Hybrid Harris Hawks Optimization with Differential Evolution for Data Clustering

Laith Abualigah, Mohamed Abd Elaziz, Mohammad Shehab, Osama Ahmad Alomari, Mohammad Alshinwan, Hamzeh Alabool, and Deemah A. Al-Arabiat

1 Introduction

The developments in technologies such as cloud computing, the Internet of things (IoT), and others lead to growing the volume of the collected data [29]. In this context, the techniques of data mining have more attention which is used to discover and extract information from this data. From the existing data mining methods, the clustering is considered as the most popular unsupervised technique used to split the unlabeled data into a set of clusters/groups of similar objects based on specific criteria [100]. Each cluster contains a set of objects that have the same features while different from the objects belong to other groups [27, 61].

Therefore, the cluster techniques have been applied to different applications such as pattern recognition [74, 87], image processing [90, 101, 103, 115, 116], and information retrieval [122, 130, 133]. In addition, there are several clustering methods have been developed and applied to other applications [3, 31, 59, 108, 123].

In general, there are two main categories of the clustering techniques, the first category is hierarchical which is an incremental approach that generates a binary hierarchical tree called a dendrogram. Initially, this technique is to determine the dissimilarity between the objects by computing distance or divergence [6, 14]. Next,

L. Abualigah (✉) · M. Alshinwan
Faculty of Computer Sciences and Informatics, Amman Arab University, Amman 11953, Jordan

M. Abd Elaziz
Faculty of Science, Department of Mathematics, Zagazig University, Zagazig 44519, Egypt

M. Shehab
Information Technology, The World Islamic Sciences and Education University, Amman, Jordan

O. Ahmad Alomari
Faculty of Engineering and Architecture, Istanbul Gelisim University, Istanbul, Turkey

H. Alabool · D. A. Al-Arabiat
College of Computing and Informatics, Saudi Electronic University, Abha, Saudi Arabia

D. Oliva et al. (eds.), *Metaheuristics in Machine Learning: Theory and Applications*,
Studies in Computational Intelligence 967,
https://doi.org/10.1007/978-3-030-70542-8_12

this technique combines the objects that met a specific aggregation criterion, so the number of clusters is created from this step. In addition, the previous step is applied to the objects and classes until reached a specific number of groups. This clustering technique has some limitations such as it needs more computational time especially when the number of objects increased.

The second category of clustering technique is the partitioning [7, 127] which aims to minimize the within-cluster inertia to create a set of clusters, each of them consists of similar objects. The popular algorithms that belong to this category are the K-means, and fuzzy C-means [107]. Both of these two algorithms aim to achieve the criteria of the partitioning category (i.e., minimize within-cluster variation), however, FCM has some advantage over the K-means [56].

These clustering techniques that belong to the partitioning category have some shortcomings, such as being trapped in the local area. Its convergence speed becomes slow, and this will affect the quality of the final solutions [72]. To address these limitations, the metaheuristic (MH) algorithms are used. These MH methods emulate different characteristics from nature that various from biological, physical, and social behaviours. The most popular MH methods are artificial bee colony (ABC) [75], genetic algorithm (GA) [28], salp swarm algorithm [10], bat algorithm [30], krill herd algorithm [24, 41], moth-flame optimization algorithm [109], particle swarm optimization (PSO) [47], lighting search algorithm [5], cuckoo search algorithm [110, 111], group search optimizer (GSO) [4], harmony search (HS) [93], sine-cosine algorithm (SCA) [91], and Arithmetic Optimization Algorithm [9].

However, most of MH methods have several drawbacks when employed to solve the clustering problem [8]. Such as there are algorithms that have a high ability to explore the search domain but their exploitation ability is low. therefore, there is another kind of MH method depends on the hybrid between MH methods. This type of MH methods established their performance in various applications such as underwater wireless sensor networks [78, 79], image segmentation [80], wireless sensor networks [102], feature selection [57], and global optimization [11, 32, 58].

In the related circumstances, this progressed performance of the Harris Hawks optimization (HHO) algorithm [92] and provide it as a clustering approach. This improvement relies upon on the usage of the Differential Evolution (DE) [98] to enhance the exploration capacity of HHO. The DE has been utilized to enhance the performance of numerous MH algorithms which includes [48–50, 52, 53].

This paper introduced a hybrid approach, Harris Hawks optimization (HHO) algorithm with differential evolution (DE), to tackle the clustering problems. The proposed H-HHO based-clustering method used the exploration and exploitation strategies of differential evolution to ameliorate the exploration and exploitation potential approach of the Harris Hawks' optimization set of rules. The proposed clustering method starts by placing the initial penalty for a set of agents then compute each agent's penalty. The following step is to discover an excellent solution, which has an optimal fitness value. Experiments are carried out using five benchmark clustering datasets to validate the performance of the proposed method against other comparative methods. The outcomes showed that the proposed method got better outcomes in evaluation with different comparative techniques.

The main contributions of this chapter are:

- To provide a new clustering approach using an improved Harris Hawks Optimization algorithm.
- Using the Differential Evolution algorithm to enhance the exploitation experience of the Harris Hawks Optimization.
- Find the optimal number of clusters for each dataset.

The remaining of the paper is arranged as in Sect. 2 presents the relevant clustering works. Section 3 shows the proposed H-HHO based-clustering method. Section 4 presents the hybrid algorithm (H-HHO). Section 5 performs experiments results and analyses. And lastly, Sect. 6 gives the outcome and expected future work.

2 Related Works

Recently, many approaches have been suggested to form a clustering function, which is, in turn, re-organize unstructured data into coherent clusters. In [12, 16, 26, 33], β-hill climbing is adapted as an unsupervised feature selection technique to select the most prominent features to empower the K-mean clustering capability to produce more accurate clusters. A new clustering method is proposed in [13] using a hybrid PSO with genetic operators (H-FSPSOTC). In this approach, the PSO algorithm is coupled with GOs to enhance its performance in feature selection tasks and pass to the k-means clustering algorithm, the most relevant and representative subset of features. The results proved that H-FSPSOTC could be an alternative clustering method. Adaptive Dynamic Directive Operative Fractional Lion (ADDOFL) is proposed as a data clustering method in [43]. Moreover, in clustering evaluation, a novel fitness function is proposed on the basis of four kernel functions rational quadratic, inverse multi-quadratic, Gaussian Kernel, and tangential kernel. The results indicated that there is an improvement in the cluster's quality at the expense of computational complexity. Data clustering using Grasshopper Optimization Algorithm (GOA) was developed in [88]. The authors promote the clustering analysis process by using the Calinski-Harabasz index as an internal clustering evaluation metric for the solutions produced by GOA.

[21, 25] proposed an optimization technique called H-KHA that combined Krill herd (KH) and harmony search algorithms in order to improve the ability of global search. H-KHA is applied for clustering purpose. The effectiveness and performance of the proposed algorithm is tested seven standard datasets of UCI Machine Learning Repository. The experiment results ensure superiority of H-KHA over the traditional KH. [117] proposed a hybrid optimization algorithm that combined elephant herding optimization algorithm and k-means for solving data clustering problem. The testing results revealed that the proposed optimization technique is better than other methods namely, particle swarm optimization algorithm and k-means algorithm. While, in [84] magnetic optimization algorithm is applied for data clustering. The result of the proposed method is compared with 5 clustering algorithm namely, HYBRID-DE,

MIN-MAX, PSO, k-means, KFCM. The performance test result shows that magnetic optimization algorithm has a competitive result comparing with other methods. Whiles, [22] developed feature selection method based on particle swarm optimization algorithm called FSPSOTC. The FSPSOTC aims to enhance the efficiency of document clustering algorithm. The testing results shown that FSPSOTS is performed better than K-mean, FSHSTC and FSGATC algorithms in term of document clustering.

[114] developed a new meta-heuristic algorithm for data clustering called intrusive tumor growth optimization algorithm (ITGO). The testing result ensured that ITGO algorithm recorded better result than PSO, DE, and GSA algorithms. [96] resented an automatic fuzzy clustering using non-dominated sorting particle swarm optimization (AFC-NSPSO) algorithm for the purpose of categorical data. The proposed technique is tested and compared with AT-DC, DHCC, PROCAD, and MOCSG algorithms. The performance test result ensures the AFC-NSPSO has obtained more accurate result than other compared algorithm. [125] implemented memetic fuzzy whale optimization algorithm (WOP) for data clustering. The testing result proves the superiority of WOP over GA and PSO. In addition, [134] applied bat algorithm for handling data clustering problem. An extensive of simulation experiments is conducted to examine the performance of parameter setting of bat algorithm. The analysis result shows that bat algorithm is capable to solve the problem of data clustering. In this study [95], a hybrid method that combined elicit teaching learning based optimization and fuzzy c-means clustering algorithm is developed for handling data clustering problem. The experiment result of the proposed method is compared with GA, PSO and IPSO algorithm. The comparison results reveal the superiority of proposed method to other method. Furthermore, [35] used fuzzy cuckoo optimization algorithm to solve data clustering problem. The proposed algorithm is tested and then the result compared with six other algorithms namely, COAC, Black hole, CS, K-mean, PSO and GSA. The comparison results indicate that fuzzy cuckoo optimization algorithm has better performance.

[73] proposed an improved krill herd algorithm to solve the data clustering algorithm. the improvement has been done by increase the exploitation capability of krill herd algorithm. However, the effectiveness and performance of the improved krill herd algorithm are tested. The testing result was compared with thirteen optimization algorithms namely, ABC, CS ,FFA, ACO, BBO, DE, ES, GA , PBIL,PSO, SGA , LFPSO and KH. The analysis result shows that the improved krill herd algorithm can perform better than other algorithm. [20] developed a hybrid krill herd algorithm that combined objective functions and hybrid KH algorithm, called, MHKHA . The study aims to solve the problem of text document clustering analysis. The method was tested by using Nine text standard datasets. In addition the performance of the MHKHA was tested by comparing the result of the proposed algorithm with thirteen of the other competitive algorithm. The analysis result revealed that MHKHA obtained a competitive result comparing to other algorithm. In [104] Whale Optimization Algorithm has been enhanced by using multi-swarm cooperative strategies. Then, the proposed algorithm was tested and applied to solve the problem of data clustering. Authors in [71] implemented a hybrid text clustering algorithm that consist

of k-means and Lion Optimization Algorithm (kLOA). Precision, recall, F-measure, purity and entropy measurements are applied to test the proposed algorithm. in addition, the analysis result was compared with other algorithms. Finally, the obtained result shows that kLOA is perform better than other algorithm.

Also, [2] has presented an improved version of black hole algorithm by using levy flight for optimization and data clustering problems. The performance of designed algorithm was tested using six datasets. The analysis result that the proposed algorithm is applicable and can be used for data clustering. Meanwhile, [15] developed a new feature weight scheme and a dynamic dimension space for feature selection problems. It used for solving text mining clustering problem. The proposed method is tested and compared with three other algorithm namely, GA, HS, and PSO. The analysis result ensure the superiority of the new text mining clustering algorithm over other competitive algorithm. While, [85] opted cuckoo search optimization algorithm and k-modes for solving data clustering problem. The proposed method was applied and tested to prove its applicability and validity. The test result was compared with k-modes and particle swarm optimization algorithm. the comparison result ensure that the proposed method is performed better than k-modes and particle swarm optimization algorithm. [126] presented a modified Fuzzy C-Means Clustering Algorithm based on Particle Swarm Optimization for handling data clustering problem. Authors [86] proposed data mining clustering algorithm based on deep neural networks. However, the above variants of data mining clustering algorithm were tested, which proof its applicability and validity to solve data clustering problems.

