ISSN 2007-9737

A Strategy for Finding the Optimal Number of Clusters Based on the Grey Wolf Algorithm

Luis Rodríguez, Juan Barraza, Oscar Castillo*, Fevrier Valdez, Patricia Melin

Instituto Tecnologico de Tijuana /TecNM,
Mexico

luis.rodriguezm@tectijuana.edu.mx,
{jfbarraza, ocastilo, fevrier, pmelin} @tectijuana.mx

Abstract. In computational sciences literature, we can find different problems that can be solved by Optimization Algorithms. In this case, we are using an Optimization Algorithm based on wolves' behavior for a clustering problem. This study proposes an optimization approach based on the Grey Wolf Optimizer (GWO), designed to determine the optimal number of clusters by leveraging centroid data. We consider that the mathematic model used by the authors into the mentioned algorithm, it is the ideal to solve a clustering problem. Therefore, we tested the proposed algorithm with three different datasets: Iris, Wine and Diagnostic Wisconsin Breast Cancer Database, respectively, and we called the method Clustering Grey Wolf Optimizer, and we denoted as CGWO. Besides, to test the proposed method, we show a comparison of results versus the Firework Algorithm. We presented a statistical comparison between both mentioned methods, GWO and FWO, for clustering with the main objective of complementing the conclusions of this work.

Keywords. GWO, CGWO, optimization algorithms, algorithms, wolves, fireworks, iris dataset, wine dataset, WDBC dataset.

1 Introduction

Computer science [12, 26] has become important in everyday life in society that helps to improve her daily activities. This important event motivates researches to search new methods in order to improve the results in each research.

The main objective is solving optimization problems [27] based on inspirations that the natural environment solves in real word, some examples are the bio-inspired metaheuristics [11, 18, 19], such as, Gravitational Search Algorithm (GSA), Particle Swarm Optimization (PSO), Genetic Algorithm (GA), String Theory Algorithm (STA) and algorithms that we can read in literature depending of the inspiration that was research it [2, 7, 10].

In the optimization area it's important to evaluate the problem that we can solve; the optimization problems [1] can be classifications as minimize or maximize in order to improve potential solutions according to the problems or objective functions. In addition, mathematical problems such as benchmark functions are important in order to probe the optimization methods [8, 22, 23, 31].

This paper has as objective the clustering validation as multi-objective problems, in other words, we should validate clustering with intra and inter cluster that we explain in detail in a section of this paper.

To address these challenges, various optimization techniques have been developed, many of which are inspired by biological processes and natural phenomena.

These methods aim to provide robust solutions that can adapt to the dynamic and complex nature of real-world problems. As the complexity of optimization problems increases, so does the need for innovative approaches that can handle large datasets, high dimensionality, and conflicting objectives. Therefore, researchers have focused on combining metaheuristics with clustering algorithms to achieve more accurate and efficient results.

Actually, there exist a lot of variety and methods of clustering algorithms [20], that main objective and applications are to solve classification problems [9], such as in image segmentation, pattern recognition and other sciences with optimal results [30].

In addition, it is worth noting that one of the challenges in clustering algorithms [14] is determining the number of clusters [20] in an arbitrary manner.

This often leads to optimization issues, as identifying the optimal number of clusters can be difficult for certain datasets, especially when data points exhibit very similar characteristics.

In this context, the primary contribution of this research is the proposal of methods that automatically search for the optimal number of clusters based on the dataset's centroids. It is important to highlight that these methods are grounded in computational intelligence techniques.

For solve these problems we have used the Grey Wolf Optimizer (GWO) in order to solve clustering problems, this metaheuristic is adapted in order to support and find optimal solutions according of the optimization problems theses adaptations are explain in a section of this research work where each dataset is represented as dimension in the problem that allow evaluate complex problems [24, 25].

One of the most important things is to show the well performance of the any optimization algorithm, for this case, and, as we mentioned above, we adapted the Grey Wolf Optimizer [13] to finding the optimal number of clusters.

For this reason, we show the good adaptability of the GWO Algorithm to solve different problems, that is to say, in the first instance, the authors of GWO algorithm in Benchmark Mathematics functions is applied, and, in second instance, we applied the GWO in clustering problem, which indicates the robustness and great adaptability of the GWO Algorithm in different problems.

The adaptation of the GWO in clustering problem is shown in the following sections.

The structure of this paper is as follows: Section 2 provides an overview of the standard Grey Wolf Optimizer algorithm. Section 3 introduces the proposed approach, the Clustering Grey Wolf Optimizer (CGWO). Section 4 presents the experimental setup and results, while Section 5 concludes the study.

2 Conventional GWO Algorithm

Seyedali Mirjalili in 2014 original published the article on the Grey Wolf Optimizer, where the author was based on the biological study of Muro [13], in order to represent the mathematical model of the behavior of the grey wolf with the main following features: hierarchy of the members of the pack and the strategy of hunting.

Mirjalili presented the following equations:

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

where the variable D represents de distance between the best Wolf $X_p(t)$ with a random movement C and the current positions that going to evaluate $X(t)$:

$$X(t+1) = X_p(t) - AD. \tag{2}$$

Finally, the $X(t+1)$ represents the new position that the potential solution will have in the next iteration, and A represents a value that decreases over time with equation 2 and D are de distance that was calculated previously:

$$A = 2a \cdot r_1 - a, \tag{3}$$

$$C = 2 \cdot r_2. \tag{4}$$

In order to improve the potential solutions, and according to the biological inspiration, the previous equations are implemented with the three best solutions in the algorithm and are called alpha, beta, and gamma respectively. Finally, the original author presented the new positions because of on average among these three best results, and we can note this in the following equations:

$$\begin{aligned} D_\alpha &= \|C_1 \cdot X_\alpha - X\|, \\ D_\beta &= \|C_2 \cdot X_\beta - X\|, \\ D_\delta &= \|C_3 \cdot X_\delta - X\|, \end{aligned} \tag{5}$$

$$\begin{aligned} X_1 &= X_\alpha - A_1 \cdot (D_\alpha), \\ X_2 &= X_\beta - A_2 \cdot (D_\beta), \\ X_3 &= X_\delta - A_3 \cdot (D_\delta), \end{aligned} \tag{6}$$

