



An Enhanced Version of Cat Swarm Optimization Algorithm for Cluster Analysis

Hakam Singh, Chitkara University, Baddi, India

 <https://orcid.org/0000-0002-0558-6325>

Yugal Kumar, Jaypee University of Information Technology, India*

 <https://orcid.org/0000-0003-3451-4897>

ABSTRACT

Clustering is an unsupervised machine learning technique that optimally organizes the data objects in a group of clusters. In the present work, a meta-heuristic algorithm based on cat intelligence is adopted for optimizing clustering problems. Further, to make the cat swarm algorithm (CSO) more robust for partitional clustering, some modifications are incorporated in it. These modifications include an improved solution search equation for balancing global and local searches, accelerated velocity equation for addressing diversity, especially in tracing mode. Furthermore, a neighborhood-based search strategy is introduced to handle the local optima and premature convergence problems. The performance of enhanced cat swarm optimization (ECSO) algorithm is tested on eight real-life datasets and compared with the well-known clustering algorithms. The simulation results confirm that the proposed algorithm attains the optimal results more than other clustering algorithms.

KEYWORDS

Cat Swarm Optimization, Clustering, Machine Learning, Meta-Heuristics

1. INTRODUCTION

Data mining is the process of discovering useful patterns and knowledge from large volume of data (Han et al., 2011). Primarily, four types of learning approaches have been described in data mining such as supervised, unsupervised, semi-supervised and active user learning. The supervised learning analyzes the data object with respect to class labels, while, unsupervised learning analyzes the data object without consulting the class labels. The semi-supervised learning considers both supervised and unsupervised learning approaches. Whereas, in active-user learning, user can interactively label the data points with desired outputs. Clustering is an unsupervised machine learning approach that divides a set of objects into distinct clusters (Jain et al., 1999; Nanda & Panda 2014). The objects within a cluster are similar to each other and dissimilar to other clusters. The prime objective of clustering is to maximize the intra-cluster compactness and minimizing inter-cluster likeness among clusters. Last few decades, momentous research has been done in the clustering field. Several optimization techniques inspired from natural phenomenon have been reported to obtain optimal solutions for clustering task. Some of these are particle swarm optimization (Cura 2012), Magnetic optimization algorithm (Kushwaha et al., 2018), charged system search approach (Kumar & Sahoo 2014), Black hole (Hatamlou 2013), artificial bee colony algorithm (Karaboga & Ozturk 2011), ant colony optimization (Shelokar et al., 2004) and big bang big crunch algorithm (Hatamlou et al., 2011).

DOI: 10.4018/IJAMC.2022010108

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

In recent time, a heuristic algorithm based on cat intelligence has gained wide popularity among research community and adopted in several research fields like workflow scheduling in the cloud, image analysis, wireless sensor networks, data analysis etc. (Chu et al., 2006; Tsai et al., 2012; Ram et al., 2015; Wang et al., 2012;). Initially, Santosa and Ningrum (2009) applied the CSO algorithm for solving clustering problems. This algorithm works in two modes, seeking mode and tracing mode. The resting behavior of cats is described using seeking mode, while the hunting skill of cat is described using tracing mode. The seeking mode responsible for local search, whereas the tracing mode responsible for global search. It is observed that CSO algorithm have good exploration capability, but suffers from weak exploitation ability (Kumar & Sahoo 2017). Sometimes, the CSO algorithm cannot explore entire search space for an optimum solution due to lack of global best position information and results in slow convergence rate (Kumar & Singh 2018). These deficiencies can affect the performance of CSO algorithm for solving optimization problems. Hence, to make the CSO algorithm more efficient and robust, several issues are identified and resolved in this research work.

1.1 Problem Identification And Contribution

This research work focuses on convergence and diversity issues of CSO algorithm. These issues are summarized as

- Imbalanced exploration and exploitation processes.
- Lack of diversification mechanism in tracing mode.
- Slow convergence rate due to inappropriate information exchange mechanism.

To address the aforementioned issues, some modifications are incorporated in the CSO algorithm which are listed as

- The position and velocity vector equations are furnished with personal best cat information to make coordination between exploration and exploitation processes.
- Incorporation of neighborhood search strategy to explore more promising search space and avoid premature convergence problem.
- The position vector equation is furnished with chaotic maps to diversify the solution search.

The significant contributions of the work are highlighted as

- To propose an enhanced CSO i.e. ECSO algorithm for addressing the shortcomings of traditional CSO algorithm and the performance of ECSO is evaluated using on real-life clustering datasets.
- To incorporated new accelerated velocity and position update equations to handle diversity, slow convergence and lack of balance between exploration and exploitation.
- To design neighborhood-based search strategy for discovering of optimum solution and also handling local optima situation.

The objective of proposed improvements is to obtain optimal set of clusters with minimized intra cluster distance. The simulation results demonstrate the potential of ECSO algorithm in clustering field.

2. RELATED WORKS

This section describes the related works in the field of partitional clustering problems. Since past few decades, large numbers of clustering algorithms have been developed. To handle the initialization issue of K-mean algorithm, Cao et al., (2009) developed a new initialization method based on neighborhood rough set model. The proposed initialization method is integrated with K-mean

algorithm. In this work, intra cluster similarity and inter cluster similarity of an object are represented in terms of cohesion and coupling. The performance of proposed algorithm is tested on three well known datasets and compared with other two initialization algorithms. Authors claimed that proposed initialization method provides superior results than traditional methods. Han et al., (2017) introduced a new diversity mechanism in gravitational search algorithm to handle clustering problems in effective manner. The proposed diversity mechanism is inspired from the collective response of birds. This mechanism works in three steps i.e. initialization, identification (nearest neighbors) and orientation alteration. In initialization step, the candidate population is generated and forwarded to second step i.e. nearest neighbor. In second step, nearest neighbors are identified through a neighborhood strategy. In third step, the current position of a candidate solution is changed based on the nearest neighbor. The performance of proposed algorithm is tested on thirteen data sets and compared with well-known clustering algorithms. It is seen that proposed algorithm is an effective and efficient for solving clustering problems. To handle clustering problems effectively, a two-step artificial bee colony algorithm is reported in (Kumar & Sahoo 2017). In this work, authors proposed three improvements in ABC algorithm. These improvements are initial cluster centre positions, modified search equations and abandoned food source locations. The initial cluster centre positions are determined through one step K-means algorithm. A PSO based search equation is developed to discover more promising search space. The abandoned food source location is determined using Hooke and Jeeves based search method. The performance of this algorithm is tested on both artificial and real-life data sets and compared with well-known clustering algorithms. It is seen that the proposed algorithm significantly improves the performance of conventional artificial bee colony algorithm. To solve clustering effectively, Kumar and Sahoo (2015) developed a new clustering algorithm based on magnetic charge system search algorithm and particle swarm optimization algorithm. In this work, the personal best mechanism of particle swarm optimization algorithm is added into magnetic charge system search algorithm for better exploration. Further, neighbourhood strategy is developed to avoid local optima situation. The performance of MCSS-PSO algorithm is tested on ten datasets and compared with K-means GA, PSO, ACO, CSS, CCSSA and MCSS. From experimental results, it is noted that proposed algorithm provides state of art clustering results. Boushaki et al. (2018) designed a new quantum chaotic cuckoo search algorithm for data clustering. To extend the global search ability of quantum chaotic cuckoo search algorithm, a nonhomogeneous update mechanism is employed in proposed algorithm. Further, to improve convergence speed, chaotic maps are incorporated in this algorithm. The performance of proposed algorithm is compared with genetic quantum cuckoo search, hybrid cuckoo search and differential evolution, hybrid K-means and improved cuckoo search, standard cuckoo search, quantum particle swarm optimization, differential evolution, hybrid K-means chaotic particle swarm optimization and genetic algorithm. The experimental results showed that proposed algorithm performs well in comparison to other algorithms. A new clustering algorithm based on genetic algorithm and message-based similarity measure is presented in (chang et al., 2012). The message-based similarity measure contains two types of messages-responsibility and availability. These messages are exchanged between data points and cluster centers. The responsibility of data point reflects the evidence regarding cluster centers, while the availability reflects the evidence that about the appropriateness of data point with respect to cluster centers. Further, variable-length real-value chromosome representation and a set of problem-specific evolutionary operators are incorporated in GAMS. The performance of proposed algorithm is tested on both artificial and real-life data set. The simulation results showed that the algorithm obtains significant results for clustering problems. Hatamlou (2013) developed a new clustering algorithm inspired through black hole phenomenon. Similar to other clustering algorithm, black hole algorithm starts with initial population selection and objective function evaluation. The best candidate solution is acted as black hole that can attract data items/stars. The absorption of stars in the black hole depends on current location and random values. The performance of proposed algorithm is tested on six real-life datasets and it is stated that black hole clustering algorithm provides better clustering results. Zhang et al., (2010) presented an