[40] presented Big Bang Big Crunch Optimization algorithm (BB-BC) to address data clustering problems. In this approach, a memory with the limited size is incorporated with the process of BB-CC algorithm in a smart way. This is to promote the exploration and exploitation capabilities of BB-CC to produce more precise clusters. [17] studied the effectiveness and performance of Krill herd algorithm with and without genetic operators in addressing clustering problems. Moreover, [40] proposed six versions based on Krill herd algorithm to provide more precise clusters. In the context of clustering, Feature selection is an important step as it removes unwanted features, which, if they exist will mislead the clustering process. In this regard, [19] advanced Genetic Algorithm (GA) as an unsupervised feature selection method as initial stages prior clustering task, called as (FSGATC). This is to reduce "high dimensionality" of raw datasets and find the most relevant features that properly reprehensive the reduced data. Thereafter, K-means is carried on only the reduced data rather than full data to overcome the curse of dimensionality of the raw data, and steering the clustering process on the most relevant feature to yield better quality results. [81] introduced Artificial Bee Colony to fuzzy clustering for outlier detection. From the clustering point of view, a robust feature schema that accurately represents the features information leads to generate more precise clusters. [23] presented two stages of clustering method. In the first stage, a novel term weighting scheme, called length feature weight (LFW), is proposed to assign a robust weight score for all features. In the second stage, a β-hill climbing algorithm is used as a clustering technique.

[19] designed clustering method based on PSO and k-means, where PSO is used to identify the most informative subset of features, and then only the information of the

selected features included in the best subset will be allocated and represent the new data, namely, reduced data. Eventually, K-means is carried out in this reduced data. [51] proposed automatic clustering method based on the integration between atom search optimization (ASO) and sine-cosine algorithm (SCA), called (ASOSCA). The main reason behind this integration is to benefit from the complementary features of both optimization algorithms and also empower the exploitation capability of ASO. In the context of clustering, ASOSCA is harnessed to provide a precise identification of the number of centroids and their positions. ASOSCA method has resulted outstanding results when compared with other state-of-arts clustering methods. Data clustering using Shared Subscribe Hyper Simulation Optimization (SUBHSO) was developed by in [105]. SUBHSO relies on PSO in producing candidate coordinate solutions for the cluster analysis and they evaluated using the Monte Carlo simulation. The performance of the proposed method SUBHSO is tested on big dataset collected from Iran Electricity Market. The results indicated that SUBHSO produced improvement in clustering accuracy. [44] presented a methodology for data clustering that consists of two phases. In the first phase, principal component analysis is used for data-dimensionality reduction, while in the second phase, PSO is used as a clustering technique. Furthermore, other researchers used PSO with some modification to provide better quality clusters.

[34] incorporated kernel density estimation (KDE) with PSO to address the limitations of the traditional PSO. The proposed method achieves better results compared to the basic PSO clustering method and the other state-of-art clustering method when tested on high-dimensional datasets. [42] presented a honey bee mating algorithm in load profile clustering in present data situations. [1] described hybridization strategies between a couple of optimization algorithms based on data clustering. The first hybridization strategy combines ant colony optimization and particle swarm optimization, while the other hybridization strategy combines these two algorithms with the genetic algorithm. The results proved that the integration between these optimizations take the advantages from the complementary features from each other's and yield better results than individual algorithms.

[46] proposed an algorithm to clustering the data based on a modified adaptive particle swarm algorithm. To modify the traditional algorithm, a combination of the bee colony algorithm and PSO algorithm proposed, because of the fast concurrence, and the behavior of adaptive technique based on the evolutionary element. The numerical analysis of the proposed algorithm shows a significant performance.The artificial bee colony (ABC) algorithm based on arbitrary search, which leads to satisfactory for exploration but inadequate for exploitation. On the other hand, The PSO algorithm has a robust performance in global search but weak exploitation search property. [97] proposed to combine both algorithms to enhance the data clustering, authors review various hybridization approach (i.e. Cellular automata, Multistage, Component-based, Recombination and Chain). The hybridization of PSO and ABC will make a balance for exploration and exploitation processes, which enhance the avoidance of the local optima and convergence speed. Finally, the ABC algorithm implemented on the data clustering.

[77] presented a hybrid algorithm to enhance the clustering process based on the Ant Colony Optimization (ACO) algorithm, Particle Swarm Optimization (PSO) algorithm, named (ACPCO). The ACO algorithm implemented to detect centroids with the simulation of the ACO system, the PSO algorithm employed to find out the optimal cluster. [60] introduced a combination of Genetic algorithm (GA) and PSO algorithm to improve data clustering. [54] proposed a hybridization of two algorithms Imperialist Competitive Algorithm (ICA) and the Fuzzy KMeans (FKM) named ICAFKM. On the other hand, another algorithm proposed called PSOFKM. This hybridization to enhance the FKM sensitivity to initial states. A modification of Animal Migration Optimization (AMO) with an entropy-based heuristic approach called (EAMO) proposed by [68]. This algorithm determines the data entropy for each attribute of a data set, and propose an adaptive approach can automatically adjust convergence rate and global search.

In [55] a review of PSO algorithm and its variants to clustering data. In [45] introduce a new algorithm called Class Topper Optimization (CTO), which inspired from the student learning ability in the class. Besides, the proposed approach is a population-based search algorithm to solve the clustering problem. [83] investigated in the applicability of the K-means algorithm in term of unsupervised learning into Teaching-Learning-Based Optimization (TLBO) in two attempts. First attempt to find out the optimized number of directly classified obstruction in the data without previous information. The second attempt is to inspect direct classified blockage with cluster validity indices (CVIs) to validate the clustering result. The introduced algorithm named automatic clustering using TLBO (AutoTLBO), achieve the evolutionary by a combination of k-means and CVIs into TLBO. Moreover, this approach reduces the inter-cluster distance and improve the inter-cluster gaps amongst the data.

The comparison study shows robust performance comparing with metaheuristic and heuristic algorithms. On the same track, [132] proposed a new clustering algorithm called multivariant optimization algorithm (MOA) to find out the optimal solution exploration and exploitation. An improved clustering algorithm combining the cat swarm optimization (CAT) and k-harmonic means (KHM), called (ICSOKHM) proposed by [82], to solve the clustering problem. To meet the big data clustering requirements (i.e. accuracy and efficiency), a new clustering algorithm based on the incremental kernel fuzzy proposed. First of all, a summary of the big data mining algorithms and fuzzy clustering approaches presented. Then, an optimized clustering of the data mining algorithm with incremental kernel fuzzy. Finally, the proposed algorithm compared with the stKFCM algorithm to validate the performance [131]. In the metagenomics, the amino acid sequences grouped with a similarity rate producing the DNA fragment clusters, which generate big data. [118] propose an Iterative k-means and TensorFlow to decrease the execution. [119] uses the contiguous Fuzzy c-means (FCM) to cluster the vertex within logical areas. To deal with the integration issue in the FCM, the authors use the Gravitational Search Algorithm (GSA). The GSA gives the optimal number of clusters and decreases the fitness method. In the trajectory data, which are suffering from a tremendous amount of data and complexity.

[64] proposed an improved density-based K-means. Based on the density and growing the density load of relevant points, the initial clustering points selected to implement the K-means clustering. [63] introduces a hybridization algorithm combining Group Search Optimizer (GSO) and the firefly algorithm. The proposed algorithm is replacing the lowest fitness values in any iteration of the GSO algorithm with new values taken from the firefly algorithm. In the area of range based multi-target localization and tracking in the wireless sensor networks (WSNs), estimation data of each period cannot correlate with its corresponding target, and clustering examination can be used to resolve this issue. To solve this problem [124] propose a new kernel clustering approach. In order to accomplish more effective power consumption, [129] introduces a load disaggregation approach for low sampling data from smart meters depend on support vector optimization and clustering algorithm. the author proposes a combination of active time warping and k-median algorithm, which will help in identifying the running device and recovers original power consumption from an aggregate meter signal through optimized support vector regression (OSVR).

3 Preliminaries

Clustering is an unsupervised task to set up bunches of information that focuses on diverse comparative clusters. The yield of the clustering predicament can be dubious (Elite clustering) or direct (Covering Clustering). Extreme clustering shows information focuses allude to one cluster, whereas in simple clustering, information focuses can allude to different bunches, as appeared in Fig. 1. Information focuses can be outlined as multi-dimensional vectors, where dimension appears a few objects within the information [84].

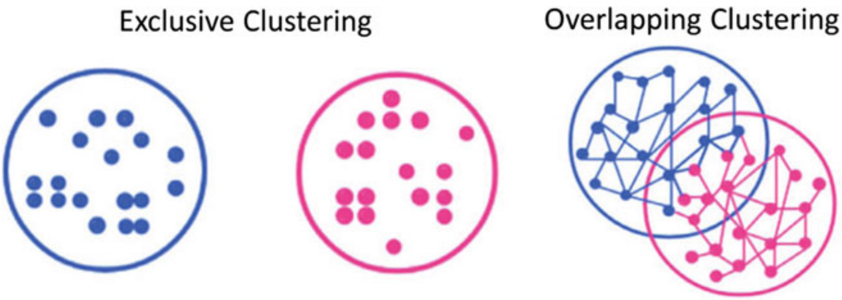


Fig. 1 Exclusive clustering vs. overlapping clustering [84]

3.1 Cluster Analysis

Mathematically, the data clustering problem can be presented in a N Euclidean space is the method of grouping N data object into K (number of groups or clusters). The goal of data clustering is to partition data point (grouping) X into K clusters $C = \{c_1, c_2, \ldots, c_j, \ldots, c_K\}$.

where,

$$X = \bigcup_{j=1}^{K} C_j \tag{1}$$

$$C_j \bigcup C_i \quad for \quad j \neq i \tag{2}$$

C_j denoted to the $i_t h$ partition in the group in such a way that data object refers to one group as alike and different to other groups as potential.

The distance test is one of the basic variables in information clustering. Ordinarily, the closeness between the two information focuses is evaluated by measuring the remove between them within the accessible look space. There are various distance tests conceivable, the foremost well known one is the Euclidean separate. The objective work is utilized to assess the nature of the clustering calculation (clusters quality) [94].

The Euclidean distance value between data object i and data object j can be determined as follows Eq. (3).

$$Des(O_i, O_j) = \sqrt{\sum_{t=1}^{m} (O_{i,t} - O_{j,t})^2} \tag{3}$$

where, j denoted to the j_{th} data object and m denoted to the length of the dimension. The most popular measure applied is the sum of intra-cluster distance [69].

Whole of intra-cluster remove degree: The separate esteem between the information objects of a cluster to the centroid of that cluster, which is decided for all clusters by utilizing Eq. (4).

$$intra_{sum} = \sum_{j=1}^{K} \left\| X_{pj} - P_j \right\| \tag{4}$$

where, j denoted to the j_{th} data object and P_j denoted to the centroid of the j_{th} cluster.

3.2 External Measure

Error Rate (ER) as a noticeable quality ratio: The percentage of misplaced data objects as described in Equation:

$$ER = \frac{Number\ of\ misplaced\ objects}{total\ number\ of\ objects\ within\ dataset} \times 100 \tag{5}$$

4 H-HHO: Hybrid Harris Hawks Optimization with Differential Evolution

This section presents the improvement processes of the basic HHO algorithm by taking the advantages of the DE algorithm by merging them together. The outcome of this hybridization summarizes in production a new proposed method, called H-HHO. The basic HHO as all metaheuristic algorithms, it not ideal to solve all problems in various fields. Consequently, it should be modified to be suitable. For instance, the basic HHO algorithm has a limitation in the local search technique. So, we utilized the DE to improve the exploitation search of the basic HHO algorithm. As well as enhance the convergence rate.