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3}. \tag{7}$$

It is worth noting that this average could be enhanced by employing a weighted average or by

ISSN 2007-9737

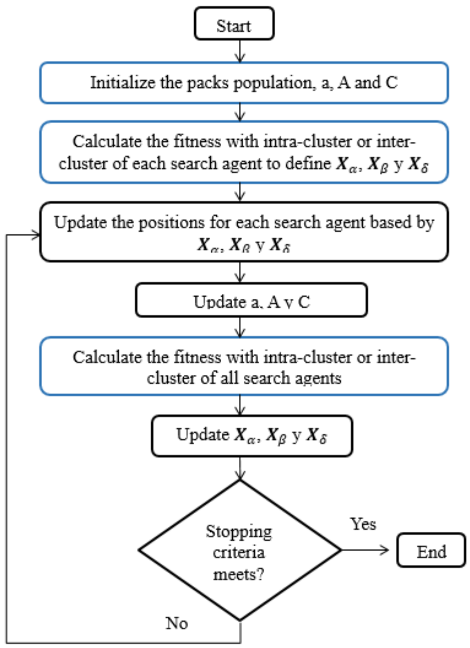


Fig. 1. Flow chart of CGWO

developing an alternative computational intelligence approach, such as fuzzy logic, to further optimize the algorithm's performance [3, 6, 16].

3 Proposed GWO for Clustering

In these sections we show a variant of the GWO Algorithm, we modified a conventional GWO algorithm in a clustering problem, i.e., with the main goal of finding the optimal number of clusters and we denoted as CGWO (Clustering Grey Wolf Optimization).

CGWO is an adaptation of the Grey Wolf Optimizer (GWO), originally introduced by S. Mirjalili in 2014 [13]. While the original GWO was developed to minimize or identify global optima in mathematical benchmark functions, this paper presents a modified strategy that applies the algorithm to determine the optimal number of clusters. In most clustering algorithms the number of clusters is manually initialized, and this process could be tedious for researchers. For this reason, we have adapted the GWO algorithm into the mentioned optimization problem, i.e., in contrast to initializing the number of clusters every execution algorithm manually in different databases, the GWO algorithm is in charge to finding the optimal number of clusters based on the data features of every database.

This approach not only facilitates the clustering process but also improves efficiency in data segmentation. By allowing the GWO algorithm to automatically determine the appropriate number of clusters, manual intervention is avoided, and execution time is optimized, which is particularly useful when working with large datasets or databases whose structure is unknown. Moreover, the adaptability of the GWO ensures that the results are more accurate and aligned with the intrinsic features of the data.

Following in the optimization algorithm line, a clustering optimization method performs optimally when the number of clusters for a given dataset is known. However, challenges emerge when this number is unknown. Therefore, the adaptation of the Grey Wolf Optimizer (GWO) to determine the optimal number of clusters is well justified.

The adaptation of the GWO algorithm mentioned above is explained in detail in the following five steps:

- Obtain the quality of wolves' solution (calculate the fitness to obtain alpha, beta y delta wolves).
- Update the positions based on alpha, beta y delta.
- Evaluate the fitness of all wolves' solutions.
- Update alpha, beta y delta wolves and a, A and C parameters.

For more details, in Figure 1 we illustrate the flow chart of the Clustering Grey Wolf Algorithm (CGWO):

The steps mentioned above in CGWO explain the detail below:

Step 1.- Initialize packs population:

$$\mathrm{Pack}_{ij} = \mathrm{lb}_j + \left(\mathrm{ub}_j - \mathrm{lb}_j\right) * \mathrm{r}_{ij}, \quad \mathrm{i} = 1,2,3, \ldots \mathrm{n}. \tag{8}$$

Here, ub and lb represent the upper and lower limits for individual *i* in feature *j*. The dimensionality of the wolves is determined by the number of features in the given dataset, and r is a randomly generated value within the range [0,1].

ISSN 2007-9737

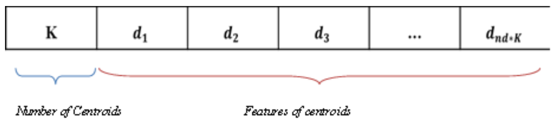
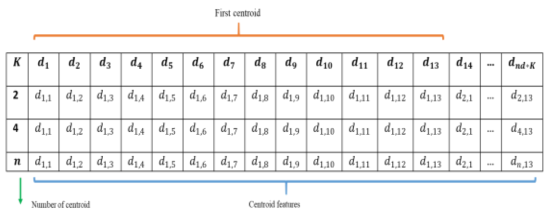
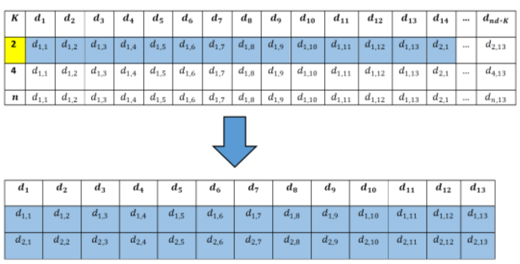