artificial bee colony algorithm for data clustering. In this work, bees are categorised into three categories onlooker bees, employed bees and scout bees. The onlooker bees and employed bees are responsible for global search, while scout bees are responsible for local search. Further, Deb's rule is incorporated to redirect search in solution space. The performance is tested on three real-life datasets and compared with other clustering algorithms. It is revealed that the proposed algorithm provides good quality results. Taherdangkoo et al., (2013) presented a new clustering based on blind naked mole-rats algorithm for data clustering. This algorithm inspires through blind naked mole-rat colonies that can effectively search food source and protect colony from attacks. The algorithm starts with the population of blind-naked mole rats and searches the whole problem space for optimal solution in random fashion. In next iterations, employed mole rats start movement to target food source and their neighbours. The performance of proposed algorithm is tested on six real-life datasets and compared with other well-known clustering algorithms. Authors claimed that blind naked mole-rats algorithm provides higher accuracy with faster convergence speed. Hatamlou (2012) developed a novel binary search algorithm to obtain high quality clusters with better convergence speed. In this work, initial seed points are generated from different location and data objects are assigned to nearest one. Further, objective function is evaluated, if value of current objective function is lesser then previous objective function value, the search will proceed in same direction otherwise it proceeds in opposite direction. The performance of proposed algorithm is tested on six benchmark datasets and compared with K-means, GA, SA, TS, ACO, HBMO and PSO. It is seen that proposed algorithm effectively solves clustering problems. Bijari et al., (2018) presented a memory-enriched big bang–big crunch algorithm for clustering. The BB-BC algorithm works in two phases- big bang and big crunch phase. In big bang phase, random points are generated near to initial point. In big crunch phase, these points are optimized in single one. Further, a memory concept with limited size is integrated to make better trade-off between exploration and exploitation. The performance of algorithm is tested on six data sets and compared with well-known algorithms like GA, PSO, GWO and original BB-BC. The simulation results show that BB-BC algorithm provides superior clustering results than other algorithms. To solve clustering search space problems Wang et al., (2016) presented a hybrid version of flower pollination algorithm with bee pollinator. In this work, discard pollen operator of artificial bee colony is used to enhance population diversity and global search ability of flower pollination algorithm. Further, elite based mutation and crossover operator are applied to enhance local search mechanism. Several artificial and real datasets are considered to evaluate the performance of proposed algorithm. The simulation results are compared with k-means, PSO, DE, CS, ABC, FPA algorithms. Authors stated that proposed algorithm is one efficient and effective algorithm for solving clustering problems. Hatamlou and Hatamlou (2013) designed a two-stage clustering approach to overcome the draw backs of particle swarm optimization like local optima and slow convergence speed. The proposed approach works in two-stages, in first stage initial candidate solution is generated through PSO algorithm. In second stage, quality of solution is improved with help of heuristic search algorithm. The performance of proposed algorithm is tested on seven benchmarks and compared with K-means, PSO GSA, BB-BC methods. It is seen that the proposed algorithm determines good quality clusters. A hybrid version of artificial bee colony algorithm with genetic algorithm is presented in (Yan et al., 2012). The primary objective of this work is to enhance the information exchange mechanism between bees. To achieve the same, the crossover operator of genetic algorithm is incorporated in artificial bee colony algorithm. The performance of proposed algorithm is evaluated on six benchmark datasets and compared with other ABC, CABC, PSO, CPSO and GA clustering algorithms. The simulation results of proposed algorithm are better than other algorithms being compared. A hybrid version of ant clustering algorithm with k-harmonic means is presented for handling clustering problems (Jiang et al., 2010). The proposed algorithm contains the merits of both algorithms i.e. initialization ability of K-harmonic means and local optima ability of ant algorithm. The performance of proposed algorithm is tested on five benchmark datasets and compared with KHM and ACA. It is stated that proposed algorithm achieves better clustering results than compared algorithms, but runtime is slightly longer.

Jiang and Wang (2014) developed a PSO based clustering algorithm. In this work, a cooperative co-evolution method is integrated with PSO to enhance the convergence rate and population diversity. The cooperative co-evolution method works as decomposer and PSO algorithm acts as optimizer. The performance of proposed algorithm is tested on some of real-life datasets and compared with PSO, SRPSO, ACO, ABC and DE, and K-means algorithms. The Simulations results showed that proposed algorithm performs well with most of datasets. To solve clustering problems effectively, an improved version of CSO algorithm is reported in (Kumar & Singh 2018). In this work, new search equations are proposed to explore search space in efficient manner. Further, a local search method is also incorporated in CSO algorithm for handling local optima problem. Simulations result showed that proposed algorithm provides effective clustering results.

3. CAT SWARM OPTIMIZATION ALGORITHM

Chu and Tsai (2006) designed a new algorithm by observing the behavior of cats, named as CSO algorithm. This algorithm works in two modes, seeking and tracing. The seeking mode responsible for local search. The tracing mode responsible for global search. Further, Santosa and Ningrum (2009) explored the capabilities of CSO algorithm in clustering field. The working of CSO algorithm is described as

3.1 Seeking Mode

This mode describes the resting behavior of cats. In seeking-mode, a cat visits different location for target identification and stay alert. The movement of cat correspond to explore the entire search field for good candidate solution. The seeking mode is accountable for explorative operations i.e. local search method.

3.2 Tracing Mode

This mode describes the hunting skills of a cat. When a cat hunts the target, the position and velocity of cat changes. This change in velocity and position vectors are computed using equation 1-2.

$$V_{j \text{ new}}^d = w \times V_j^d + c \times r (X_{j \text{ best}}^d - X_j^d) \quad (1)$$

Where, $V_{j \text{ new}}^d$, V_j^d are new and old velocity values of j^{th} cat in the d^{th} dimension, w is a weight factor between 0 and 1, r is a random value, c is user defined parameter, $X_{j \text{ best}}^d$ represent best position and X_j^d represent current position of the j^{th} cat and $d = 1, 2, \dots D$.

$$X_{j \text{ new}}^d = X_j^d + V_j^d \quad (2)$$

Where, $X_{j \text{ new}}^d$ represents the updated position of the j^{th} cat in d^{th} dimension.

4. ENHANCED CSO ALGORITHM FOR CLUSTERING

This section describes the working of enhanced CSO algorithm in clustering field. It is observed that CSO algorithm efficiently explores the search space, but suffers with weak exploitation. In turn, local and global searches become imbalanced (Kumar & Sahoo 2017; Kumar & Singh 2018). Furthermore, CSO algorithm suffers with local optima and sometimes converges on premature results. To address aforementioned issues, some modifications are proposed. The detailed description of these improvements are given in subsection 4.1.

4.1 Proposed Improvements

4.1.1 Modified Search Space Methods

In this work, the searching equations i.e. position and velocity equations of CSO are modified to guide the search in positive direction. The position updated equation of tracing mode is improved using personal best position information (vector that keep track of personal best position). Further, chaotic map also integrates into position update equation for generating more diversified population.

$$X_{j \text{ new}}^{d+1} = \lambda \times X_j^d + c_n \times P_g + V_j^d \quad (3)$$

$$\lambda = \ln(\max(x_j) - \min(x_j)) / 2$$

Where, $X_{j \text{ new}}^{d+1}$ is new centroid or cats' position, λ is a constant, $\max(x_j), \min(x_j)$ are maximum and minimum values of i^{th} cat, c_n in logistic chaotic map, X_j^d, V_j^d are position and velocity values of j^{th} cat and P_g are personal global best position.

The CSO algorithm uses velocity value to update the position of cat. Henceforward, to add more diversity in solution space, the velocity equation is modified as

$$V_{j \text{ new}}^{d+1} = V_j^d + c_n (P_g - X_j^d) \quad (4)$$

Where, $V_{j \text{ new}}^{d+1}$ is new velocity value of j^{th} cat, V_j^d is old velocity value of j^{th} cat

4.1.2 Neighborhood Search Strategy

To enhance the search ability of the CSO algorithm, a neighborhood strategy based on "step division method" is designed in this work. This method provides more probabilities for finding efficient solutions and enables to handle local optima state.