4.1 Harris Hawks Optimization Algorithm

This segment presents the exploratory and exploitative perspectives of the fundamental Harris Hawks of prey optimization (HHO) calculation. It was propelled by investigating prey, astonish jump, and different assaulting plans of Harris Hawks of prey. This calculation, HHO, is considered as a population-based procedure, gradient-free optimization strategy; subsequently, it can be utilized to illuminate any optimization issue subject to a customary detailing. Figure 2 presents all stages of Harris Hawks of the prey optimization algorithms, which are spoken to within the taking after subsections. The Pseudo-code of the Harris Hawks optimization calculation is detailed in Calculation 3.

4.1.1 Exploration Strategy

In this segment, the investigation procedure of Harris Hawks optimization calculation is given. In case we examine the characteristics (nature)of Harris' Hawks of prey, they can seek after and distinguish the prey by their gigantic eyes, but sometimes the casualty can not be found essentially. Subsequently, the Hawks of prey hold up, recognize, and observe the leave put to capture prey possibly following a few hours.

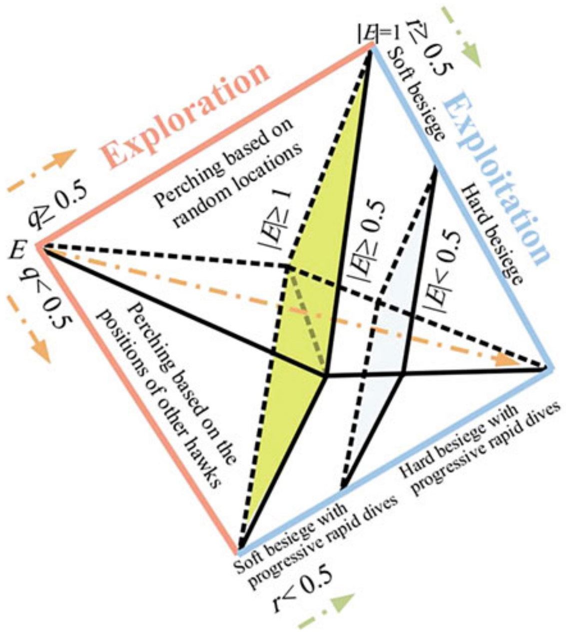


Fig. 2 Different phases of Harris Hawks optimization (HHO) algorithm [67]

Algorithm 1 Pseudo-code of HHO algorithm

Inputs: The population size N and maximum number of iterations T
Outputs: The location of rabbit and its fitness value
Initialize the random population $X_i (i = 1, 2, \ldots, N)$
while (stopping condition is not met) **do**
  Calculate the fitness values of hawks
  Set X_{rabbit} as the location of rabbit (best location)
  for (each hawk (X_i)) **do**
    Update the initial energy E_0 and jump strength J ▹ E_0=2rand()-1, J=2(1-rand())
    Update the E using Eq. (3)
    if ($|E| \geq 1$) **then** ▹ Exploration phase
      Update the location vector using Eq. (1)
    if ($|E| < 1$) **then** ▹ Exploitation phase
      if ($r \geq 0.5$ and $|E| \geq 0.5$) **then** ▹ Soft besiege
        Update the location vector using Eq. (4)
      else if ($r \geq 0.5$ and $|E| < 0.5$) **then** ▹ Hard besiege
        Update the location vector using Eq. (6)
      else if ($r < 0.5$ and $|E| \geq 0.5$) **then** ▹ Soft besiege with progressive rapid dives
        Update the location vector using Eq. (10)
      else if ($r < 0.5$ and $|E| < 0.5$) **then** ▹ Hard besiege with progressive rapid dives
        Update the location vector using Eq. (11)
Return X_{rabbit}

Fig. 3 The Pseudo-code of the Harris Hawks optimization algorithm [67]

In this calculation, the Harris' Hawks of prey are the competitor arrangements (candidate), and the most excellent arrangement in each step is recognized as the aiming prey or roughly the ideal (about). In HHO, the Harris' Hawks rest arbitrarily on any area and hold up to uncover a casualty based on two approaches.

In the event that we look at an break even with opportunity q for each roosting approach, the land-based on the areas of other family individuals, positions, (to be close sufficient to them when locks in) and the creature, which is displayed in Eq. (6) for the condition of $q < 0.5$, or roost on arbitrary tall trees (irregular focuses interior the group's base extend), which is spoken to in Eq. (6) for condition of $q \geq 0.5$.

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1|X_{rand}(t) - 2r_2| & q \geq 0.5 \\ (X_{rabbit}(t)) - r_3(LB + r_4(UB - LB)) & q < 0.5 \end{cases} \tag{6}$$

where $X(t+1)$ is the vector of Hawks (solution' positions) for the another cycle when the current emphasis is t, $X_{rabbit}(t)$ is the position of rabbit, $X(t)$ is the current values vector of birds of prey. r_1, r_2, r_3, r_4, and q are irregular values inside (0,1), which are reestablished for each emphasis. LB and UB indicated to the upper bounds and lower bounds of the issue factors. $X_{rand}(t)$ could be a haphazardly chosen sell from the candidate arrangements (populace), and X_m is the normal position esteem of the current arrangements on of Hawks.

A straightforward outline to create arbitrary areas inside the group's base extend (LB, UB). The primary control produces arrangements concurring to a irregular place and other birds of prey. Within the second control of Eq. (6), the variety of the finest area so distant and the medium position of the bunch furthermore a randomly-calculated portion based on arrangement of factors, whereas r_3 could be a scaling coefficient to extend the arbitrary nature of run the show advance once r_4 gets near rates to 1, and comparative recurrence designs may happen. In this control, a randomly-calculated development separate is included to the LB. Next, a random-calculated coefficient is considered for the portion to provide more expansion patterns and investigate different districts of the look space. It is sensible to form different upgrading rules, but the foremost direct run the show is utilized, which can imitate the practices of falcons. The medium position of falcons is accomplished utilizing Eq. (7).

$$X_m(t) = \frac{1}{N}\sum_{i=1}^{N} X_i(t) \tag{7}$$

where $X_i(t)$ denoted to the place of each hawk in the $t_t h$ iteration and N indicated to the total number of Hawks. It is sensible to get the average location in various ways, but the most straightforward rule is utilized.

4.1.2 Transition from Exploration Stage to Exploitation Stage

The Harris Hawks optimization can move from investigation look procedure to misuse look technique and following, switch between other exploitative behaviors based on the getting away control of the prey. The capacity of prey decreases impressively when getting away response. To speak to this reality, the control of prey is decided by Eq. (8).

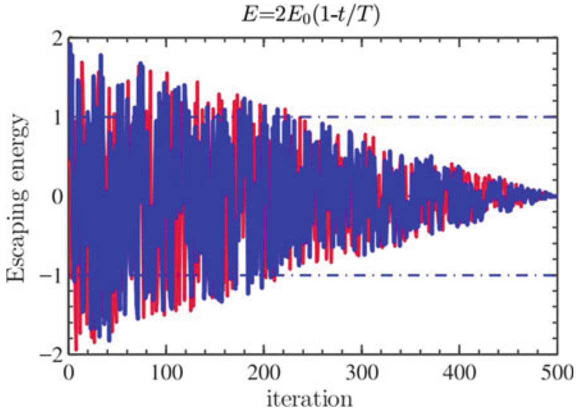


Fig. 4 The behavior of E through 2 runs and 500 iterations [67]

$$E = 2E_0(1 - \frac{t}{T}) \tag{8}$$

where E implies the getting away control of the prey, T signified to the greatest emphasess number, and E_0 signified to the initial state of its control. Within the Harris Hawks of prey optimization calculation, E_0 stochastic alterations within the interim -1 to 1 at each emphasis. When the E_0 diminishes its esteem from to -1, the rabbit is truly diminishing, whereas amid the esteem of E_0 develops from to 1, it demonstrates that the rabbit is fortification. The dynamic getting away potential E incorporates a decreasing point through the cycles. When the getting away control $|E| \geq 1$, the Hawks look different locales to look at a rabbit area. Subsequently, the Harris Hawks optimization produces the investigation period, and when $|E| < 1$, the calculation tries to utilize the neighborhood of the current arrangements through the abuse activities. In brief, investigation happens when $|E| \geq 1$, and misuse happens within the another rounds when $|E| < 1$. The time-dependent work of E is advance illustrated in Fig. 4.

4.1.3 Exploitation Stage

In this stage, Harris' Hawks make the surprise pounce by hitting the planned prey discovered in the early stage, as called in [39]. Nevertheless, prey usually tries to avoid aggressive circumstances. Therefore, several chasing behaviors happen in real circumstances. Based on the escaping conventions of the prey and chasing policies of Harris' Hawks, four potential approaches are introduced in the Harris hawk's optimizer to reduce the attacking plane.

The preys persistently attempt to maintain a strategic distance from undermining circumstances. Expect that r is the opportunity of prey in effectively maintaining a strategic distance from ($r < 0.5$) or not effectively dodging ($r < 0.5$) ere wonderment jump. Anything the prey does, the birds of prey will make a difficult or delicate assault to seize the prey. It demonstrates that they will encompass the prey from distinctive ranges delicately or difficult, depending on the reviewed control of the prey. In genuine circumstances, the birds of prey gotten to be closer and closer to the anticipated prey to improve their openings in agreeably hitting the rabbit by making the shock jump. After a few minutes, the escaping prey will squander increasingly control; at that point, the birds of prey increment the blockade strategy to capture the depleted prey easily.

To make this approach and permit the Harris Hawks optimization to move among delicate and difficult attack forms, the E parameter esteem is utilized. In this regard, when $|E| \leq 0.5$, the delicate assault issues, and whereas $|E| < 0.5$, the difficult assault happens.

1—Soft Besiege

When $r \leq 0.5$ and $|E| \leq 0.5$, the rabbit however has adequate control and tries to dodge by a few arbitrary deluding boundaries, but inevitably, it can not. Amid these endeavors, the Harris' falcons encompass it delicately to urge the rabbit more utilized and after that make the astonish jump. The following rules model this behavior:

$$X(t+1) = \Delta X(t) - E|JX_{rabbit}(t) - X(t)| \tag{9}$$

$$\Delta X(t) = X_{rabbit}(t) - X(t) \tag{10}$$

where $\Delta X(t)$ is the distinction among the position solution of the rabbit and the current position in t_{th} iteration, r_5 is a random value within (0,1), and $J = 2(1 - r_5)$ denoted to the random value of the jump power of the rabbit during the escaping scheme. The J value switches randomly at each iteration to mimic the nature of rabbit movements.

2—Hard Besiege

When $r \geq 0.5$ and $|E| < 0.5$, the prey is so tiring, and it has moo rising control. Other than, the Harris' Hawks continuously encompass the anticipated prey to perform the shock jump at long last. In this circumstance, the current positions are reestablished utilizing Condition (11).

$$X(t+1) = X_{rabbit}(t) - E|\Delta X(t)| \tag{11}$$

An easy illustration of this step with one hawk is represented in Fig. 5.