Fig. 2. Representation of the CGWO Wolves



K	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	...	d_{nd*K}
2	$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	$d_{1,4}$	$d_{1,5}$	$d_{1,6}$	$d_{1,7}$	$d_{1,8}$	$d_{1,9}$	$d_{1,10}$	$d_{1,11}$	$d_{1,12}$	$d_{1,13}$	$d_{2,1}$	...	$d_{2,13}$
4	$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	$d_{1,4}$	$d_{1,5}$	$d_{1,6}$	$d_{1,7}$	$d_{1,8}$	$d_{1,9}$	$d_{1,10}$	$d_{1,11}$	$d_{1,12}$	$d_{1,13}$	$d_{2,1}$	...	$d_{4,13}$
n	$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	$d_{1,4}$	$d_{1,5}$	$d_{1,6}$	$d_{1,7}$	$d_{1,8}$	$d_{1,9}$	$d_{1,10}$	$d_{1,11}$	$d_{1,12}$	$d_{1,13}$	$d_{2,1}$	...	$d_{n,13}$

Fig. 3. Matrix representation of the wolves in CGWO



K	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	...	d_{nd*K}
2	$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	$d_{1,4}$	$d_{1,5}$	$d_{1,6}$	$d_{1,7}$	$d_{1,8}$	$d_{1,9}$	$d_{1,10}$	$d_{1,11}$	$d_{1,12}$	$d_{1,13}$	$d_{2,1}$	...	$d_{2,13}$
4	$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	$d_{1,4}$	$d_{1,5}$	$d_{1,6}$	$d_{1,7}$	$d_{1,8}$	$d_{1,9}$	$d_{1,10}$	$d_{1,11}$	$d_{1,12}$	$d_{1,13}$	$d_{2,1}$	...	$d_{4,13}$
n	$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	$d_{1,4}$	$d_{1,5}$	$d_{1,6}$	$d_{1,7}$	$d_{1,8}$	$d_{1,9}$	$d_{1,10}$	$d_{1,11}$	$d_{1,12}$	$d_{1,13}$	$d_{2,1}$	...	$d_{n,13}$

d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}
$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	$d_{1,4}$	$d_{1,5}$	$d_{1,6}$	$d_{1,7}$	$d_{1,8}$	$d_{1,9}$	$d_{1,10}$	$d_{1,11}$	$d_{1,12}$	$d_{1,13}$
$d_{2,1}$	$d_{2,2}$	$d_{2,3}$	$d_{2,4}$	$d_{2,5}$	$d_{2,6}$	$d_{2,7}$	$d_{2,8}$	$d_{2,9}$	$d_{2,10}$	$d_{2,11}$	$d_{2,12}$	$d_{2,13}$

Fig. 4. Vector to matrix conversion example in CGWO

To provide a clearer understanding, an example using the Wine dataset is presented, which consists of 178 samples, each characterized by 13 features.

In Figure 3 we show an example of the vector of solution in Clustering Grey Wolf Optimizer, that it to say, an example of the wolfs representation as possible solution is presented. The vector solution in two parts is divided, the first part (first dimension) represents by means of an integer number the number of centroids and, the second part (remainder of the vector) represents by means of a real number the features of each data for the given dataset. The example is listed below:

- First part of vector solution is K.
- Second part of vector solution are: $\mathrm{d}_1, \mathrm{d}_2, \mathrm{d}_3, \dots, \mathrm{d}_{\mathrm{nd}*\mathrm{K}}$.

Where nd are the features of each data.

As we mentioned above, the data of the wine dataset has thirteen features per data, therefore, the features of the first centroid are representing by the cells two $d_{1,1}$, three $d_{1,2}$, four $d_{1,3}$, five $d_{1,4}$, six $d_{1,5}$, seven $d_{1,6}$, eight $d_{1,7}$, nine $d_{1,8}$,ten $d_{1,9}$, eleven $d_{1,10}$, twelve $d_{1,11}$, thirteen $d_{1,12}$ and fourteen $d_{1,13}$.

The need to convert the vector into a matrix to avoid problems in processing computation based on mathematical calculus is notable, therefore, in Figure 4 we show a conversion from matrix to vector as an example.

As we can note in Figure 4, the blue cells represent the real numbers of the centroid 1 and centroid 2 for a wolf in CGWO.

In order to reduce the adverse effects associated with high vector dimensionality (directly influenced by the number of clusters K) this study employs two well-established statistical heuristics: Sturges Rule and the Square Root Rule. These approaches are utilized to estimate the upper bound for the number of centroids, denoted as $\mathrm{K}_{\mathrm{upper}}$ or K_{max}.

Continuing with the Wine dataset example, the equation of the first one rule mentioned above, in Equation 9 is shown:

$$\underline{\mathrm{K}_{\mathrm{upper}}} = 1 + 3.322 * \log \mathbf{178}. \qquad (9)$$

In Equation 9 we demonstrate the example with Wine dataset, where the number in bold represents the number of data on dataset mentioned.

As we know and we mentioned in previous sections, the number of K needs to be an integer number, thus, we decided to implement the floor of the total result in the Equation to represent the $\mathrm{K}_{\mathrm{upper}}$ as integer number. In this case, the total result of Equation 9 is equal to 9.7623, and after applying the floor rounding rule at the total result, the real total value of Equation is equal to 9.

Then, we have a range between 2 and 9 number of centroids K, that is to say, CGWO algorithm will find the optimal number of centroids from $\mathrm{K}_{\mathrm{lower}}$=2 to $\mathrm{K}_{\mathrm{upper}}$=9, where $\mathrm{K}_{\mathrm{lower}}$ is the number recommended by researches and $\mathrm{K}_{\mathrm{upper}}$ is obtained with the rule.