The proposed neighborhood strategy works in two phases, identification and evaluation. In neighborhood strategy, first step is to identify neighboring data points of cluster centers. In this work, we have chosen n neighboring data points (best cats' positions) and randomly nominated k data points for evaluation. Let X_i represents (best cats' positions) and $X_{k, \text{best cat}}$ represents nominated neighboring data points in figure 1(a). In evaluation phase, the nominated neighboring data points are used to determine the new data points i.e. to produce single output using multiple inputs. We have chosen $n = 10$ neighboring data points (best cats' positions) and randomly nominated $k = 3$ data points for evaluation as see in figure 1(b). The new neighboring data point is obtained through equation 5

$$X_{\text{neigh}}^{\text{new}} = \bar{X}_{\text{neigh}} + \sum d'x / n \times c \quad (5)$$

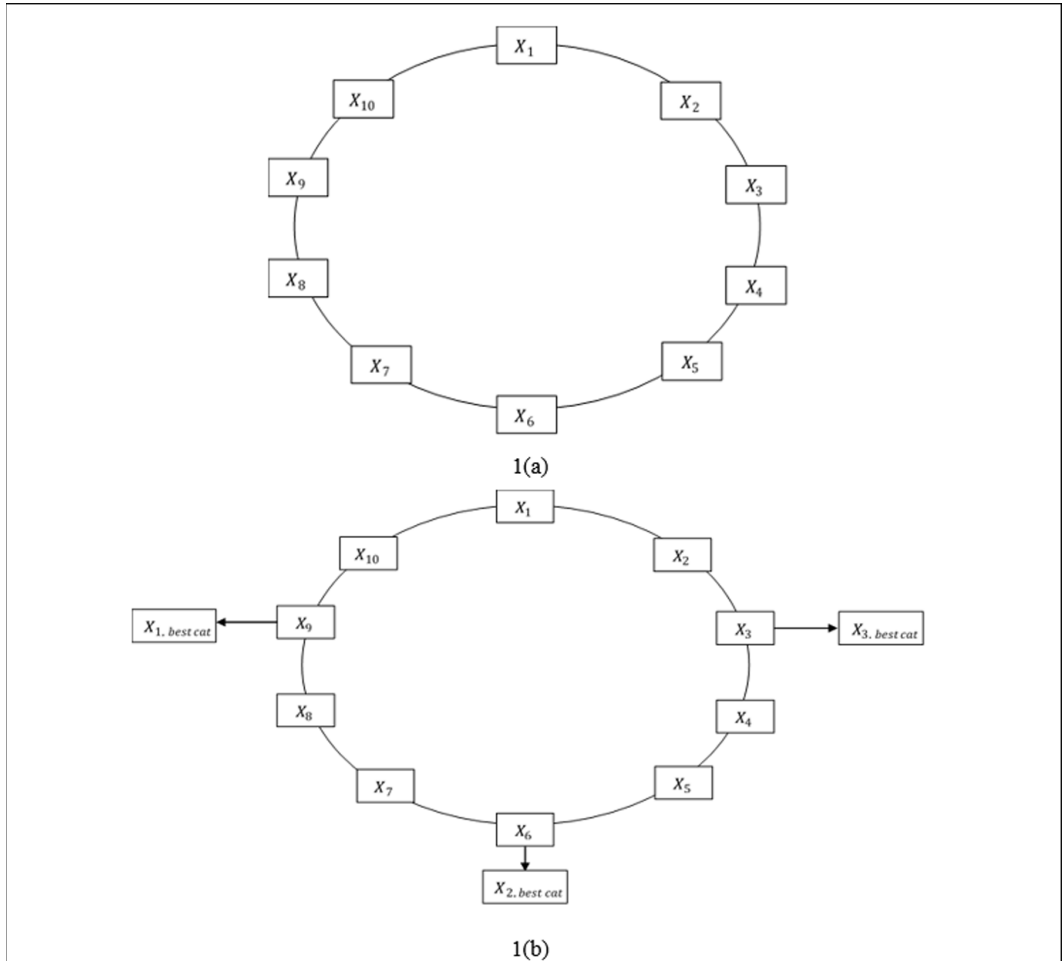
$X_{\text{neigh}}^{\text{new}}$ is new neighborhood, \bar{X} is mean of k number of best cats, $\sum dx$ is sum of k best cats' positions, c is common factor taken from best cats' position, and $d'x = dx / c$

The steps of neighborhood search are given below.

Algorithm 1: Neighborhood Search Strategy

Begin

Figure 1 (a). Circle representation of n best cats. Figure 1 (b) Represents k -neighborhood structure, $n = 10$ and $k = 3$



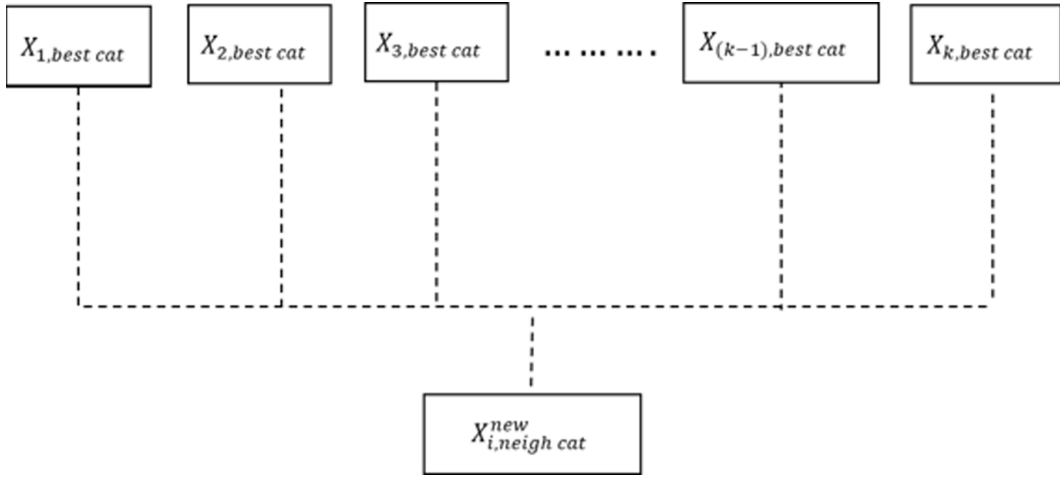
Step 1: Select n neighboring data points (best positions of cats)
 Step 2: Randomly nominated k data points for evaluation.
 Step 3: Evaluate the k -neighbourhood data points using equation
 Step 4: Obtain new solution $X_{i, \text{neigh cat}}^{\text{new}}$
 End

4.2 Enhanced Cat Swarm Optimization Clustering Algorithm

The proposed algorithm is adopted for determining the optimized clustering results. The Euclidean distance can be taken as a distance measure. The basic steps of proposed ECSO algorithm for data clustering are described as.

Step 1 Initialization: In this segment, basic algorithmic parameters values, like number of cats (K i.e. centroids), SMP, SRD, number of iterations are initialized. Moreover, the initial positions of cats are selected from dataset where, $d \in \{1, 2 \dots D\}$. The position of cats represents the initial cluster centers.

Figure 2. Evaluation of k number of best cats



$$\left\{ \begin{array}{c} C_{11}, C_{12}, C_{13}, C_{14}, \dots, C_{1D} \\ C_{21}, C_{22}, C_{23}, C_{24}, \dots, C_{2D} \\ C_{31}, C_{32}, C_{33}, C_{34}, \dots, C_{3D} \\ \vdots \\ C_{K1}, C_{K2}, C_{K3}, C_{K4}, \dots, C_{KD} \end{array} \right\}$$

Step 2 Velocity Determination: After initial position selection, the next algorithmic step is to evaluate the velocity of each cat, which is given as below.

$$\left\{ \begin{array}{c} V_1 \\ V_2 \\ V_3 \\ \vdots \\ V_K \end{array} \right\}$$

Step 3 Objective Function Evaluation: In-order to perform clustering, Euclidean distance is taken as objective function in this experiment. The objective function value is used to assign data object to respective clusters.

$$\text{minimize } f(X, C) = \sum_{k=1}^K \sum_{x \in D_i} \min X_i - C_j^2 \quad (6)$$

Step 4 Global Best Position Identification: To obtain global best position, the objective function values are examined using a fitness function which is described in equation 7.

$$F(C_k) = \sum_{k \in K} \frac{SSE(C_k)}{SSE(C_k)_{\min} + SSE(C_k)_{\max}} \quad (7)$$

Step 5 Seeking Mode: In this mode, n number of identical positions are generated of each cat and also computes shifting value for each duplicate position of cat. After determine the shifting bit value, these values are randomly add and subtract from current position of cat to determine the update position of a cat. The value of fitness function is computed for the updated position of cat and stores the best values of cat and known as seeking mode best position of cats. The seeking mode operations can be summarized as

- Define the replicated copies (T) of the i^{th} cat.
- Add/Subtract the SRD values from cats' current position and replace the old values.
- Compute fitness function.
- Select and update best position of the i^{th} cat.

Step 6 Tracing Mode: In tracing mode, the new cat's position is obtained using equation 3. Further, the velocity rate of cats is also updated in this phase using equation 4. The local optima condition is also checked in this step, if local optima problem arise, apply neighborhood strategy (Algorithm 1).

Step 7 Termination Condition: This step corresponds to check the termination condition. Stop the execution of algorithm, if termination condition is met, otherwise, steps 3 to 7 can be repeated.

5. RESULT AND DISCUSSION

This section describes the experimental results of ECSO clustering algorithm. The algorithm is implemented in MATLAB 2016a environment and tested on eight real-life clustering problems. The detailed description of datasets is provided in Table 1. The simulation results are presented an average of thirty independent runs and each run executes hundred times.

