Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi
Pamukkale University Journal of Engineering Sciences

An enhanced version of honey badger algorithm for data clustering problems

Veri kümeleme problemleri için bal porsuğu algoritmasının geliştirilmiş bir sürümü

Harun Gezici[11]

[1]Department of Electronics and Automation, Vocational School of Technical Sciences, Kırklareli University, Kırklareli, Türkiye.
harun.gezici@klu.edu.tr

Received/Geliş Tarihi: 07.01.2025 Revision/Düzeltme Tarihi: 07.08.2025 doi: 10.5505/pajes.2025.99402
Accepted/Kabul Tarihi: 20.08.2025 Research Article/Araştırma Makalesi

Abstract

This study proposes an improved version of the Honey Badger Algorithm (HBA) for solving clustering problems, called the Clustering Honey Badger Algorithm (CHBA). The main enhancement involves modeling the smell intensity using an exponential decay function instead of the inverse square law. This modification reduces the likelihood of getting trapped in local optima and improves the algorithm's exploratory behavior. CHBA was compared against six state-of-the-art meta-heuristic algorithms, including the original HBA, on seven benchmark clustering datasets. The evaluation was based on five common external performance metrics: accuracy, F-score, precision, sensitivity, and intra-cluster distance. According to the results, CHBA achieved the highest performance on datasets such as Cancer (94.86% accuracy), Iris (93.94% accuracy), and Ecoli (84.52% accuracy). Furthermore, Friedman test results showed that CHBA consistently ranked first in all performance metrics, with p-values less than 0.005, indicating statistically significant superiority. These findings demonstrate that CHBA is a competitive and reliable clustering algorithm, especially in complex and imbalanced data scenarios.

Keywords: Honey badger algorithm, Clustering problem, Meta-Heuristic algorithm, Swarm intelligence.

Öz

Bu çalışma, kümeleme problemlerinin çözümüne yönelik olarak Bal Porsuğu Algoritmasının (HBA) geliştirilmiş bir versiyonu olan Kümeleme Bal Porsuğu Algoritması (CHBA)'yı önermektedir. Yapılan temel iyileştirme, avın koku yoğunluğunu modellemek için kullanılan ters kare yasası yerine eksponansiyel azalma fonksiyonunun uygulanmasıdır. Bu sayede algoritmanın yerel minimumlara takılma olasılığı azaltılmış ve keşif yeteneği artırılmıştır. CHBA, yedi farklı kümeleme veri kümesi üzerinde, orijinal HBA dahil altı güncel meta-sezgisel algoritma ile karşılaştırılmıştır. Karşılaştırmalar doğruluk, F-skor, keskinlik, duyarlılık ve küme içi mesafe olmak üzere beş yaygın dış performans metriğine göre yapılmıştır. Elde edilen sonuçlara göre, CHBA, özellikle Cancer (%94,86 doğruluk), Iris (%93,94 doğruluk) ve Ecoli (%84,52 doğruluk) veri kümelerinde en yüksek başarıyı göstermiştir. Ayrıca, tüm performans metrikleri için yapılan Friedman testinde CHBA'nın ortalama sıralama değeri en düşük algoritma olduğu ve p-değerlerinin tümünde <0.005 olduğu görülmüştür. Bu bulgular, CHBA'nın karmaşık ve dengesiz veri kümelerinde kullanılabilecek rekabetçi ve güvenilir bir kümeleme algoritması olduğunu göstermektedir.

Anahtar kelimeler: Bal porsuğu algoritması, Kümeleme problemleri, Meta sezgisel algoritmalar, Sürü zekâsı.

1 Introduction

In data mining, clustering seems to be a significant data analysis method. This technique, which is named unsupervised learning, aims to discover the structure of data without predefined class labels [1]. Clustering aims to assemble objects with similar characteristics within the same group while it assembles objects with different characteristics within different ones. This approach plays a significant role upon making decisions by revealing out hidden patterns and structures within the dataset, predicting and diagnosing future values [2].

Recently, techniques of clustering have been widely used in various research areas. The potential of clustering techniques has been proven in areas like web analysis [3], management [4], data science [5], medical diagnosis [6], image segmentation [7], text mining, networks of wireless sensors [8], and financial analysis [9]. Specifically, text clustering, specifically, stands out as an important technique in dividing large sets of text documents into subsets with similar characteristics [10].