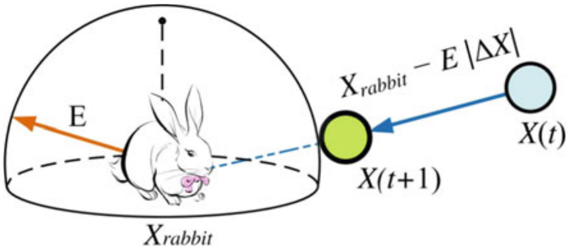


Fig. 5 An easy illustration of overall vectors in the state of tough besiege [67]

3—Soft Besiege with Progressive Rapid Dives

When however $|E| \geq 0.5$ but $r < 0.5$, the rabbit has adequate control to elude effectively, and still, a delicate assault is gathered ere the shock jump. This strategy is more brilliantly than the prior occurrence.

To numerically speak to the getting away models of the prey and jump activities (as called in [39]), the exact flight (LF) hypothesis is utilized within the Harris Hawks optimization. The exact flight is utilized to mimic the genuine crisscross deceptive developments of preys (identity rabbits) amid the getting away state and bizarre, unexpected, and quick plunges of birds of prey adjacent the getting away prey. Falcons make a few groups quick plunges adjacent the rabbit and endeavor to continuously move forward their area and bearings concerning the equivocal developments of prey. Genuine estimations moreover suggest this instrument in other competitive positions in nature. It has been strengthened that exact flight-based developments are the ideal looking strategies for foragers/predators in non-destructive scavenging prerequisites [70, 120]. Also, it has been recognized that the LF-based designs can be distinguished within the chasing developments of creatures, comparative to monkeys and sharks [62, 112, 121]. So, the exact flight-based developments were utilized inside this state of the Harris Hawks optimization.

Persuaded by genuine behaviors of Hawks, it is expected that they can continuously select the leading doable plunge close the prey when they need to catch the prey in competitive circumstances. In like manner, to create a soft besiege, they expected that the Hawks may judge (choose) their ensuing move based on the another run the show in Condition (12).

$$Y = X_{rabbit}(t) - E|JX_{rabbit}(t) - X(t)| \tag{12}$$

At that point, they analyze the sensible result of such a alter to the past drop to find that will it be a substantial jump or not. On the off chance that it was not doable (when they recognize that the prey is making more equivocal activities), they too start to work sporadic, unpleasant, and quick jumps when comparing the rabbit. It is expected that they will hop based on the exact flight-based models utilizing the following rule [By Eq. (13)]:

$$Z = Y + S \times LF(D) \tag{13}$$

where D is the length of dilemma (dimension), and S is random vector values by size $1 \times D$, and LF is the levy flight, which is determined using Eq. (14) [128].

$$LF(x) = 0.01 \times \frac{\mu \times \sigma}{|\upsilon|^{\frac{1}{\beta}}} \tag{14}$$

where,

$$\sigma = \left(\frac{\Gamma(1+\beta) \times sin(\frac{\pi\beta}{2})}{\Gamma(1+\frac{\beta}{2}) \times \beta \times 2^{\frac{\beta-1}{2}}} \right) \tag{15}$$

where μ, υ are random values between 0 and 1, β is a default fixed value to 1.5.

Hence, the final procedure for updating the positions of solutions in the soft besiege state can be determined by Eq. (16).

$$X(t+1) = \begin{cases} Y \ if \ \ F(Y) < F(x(t)) \\ Z \ if \ \ F(Z) < F(x(t)) \end{cases} \tag{16}$$

where Y value and Z value are determined by using Eqs. (12) and (13).

A straightforward illustration of this activity for one hawk is outlined in Fig. 6. Note that the area history of require flight-based jump movement designs through a few emphasess are too detailed and displayed in this case. The colored dots are the area follows of exact flight-based models in one trial, and after, that the Harris Hawks optimization gives to the area Z. In each activity, as it were the more favorable position Y or Z will be chosen as the consequent area. This approach is utilized in all arrangements.

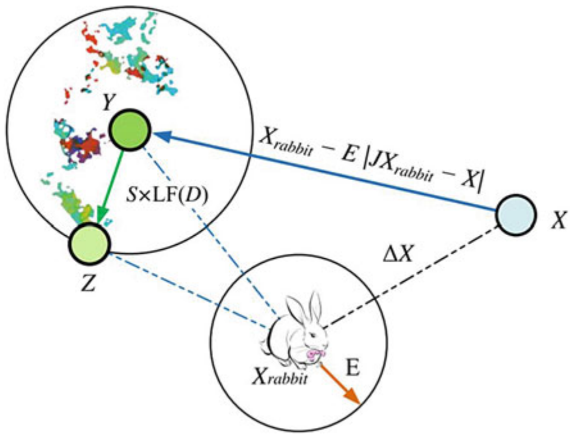


Fig. 6 The pattern of overall vectors in the state of soft besiege with continuous speedy dives [67]

4—Hard besiege with progressive rapid dives

When $|E| < 0.5$ and $r < 0.5$, the rabbit has not adequate control to turn away, and a hard besiege is gathered some time recently the astonish jump to capture and slaughter the prey. The position of this activity within the prey plane is similar to that within the soft besiege, but this time, the falcons endeavor to decrease the remove of their medium position with the getting away prey. Appropriately, the taking after run the show is made in difficult assault circumstance [see Eq. (17)]:

$$X(t+1) = \begin{cases} Y \ if \ F(Y) < F(X(t)) \\ Z \ if \ F(Z) < F(X(t)) \end{cases} \tag{17}$$

where Y value and Z value are determined by using new rules in Eqs. (18) and (19).

$$Y = X_{rabbit}(t) - E|JX_{rabbit} - X_m(t)| \tag{18}$$

$$Z = Y + S \times LF(D) \tag{19}$$

where $X_m(t)$ is achieved using Eq. (7). A simple illustration of this measure is shown in Fig. 7. Note that the colored specks are the range impressions of exact flight-based models in one trial, and as it were Y or Z will be the convenient area for the current iteration.

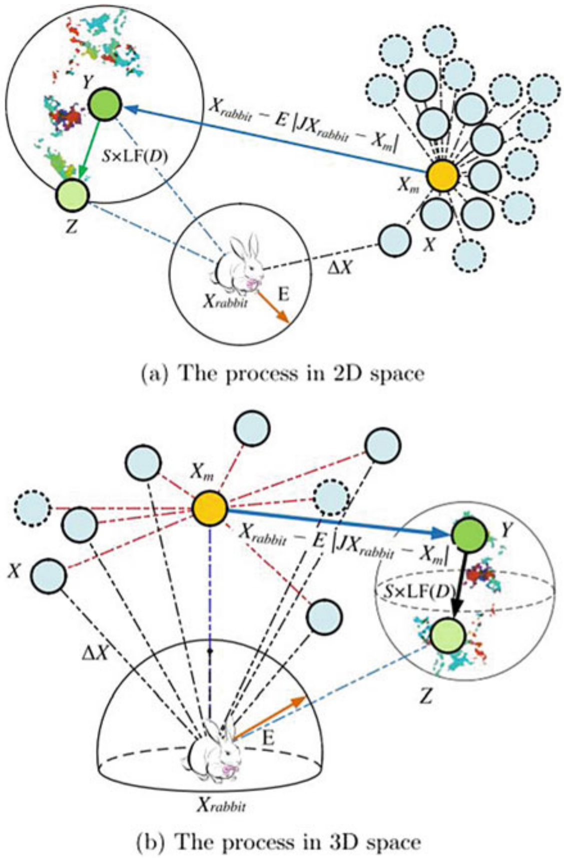
4.1.4 Differential Evolution (DE)

Differential evolution (DE) algorithms are used metaheuristics for global optimization. As well as it locates under the population-based search techniques which used to minimize an objective function. DE is one of the commonly used algorithms because it has various attractive advantages. For instance, it has a few control parameters with a strong structure and small code. Also, it's significantly robust and faster for fixing constrained and unconstrained Optimization Problems. DE improved its efficiency in solving problems in different areas, such as the chemical field [36], power and energy [38], and engineering [37].

The optimization processes of DE consists of three main operations (similar to GA): Mutation, Crossover, and selection. As shown in Fig. 8, the DE algorithm starts with generating new solutions where each solution in the population is created by appointing random values related with the decision parameter. In each iteration (t), each individual $X_i(t)$ in the population is changed based on the Donor vector $V_i(t)$. Where the $V_i(t)$ for each individual of *ith* is created based on the r_1, r_2, and r_3th vectors, these vectors are selected randomly from the population. The following equation shows the process for the *jth* component of each vector.

$$v_{i,j}(t+1) = x_{r_i.G}(t) + F.(x_{r_2,j}(t) - x_{r_3,j}(t)) \tag{20}$$

Fig. 7 The pattern of overall vectors in the state of tough besiege with continuous fast dives in 2D and 3D space [67]



F is a real and fixed factor which monitors the amplification of the differential variation $(x_{r_2,j}(t) - x_{r_3,j}(t))$.

For raising the diversity of the parameter vectors, the vector $U = (u_1, u_2, \ldots, u_D)$ by

$$u_j = \begin{cases} v_j \quad forj = \langle n \rangle_D, \langle n+1 \rangle_D, \langle n+L-1 \rangle_D \\ (x_{j,G})_j \qquad\qquad otherwise \end{cases} \tag{21}$$

where the brackets $\langle\rangle_D$ refers to the modulo function with modulus D.

In case there is a new vector in the population, it will compare G with x_i. If the objective function value of u_j is smaller than x_i, G, then the G+1 is set to u_j. Else, it will keep the x_i, G.

For the mutation operator, the generation of a mutant vector is calculated by the following equation.

$$v_{i,G} = \left\{v_{1i,G}, v_{2i,G}, \ldots, v_{Di,G}\right\} \tag{22}$$

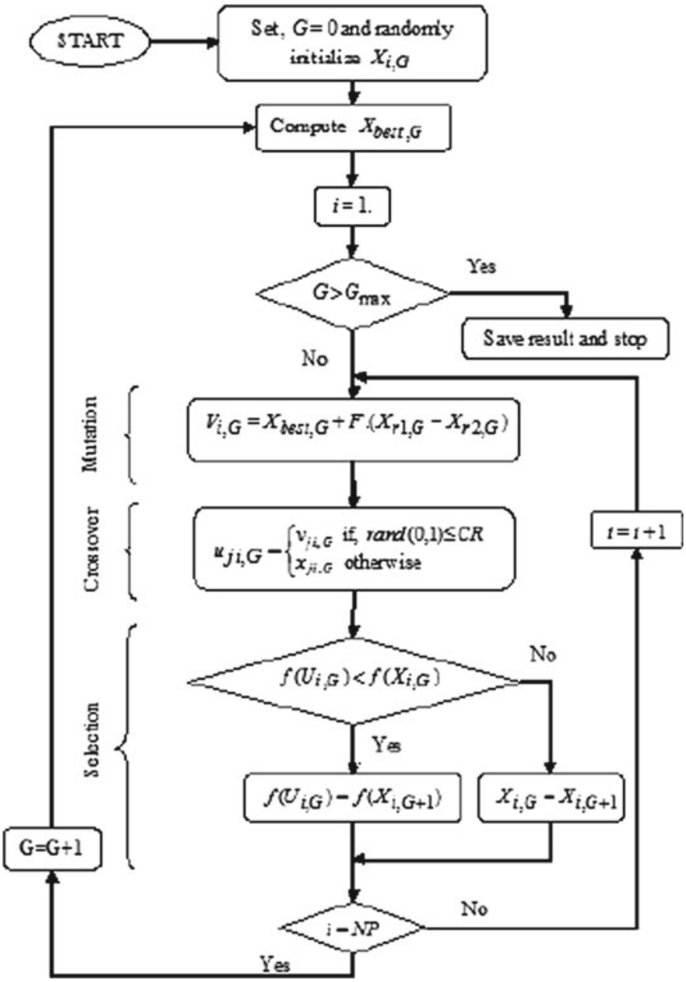


Fig. 8 Flowchart of the differential evolution

for each target vector $x_{i,G}$. In this work, the DE strategy used is

$$v_{i,G} = x_{best,G} + F.(x_{r1,G} - x_{r2,G}) \tag{23}$$

where $r_1, r_2 \in [1, NP]$, $r_1 \neq r_2 \neq i$. $F \in [0, 2]$, and $x_{best,G}$ is the vector which has best fitness at G^{th} generation.