5.2 Performance Evaluation and Results

This subsection presents a comparative analysis on ECSO, CSO, ACO, PSO and K-means clustering algorithms. Table 2 presents the performance comparison of ECSO and other clustering algorithm using average intra cluster distance and f-measure parameters. It is observed that proposed algorithm attains minimum intra cluster distance except glass dataset. Furthermore, a statistical test (Quade) is also conducted to prove its significance in clustering filed. The results of statistical tests are reported in subsection 5.3.

Figures 3(a-e) shows the clustering of data objects using ECSO algorithm. The figure 3(a) depicts the clustering of iris dataset using ECSO algorithm. The data objects of iris dataset are grouped into three clusters (setosa, versicolour and virginica). Figure 3(b) shows the clustering of the cancer dataset's (cell size, cell shape and bare nuclei) objects. The data objects of cancer dataset are grouped into two clusters as benign and malignant. Figure 3(c) display the clustering of the CMC dataset using ECSO algorithm. The CMC dataset is divided into three clusters as 'Cluster No use1', 'Cluster Long Term2' and 'Cluster Short Term3'. The figure 3(d) shows the clustering of wine dataset (alcohol, malic acid and ash). Wine dataset is divided into three clusters as 'Wine Type 1', 'Wine Type 2' and 'Wine Type 3'. Figure 3(e) shows the clustering glass dataset.

Figures 4(a-h) shows the convergence behavior of ECSO and other clustering algorithm. In this graphical illustration X-axis represents the number of iteration and Y-axis represents the intra-cluster distance values. From graphical illustration it is find that ECSO provides better convergence speed except the glass dataset.

Table 1. Description of datasets

Dataset	Clusters (K)	Attributes (D)	Data instances (N)	Description
Iris	3	4	150	Fisher's iris data
Wine	3	13	178	Wine data
CMC	3	9	1473	Contraceptive method choice
Cancer	2	9	683	Cancer
Glass	6	9	214	Glass identification data
Thyroid	3	5	215	Thyroid
Vowel	6	3	871	Indian Telugu vowel
LD	2	6	345	Liver Disorder Data

5.3 Statistical Analysis

This subsection illustrates the statistical results of ECSO and other clustering algorithms. In this work, Quade test is adopted to confirm the existence of proposed ECSO algorithm. Quade test is an advanced nonparametric test that can give weightage to the size of datasets. Two hypotheses are generated to check the significant difference between performances of ECSO and other clustering algorithms. These hypotheses are interpreted as hypothesis (H_0) and (H_1). Hypothesis (H_0) means algorithms are not different, hypothesis (H_1) means at least one i.e. newly proposed algorithm is different from others. Tables 3-6 summarize the results of Quade test using intra cluster distance and f-measure.

- **Quade test on average intra-cluster distance**

Tables 3-4 presents the results obtained from Quade test on intra-clustering distance. It is observed that the proposed ECSO algorithm obtains first rank in most cases except glass dataset. Further, it is also seen that K- means algorithm exhibits worst performance using intra cluster distance parameter. The critical value for Quade test is 2.485142. The p-value for Quade test is 5.727e-6. Hence, the hypothesis is strongly rejected at the significance level 0.05. So, substantial difference is occurred between the performances of ECSO and other clustering algorithms. Hence, the ECSO algorithm is statistically as well as experimentally better than other clustering algorithms.

- **Statistical test on F-Measure parameter**

Tables 5-6 presents the results obtained from Quade test on f-measure parameter. The critical for Quade test is 2.485142, which is smaller than significant level 0.05. The p-value for Quade test is 7.59E-04 that rejects the null hypothesis. Finally, it is stated that the performances of ECSO and other algorithms are significantly different.

7. CONCLUSION AND FUTURE SCOPE

In this work, an enhanced version of CSO algorithm is developed to solve clustering problems. Several shortcomings are associated with traditional, CSO algorithm such as diversity mechanism, local optima and imbalanced local and global searches. These shortcomings are addressed in this work using-

- Solution search (position update and velocity) equations are modified for better tradeoff between local and global searches.

Table 2. Performance comparison of ECSO and other well-known clustering algorithms

Dataset	Parameters	Algorithms					
		K-means	GA	PSO	ACO	CSO	Proposed CSO
Iris	Best	97.12	113.98	96.48	96.89	96.94	96.02
	Avg.	112.44	125.19	98.56	98.28	97.86	97.14
	Worst	122.46	139.77	99.67	99.34	98.58	98.32
	F-measure	0.781	0.774	0.780	0.779	0.776	0.789
Cancer	Best	2989.46	2999.32	2978.68	2983.49	2985.16	2964.61
	Avg.	3248.25	3249.46	3116.64	3178.09	3124.15	3043.09
	Worst	3566.94	3427.43	3358.43	3292.41	3443.56	3279.01
	F-measure	0.832	0.819	0.826	0.829	0.831	0.832
CMC	Best	5828.25	5705.63	5792.48	5756.42	5712.78	5672.16
	Avg.	5903.82	5756.59	5846.63	5831.25	5804.52	5751.35
	Worst	5974.46	5812.64	5936.14	5929.36	5921.28	5912.25
	F-measure	0.337	0.324	0.333	0.332	0.334	0.336
Wine	Best	16768.18	16530.53	16483.61	16448.35	16431.76	16314.02
	Avg.	18061.24	16530.53	16417.47	16530.53	16395.18	16340.89
	Worst	18764.49	16530.53	16594.26	16616.36	16589.54	16596.76
	F-measure	0.519	0.515	0.516	0.522	0.521	0.525
Glass	Best	222.43	272.37	267.56	261.22	256.53	259.62
	Avg.	246.51	282.32	275.71	273.46	278.44	266.89
	Worst	258.38	291.77	284.52	293.08	282.27	275.41
	F-measure	0.426	0.333	0.412	0.402	0.416	0.426
Thyroid	Best	13956.83	11576.29	10354.56	10085.82	10585.91	10289.29
	Avg.	14133.14	12218.82	11149.70	10758.13	10687.56	10408.10
	Worst	14642.21	13254.39	13172.86	12134.82	11934.34	11548.10
	F-measure	0.731	0.763	0.778	0.781	0.774	0.783
Vowel	Best	152422.26	152234.73	151976.01	149395.60	152436.58	148826.20
	Avg.	159642.89	159353.49	157999.82	158458.14	158956.81	157523.06
	Worst	161236.81	165991.65	158121.18	160632.82	160539.82	165929.56
	F-measure	0.652	0.643	0.645	0.648	0.646	0.652
LD	Best	11397.83	532.48	209.15	224.76	231.54	216.76
	Avg.	11,673.12	543.69	236.47	241.23	240.16	232.01
	Worst	12043.12	563.26	239.11	256.44	261.06	242.43
	F-measure	0.467	0.482	0.491	0.487	0.485	0.492

- A neighborhood based solution is designed for handling the local optima situation effectively.
- Diversity mechanism of traditional CSO is improved using chaotic maps.

Figure 3(a).

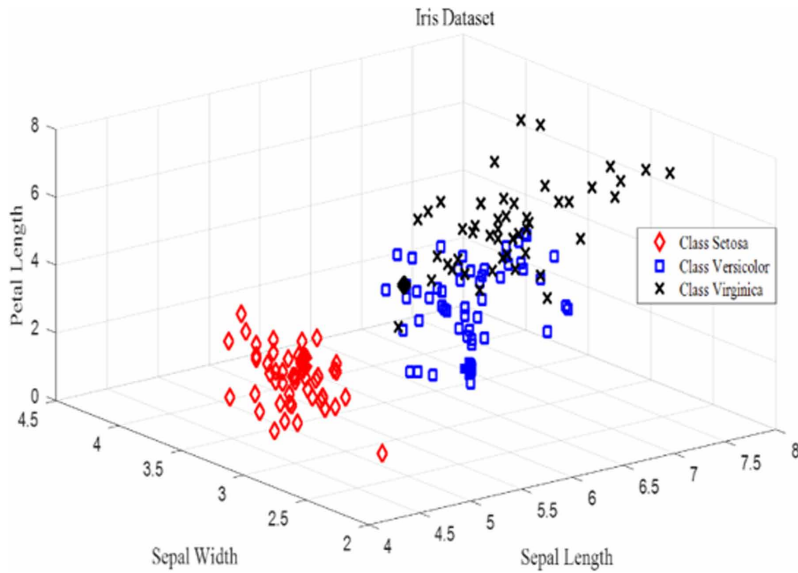
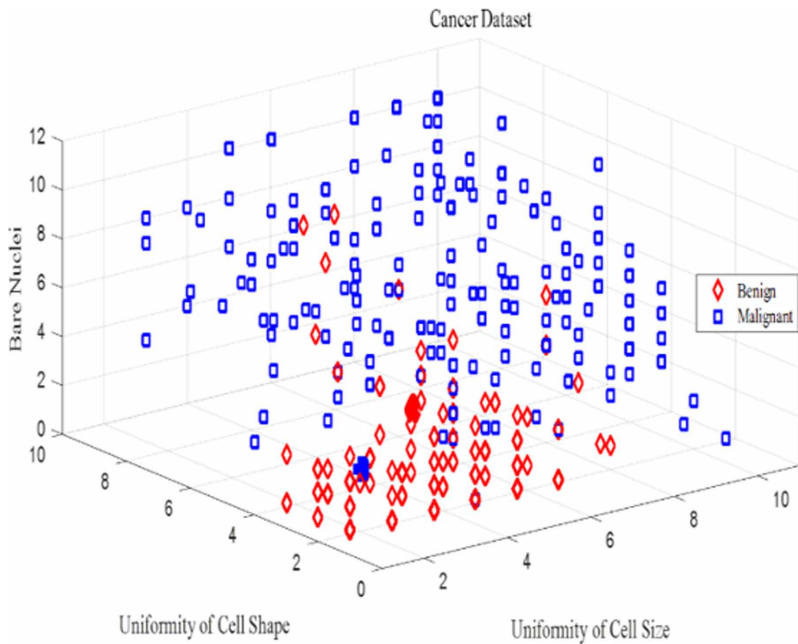


Figure 3(b).