Similar to the approach used in Equation 9, Equation 10 applies the floor function to the resulting value in order to obtain an integer that defines the number of centroids K. Accordingly, the second heuristic, the square root of N is illustrated using the Wine dataset as an example in Equation 10:

ISSN 2007-9737

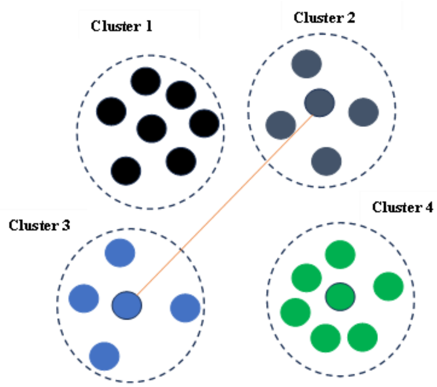


Fig. 5. Inter-cluster distance

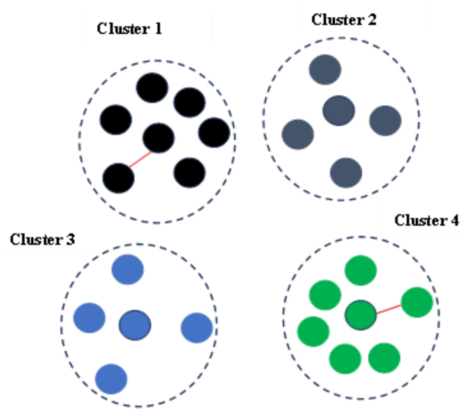


Fig. 6. Intra-cluster distance

$$\underline{K_{upper}} = \sqrt{\mathbf{178}}. \qquad (10)$$

In this case, the total result of Equation 10 is equal to 13.3416, and after applying the floor rounding rule to obtain K_{upper}, the real total value is equal to 13, in such a way that the CGWO algorithm could find the optimal value of centroid in a range between 2 and 13.

This upper bound K_{upper} serves as a crucial parameter to limit the search space, making the optimization process more efficient and computationally feasible. By narrowing the range of possible cluster numbers, CGWO can converge more quickly while still exploring diverse clustering configurations within a meaningful boundary.

Both Sturges Rule and the Square Root Rule serve as heuristic methods to estimate the appropriate number of clusters based on the total number of data points within a given dataset. In the context of CGWO, these rules allow for an approximate determination of the number of clusters derived directly from the dataset size.

These estimation rules not only guide the algorithm in defining the initial search space but also provide flexibility to adapt across different datasets with varying sizes and features. By using data-driven heuristics like Sturges [21] and the square root of N, CGWO becomes more generalizable and scalable, which is particularly useful when working with unknown or high-dimensional data distributions.

Steps 2 and 4.- Obtain the quality of wolf's solutions and evaluate the fitness of all wolves' solutions.

The following step in CGWO is to evaluate the quality of each wolf's solution. In this clustering problem, we applied two basics metrics to obtain the performance of each cluster in CGWO Algorithm. Firstly, we implemented the Inter-cluster metric, which is shown in Equation 11:

$$\mathrm{InterCluster} = \sum_{\substack{i,j=1, \\ i \neq j}}^{n} \mathrm{distance}\,(K_i, K_j). \qquad (11)$$

Inter-cluster metric refers to obtaining the distance between two or more different centroids [28]. In Equation 11 i and j represents the number identification of centroid, on this way, obtain the distance between two centroids represents by K_i and K_j, and n is the maximum number of centroids in every individual solution into the CGWO as we show in Figure 5.

The orange line in Figure 5 represents the inter-cluster distance.

Secondly, we have implemented the Intra-cluster metric to obtain the quality of the clusters, which refers to the distance value between the centroid and their data points belonging at the centroid. Equation 12 shows the Intra-cluster metric:

$$\mathrm{IntraCluster} = \sum_{i=1}^{n} \mathrm{distance}(k_i, K). \qquad (12)$$

where k_i are the data that belong to centroid K.

The Intra-cluster metric in Figure 6 is shown where the red line represents the Intra-cluster distance [17].

ISSN 2007-9737

Table 1. Minimum distance with Inter-cluster in CGWO

Dataset	K optimal	Square Root of N		Sturges	
		Mean	STD	Mean	STD
WDBC	2	3.18	1.16	**2.27**	0.63
IRIS	3	**2.42**	0.61	2.33	0.60
WINE	3	**2.73**	0.84	2.45	0.67

Table 2. Maximum distance with Inter-cluster in CGWO

Dataset	K optimal	Square Root of N		Sturges	
		Mean	STD	Mean	STD
WDBC	2	22.8	1.08	9.48	0.71
IRIS	3	11.1	0.87	7.45	0.46
WINE	3	12.2	0.87	7.48	0.67

Table 3. Minimum distance with Intra-cluster in CGWO

Dataset	K optimal	Square Root of N		Sturges	
		Mean	STD	Mean	STD
WDBC	2	8.94	6.60	3.76	1.80
IRIS	3	4.55	2.25	**3.73**	1.74
WINE	3	5.52	3.19	4.09	1.86

Table 4. Maximum distance with Intra-cluster in CGWO

Dataset	K optimal	Square Root of N		Sturges	
		Mean	STD	Mean	STD
WDBC	2	4.06	1.56	**3.15**	1.06
IRIS	3	3.67	2.10	4.06	1.62
WINE	3	4.12	1.92	**3.70**	1.42

It is important to mention that to evaluate the fitness of all wolves' solutions as in step 4 is described, we have used the two metrics mentioned recently, inter and intra-cluster, respectively with minimum and maximum distance.

Steps 3 and 5 refers to updating alpha, beta y delta wolves and a, A and C parameters. For this problem we used the hierarchical method implemented in a modification of the GWO Algorithm [15].

And to update A and C parameters, we used the equation used by author in the conventional GWO Algorithm. In the following section we show the performance of the applied Clustering Grey Wolf Optimizer Algorithm in this clustering problem.

4 Results and Discussion

In this Section we present the experiments of the proposed Clustering Grey Wolf Algorithm (CGWO); with the aim to evaluate the performance and efficiency in the clustering problem mentioned.

In this instance, the proposed method was implemented using two static metrics to determine the maximum number of centroids: Sturges' law and the square root of N. Additionally, we utilized the maximum and minimum distance measures to find the optimal solution, validated through both inter-cluster and intra-cluster criteria.