Clustering algorithms could be divided into five main categories according to their working mechanisms: partitional, hierarchical, density-based, graph-based, and optimization-based algorithms [1]. Among these algorithms, partitional algorithms are particularly popular because of their linear time complexity [11]. However, these algorithms have disadvantages due to being sensitive to initial cluster centers, having difficulty in partitioning overlapping data, and having performance degradation in high-dimensional datasets [12], [13].

Lately, researchers prefer Meta Heuristic Algorithms (MHA) more and more in order to solve clustering problems. These algorithms have more competitive and effective results compared to conventional methods [14]. MHAs are developed being inspired by various sources such as human and animal behaviors, evolution mechanism of the nature, laws of physics [15]

Various MHAs like Artificial Bee Colony [16], Teacher Learning Based Optimization [17], Artificial Chemical Reaction Optimization [18], Cuckoo Search [19], Cat Swarm Optimization [20], Grey Wolf Optimizer (GWO) [21], Krill Herd Algorithm [22], and Water Flow Optimizer [14] are successfully employed in order to solve clustering problems. These algorithms have various internal mechanisms to have satisfactory solutions and

[1]Corresponding author/Yazışılan Yazar

they search the solution space through both the local and global searching strategies.

The popularity of MHAs comes from their advantages like being less dependent on the dimension of the problem, solution space, limitations, and variants. In addition to that, they also have advantages like being able to adapt themselves depending on the problem area and having effective mechanisms in solving combinatorial and non-linear problems [18]. However, these algorithms have some weak points as well. For example, in some cases, they may be caught by local optimum or have difficulty in finding the global optimum due to their slow convergence speed or homogeneous searching behaviour [23], [24].

In order to overcome such difficulties, researchers apply hybrid approaches and adaptive strategies. For example, hybrid approaches which combine global and local search aim to create a balance between exploration and exploitation [25]. As evolutionary hybrid algorithms integrate local searching strategies, population management, and learning strategies, they enable an effective optimization framework. [26].

In brief, clustering techniques in data mining, especially with the use of MHAs, increasingly play an important role in analysing complex and large-scale datasets. Researches in this area focus on improving the performance of algorithms, developing new hybrid approaches, and enabling more effective solutions in various application domains. Future works are likely to focus on further development of these algorithms, big data analysis, internet of things, and artificial intelligence applications. [27], [28]

In this study, an improved version of the Honey Badger Algorithm (HBA) is proposed in order to solve clustering problems. This improved version is named Clustering Honey Badger Algorithm (CHBA). The improvement process focuses on smell intensity of the prey. HBA consists of a digging phase and a honey phase. Smell intensity is an important parameter of the digging phase. In HBA, the smell intensity is modelled by inverse square law. It means that the smell will fade proportionally with distance. However, in the nature, besides distance, there are also external factors like wind or rain. In order to model all these external factors, this study proposes exponential decay method. CHBA, is compared to six MHAs. For comparison, six clustering datasets are employed. The data allows us to observe that exponential decay method improves the premature convergence problem of HBA.

The structure of the article is as follows: In section 2, information on studies regarding clustering problems are given. HBA and CHBA are introduced respectively in section 3 and 4. In section 5, algorithms' performance evaluation criteria are presented. Section 6 gives experimental results of CHBA and competitor algorithms. Section 7 discusses the results.

2 Literature

Data clustering has a significant part in big data analysis. Conventional clustering algorithms may go through difficulties like being caught by local minimum, slow convergence, or being overdependent on initial center selection when they face complex datasets. Therefore, researchers aim to overcome such problems and to increase the performance of clustering by using MHAs.

Combining Particle swarm optimization (PSO) and Fuzzy c-means (FCM) algorithm is a common approach to increase the clustering performance. Tiwari et al. [29] have succeeded to overcome the local minimum problem of FCM by developing a hybrid algorithm named PSO-FCM. This algorithm has outstanding results in complex image and multimedia data. Similarly, Al-Behadili [30] has a better balance between exploration and exploitation by combining Firefly Algorithm (FA) and Variable Neighborhood Search (VNS). This approach has improved the limited exploitation ability of FA and enabled more compact clusters.

In order to reduce the dependency of the K-Means algorithm on the initial center selection, various approaches have been proposed. Xia and Liu [31] developed a K-Means algorithm that is optimized by genetic algorithm and they had a great accuracy rate of 98.67% on National Basketball Association (NBA) scoring data. This optimization reduced the number of iterations of the algorithm as well. Singh and Kumar [32] aimed to create a balance between local and global mechanisms by presenting a meta-heuristic clustering algorithm based on Cat Algorithm. This algorithm increased diversity by using an improved solution searching equation and an accelerated speed equation.