Furthermore, the performance of proposed CSO algorithm is examined on eight real-life datasets and compared with well-known clustering algorithms. From experimental results, it is observed that ECSO algorithm delivers good quality results and also capable to handle local optima situation. In future, the ECSO can be used for addressing multi objective clustering problems. This algorithm will be adopted for feature selection and generating rules for association mining task. Several other optimization problem like resource scheduling, parameter optimization of ANN and SVM will be taken into consideration using ECSO algorithm.

Figure 3(c).

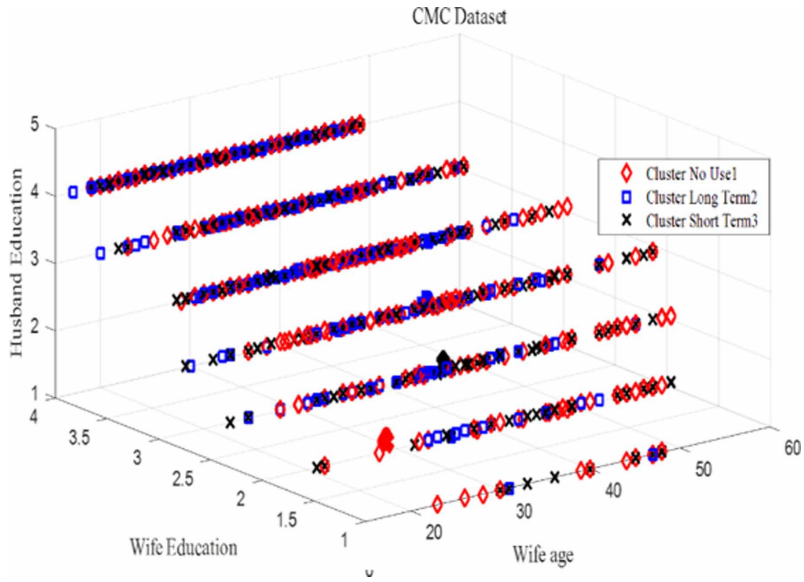


Figure 3(d).

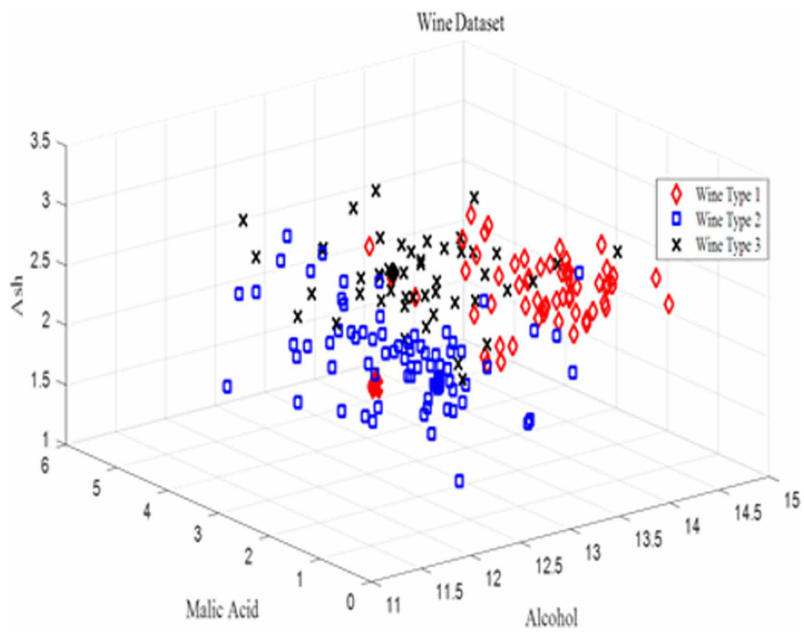


Figure 3(e).

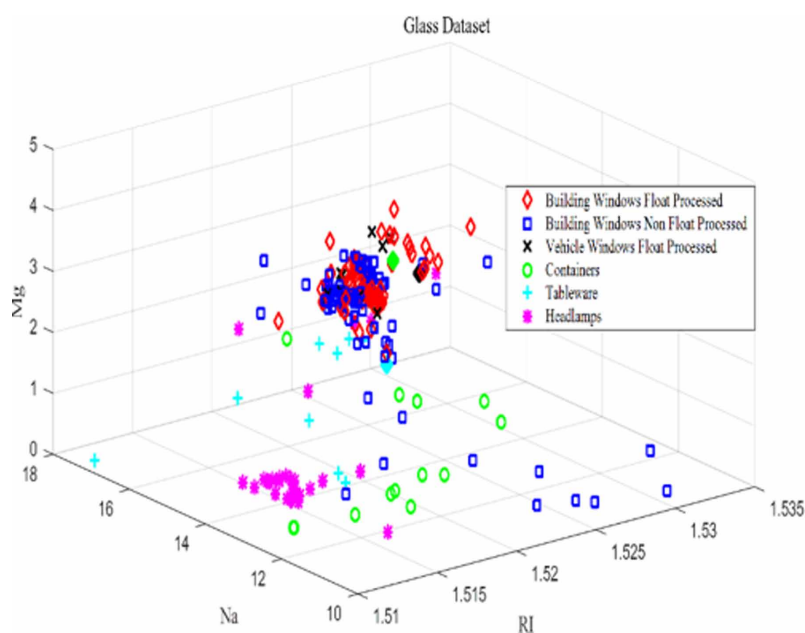


Figure 4(a).

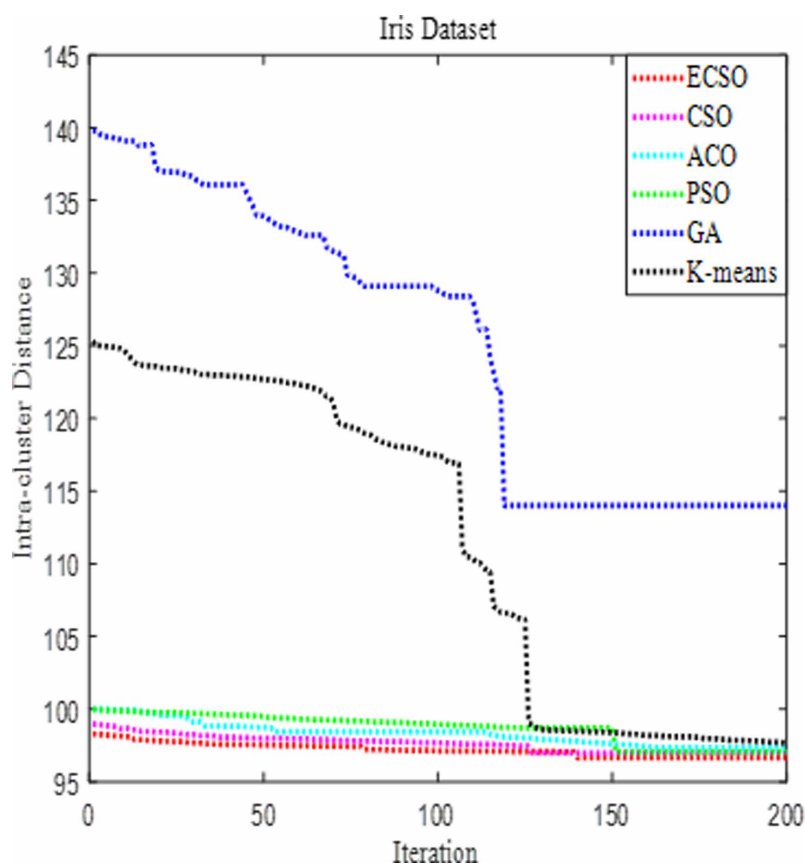


Figure 4(b).