To ensure a comprehensive evaluation, we conducted the experiments under different conditions, varying the number of clusters and adjusting the algorithm's parameters. This helped us assess the robustness of the Clustering Grey Wolf Algorithm in different scenarios and its ability to adapt to different types of data distributions.

Three different datasets were used: Iris, Wine and WDBC dataset. The results shown in the following Tables were obtained with the 30 independents execution of the Clustering Grey Wolf Algorithm.

The execution parameters of the proposed method and datasets are listed below:

CGWO Parameters:

- Wolves = 30.
- Iterations = 500.
- a = [0, 2].
- r_1 = [0, 1].
- r_2 = [0, 1].
- A y C = [0, 2].

Dataset parameters:

- IRIS: Contains 150 samples, each with 4 attributes.
- WINE: Contains 178 samples, each with 13 attributes.
- WDBC: Contains 569 samples, each with 30 attributes.

Table 5. Square root of N with Inter-cluster and minimum distance

Dataset	K optimal	Square Root of N			
		CGWO		FWAC	
		Mean	ABS	Mean	ABS
WDBC	2	3.18	**1.18**	6.77	4.77
IRIS	3	2.42	**0.58**	4.03	1.03
WINE	3	2.73	**0.27**	4.03	1.03

Table 6. Sturges with Inter-cluster and minimum distance

Dataset	K optimal	Sturges			
		CGWO		FWAC	
		Mean	ABS	Mean	ABS
WDBC	2	2.27	**0.27**	4.45	2.45
IRIS	3	2.33	0.67	2.81	**0.19**
WINE	3	2.45	0.55	3.39	**0.39**

Table 7. Square root of N with Inter-cluster and maximum distance

Dataset	K optimal	Square Root of N			
		CGWO		FWAC	
		Mean	ABS	Mean	ABS
WDBC	2	22.8	20.8	9.61	**7.61**
IRIS	3	11.1	8.15	4.1	**1.1**
WINE	3	12.2	9.24	4.35	**1.35**

Table 8. Sturges with Inter-cluster and maximum distance

Dataset	K optimal	Sturges			
		CGWO		FWAC	
		Mean	ABS	Mean	ABS
WDBC	2	9.48	7.48	3.87	**1.87**
IRIS	3	7.45	4.45	3.1	**0.1**
WINE	3	7.48	4.48	3.48	**0.48**

Tables 1, 2 3 and 4 shows the experimental results with CGWO; within the tables we can observe the database name, k optimal for each data base, the mean and standard deviation using square root of N or Sturges law as upper k number bound with inter-cluster or intra-cluster using maximum or minimum distance, respectively.

Tables 1 and 2 show the data results using Inter-cluster as cluster validation, in Table 1 with minimum distance and Table 2 with maximum distance.

he best result for WDBC and Iris dataset with inter-cluster validation was obtained using Sturges law as upper k number bound and minimum distance in cluster validation. For Wine dataset was obtained using square root of n as upper k number bound and minimum distance in cluster validation too.

The results of CGWO using intra-cluster as cluster validation in Table 3 and 4 are presented, in Table 3 with minimum distance and in Table 4 with maximum distance.

The results of Tables 3 and 4 show that the convenient combination is using intra-cluster minimum distance with Sturges law for Iris dataset, and, for the WDBC and Wine dataset the convenient combination is using intra-cluster maximum distance with Sturges Law.

4.1 Comparison between CGWO and FWAC with Absolute Error

In this section we present a comparison between Clustering Grey Wolf Optimizer and Fireworks Algorithm for Clustering [4-6]. We are comparing an absolute error, which is obtained with the subtracting between mean and k optimal of each database. The comparison is using the variants mentioned above, i.e., upper k number bound, cluster validation and minimum and maximum distance, respectively [17,29].

As we can see, from Tables 5 to 12 the ABS (Absolute error) in bold shows the best performance of a variant given algorithm.

In Tables 5 and 6 CGWO has a better performance than FWAC in 4 of 6 possible variants.

In Tables 7 and 8 where we used Inter-cluster with maximum distance; the performance of CGWO was not satisfactory, i.e., the 6 possible variants, FWAC obtained 6 better performances than CGWO.

The comparison using Intra-cluster as cluster validations from Table 9 to 12 are presented.

We can note that the better performance was obtained with CGWO, i.e., in Tables 9 and 10 CGWO shows a better performance in 5 of 6 possible variants and, in Tables 11 and 12 shows better results in 6 of 6 possible variants for CGWO is shown.

Table 9. Square root of N with Intra-cluster and minimum distance

Dataset	K optimal	Square Root of N			
		CGWO		FWAC	
		Mean	ABS	Mean	ABS
WDBC	2	8.94	**6.94**	11.1	9.1
IRIS	3	4.55	**1.55**	5.13	2.13
WINE	3	5.52	**2.52**	8.94	5.94

Table 10. Sturges with Intra-cluster and minimum distance

Dataset	K optimal	Sturges			
		CGWO		FWAC	
		Mean	ABS	Mean	ABS
WDBC	2	3.76	**1.76**	6.16	4.16
IRIS	3	3.73	0.73	3.45	**0.45**
WINE	3	4.09	**1.09**	5.61	2.61

Table 11. Square root of N with Intra-cluster and maximum distance

Dataset	K optimal	Square Root of N			
		CGWO		FWAC	
		Mean	ABS	Mean	ABS
WDBC	2	4.06	**2.06**	14.9	12.9
IRIS	3	3.67	**0.67**	7	4
WINE	3	4.12	**1.12**	8.71	5.71

Table 12. Sturges with Intra-cluster and maximum distance

Dataset	K optimal	Sturges			
		CGWO		FWAC	
		Mean	ABS	Mean	ABS
WDBC	2	3.15	**1.15**	6.68	4.68
IRIS	3	4.06	**1.06**	5.06	2.06
WINE	3	3.7	**0.7**	5.13	2.13

4.2 Z-Test Comparison between CGWO and FWAC

In this section we show a hypothesis test comparison between CGWO and FWAC with the aim of reaffirming the best performance algorithm in a forceful manner.