Equation 26 represents the crossover operation trial vector.

$$U_{i,G} = \{u_{1i,G}, u_{2i,G}, \ldots, u_{Di,G}\} \tag{24}$$

Based on the target vector $x_{i,G}$ and the mutant $v_{i,G}$, the crossover can be defined as the following equation.

$$u_{ji,G} = \begin{cases} v_{ji,G} \ \mathit{if} \ \ rand(0,1) \leq CR \\ x_{ji,G} \qquad otherwise \end{cases} \tag{25}$$

where CR (Crossover factor) is constant in the range of (1, 0). The common value of CR = 0.2.

Finally, the selection operator defines by comparing the value of the objective function for each $f(U_{i,G})$ and $f(x_{i,G})$. The vector with smaller value of objective function will continue to the next generation. The following equation describes the selection operation.

$$x_{i,G+1} = \begin{cases} U_{i,G} \ \mathit{if} \ f(U_{i,G}) < f(x_{i,G}) \\ x_{i,G} \qquad otherwise \end{cases} \tag{26}$$

5 Experiments Results and Discussions

5.1 Experimental Setup

Experiments were executed using a personal computer with 8G RAM. The proposed algorithm was programmed in MATLAB R2015a (8.5.0.197613). The experimental outcomes are means of 15 runs of the simulation. The termination condition is set 1000 max number iteration.

We experiment several various methods on five different datasets (i.e., 1—Iris dataset, 2—Wine dataset, 3—WBreast-Cancer dataset, 4—Contraceptive method choice (CMC) dataset, and 5—Glass dataset as shown in Table 1) that cover examples of datasets of the low, medium, and high dimensions space. All these datasets, which we presented in this chapter, are available in the UCI repository at https://archive.ics.uci.edu/ml/index.php.

The descriptions of the used datasets are presented as follows:

- The Iris dataset includes three classes, where each type belongs to a variety of iris plants. In the iris dataset, there are four characteristics, which are sepal width in cm, sepal width in cm, petal length in cm, and petal width in cm.
- The wine test dataset includes chemical studies of wines obtained from three various cultivars, with thirteen attributes, specifically, malic acid, ash, the alkalinity of ash, non-flavonoid phenols, magnesium, proanthocyanins, total phenols, flavonoids, color intensity, hue, alcohol, OD280/OD315 of diluted wines, and praline.

Table 1 The main components of the selected test datasets.

Data set	Number of centers (clusters)	Number of attributes (features)	Number of data objects
Iris	3	4	150 (50, 50, 50)
Wine	3	13	178 (59, 71, 48)
WBreast-cancer	2	9	699 (458, 241)
CMC	3	9	1473 (629, 334, 510)
Glass	9	6	214 (70, 17, 76, 13, 9, 29)

- The dataset was a description of the Wisconsin breast cancer benchmark, including 683 cases having nine components. They included: Clump Thickness, Cell Size Uniformity, Cell Shape Uniformity, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nuclei, and Mitoses. Each of the cases was reasonable for one class, either favorable or dangerous.
- The CMC test dataset is fonded in 1987 by the National Indonesia Contraceptive Prevalence Survey. The objects are married women who either were not pregnant or did not recognize if they were at the time of the interview. The problem requires divining the choice of the modern contraceptive means of a woman based on her demographic and socioeconomic features.
- The glass classification dataset includes six various types of glassful, which are float-processed vehicle windows, float-processed building windows, containers, tableware, non-float-processed building windows, and headlamps. There are nine characteristics, namely, aluminum, refractive index, barium, sodium, magnesium, silicon, potassium, calcium, and iron.

5.2 Comparisons for Data Clustering

We take five data clustering datasets with various dimensions from the repository of the machine learning, namely, Iris dataset, Wine dataset, WBreast-Cancer dataset, Contraceptive method choice (CMC) dataset, and Glass dataset. These datasets have been utilized to solve the data clustering using several different other optimization algorithms.

Table 2 clarifies the wellness values obtained by the comparative optimization calculations. For Iris datasets, TLBO, DE, GSA, and ITGO made way better than other comparative strategies (i.e., K-means, H-HHO) concurring to the min intra-cluster remove values. ITGO made the finest comes about compared with other calculations (i.e., K-means, TLBO, DE, GSA, and H-HHO) concurring to max intra-cluster remove values. ITGO made the finest comes about compared with other calculations (i.e., K-means, TLBO, DE, GSA, and H-HHO) concurring to Implies intra-cluster separate values. As well as, ITGO made the most excellent comes about compared with other algorithms (i.e., K-means, TLBO, DE, GSA, and H-HHO) concurring to Std intra-cluster remove values. At long last, for this dataset, ITGO made the most excellent comes about compared with other calculations (i.e., K-means, TLBO, DE, GSA, and H-HHO) concurring to Middle intra-cluster separate values.

For Wine datasets (see Table 2), H-HHO made the finest comes about compared with other calculations (i.e., K-means, TLBO, DE, GSA, and ITGO) concurring to min intra-cluster remove values. Moreover, H-HHO made the leading comes about compared with other calculations (i.e., K-means, TLBO, DE, GSA, and ITGO) agreeing to max intra-cluster separate values. H-HHO made the most excellent results compared with other calculations (i.e., K-means, TLBO, DE, GSA, and ITGO) concurring to Implies intra-cluster separate values, as well. ITGO made the leading comes about compared with other calculations (i.e., K-means, TLBO, DE, GSA, and H-HHO) concurring to Std intra-cluster remove values. At long last, for this

Table 2 The sum of intra-cluster distances achieved by several comparative algorithms using different datasets.

Datasets	Criterion	K-means [66]	TLBO [106]	DE [113]	GSA [65]	ITGO [114]	H-HHO
Iris	Min	97.3259	**96.6554**	**96.6554**	**96.6554**	**96.6554**	97.2352
	Max	123.8497	125.0844	124.5916	127.6676	**96.9682**	122.3654
	Means	104.8645	103.6115	100.2417	98.2060	**96.7117**	98.2954
	Std	11.8063	11.0483	8.8373	6.9345	**0.1065**	4.3658
	Median	97.3462	96.6686	96.6913	96.6554	**96.6554**	98.6985
Wine	Min	16,555.6794	16,295.1447	16,292.1888	17,123.2430	16,292.1904	**16,121.5714**
	Max	19,436.9520	16,461.2366	16,333.0285	22,259.2894	16,294.1875	**18,457,5415**
	Means	16,962.0423	16,329.2662	16,296.0207	19,767.6886	16,293.1343	**16,291,2198**
	Std	733.8995	52.4949	9.0831	1194.7086	**0.7129**	19.5451
	Median	16,555.6794	16,306.1358	16,293.8157	20,029.2376	16,292.6767	**16,290.2468**
Cancer	Min	2986.9613	2964.3870	**2964.3869**	**2964.3869**	**2964.3869**	**2964.3869**
	Max	5115.4278	3542.0871	2974.8663	3071.7828	**2964.3874**	2998.6541
	Means	3031.7678	3097.1951	2965.0402	2980.4701	**2964.3871**	2987.2560
	Std	0.7485	195.5960	2.3327	27.7893	**0.0001**	19.2654
	Median	2998.4278	2970.8433	2964.3870	2968.6077	2964.3871	**2964.3869**
CMC	Min	5542.1821	5532.2446	5532.1847	5532.1847	5532.1881	**5532.1880**
	Max	5545.2928	5598.0817	7018.0412	5532.9950	**5532.2435**	5554.2695
	Means	5543.7253	5542.5356	5885.6379	5532.2642	**5532.2008**	5541.2974
	Std	1.5841	18.5901	627.5897	0.2123	**0.0127**	1.2954
	Median	5545.0497	5532.7779	5532.1848	5532.1847	5532.1964	**5532.1880**
Glass	Min	215.7316	218.3590	382.3724	214.4027	**213.4988**	215.6584
	Max	250.0580	303.1350	410.2468	265.2572	241.1775	**240.1541**
	Means	228.7861	253.1780	405.0525	239.0503	**228.5708**	231.245
	Std	7.6414	18.3250	9.2617	16.8698	10.2850	**7.63541**
	Median	**218.6841**	249.1153	410.2468	236.1694	234.1775	221.6584

dataset, H-HHO made the finest comes about compared with other calculations (i.e., K-means, TLBO, DE, GSA, and ITGO) agreeing to Median intra-cluster remove values.

For Cancer datasets (see Table 2), H-HHO, ITGO, GSA, and DE made the finest comes about compared with other calculations (i.e., K-means and ITGO), agreeing to min intra-cluster remove values. ITGO made the leading comes about compared with other calculations (i.e., K-means, ITGO, ITGO, GSA, DE, and H-HHO) agreeing to max intra-cluster separate values. AS well as, ITGO made the leading comes about compared with other calculations (i.e., K-means, ITGO, ITGO, GSA, DE, and H-HHO) concurring to Implies intra-cluster remove values. Moreover, ITGO made the finest comes about compared with other calculations (i.e., K-means, ITGO, ITGO, GSA, DE, and H-HHO) concurring to Std intra-cluster remove values. At long last, for this dataset, H-HHO made the most excellent comes about compared with other calculations (i.e., K-means, TLBO, DE, GSA, and ITGO) concurring to Meadian intra-cluster remove values.

For CMC datasets (see Table 2), H-HHO made the leading comes about compared with other calculations (i.e., K-means, TLBO, DE, GSA, and ITGO) agreeing to min intra-cluster remove values. Moreover, H-HHO made the finest comes about compared with other calculations (i.e., K-means, TLBO, DE, GSA, and ITGO) agreeing to max intra-cluster separate values. H-HHO made the most excellent results compared with other calculations (i.e., K-means, TLBO, DE, GSA, and ITGO) agreeing to Implies intra-cluster separate values, as well. ITGO made the finest comes about compared with other calculations (i.e., K-means, TLBO, DE, GSA, and H-HHO) agreeing to Std intra-cluster separate values. At last, for this dataset, H-HHO made the finest comes about compared with other calculations (i.e., K-means, TLBO, DE, GSA, and ITGO) agreeing to Meadian intra-cluster remove values.

For Glass datasets (see Table 2), ITGO made the leading comes about compared with other calculations (i.e., K-means, TLBO, DE, GSA, and H-HHO) agreeing to min intra-cluster remove values. Too, H-HHO made the leading comes about compared with other calculations (i.e., K-means, TLBO, DE, GSA, and ITGO) agreeing to max intra-cluster remove values. ITGO made the most excellent results compared with other calculations (i.e., K-means, TLBO, DE, GSA, H-HHO, and ITGO) concurring to Implies intra-cluster separate values. H-HHO made the most excellent comes about compared with other calculations (i.e., K-means, TLBO, DE, GSA, and H-HHO) agreeing to Std intra-cluster separate values. At last, for this dataset, K-means made the finest comes about compared with other calculations (i.e., H-HHO, TLBO, DE, GSA, and ITGO) concurring to Meadian intra-cluster values.