In order to evaluate the structures of data clusters and to deal with categorical data, Kuo et al. [33]proposed Possibilistic Fuzzy K-Modes (PFKM) algorithm. This algorithm is further improved by integrating it with the Sine Cosine Algorithm (SCA), Genetic Algorithm (GA) and PSO. The results suggest that especially SCA-PFCM outperforms other algorithms.

In order to overcome the problem of being caught by local minimum trap, Kushwaha et al. [34] developed Electromagnetic Field Optimization (EFO) algorithm. The pulling and pushing mechanisms of EFO help the algorithm preserve its diversification and reduce its dependence on initial cluster center selection. This approach, especially in terms of Rand index (RI), Normalized Mutual Information (NMI) and Purity metrics, has a better performance compared to competitor algorithms.

Hashemi et al. [35] used an improved PSO algorithm for the purpose of reducing the calculation time of big data clustering optimization. This algorithm is an out product of hybridizing multi-start pattern reduction mechanism with PSO. This mechanism includes both a reduction operator that reduces the calculation time and a multi-start operator that increases the population diversity and prevents local minimum. Results suggest that this approach significantly reduces the execution time of clustering.

Hybrid approaches are generally employed in order to improve the clustering performance. For example, Mohammadi and Mobarakeh [36] developed a hybrid algorithm named FA-SOM by combining Self-Organized Map (SOM) and FA. This algorithm calculates the initial cluster centers with the help of FA. The cluster centers determined by FA are used to calculate the initial weight of SOM. This method has lower Sum of Squared Error (SSE) and standard deviation.

The K-means Clustering-based Grey Wolf Optimizer (KCGWO), developed by Premkumar et al. [21], hybridized the traditional GWO with the K-means algorithm and added dynamic weight factors to enhance the exploration and exploitation capabilities of conventional GWO. This approach significantly improved clustering performance by solving GWO's premature convergence and local minimum trap problems. While KCGWO uses K-means concepts to refine initial solutions, it enhances diversity by adding a dynamic weight factor to maintain the balance between exploration and exploitation throughout the optimization process. Comprehensive evaluations on ten numerical test functions and eight real-world datasets demonstrated that KCGWO exhibits 34% better performance compared to the original GWO.

Recent years have witnessed significant developments in clustering and routing protocols based on meta-heuristic algorithms for Wireless Sensor Networks (WSNs). The Meta-heuristic Optimized Cluster head selection-based Routing Algorithm for WSNs (MOCRAW) protocol, proposed by Chaurasia et al. [37], has been developed to improve energy efficiency and network lifetime by utilizing the capabilities of the Dragonfly Algorithm MHAs. The MOCRAW protocol employs two sub-processes: the Cluster Head Selection Algorithm for optimal cluster head selection and the Route Search Algorithm for optimal route discovery. The protocol leverages the exploration and exploitation capabilities of the Dragonfly Algorithm to optimize parameters such as node density, residual energy, and intra-cluster distance, while performing optimal path discovery through levy distribution.

Dynamic neighborhood-based approaches too have been used to improve clustering performance. Zeng et al. [38] developed a PSO variant named dynamic-neighborhood-based switching PSO (DNSPSO) which uses a dynamic neighborhood strategy. The purpose of this algorithm is to remove the premature convergence problem by determining the best individual and global positions. In addition, the diversity of PSO is increased by using a new learning strategy and differential evolution method.

Finally, Singh [39] employed the Harris Hawk's Optimization (HHO) algorithm for data clustering problems and improved the search pattern of the algorithm by using chaotic sequence numbers. This approach reduces the dependency on random numbers and shows superior performance compared to six state-of-the-art techniques when tested on twelve comparison datasets.

While existing literature demonstrates that MHAs can provide potential solutions for clustering problems, several critical gaps remain unaddressed. Recent studies indicate that traditional smell intensity modeling in nature-inspired algorithms, particularly in HBA, relies on oversimplified inverse square law assumptions. However, real-world environmental factors such as atmospheric turbulence, humidity gradients, and wind patterns create non-linear intensity decay patterns that have not been adequately modeled in clustering contexts.

The proposed CHBA addresses this fundamental limitation by introducing exponential decay modeling for smell intensity, representing a comprehensive approach to incorporate realistic environmental factors in honey badger-based clustering algorithms. Furthermore, while existing studies focus on HBA's premature convergence and local minimum trap problems, the theoretical deficiency in smell intensity modeling, which is the root cause of these problems, has been overlooked.