We use the Z-test as statistical test for comparison with a significance level of 0.05, the alternative hypothesis stating that μ_1 is greater than μ_2 and, of course, the null hypothesis tells us that the μ_1 is lower than μ_2.

The main goal to apply for the z-test is to show the analysis of the proposed method, in this case, CGWO Algorithm. We will prove if there is significant evidence of the CGWO results being better of the FWAC Algorithm [4]; so, there is significant evidence if the value of Z score is less than -1.645, therefore we can say there exists enough evidence to affirm that CGWO Algorithm was better results than FWAC Algorithm is obtained.

Additionally, this analysis is critical for determining whether the CGWO Algorithm can be considered superior not just in isolated tests, but across various experimental conditions. This would provide a more robust argument for the use of CGWO in real-world applications over FWAC [4].

The elements of the Z-test are listed below:

- $\mu_1 = CGWO\ Mean.$
- $\sigma_1 = CGWO\ Standard\ deviation.$
- $\mu_2 = FWAC\ Mean.$
- $\sigma_2 = FWAC\ Standard\ deviation.$
- $Confidence\ Interval = 95\%.$
- $\alpha = 5\%.$
- $Critical\ value = -1.645$

Figure 7 shows the critical value of the rejection region. The critical value is negative because the graphic is on the left tail chart.

Alternative and null hypothesis are enlisted below:

- $H_0: \mu_1 \geq \mu_2$
- $H_1: \mu_1 < \mu_2 (claim)$

Equation 13 demonstrates the z value, which is the value that allows accept or rejected the null hypothesis:

$$Z = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}. \quad (13)$$

The results of the statistical Z-Test apply from Tables 13 to 20 are shown, which are obtained of comparison between CGWO and FWAC Algorithm [10]. The best results of Z-value with bold letters are marked.

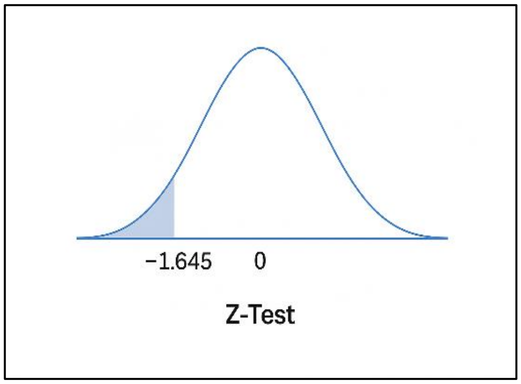


Fig. 7. Critical value for Z-test (left tail chart)

Table 13. Z-test with Inter-cluster and minimum distance

	Square Root of N				
Dataset	CGWO		FWAC		Z-value
	Mean	STD	Mean	STD	
WDBC	3.18	1.16	6.77	5.44	**-3.59**
IRIS	2.42	0.61	4.03	2.01	**-4.26**
WINE	2.73	0.84	4.03	2.03	**-3.29**

Table 14. Z-test with Inter-cluster and minimum distance

	Sturges				
Dataset	CGWO		FWAC		Z-value
	Mean	STD	Mean	STD	
WDBC	2.27	0.63	4.45	2.51	**-4.69**
IRIS	2.33	0.6	2.81	0.95	**-2.37**
WINE	2.45	0.67	3.39	1.69	**-2.87**

Table 15. Z-test with Inter-cluster and maximum distance

	Square Root of N				
Dataset	CGWO		FWAC		Z-value
	Mean	STD	Mean	STD	
WDBC	22.8	1.08	9.61	5.69	12.7
IRIS	11.1	0.87	4.1	1.68	20.7
WINE	12.2	0.87	4.35	2.18	18.7

Tables 13 and 14 show the Z-test of the Inter-cluster with minimum distance. In the six z-test the evidence is significant, i.e., when used square root of N or Sturges law as upper k number bound, CGWO obtained better results than FWAC in the three different Databases.

On the contrary, the results were not satisfactory with Inter-cluster and maximum distance. Tables 15 and 16 present the results; FWAC is better than CGWO in the six Z-test applied.

For the intra-cluster with minimum distance 3 of 6 Z-tests shows the significant evidence; the results are shown in Tables 17 and 18.

Tables 19 and 20 present 6 of s6 Z-tests with significant evidence of the CGWO when we are using Intra-cluster with maximum distance.

Based on the statistical Z-tests results, we can make the conclusion that the results of Clustering Grey Wolf Optimizer (CGWO) are better than Fireworks Algorithm for Clustering (FWAC); we achieved a 63 % of significant evidence in the Z- tests.

5 Conclusions

In this work, we have presented a variant of Grey Wolf Optimizer (GWO), that is to say, we adapted the conventional GWO to clustering problems, which was named as Clustering Grey Wolf Optimizer and denoted as CGWO.

To carry out the clustering task, CGWO uses the natural hierarchy and leadership-based movement of wolves to search the solution space efficiently.

The algorithm assigns cluster centers and iteratively adjusts them to minimize intra-cluster distance while maximizing inter-cluster separation. This bio-inspired mechanism allows CGWO to escape local optima and improve the global search capacity of the optimization process.

Additionally, we incorporated various upper bound strategies for the number of clusters (k), including Sturges rule, square root of N, and maximum/minimum distances. These heuristics enabled the algorithm to explore a diverse set of clustering configurations, enhancing its adaptability across different types of datasets and dimensional complexities.