The test results in terms of the utilized Error rate are displayed in Table 3. As recorded in this table, the proposed H-HHO got the leading normal blunder rate esteem, which is better than the other comparison calculations, particularly within the Iris, cancer, CMC, and Glass benchmarks. Hence, the proposed H-HHO calculation may be a more successful optimization calculation than other comparative strategies in understanding clustering issues. Other than, the comparison between the H-HHO

Table 3 Error rate results for five clustering datasets

Dataset	Statistics	GA [89]	PSO [99]	HS [21]	KHA [18]	H-GA [19]	H-PSO [76]	H-HHO
Iris	Best	10.666	10.667	10.509	9.430	9.765	9.666	**9.332**
	Average	21.652	15.867	21.054	22.658	21.100	**15.800**	20.866
	Worst	43.333	43.447	44.286	**42.548**	44.667	44.333	43.333
	Rank	6	2	4	7	5	1	3
Wine	Best	29.310	29.775	29.865	**29.213**	29.654	29.775	29.653
	Average	34.270	32.051	32.568	32.303	30.989	**30.871**	33.564
	Worst	47.753	44.449	44.467	47.191	44.001	43.888	**43.584**
	Rank	7	3	5	4	2	1	6
Cancer	Best	39.510	40.775	40.111	39.256	40.254	39.775	**39.470**
	Average	44.270	43.051	42.054	42.543	41.214	42.125	**39.119**
	Worst	47.753	45.455	45.640	47.191	46.214	46.758	**45.365**
	Rank	7	6	3	5	2	4	1
CMC	Best	54.656	54.101	55.430	53.936	53.124	53.201	**53.165**
	Average	56.697	55.899	56.001	56.056	55.142	54.204	**54.109**
	Worst	57.296	56.486	57.906	56.999	56.214	**55.333**	55.693
	Rank	7	4	5	6	3	2	1
Glass	Best	42.991	43.925	41.162	38.318	35.249	41.589	**34.242**
	Average	51.028	46.262	**43.054**	43.925	44.219	51.617	44.002
	Worst	56.075	52.804	**46.255**	50.476	51.985	56.075	51.445
	Rank	6	5	1	2	4	7	3
Average	Rank	33	20	18	24	16	15	**14**
Final	Rank	7	5	4	6	3	2	**1**

Table 4 The best cluster centroids of Iris dataset acquired by H-HHO

Center 1	Center 2	Center 3
6.021547854	6.521453625	5.014586595
2.654325869	3.002589875	3.321562545
4.326515492	5.526971637	1.369856985
1.365265452	2.032629663	1.112656944

and diverse calculations has values particular from blunder rate values; subsequently, the H-HHO calculation exceptionally effectively works a neighborhood look.

Tables 4, 5, 6, 7 and 8 give the best center of clusters for the used five clustering datasets. Also, to the previously given comparisons in Table 3, the optimization algorithms have been analyzed statistically using the Friedman test to conclude whether there is a significant improvement in the results of the proposed clustering algorithm, hybrid Harris Hawks optimization with differential evolution for data clustering, (H-HHO). Table 3 shows the ranking of the comparative optimization algorithms based

Table 5 The best cluster centroids of Wine dataset acquired by H-HHO

Center 1	Center 2	Center 3
11.758658645	13.654682645	12.576554545
2.9617831254	2.6843234562	1.735767464
2.5451564454	2.4686646232	2.375755991
19.645484545	20.864864686	15.13245345
99.464684564	93.259888224	106.2785705
2.3546845465	2.6451844525	2.827204596
1.6545454324	1.4545435645	3.054445707
0.3544654548	0.2547545348	0.257084030
1.3653365440	1.8653345667	1.025845521
4.5484538428	5.4378769653	5.645832970
0.8548448351	0.9756356867	2.545646543
2.8432684308	2.8635654685	3.435556875
676.18448464	472.65456565	1156.786758

Table 6 The best cluster centroids of Cancer dataset acquired by H-HHO

Center 1	Center 2
2.5484516545	7.7875412582
1.8418518945	6.8754121468
1.6514415414	6.3146841345
1.5842165459	5.8154354138
1.8554821845	5.3415441564
2.8414641648	7.3144884874
2.3148941564	6.1435415351
1.4464164564	6.6241534843
1.0324187842	2.6843641137

Table 7 The best cluster centroids of CMC dataset acquired by H-HHO

Center 1	Center 2	Center 3
23.618464548	34.671653484	43.541648415
3.5415146548	3.2316589369	3.0715416548
3.2165487521	3.2458753245	3.4871564561
1.2523445378	3.2587578472	4.6841348465
1.6515454314	0.2547427674	0.0548148088
0.8724512515	1.8467595754	0.0354514681
2.5144337578	2.9567365867	1.5045645404
2.3174683715	3.6348452635	3.8748941846
0.6471548424	0.3568843567	0.2878991862

Table 8 The best cluster centroids of Glass dataset acquired by H-HHO

Center 1	Center 2	Center 3	Center 4	Center 5	Center 6
1.7841684146	1.5148413501	1.6514654840	1.0556410487	1.2863460436	1.3871843854
14.425682682	17.384714354	15.316786411	19.540404345	15.245624522	18.381368737
0.6787181314	3.5641348148	3.6834148744	2.3064341334	0.2879304536	0.5641387044
1.3878145438	1.3671684874	1.6371504189	2.3541306844	1.5416843844	2.6154184151
75.341841538	75.387184384	79.574101341	76.848048434	73.846887160	79.978484084
0.4154815443	0.5414544852	1.6384104874	2.3410431448	1.3178043184	0.6841044368
12.654056789	9.5468140438	8.1564105489	7.3841848403	14.769549129	8.5410489703
1.5341048413	0.6878981468	0.6810434840	1.4184041348	0.8718648734	1.5641086843
0.5781408044	1.5414841436	0.5741048148	1.3687143464	0.3678314841	0.1841406440

on them. The proposed method got the best final ranking overall, the used datasets in comparison with the other comparative method in that table. Which confirmed the ability of the proposed algorithm in solving the data clustering problem and avoid any exploration and exploitation problems during the algorithm working.

6 Conclusion and Future Work

This paper introduced a new data clustering problem using the hybrid Harris Hawks optimization (HHO) algorithm with the differential evolution (DE) called H-HHO. The H-HHO algorithm utilizes the advantages of the local search strategy of differential evolution, and a few parameters of the Harris Hawks optimization algorithm. The performance of the proposed H-HHO algorithm is assessed using five UCI benchmarks. A comparison with several different algorithms illustrated the better performance of the proposed algorithm (H-HHO) in terms of the Error rate and the fitness value. In future work, we propose to do the following:

- The proposed approach is employed to solve other useful engineering design problems, such as benchmark, scheduling, path planning, document clustering, and constrained optimization problems.
- The performance of the Harris Hawks optimization algorithm can be improved by hybridizing it with other optimization approaches.

References

1. M.F. Abadi, H. Rezaei, Data clustering using hybridization strategies of continuous ant colony optimization, particle swarm optimization and genetic algorithm. J. Adv Math. Comput. Sci. **9**, 336–350 (2015)
2. H.A. Abdulwahab, A. Noraziah, A.A. Alsewari, S.Q. Salih, An enhanced version of black hole algorithm via levy flight for optimization and data clustering problems. IEEE Access **7**, 142085–142096 (2019)
3. L. Abualigah, Group search optimizer: a nature-inspired meta-heuristic optimization algorithm with its results, variants, and applications. Neural Comput. Appl., pp. 1–24 (2020)
4. L. Abualigah, Multi-verse optimizer algorithm: a comprehensive survey of its results, variants, and applications. Neural Comput. Appl., pp. 1–21 (2020)
5. L. Abualigah, M. Abd Elaziz, A.G. Hussien, B. Alsalibi, S.M. Jafar Jalali, A.H. Gandomi, Lightning search algorithm: a comprehensive survey. Appl. Intell., pp. 1–24 (2020)
6. L. Abualigah, B. Alsalibi, M. Shehab, M. Alshinwan, A.M. Khasawneh, H. Alabool, A parallel hybrid krill herd algorithm for feature selection. Int. J. Mach. Learn. Cybern., pp. 1–24 (2020)
7. L. Abualigah, A. Diabat, A comprehensive survey of the grasshopper optimization algorithm: results, variants, and applications. Neural Comput. Appl., pp. 1–24 (2020)
8. L. Abualigah, A. Diabat, Z.W. Geem, A comprehensive survey of the harmony search algorithm in clustering applications. Appl. Sci. **10**(11), 3827 (2020)
9. L. Abualigah, A. Diabat, S. Mirjalili, M. Abd Elaziz, A.H. Gandomi. The arithmetic optimization algorithm. Comput. Methods Appl. Mech. Eng. (2020)
10. L. Abualigah, M. Shehab, M. Alshinwan, H. Alabool, Salp swarm algorithm: a comprehensive survey. Neural Comput. Appl. (2019)
11. L. Abualigah, M. Shehab, M. Alshinwan, S. Mirjalili, M. Abd Elaziz, Ant lion optimizer: a comprehensive survey of its variants and applications. Arch. Comput. Methods Eng. (2020)
12. L.M. Abualigah, E.S. Hanandeh, A. Tajudin, M.A. Khader, S.K. Shandilya, β-hill climbing technique for improving the text document clustering problem. Curr. Med. Imag. (2018)
13. L.M. Abualigah, A.T. Khader, Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering. J. Supercomput. **73**(11), 4773–4795 (2017)
14. L.M. Abualigah, A.T. Khader, M. Azmi Al-Betar, Multi-objectives-based text clustering technique using k-mean algorithm, in *2016 7th International Conference on Computer Science and Information Technology (CSIT)* (IEEE, 2016), pp. 1–6
15. L.M. Abualigah, A.T. Khader, M.A. Al-Betar, O.A. Alomari, Text feature selection with a robust weight scheme and dynamic dimension reduction to text document clustering. Expert Syst. Appl. **84**, 24–36 (2017)
16. L.M. Abualigah, A.T. Khader, M.A. Al-Betar, Z.A.A. Alyasseri, O.A. Alomari, E.S. Hanandeh. Feature selection with β-hill climbing search for text clustering application, in *2017 Palestinian International Conference on Information and Communication Technology (PICICT)* (IEEE, 2017), pp. 22–27
17. L.M. Abualigah, A.T. Khader, M.A. Al-Betar, Mohammed A Awadallah, A krill herd algorithm for efficient text documents clustering, in *2016 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)* (IEEE, 2016), pp. 67–72
18. L.M. Abualigah, A.T. Khader, M.A. AlBetar, cc Hanandeh. A new hybridization strategy for krill herd algorithm and harmony search algorithm applied to improve the data clustering, in *1st EAI International Conference on Computer Science and Engineering* (European Alliance for Innovation (EAI), 2016), p. 54
19. L.M. Abualigah, A.T. Khader, M.A. AlBetar, E.S. Hanandeh. Unsupervised text feature selection technique based on particle swarm optimization algorithm for improving the text clustering, in *1st EAI International Conference on Computer Science and Engineering* (European Alliance for Innovation (EAI), 2016), p. 169