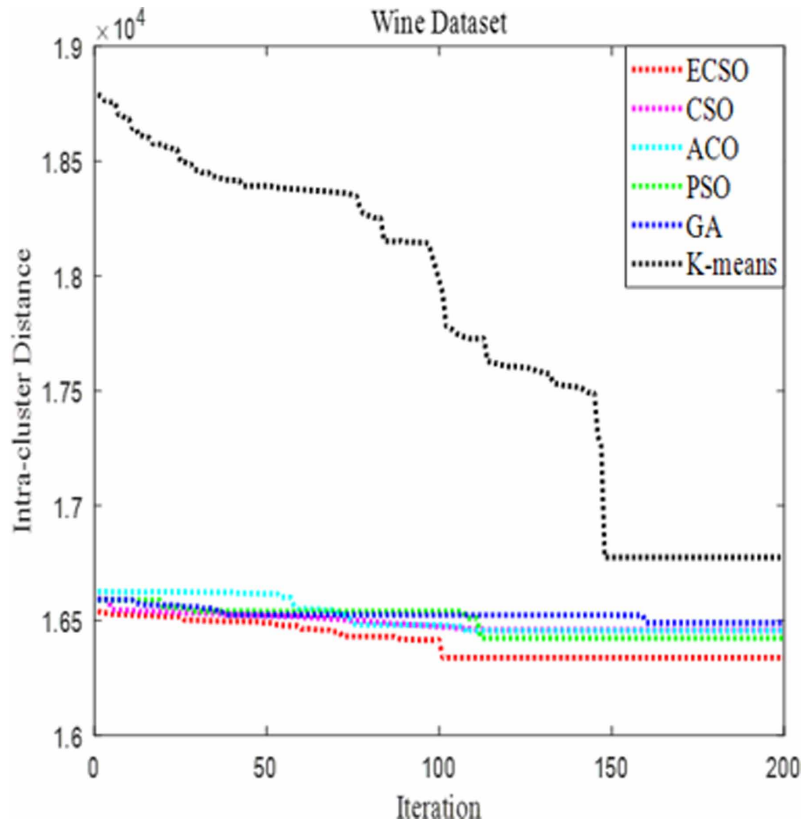


Figure 4(c).

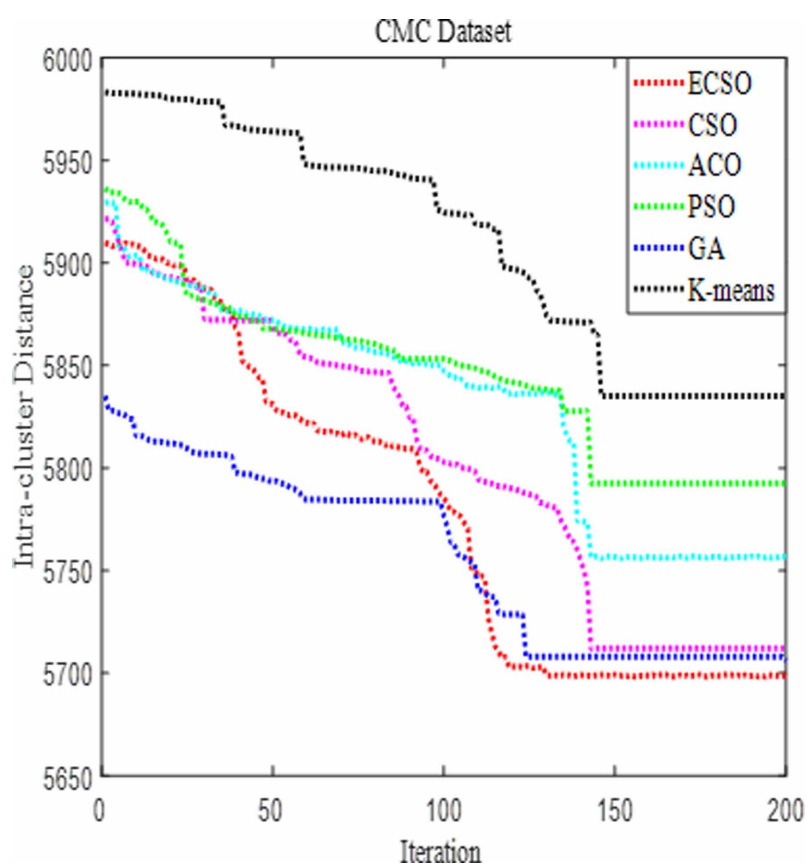


Figure 4(d).

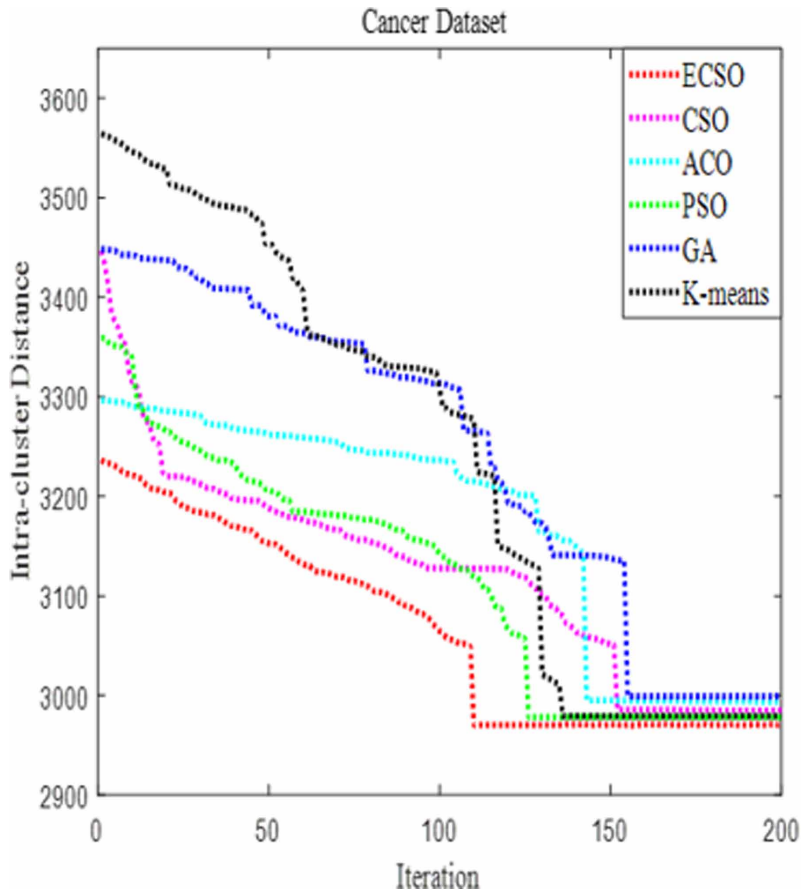


Figure 4(e).

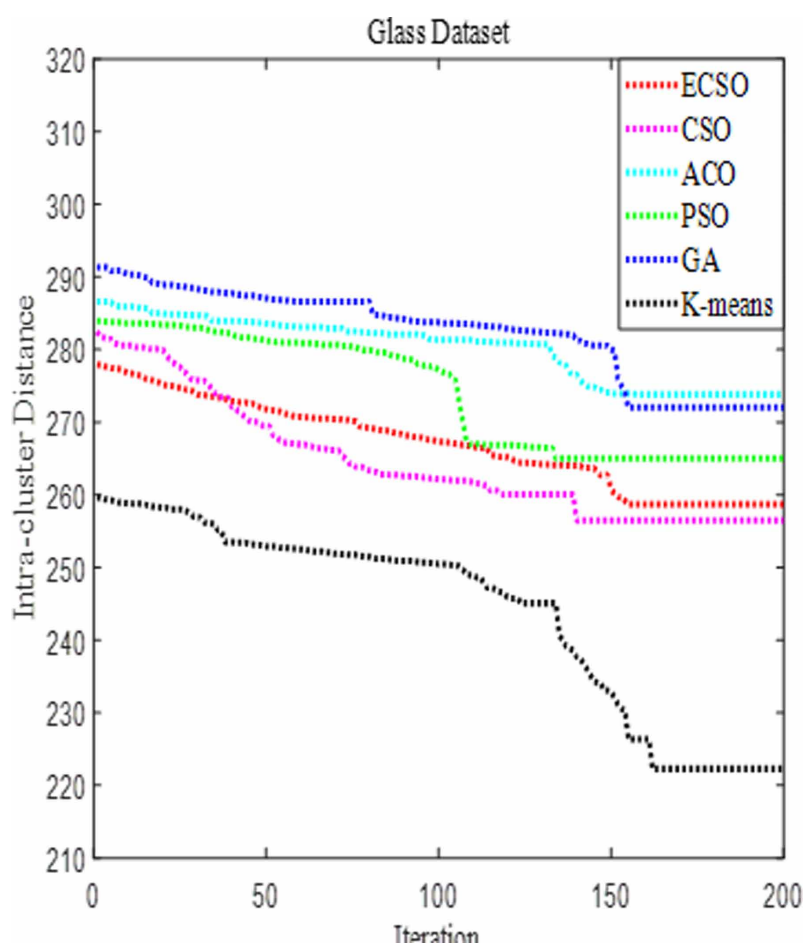


Figure 4(f).