In summary, we can say that the main goal of this paper was achieved because as we mentioned, we adapted the conventional GWO for finding the k optimal number and we tested the

Table 16. Z-test with Inter-cluster and maximum distance

Dataset	Sturges				
	CGWO		FWAC		Z-value
	Mean	STD	Mean	STD	
WDBC	9.48	0.71	3.87	1.59	17.9
IRIS	7.45	0.46	3.1	1.22	18.5
WINE	7.48	0.67	3.48	1.36	14.6

Table 17. Z-test with Intra-cluster and minimum distance

Dataset	Square Root of N				
	CGWO		FWAC		Z-value
	Mean	STD	Mean	STD	
WDBC	8.94	6.6	11.1	6.98	-1.25
IRIS	4.55	2.25	5.13	2.13	-1.04
WINE	5.52	3.19	8.94	3.23	**-4.19**

Table 18. Z-test with Intra-cluster and minimum distance

Dataset	Sturges				
	CGWO		FWAC		Z-value
	Mean	STD	Mean	STD	
WDBC	3.76	1.8	6.16	2.41	**-4.44**
IRIS	3.73	1.74	3.45	1.59	0.66
WINE	4.09	1.86	5.61	2.16	**-2.96**

Table 19. Z-test with Intra-cluster and maximum distance

Dataset	Square Root of N				
	CGWO		FWAC		Z-value
	Mean	STD	Mean	STD	
WDBC	4.06	1.56	14.9	6.21	**-9.46**
IRIS	3.67	2.1	7	2.66	**-5.47**
WINE	4.12	1.92	8.71	2.8	**-7.52**

performance of CGWO Algorithm in three different databases, and we compared the performance of CGWO against FWAC.

Based on the results, the performance of Clustering Grey Wolf Optimizer was satisfactory because results are better than Fireworks Algorithm for Clustering. CGWO obtained better results in 15 of 24 total possible variants, which is equivalent to 63% of variants.

We can separate the comparison according to cluster validation, i.e., Intra-cluster and Inter-cluster. In the results of Inter-cluster validation, CGWO obtained better results using minimum distance and square root of N as upper k number bound; on the contrary, when we use maximum distance, CGWO only shows better results than FWAC in 1 of 3 possible variants using Sturges rule as upper k number bound.

Otherwise, in the results of Intra-cluster validation, CGWO presents better performance in 11 of 12 possible variants, therefore, if we use minimum or maximum distance in Intra-cluster does not affect because CGWO behaves better than FWAC.

We can conclude each database used. For WDBC database the best performance of the CGWO Algorithm is using Inter-cluster with minimum distance and using Sturges law as upper k number bound, the results mark means of 2.27 and 0.63 of standard deviation, getting like this 0.27 of an absolute error according to 2 k optimal for WDBC database.

For Iris database with 3 k optimal, CGWO was obtained 2.42 and 0.61 of mean and standard deviation, respectively; getting an error of 0.58, the results were obtained using Inter-cluster and square root of N. And, for the Wine database we obtained an absolute error of 0.27, which is based on the difference between 3 k optimal and 2.73 as mean of CGWO, the standard deviation was 0.84.

On the other side, based on the results in tables 13, 14, 19 and 20, CGWO consistently outperforms FWAC across all datasets and both clustering heuristics.

The lower mean intra-cluster with minimum distance and inter-cluster with maximum distances achieved by CGWO indicate better compactness in clustering. Negative Z-values marked with bold letter further confirm the statistical significance of CGWO's superior performance. These findings validate CGWO as a more effective clustering algorithm than FWAC. On the contrary, the results of Tables 15, 16, 17 and 18 was not satisfactory; In Tables 15 and 16 CGWO algorithm presents a performed below expectations. In Tables 17 and 18, the CGWO algorithm only presents a 50 percent better performance than FWAC Algorithm. It is very important to mention that we used the complete number of features for each database, that is to say, in Diagnostic Wisconsin Breast Cancer Database we use 30 features; in Iris

Table 20. Z-test with Intra-cluster and maximum distance

	Sturges				
Dataset	CGWO		FWAC		Z-value
	Mean	STD	Mean	STD	
WDBC	3.15	1.06	6.68	2.45	**-7.36**
IRIS	4.06	1.62	5.06	1.97	**-2.18**
WINE	3.7	1.42	5.13	1.63	**-3.68**

database 4 variables are used, and we use 13 features in the Wine database.

To ensure consistency and comparability, all datasets were preprocessed using standard normalization techniques before applying the Clustering Grey Wolf Optimizer (CGWO). This step was crucial to avoid bias in the optimization process due to differences in the scale or distribution of the features.

Further work, we could apply the GWO in another optimization problems, such as, control or classification problems. Or else, we can try to adjust dynamically parameters with Fuzzy Logic and/or Interval Type 2 Fuzzy Logic in the Clustering Grey Wolf Optimizer; in addition, we can prove the performance of the CGWO in another databases.

Acknowledgments

This work has been funded by TecNM and Tijuana Institute of Technology.