20. L.M. Abualigah, A.T. Khader, E.S. Hanandeh, A combination of objective functions and hybrid krill herd algorithm for text document clustering analysis. Eng. Appl. Artif. Intell. **73**, 111–125 (2018)
21. L.M. Abualigah, A.T. Khader, E.S. Hanandeh, Hybrid clustering analysis using improved krill herd algorithm. Appli. Intell. **48**(11), 4047–4071 (2018)
22. L.M. Abualigah, A.T. Khader, E.S. Hanandeh, A new feature selection method to improve the document clustering using particle swarm optimization algorithm. J. Comput. Sci. **25**, 456–466 (2018)
23. L.M. Abualigah, A.T. Khader, E.S. Hanandeh. A novel weighting scheme applied to improve the text document clustering techniques, in *Innovative Computing, Optimization and Its Applications* (Springer, 2018), pp. 305–320
24. L.M. Abualigah, A.T. Khader, E.S. Hanandeh. Modified krill herd algorithm for global numerical optimization problems, in *Advances in Nature-Inspired Computing and Applications* (Springer, 2019), pp. 205–221
25. L.M. Abualigah, A.T. Khader, E.S. Hanandeh, A.H. Gandomi, A novel hybridization strategy for krill herd algorithm applied to clustering techniques. Appl. Soft Comput. **60**, 423–435 (2017)
26. L.M. Abualigah, A.M. Sawaie, A.T. Khader, H. Rashaideh, M.A. Al-Betar, M. Shehab, β-hill climbing technique for the text document clustering in *New Trends in Information Technology (NTIT)–2017* (2017), p. 60
27. L.M. Qasim Abualigah, *Feature Selection and Enhanced Krill Herd Algorithm for Text Document Clustering* (Springer, 2019)
28. L.M. Qasim Abualigah, E.S. Hanandeh, Applying genetic algorithms to information retrieval using vector space model. Int. J. Comput. Sci. Eng. Appl. **5**(1), 19 (2015)
29. Z.A. Al-Sai, L.M. Abualigah, Big data and e-government: a review, in *2017 8th International Conference on Information Technology (ICIT)* (IEEE, 2017), pp. 580–587
30. O.A. Alomari, A.T. Khader, M.A.Z.M.I. Al-Betar, L.M. Abualigah, Mrmr ba: a hybrid gene selection algorithm for cancer classification. J. Theor. Appl. Inf. Technol. **95**(12), 2610–2618 (2017)
31. O.A. Alomari, A.T. Khader, M.A. Al-Betar, L.M. Abualigah, Gene selection for cancer classification by combining minimum redundancy maximum relevancy and bat-inspired algorithm. Int. J. Data Min. Bioinf. **19**(1), 32–51 (2017)
32. Bisan Alsalibi, Laith Abualigah, A.T. Khader, A novel bat algorithm with dynamic membrane structure for optimization problems. Appl. Intell., pp. 1–26 (2020)
33. H.N. Alshaer, M.A. Otair, L. Abualigah, M .Alshinwan, A.M. Khasawneh, Feature selection method using improved chi square on arabic text classifiers: analysis and application. Multimedia Tools Appl., pp. 1–18 (2020)
34. M. Alswaitti, M. Albughdadi, N.A.M. Isa, Density-based particle swarm optimization algorithm for data clustering. Expert Syst. Appl. **91**, 170–186 (2018)
35. E. Amiri, S. Mahmoudi, Efficient protocol for data clustering by fuzzy cuckoo optimization algorithm. Appl. Soft Comput. **41**, 15–21 (2016)
36. R. Angira, B.V. Babu, Optimization of non-linear chemical processes using modified differential evolution, in *IICAI* (2005), pp. 911–923
37. R. Angira, B.V. Babu, Optimization of process synthesis and design problems: a modified differential evolution approach. Chem. Eng. Sci. **61**(14), 4707–4721 (2006)
38. B.V. Babu, R. Angira, Optimization of water pumping system using differential evolution strategies, in *Proceedings of The Second International Conference on Computational Intelligence, Robotics, and Autonomous Systems* (Singapore, 2003)
39. J.C. Bednarz, Cooperative hunting harris' hawks (parabuteo unicinctus). Science **239**(4847), 1525–1527 (1988)
40. K. Bijari, H. Zare, H. Veisi, H. Bobarshad, Memory-enriched big bang-big crunch optimization algorithm for data clustering. Neural Comput. Appl. **29**(6), 111–121 (2018)
41. A.L. Bolaji, M.A. Al-Betar, Mohammed A. Awadallah, A.T. Khader, L.M. Abualigah, A comprehensive review: Krill herd algorithm (kh) and its applications. Appl. Soft Comput. **49**, 437–446 (2016)

42. T. Chakaravarthy, K. Kalyani, A brief survey of honey bee mating optimization algorithm to efficient data clustering. Indian J. Sci. Technol. **8**(24) (2015)
43. S. Chander, P. Vijaya, P. Dhyani, Multi kernel and dynamic fractional lion optimization algorithm for data clustering. Alexandria Eng. J. **57**(1), 267–276 (2018)
44. E. Cuevas, E.B. Espejo, A.C. Enríquez, Clustering representative electricity load data using a particle swarm optimization algorithm, in *Metaheuristics Algorithms in Power Systems* (Springer, 2019), pp. 187–210
45. P. Das, D.K. Das, S. Dey, A new class topper optimization algorithm with an application to data clustering. IEEE Trans. Emerg. Top. Comput. (2018)
46. G. Duan, W. Hu, Z. Zhang, A novel data clustering algorithm based on modified adaptive particle swarm optimization. Int. J. Sig. Process Image Process Pattern Recogn. **9**(3), 179–188 (2016)
47. R. Eberhart, J. Kennedy, A new optimizer using particle swarm theory, in *Proceedings of the Sixth International Symposium on Micro Machine and Human Science, 1995. MHS'95* (IEEE, 1995), pp. 39–43
48. M.A. Elaziz, A.A. Ewees, R.A. Ibrahim, S. Lu, Opposition-based moth-flame optimization improved by differential evolution for feature selection. Math. Comput. Simul. **168**, 48–75 (2020)
49. M.A. Elaziz, L. Li, K.P.N. Jayasena, S. Xiong, Multiobjective big data optimization based on a hybrid salp swarm algorithm and differential evolution. Appl. Math. Modell. (2019)
50. M.A. Elaziz, S. Mirjalili, A hyper-heuristic for improving the initial population of whale optimization algorithm. Knowl.-Based Syst. **172**, 42–63 (2019)
51. M.A. Elaziz, N. Nabil, A.A. Ewees, S. Lu, Automatic data clustering based on hybrid atom search optimization and sine-cosine algorithm, in *2019 IEEE Congress on Evolutionary Computation (CEC)* (IEEE, 2019), pp. 2315–2322
52. M.A. Elaziz, S. Xiong, K.P.N. Jayasena, L. Li, Task scheduling in cloud computing based on hybrid moth search algorithm and differential evolution. Knowl.-Based Syst. **169**, 39–52 (2019)
53. M.E.A. Elaziz, A.A. Ewees, D. Oliva, P. Duan, S. Xiong, A hybrid method of sine cosine algorithm and differential evolution for feature selection, in *International Conference on Neural Information Processing* (Springer, 2017), pp. 145–155
54. H. Emami, F. Derakhshan, Integrating fuzzy k-means, particle swarm optimization, and imperialist competitive algorithm for data clustering. Arab. J. Sci. Eng. **40**(12), 3545–3554 (2015)
55. A.A.A. Esmin, R.A. Coelho, S. Matwin, A review on particle swarm optimization algorithm and its variants to clustering high-dimensional data. Artif. Intell. Rev. **44**(1), 23–45 (2015)
56. F.T. Evers, F. Höppner, F. Klawonn, R. Kruse, T. Runkler, *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition* (Wiley & Sons, New York, 1999)
57. A.A. Ewees, M.A. El Aziz, A.E. Hassanien, Chaotic multi-verse optimizer-based feature selection. Neural Comput. Appl., pp. 1–16 (2017)
58. A.A. Ewees, M.A. Elaziz, E.H. Houssein, Improved grasshopper optimization algorithm using opposition-based learning. Expert Syst. Appl. (2018)
59. C. Fan, F. Xiao, Z. Li, J. Wang, Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review. Energy Build. **159**, 296–308 (2018)
60. M.D. Gandamalla, S. Maddala, K.V.N. Sunitha, A novel genetic algorithm and particle swarm optimization for data clustering, in *Information Systems Design and Intelligent Applications* (Springer, 2016), pp. 199–208
61. A.H. Gandomi, A.H. Alavi, Krill herd: a new bio-inspired optimization algorithm. Commun. Nonlinear Sci. Numer. Simul. **17**(12), 4831–4845 (2012)
62. A.O. Gautestad, I. Mysterud, Complex animal distribution and abundance from memory-dependent kinetics. Ecol. Complex. **3**(1), 44–55 (2006)
63. G. George, L. Parthiban, Multi objective hybridized firefly algorithm with group search optimization for data clustering, in *2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)* (IEEE, 2015), pp. 125–130

64. M.-W. Hao, H.-L. Dai, K. Hao, C. Li, Y.-J. Zhang, H.-N. Song, Optimization of density-based k-means algorithm in trajectory data clustering, in *International Wireless Internet Conference* (Springer, 2017), pp. 440–450
65. A. Hatamlou, S. Abdullah, H. Nezamabadi-Pour, Application of gravitational search algorithm on data clustering, in *International Conference on Rough Sets and Knowledge Technology* (Springer, 2011), pp. 337–346
66. A. Hatamlou, S. Abdullah, H. Nezamabadi-Pour, A combined approach for clustering based on k-means and gravitational search algorithms. Swarm Evol. Comput. **6**, 47–52 (2012)
67. A.A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, H. Chen, Harris hawks optimization: Algorithm and applications. Future Gener. Comput. Syst. **97**, 849–872 (2019)
68. L. Hou, J. Gao, R. Chen, An information entropy-based animal migration optimization algorithm for data clustering. Entropy **18**(5), 185 (2016)
69. E.R. Hruschka, R.J.G.B. Campello, A.A. Freitas et al., A survey of evolutionary algorithms for clustering. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **39**(2), 133–155 (2009)
70. N.E. Humphries, N. Queiroz, J.R.M. Dyer, N.G. Pade, M.K. Musyl, K.M. Schaefer, D.W. Fuller, J.M. Brunnschweiler, T.K. Doyle, J.D.R. Houghton et al., Environmental context explains lévy and brownian movement patterns of marine predators. Nature **465**(7301), 1066 (2010)
71. G. Jagatheeshkumar, S. Selva Brunda, Lion optimization algorithm based k-means for textual data clustering. Int. J. Pure Appl. Math. **117**(22), 167–171 (2017)
72. A.K. Jain, Data clustering: 50 years beyond k-means. Pattern Recogn. Lett. **31**(8), 651–666 (2010)
73. R. Jensi, G.W. Jiji, An improved krill herd algorithm with global exploration capability for solving numerical function optimization problems and its application to data clustering. Appl. Soft Comput. **46**, 230–245 (2016)
74. M.R. Kalhori, M.F. Zarandi, Interval type-2 credibilistic clustering for pattern recognition. Pattern Recogn. **48**(11), 3652–3672 (2015)
75. D. Karaboga, B. Akay, A comparative study of artificial bee colony algorithm. Appl. Math. Comput. **214**(1), 108–132 (2009)
76. J. Kennedy, R. Eberhart, Particle swarm optimization (pso), in *Proceedings of IEEE International Conference on Neural Networks* (Australia, Perth, 1995), pp. 1942–1948
77. D. Kharche, A. Thakare, Acpso: Hybridization of ant colony and particle swarm algorithm for optimization in data clustering using multiple objective functions, in *2015 Global Conference on Communication Technologies (GCCT)* (IEEE, 2015), pp. 854–859
78. A.M. Khasawneh, L. Abualigah, M. Al Shinwan, Void aware routing protocols in underwater wireless sensor networks: variants and challenges, in *Journal of Physics: Conference Series*, vol. 1550 (IOP Publishing, 2020), p. 032145
79. A.M. Khasawneh, O. Kaiwartya, L.M. Abualigah, J. Lloret, et. al., Green computing in underwater wireless sensor networks pressure centric energy modeling. IEEE Syst. J. (2020)
80. S. Kotte, R.K. Pullakura, S.K. Injeti, Optimal multilevel thresholding selection for brain mri image segmentation based on adaptive wind driven optimization. Measurement **130**, 340–361 (2018)
81. A. Kumar, Optimization of artificial bee colony algorithm for clustering in data mining. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **5**(7) (2015)
82. Y. Kumar, G. Sahoo, A hybrid data clustering approach based on improved cat swarm optimization and k-harmonic mean algorithm. AI Commun. **28**(4), 751–764 (2015)
83. R.R. Kurada, K.P. Kanadam, A novel evolutionary automatic data clustering algorithm using teaching-learning-based optimization. Int. J. Intell. Syst. Appl. **10**(5), 61 (2018)
84. N. Kushwaha, M. Pant, S. Kant, V.K. Jain, Magnetic optimization algorithm for data clustering. Pattern Recogn. Lett. **115**, 59–65 (2018)
85. K. Lakshmi, N. Karthikeyani Visalakshi, S. Shanthi, S. Parvathavarthini, Clustering categorical data using k-modes based on cuckoo search optimization algorithm. ICTACT J. Soft Comput. 8(1) (2017)