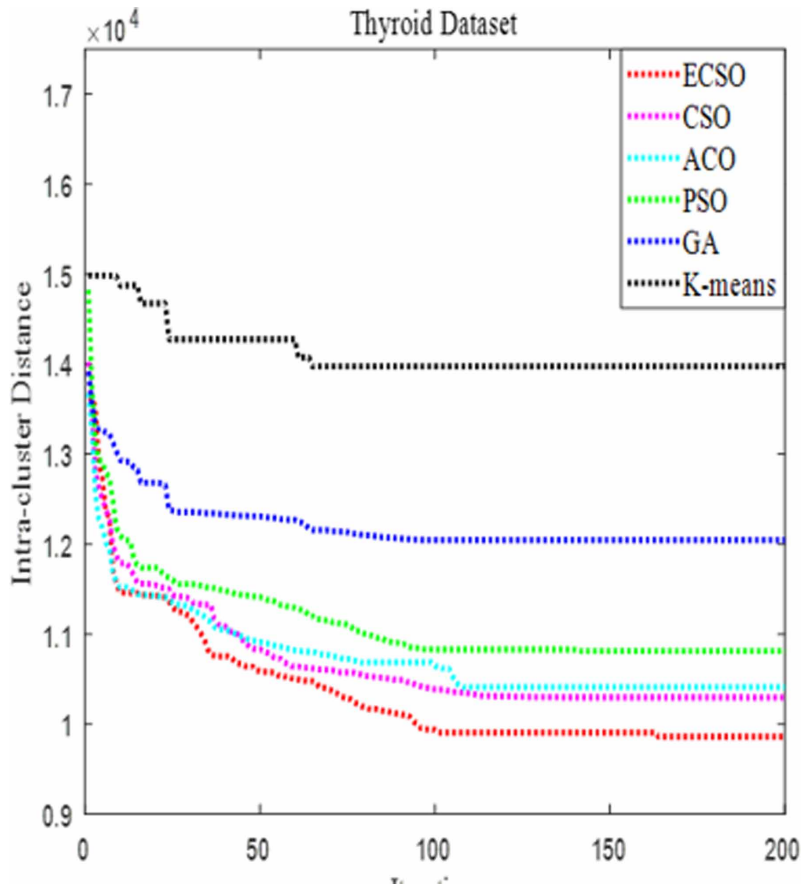


Figure 4(g).

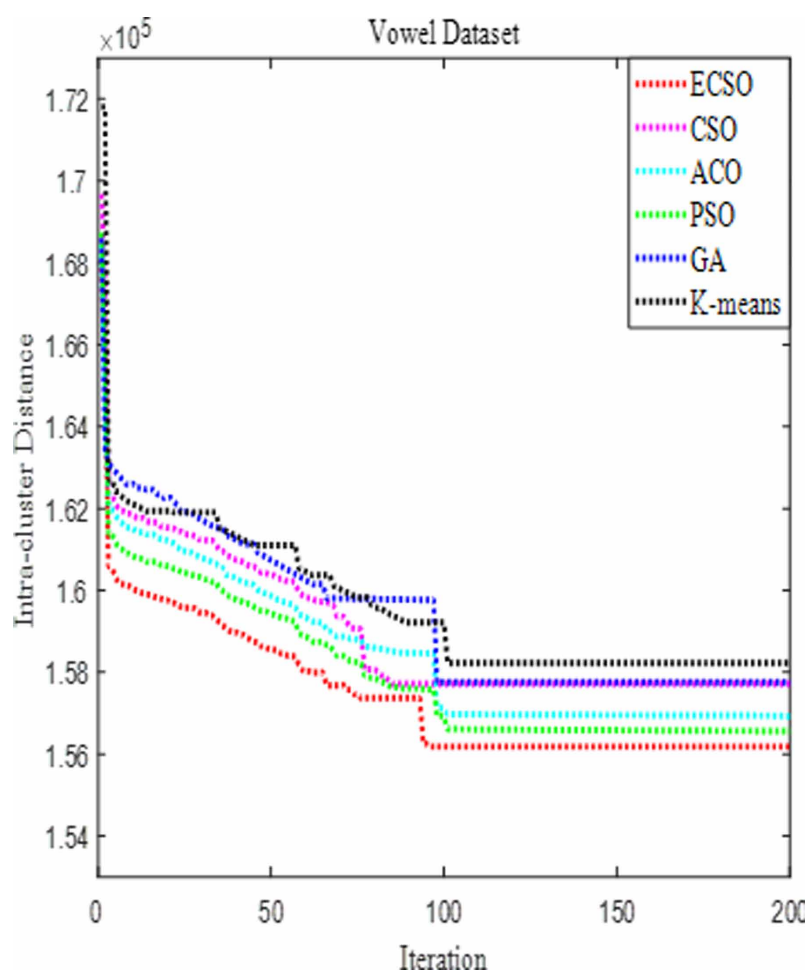


Figure 4(h).

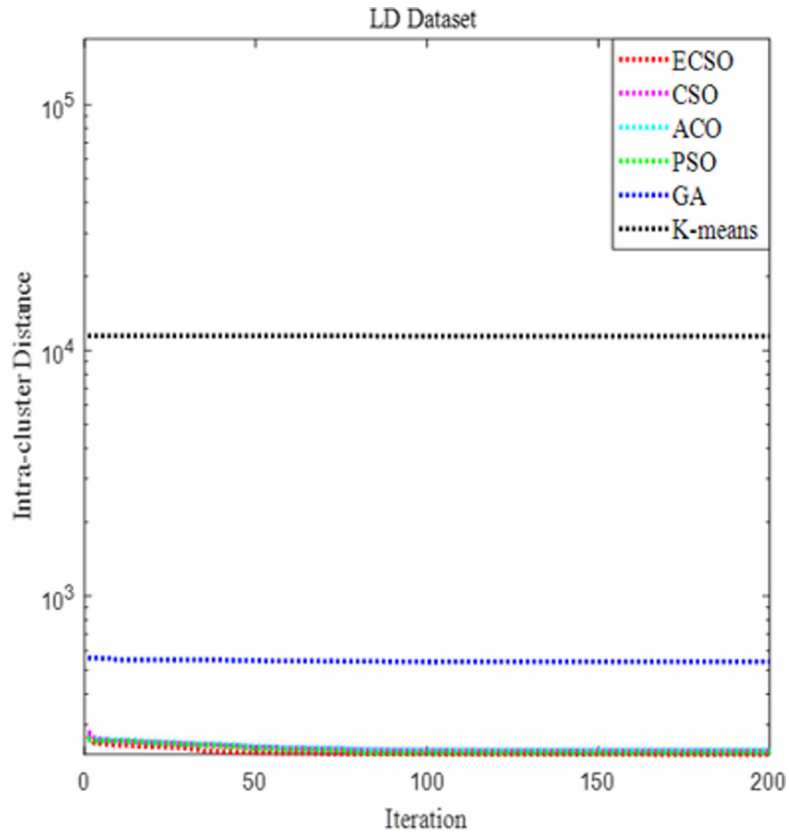


Table 3. Relative size of datasets for ECSO and other clustering algorithms using Quade test

Dataset /Algorithm	K-Means	GA	PSO	ACO	CSO	ECSO
Iris	1.5	2.5	0.5	-0.5	-1.5	-2.5
Cancer	6	10	-6	2	-2	-10
CMC	7.5	-4.5	4.5	1.5	-1.5	-7.5
Wine	12.5	5	-2.5	5	-7.5	-12.5
Glass	-5	5	1	-1	3	-3
Thyroid	17.5	10.5	3.5	-3.5	-10.5	-17.5
Vowel	15	9	-9	-3	3	-15
LD	20	12	-12	4	4	-20
Relative Size	75	49.5	-20	4.5	-21	-88

Table 4. Shows the algorithms ranking compute by Quade test

Datasets/Algorithm	K-Means	GA	PSO	ACO	CSO	ECSO	Relative Size of datasets
Iris	5	6	4	3	2	1	1
Cancer	5	6	2	4	3	1	4
CMC	6	2	5	4	3	1	3
Wine	6	4.5	3	4.5	2	1	5
Glass	1	6	4	3	5	2	2
Thyroid	6	5	4	3	2	1	7
Vowel	6	5	2	3	4	1	6
LD	6	5	2	4	3	1	8
Sum	41	39.5	26	28.5	24	9	36
Relative Size	5.13	4.94	3.25	3.56	3	1.13	4.5

Table 5. Datasets relative size comparison using Quade test

Dataset /Algorithm	K-Means	GA	PSO	ACO	CSO	ECSO
Iris	7.5	12.5	-2.5	2.5	7.5	-12.5
Cancer	-6	7.5	4.5	1.5	-1.5	-6
CMC	-10	10	2	6	-2	-6
Wine	1	5	3	-3	-1	-5
Glass	-16	20	4	12	-4	-16
Thyroid	17.5	10.5	-3.5	-10.5	3.5	-17.5
Vowel	-2	2.5	1.5	-0.5	0.5	-2
LD	15	9	-9	-3	3	-15
Relative Size	-8	77	0	5	6	-80