References

1. **Abdulmajeed, N. H., Ayob, M. (2014).** A Firework Algorithm for Solving Capacitated Vehicle Routing Problem. International Journal of Advancements in Computing Technology, (IJACT), Vol. 6, No. 1, pp. 79-86.
2. **Aladwan, F., Alshraideh, M., Rasol, M. (2015).** A Genetic Algorithm Approach for Breaking of Simplified Data Encryption Standard. International Journal of Security and Its Applications, Vol. 9, No. 9, pp. 295–304.
3. **Barraza, J., Melin, P., Valdez, F., González, C.I. (2017).** Fuzzy fireworks algorithm based on a sparks dispersion measure. Algorithms, Vol. 10, No. 3, pp. 83.
4. **Barraza, J., Valdez, F., Melin, P., González, C. (2019).** Optimal number of clusters finding using the fireworks algorithm. Hybrid intelligent systems in control, pattern recognition and medicine, Springer International Publishing, pp. 83–93.
5. **Barraza, J., Rodriguez, L., Valdez, F., Melin, P., Castillo, O. (2024).** A Hybrid Fuzzy Fireworks and Grey Wolf Metaheuristic Algorithm. New Directions on Hybrid Intelligent Systems Based on Neural Networks, Fuzzy Logic, and Optimization Algorithms, Springer Nature Switzerland, pp. 195– 210.
6. **Barraza, J., Valdez, F., Melin, P., González, C. (2020).** Interval type 2 fuzzy fireworks algorithm for clustering. Handbook of Research on Fireworks Algorithms and Swarm Intelligence, IGI Global, pp. 195–211.
7. **Bonabeau, E., Dorigo, M., Theraulaz**, **G. (1999).** Swarm intelligence: from natural to artificial systems: OUP USA.
8. **Can Y. U., Alatas, B. (2015).** Physics based metaheuristic algorithms for global optimization. American Journal of Information Science and Computer Engineering, Vol. 1, pp. 94–106.
9. **Díaz-Ramírez, A., Diaz-Escobar, J., Quintero-Rosas, V., Moncada-Sánchez, R. (2024).** Classification of Fall Events in the Elderly Using a Thermal Sensor and Machine Learning Techniques. Computación y Sistemas, Vol. 28, No. 4, pp. 1773–1781.
10. **Dorigo, M., Birattari, M., Stutzle T. (2006).** Ant colony optimization. Comput Intell Magaz, pp. 28– 39.
11. **Gao, Z. M., Zhao, J. (2019).** An improved grey wolf optimization algorithm with variable weights. Computational intelligence and neuroscience, Vol. 2019, No. 1, 2981282.
12. **Jukan, A., Masip-Bruin, X., Amla, N. (2017).** Smart computing and sensing technologies for animal welfare: A systematic review. ACM Computing Surveys (CSUR), Vol. 50, No. 1, pp. 1– 27.

13. **Mirjalili, S., Mirjalili, S. M., Lewis, A. (2014).** Grey wolf optimizer. Advances in engineering software, Vol. 69, pp. 46–61.
14. **Nerurkar, P., Shirkeb, A., Chandanec, M., Bhirudd, S. (2018).** A novel Heuristic for evolutionary clustering. Procedia Computer Science, Vol. 125, pp. 780–789.
15. **Rodriguez, L., Castillo, O., Soria, J., Melin, P., Valdez, F., Gonzalez, C., Martinez, G., Soto J. (2017).** A fuzzy hierarchical operator in the grey wolf optimizer algorithm. Appl Soft Comput., Vol. 57, pp. 315–328.
16. **Rodríguez, L., Castillo, O., Soria, J. (2016).** Grey wolf optimizer with dynamic adaptation of parameters using fuzzy logic. IEEE Congress on Evolutionary Computation (CEC), pp. 3116–3123. Doi: 10.1109/CEC.2016.7744183.
17. **Rubio, E., Castillo, O., Valdez, F., Melin, P., González, C. I., Martinez, G. (2017).** An extension of the fuzzy possibilistic clustering algorithm using Type-2 fuzzy logic techniques. Adv. Fuzzy Syst., pp. 7094046:1–7094046:23.
18. **Sánchez, D., Melin, P., Castillo O. (2020).** Comparison of particle swarm optimization variants with fuzzy dynamic parameter adaptation for modular granular neural networks for human recognition. Journal of Intelligent & Fuzzy Systems, Vol. 38, No. 3, pp. 3229–3252.
19. **Saremi, S., Mirjalili, S. Z., Mirjalili, S. M. (2015).** Evolutionary population dynamics and grey wolf optimizer, Neural Computing and Applications, Vol. 26, No. 5, pp. 1257–1263.
20. **Soler, J., Tencé, F., Gaubert, L., Buche, C. (2013).** Data clustering and similarity. Proceedings of the Twenty-Sixth International Florida Artificial Intelligence Research Society Conference, pp. 492–495.
21. **Sturges. H.A. (1926).** The choice of a class interval. J. Am. Stat. Assoc., Vol. 21, No. 153, pp. 65–66.
22. **Tan, Y., Zhu, Y. (2010).** Fireworks Algorithm for Optimization. Springer-Verlag Berlin Heidelberg, pp. 355–364.
23. **Valdez, F., Castillo, O., Melin, P. (2025).** A Review on Enhanced Evolutionary Algorithms Based on Extensions of Fuzzy Logic Systems. Computación y Sistemas, Vol. 29, No. 2.
24. **Valdez, F., Melin, P., Castillo, O. (2014).** Modular neural networks architecture optimization with a new nature inspired method using a fuzzy combination of particle swarm optimization and genetic algorithms. Information Sciences, Vol. 270, pp. 143–153.
25. **Valdez, F., Barraza, J., Melin, P. (2023).** Hybrid Competitive Learning Method Using the Fireworks Algorithm and Artificial Neural Networks. Springer.
26. **Vasant, P., Zelinka, I., Weber, G.W. (2019).** Intelligent computing and optimization. pp. 575, Berlin: Springer International Publishing.
27. **Wolpert, D.H. (1997).** No free lunch theorems for optimization. Evolut Comput, IEEE Trans, pp. 67– 82.
28. **Wu. B.Y. (2006).** On the intercluster distance of a tree metric. Theoret. Comput. Sci., Vol. 369, pp. 136–141.
29. **Xu, R., Wunsch, D. (2008).** Clustering. John Wiley & Sons.
30. **Zadeh, L.A. (1989).** Knowledge Representation in Fuzzy Logic. IEEE Trans. Knowl. Data Eng., Vol. I, pp. 89–0084.
31. **Zheng, Y., Song Qin, Chen, S.-Y (2013).** Multiobjective fireworks optimization for variable-rate fertilization in oil crop production. Applied Soft Computing, Vol. 13, pp. 4253–4263.

Article received on 07/08/2025; accepted on 12/09/2025.
**Corresponding author is Oscar Castillo.*