86. R. Li, Y. Wang, Optimization of physical education teaching mode in colleges based on data mining clustering algorithm. Int. J. Database Theor. Appl. **9**(10), 151–162 (2016)
87. D. Liu, M. Jiang, X. Yang, H. Li, Analyzing documents with quantum clustering: A novel pattern recognition algorithm based on quantum mechanics. Pattern Recogn. Lett. **77**, 8–13 (2016)
88. S. Łukasik, P.A. Kowalski, M. Charytanowicz, P. Kulczycki, Data clustering with grasshopper optimization algorithm, in *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)* (IEEE, 2017), pp. 71–74
89. U. Maulik, S. Bandyopadhyay, Genetic algorithm-based clustering technique. Pattern Recogn. **33**(9), 1455–1465 (2000)
90. R. Mehmood, G. Zhang, R. Bie, H. Dawood, H. Ahmad, Clustering by fast search and find of density peaks via heat diffusion. Neurocomputing **208**, 210–217 (2016)
91. S. Mirjalili, Sca: a sine cosine algorithm for solving optimization problems. Knowl.-Based Syst. **96**, 120–133 (2016)
92. S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, H. Chen, *Harris Hawks Optimization: Algorithm and Applications* (2019)
93. L.M. Abualigah, M. Al-Diabat, M. Al Shinwan, K. Dhou, B. Alsalibi, E.S. Hanandeh, M. Shehab, Hybrid harmony search algorithm to solve the feature selection for data mining applications, in *Recent Advances in Hybrid Metaheuristics for Data Clustering* (2020), pp. 19–37
94. S.J. Nanda, G. Panda, A survey on nature inspired metaheuristic algorithms for partitional clustering. Swarm Evol. Comput. **16**, 1–18 (2014)
95. J. Nayak, B. Naik, D.P. Kanungo, H.S. Behera, A hybrid elicit teaching learning based optimization with fuzzy c-means (etlbo-fcm) algorithm for data clustering. Ain Shams Eng. J. **9**(3), 379–393 (2018)
96. T.P. Nguyen, R.J. Kuo, Automatic fuzzy clustering using non-dominated sorting particle swarm optimization algorithm for categorical data. IEEE Access. **7**, 99721–997234 (2019)
97. V. Patel, A. Tiwari, A. Patel, A comprehensive survey on hybridization of artificial bee colony with particle swarm optimization algorithm and abc applications to data clustering, in *Proceedings of the International Conference on Informatics and Analytics* (ACM, 2016), p. 65
98. A.K. Qin, P.N. Suganthan, Self-adaptive differential evolution algorithm for numerical optimization, in *2005 IEEE Congress on Evolutionary Computation*, vol. 2 (IEEE, 2005), pp. 1785–1791
99. S. Rana, S. Jasola, R. Kumar, A review on particle swarm optimization algorithms and their applications to data clustering. Artif. Intell. Rev. **35**(3), 211–222 (2011)
100. H. Rashaideh, A. Sawaie, M.A. Al-Betar, L.M. Abualigah, M.M. Al-Laham, M. Ra'ed, M. Braik, A grey wolf optimizer for text document clustering. J. Intell. Syst
101. A. Rodriguez, A. Laio, Clustering by fast search and find of density peaks. Science **344**(6191), 1492–1496 (2014)
102. M. Safaldin, M. Otair, L. Abualigah, Improved binary gray wolf optimizer and svm for intrusion detection system in wireless sensor networks. J. Ambient Intell. Hum. Comput., pp. 1–18 (2020)
103. S. Saha, A.K. Alok, A. Ekbal, Brain image segmentation using semi-supervised clustering. Expert Syst. Appl. **52**, 50–63 (2016)
104. R.K. Saidala, N. Devarakonda, Multi-swarm whale optimization algorithm for data clustering problems using multiple cooperative strategies. Int. J. Intell. Syst. Appl. **10**(8), 36 (2018)
105. M. Salami, F.M. Sobhani, M.S. Ghazizadeh, Shared subscribe hyper simulation optimization (subhso) algorithm for clustering big data-using big databases of iran electricity market. Appl. Comput. Syst. **24**(1), 49–60 (2019)
106. S.C. Satapathy, A. Naik, Data clustering based on teaching-learning-based optimization, in *International Conference on Swarm, Evolutionary, and Memetic Computing* (Springer, 2011), pp. 148–156
107. A. Saxena, M. Prasad, A. Gupta, N. Bharill, O.P. Patel, A. Tiwari, M.J. Er, W. Ding, C.-T. Lin, A review of clustering techniques and developments. Neurocomputing **267**, 664–681 (2017)

108. N. Sharma, P. Makhija, *A Review on Optimizing Clustering Technique for Data Stream Using Genetic Algorithm* (2018)
109. M. Shehab, L. Abualigah, H. Al Hamad, H. Alabool, M. Alshinwan, A.M. Khasawneh, Moth–flame optimization algorithm: variants and applications. Neural Comput. Appl., pp. 1–26 (2019)
110. M. Shehab, M.S. Daoud, H.M. AlMimi, L.M. Abualigah, A.T. Khader, Hybridising cuckoo search algorithm for extracting the odf maxima in spherical harmonic representation. Int. J. Bio-Inspired Comput. **14**(3), 190–199 (2019)
111. M. Shehab, A.T. Khader, M.A. Al-Betar, L.M. Abualigah, Hybridizing cuckoo search algorithm with hill climbing for numerical optimization problems, in *2017 8th International Conference on Information Technology (ICIT)* (IEEE, 2017), pp. 36–43
112. M.F. Shlesinger, Levy flights: variations on a theme. Phys. D Nonlinear Phenom. **38**(1–3), 304–309 (1989)
113. R. Storn, K. Price, Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces. J. Glob. Optim. **11**(4), 341–359 (1997)
114. D. Tang, S. Dong, L. He, Y. Jiang, Intrusive tumor growth inspired optimization algorithm for data clustering. Neural Comput. Appl. **27**(2), 349–374 (2016)
115. N.D. Thanh, M. Ali et al., A novel clustering algorithm in a neutrosophic recommender system for medical diagnosis. Cogn. Comput. **9**(4), 526–544 (2017)
116. N.T. Thong et al., Hifcf: an effective hybrid model between picture fuzzy clustering and intuitionistic fuzzy recommender systems for medical diagnosis. Expert Syst. Appl. **42**(7), 3682–3701 (2015)
117. E. Tuba, D. Dolicanin-Djekic, R. Jovanovic, D. Simian, M. Tuba, Combined elephant herding optimization algorithm with k-means for data clustering, in *Information and Communication Technology for Intelligent Systems* (Springer, Berlin, 2019), pp. 665–673
118. J. Vanegas, I. Bonet, Clustering algorithm optimization applied to metagenomics using big data, in *Conference on Information Technologies and Communication of Ecuador* (Springer, 2018), pp. 182–192
119. R. Venkat, P. Srinivasulu, Clustering of data using fuzzy c-means (fcm) algorithm with aid of gravitational search optimization, in *2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)* (IEEE, 2017), pp. 232–234
120. G.M. Viswanathan, V. Afanasyev, S.V. Buldyrev, E.J. Murphy, P.A. Prince, H. Eugene Stanley, Lévy flight search patterns of wandering albatrosses. Nature **381**(6581), 413 (1996)
121. G.M. Viswanathan, V. Afanasyev, Sergey V. Buldyrev, M.G.E. Shlomo Havlin, E.P. Da Luz, H.Eugene Raposo, Stanley, Lévy flights in random searches. Phys. A Stat. Mech. Appl. **282**(1–2), 1–12 (2000)
122. D.V. Walker, G.J. Wills, *Statistical Preparation of Data Using Semantic Clustering*, 26 Sept 2019. US Patent App. 16/440,049
123. R. Wang, W. Ji, M. Liu, X. Wang, J. Weng, S. Deng, S. Gao, C.A. Yuan, Review on mining data from multiple data sources. Pattern Recognition Letters. **15**(109), 120–128 (2018)
124. T. Wang, X. Wang, Z. Zhao, Z. He, T. Xia, Measurement data classification optimization based on a novel evolutionary kernel clustering algorithm for multi-target tracking. IEEE Sens. J. **18**(9), 3722–3733 (2018)
125. Z.-X. Wu, K.-W. Huang, J.-L. Chen, C.-S. Yang, A memetic fuzzy whale optimization algorithm for data clustering, in *2019 IEEE Congress on Evolutionary Computation (CEC)* (IEEE, 2019), pp. 1446–1452
126. Y. Xianfeng, L. Pengfei, Tailoring fuzzy c-means clustering algorithm for big data using random sampling and particle swarm optimization. Int. J. Database Theor. Appl. **8**(3), 191–202 (2015)
127. R. Xu, D. Wunsch, Survey of clustering algorithms. IEEE Trans. Neural Netw. **16**(3), 645–678 (2005)
128. X.-S. Yang. *Nature-Inspired Metaheuristic Algorithms* (Luniver press, 2010)

129. Q. Yuan, H. Wang, W. Botao, Y. Song, H. Wang, A fusion load disaggregation method based on clustering algorithm and support vector regression optimization for low sampling data. Future Internet **11**(2), 51 (2019)
130. C. Zhang, F. Tao, X. Chen, J. Shen, M. Jiang, B. Sadler, M. Vanni, J. Han, *Taxogen: Constructing topical concept taxonomy by adaptive term embedding and clustering* (Proc, KDDI, 2018)
131. C. Zhang, L. Hao, L. Fan, Optimization and improvement of data mining algorithm based on efficient incremental kernel fuzzy clustering for large data. Cluster Comput. **22**(2), 3001–3010 (2019)
132. Q.-H. Zhang, B.-L. Li, Y.-J. Liu, L. Gao, L.-J. Liu, X.-L. Shi, Data clustering using multivariant optimization algorithm. Int. J. Mach. Learn. Cybern. **7**(5), 773–782 (2016)
133. C.T. Zheng, C. Liu, H.S. Wong, Corpus-based topic diffusion for short text clustering. Neurocomputing **275**, 2444–2458 (2018)
134. L.-F. Zhu, J.-S. Wang, Data clustering method based on bat algorithm and parameters optimization. Eng. Lett. **27**(1), 241–250 (2019)