Table 6. Shows the algorithms ranking compute by Quade test

Datasets/Algorithm	K-Means	GA	PSO	ACO	CSO	ECSSO	Relative Size of datasets
Iris	2	6	3	4	5	2	5
Cancer	1.5	6	5	4	3	1.5	3
CMC	1	6	4	5	3	1	4
Wine	4	6	5	2	3	4	2
Glass	1.5	6	4	5	3	1.5	8
Thyroid	6	5	3	2	4	6	7
Vowel	1.5	6	5	3	4	1.5	1
LD	6	5	2	3	4	6	6
Sum	23.5	46	31	28	29	10.5	36
Relative Size	2.94	5.75	3.88	3.5	3.63	1.31	4.5

Table 7. Abbreviations

ABC:	Artificial Bee Colony
ACO:	Ant Colony Optimization
BB-BC	Big Bang-Big Crunch
CDC:	Count to Dimensions
CSO:	Cat Swarm Optimization
CSS:	Charge System Search
DE:	Differential Evolution
ECSSO:	Enhanced Cat Swarm Optimization
FPA:	Flower Pollination Algorithm
GA:	Genetic Algorithm
GWO:	Grey Wolf Optimization
HABC:	Hybrid Artificial Bee Colony Algorithm.
HBMO:	Honey Bee Mating Optimization
KHM:	K-harmonic Means
MCSS:	Magnetic Charge System Search
MOCA:	Magnetic Optimization Algorithm for Data Clustering
PSO:	Particle Swarm Optimization
SA:	Simulated Annealing
SMP:	Seeking Memory Pool
SRD:	Seeking Range of Selected Dimension
TS:	Tabu Search

REFERENCES

- Bijari, K., Zare, H., Veisi, H., & Bobarshad, H. (2018). Memory-enriched big bang–big crunch optimization algorithm for data clustering. *Neural Computing & Applications*, 29(6), 111–121. doi:10.1007/s00521-016-2528-9
- Boushaki, S. I., Kamel, N., & Bendjeghaba, O. (2018). A new quantum chaotic cuckoo search algorithm for data clustering. *Expert Systems with Applications*, 96, 358–372. doi:10.1016/j.eswa.2017.12.001
- Cao, F., Liang, J., & Jiang, G. (2009). An initialization method for the K-Means algorithm using neighborhood model. *Computers & Mathematics with Applications (Oxford, England)*, 58(3), 474–483. doi:10.1016/j.camwa.2009.04.017
- Chang, D., Zhao, Y., Zheng, C., & Zhang, X. (2012). A genetic clustering algorithm using a message-based similarity measure. *Expert Systems with Applications*, 39(2), 2194–2202. doi:10.1016/j.eswa.2011.07.009
- Chu, S. C., Tsai, P. W., & Pan, J. S. (2006, August). Cat swarm optimization. In *Pacific Rim international conference on artificial intelligence* (pp. 854–858). Springer.
- Cura, T. (2012). A particle swarm optimization approach to clustering. *Expert Systems with Applications*, 39(1), 1582–1588. doi:10.1016/j.eswa.2011.07.123
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Han, X., Quan, L., Xiong, X., Almeter, M., Xiang, J., & Lan, Y. (2017). A novel data clustering algorithm based on modified gravitational search algorithm. *Engineering Applications of Artificial Intelligence*, 61, 1–7. doi:10.1016/j.engappai.2016.11.003
- Hatamlou, A. (2012). In search of optimal centroids on data clustering using a binary search algorithm. *Pattern Recognition Letters*, 33(13), 1756–1760. doi:10.1016/j.patrec.2012.06.008
- Hatamlou, A. (2013). Black hole: A new heuristic optimization approach for data clustering. *Information Sciences*, 222, 175–184. doi:10.1016/j.ins.2012.08.023
- Hatamlou, A., Abdullah, S., & Hatamlou, M. (2011, December). Data clustering using big bang–big crunch algorithm. In *International conference on innovative computing technology* (pp. 383–388). Springer. doi:10.1007/978-3-642-27337-7_36
- Hatamlou, A., & Hatamlou, M. (2013). PSOHS: An efficient two-stage approach for data clustering. *Memetic Computing*, 5(2), 155–161. doi:10.1007/s12293-013-0110-x
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys (CSUR)*, 31(3), 264–323.
- Jiang, B., & Wang, N. (2014). Cooperative bare-bone particle swarm optimization for data clustering. *Soft Computing*, 18(6), 1079–1091. doi:10.1007/s00500-013-1128-1
- Jiang, H., Yi, S., Li, J., Yang, F., & Hu, X. (2010). Ant clustering algorithm with K-harmonic means clustering. *Expert Systems with Applications*, 37(12), 8679–8684. doi:10.1016/j.eswa.2010.06.061
- Karaboga, D., & Ozturk, C. (2011). A novel clustering approach: Artificial Bee Colony (ABC) algorithm. *Applied Soft Computing*, 11(1), 652–657. doi:10.1016/j.asoc.2009.12.025
- Kumar, Y., & Sahoo, G. (2014). A charged system search approach for data clustering. *Progress in Artificial Intelligence*, 2(2-3), 153–166. doi:10.1007/s13748-014-0049-2
- Kumar, Y., & Sahoo, G. (2015). Hybridization of magnetic charge system search and particle swarm optimization for efficient data clustering using neighborhood search strategy. *Soft Computing*, 19(12), 3621–3645. doi:10.1007/s00500-015-1719-0
- Kumar, Y., & Sahoo, G. (2017). Gaussian cat swarm optimisation algorithm based on Monte Carlo method for data clustering. *International Journal on Computer Science and Engineering*, 14(2), 198–210.
- Kumar, Y., & Sahoo, G. (2017). A two-step artificial bee colony algorithm for clustering. *Neural Computing & Applications*, 28(3), 537–551. doi:10.1007/s00521-015-2095-5

- Kumar, Y., & Singh, P. K. (2018). Improved cat swarm optimization algorithm for solving global optimization problems and its application to clustering. *Applied Intelligence*, 48(9), 2681–2697. doi:10.1007/s10489-017-1096-8
- Kushwaha, N., Pant, M., Kant, S., & Jain, V. K. (2018). Magnetic optimization algorithm for data clustering. *Pattern Recognition Letters*, 115, 59–65. doi:10.1016/j.patrec.2017.10.031
- Nanda, S. J., & Panda, G. (2014). A survey on nature inspired metaheuristic algorithms for partitioned clustering. *Swarm and Evolutionary Computation*, 16, 1–18. doi:10.1016/j.swevo.2013.11.003
- Ram, G., Mandal, D., Kar, R., & Ghoshal, S. P. (2015). Cat swarm optimization as applied to time-modulated concentric circular antenna array: Analysis and comparison with other stochastic optimization methods. *IEEE Transactions on Antennas and Propagation*, 63(9), 4180–4183. doi:10.1109/TAP.2015.2444439
- Santosa, B., & Ningrum, M. K. (2009, December). Cat swarm optimization for clustering. In *2009 International Conference of Soft Computing and Pattern Recognition* (pp. 54–59). IEEE. doi:10.1109/SoCPaR.2009.23
- Shelokar, P. S., Jayaraman, V. K., & Kulkarni, B. D. (2004). An ant colony approach for clustering. *Analytica Chimica Acta*, 509(2), 187–195. doi:10.1016/j.aca.2003.12.032
- Taherdangkoo, M., Shirzadi, M. H., Yazdi, M., & Bagheri, M. H. (2013). A robust clustering method based on blind, naked mole-rats (BNMR) algorithm. *Swarm and Evolutionary Computation*, 10, 1–11. doi:10.1016/j.swevo.2013.01.001
- Tsai, P. W., Pan, J. S., Chen, S. M., & Liao, B. Y. (2012). Enhanced parallel cat swarm optimization based on the Taguchi method. *Expert Systems with Applications*, 39(7), 6309–6319. doi:10.1016/j.eswa.2011.11.117
- Wang, R., Zhou, Y., Qiao, S., & Huang, K. (2016). Flower pollination algorithm with bee pollinator for cluster analysis. *Information Processing Letters*, 116(1), 1–14. doi:10.1016/j.ipl.2015.08.007
- Wang, Z. H., Chang, C. C., & Li, M. C. (2012). Optimizing least-significant-bit substitution using cat swarm optimization strategy. *Information Sciences*, 192, 98–108. doi:10.1016/j.ins.2010.07.011
- Yan, X., Zhu, Y., Zou, W., & Wang, L. (2012). A new approach for data clustering using hybrid artificial bee colony algorithm. *Neurocomputing*, 97, 241–250. doi:10.1016/j.neucom.2012.04.025
- Zhang, C., Ouyang, D., & Ning, J. (2010). An artificial bee colony approach for clustering. *Expert Systems with Applications*, 37(7), 4761–4767. doi:10.1016/j.eswa.2009.11.003