent Swarm-Based Algorithms for Solving Optimization Problems. In *Nature-inspired Methods for Stochastic Robust and Dynamic Optimization*, 2018.
39. website1, <https://archive.ics.uci.edu/ml/index.php>. 2020.