CHBA's stochastic exponential decay approach systematically models environmental uncertainties, dynamically optimizing the algorithm's exploration-exploitation balance. This innovation provides superior performance compared to traditional HBA and other meta-heuristic algorithms, particularly in complex and imbalanced datasets.

Future research could focus on applying CHBA to large-scale datasets, strengthening the honey phase, and optimizing transition mechanisms between digging-honey phases. Additionally, adapting the proposed exponential decay approach to other swarm intelligence algorithms could also be an important research area.

3 Honey badger algorithm

In this section, the mathematical model of HBA is introduced. HBA consists of exploration and exploitation stages. Therefore, it could be perceived as a global optimization algorithm. HBA is as follows.

Step 1: Initialization phase

While initiating HBA, the number of honey badgers should be determined (N: the number of honey badgers). The locations of honey badgers are determined according to this number. The locations are calculated by the following equation (Eqn. (1)).

$$x_i = lb_i + r_1 \times (ub_i - lb_i), \quad r_1 \; is \; a \; random \; number \; [0,1] \tag{1}$$

Where, x_i represents solution cluster with N element while i. represents the solution. In other words, i. stands for the location of the honey badger. ub_i and lb_i respectively, represent the upper and lower limitations of search space.

Step 2: Defining intensity (I)

Intensity is related to the distance between the honey badger and prey and concentration strength of the prey. I_i is the smell intensity of the prey. The higher the smell intensity is the quicker the movement is. This case is modelled by inverse square law in HBA. The smell intensity is calculated by the following equation (Eqn. (2-4)).

$$I_i = r_2 \times \frac{S}{4\pi d_i^2}, \quad r_2 \; is \; a \; random \; number \; [0,1] \tag{2}$$

$$S = (x_i - x_{i+1})^2 \tag{3}$$

$$d_i = x_{prey} - x_i \tag{4}$$

Where, S is concentration power, d_i represents the distance between the prey and the honey badger.

Step 3: Update density factor

Intensity factor (α) is used to make the transition between exploration and exploitation. α is generated depending on the iteration. The more the iteration increases, the lower the value of αgets. α is calculated by using the equation below (Eqn. (5)).

$$\alpha = C \times \exp\left(\frac{-t}{t_{max}}\right), \quad t_{max} = maximum \; number \; of \; iteration \tag{5}$$

Where t is iteration, t_{max} is maximum iteration, and C is a constant and greater than 1 ($C = 2$).

Step 4: Escaping from local optimum

In HBA, there is an F operator in order to prevent search agents from being caught by the local minimum. This F operator has the values of -1 and 1 under certain circumstances. F changes the direction of the search according to these values.

Step 5: Updating the agents' positions

This section introduces how the locations of honey badgers are updated. In HBA, locations are updated in two stages. These two stages are explained as follows.

Step 5-1: Digging phase

At this stage, to hunt, honey badgers follow a route similar to the shape of a cardioid. The mathematical model of digging process is as follows (Eqn. (6) and Enq. (7)).

$$x_{new} = x_{prey} + F \times \beta \times I \times x_{prey} + F \times r_3 \times \alpha \times d_i \times |\cos(2\pi r_4) \times [1 - \cos(2\pi r_5)]| \tag{6}$$

$$F = \begin{cases} 1, & r_6 \le 0.5 \\ -1, & r_6 > 0.5 \end{cases}, r_6 \; is \; a \; random \; number \; [0,1] \tag{7}$$

Where x_{new}, is the new location of the honey badger. x_{prey} is the location of the prey. In other words, global is the best location. β represents the ability of honey badgers to find food. ($\beta > 1, default = 6$). d_i is the distance between the prey and the honey badgers. r_3, r_4 ve r_5 are random numbers between 0 and 1. F is a flag that changes the direciton of the search.
The performance of digging process depends on smell intensity (I), the distance between the prey and the honey badger (d_i), and the impact factor that changes by the iteration (α). Moreover, F, which changes the direction of the search, has an impact upon digging performance.
Step 5-2: Honey phase
Honey badgers follow honeyguide birds. Honey phase is developed being inspired by this following process. The mathematical model of honey phase is given below (Eqn. (8)).

$$x_{new} = x_{prey} + F \times r_7 \times \alpha \times d_i, \quad r_7 \; is \; a \; random \; number \; [0,1] \tag{8}$$

Where, x_{new} is the new location of the honeybager. x_{prey} is the location of the prey. d_i, α, and F are calculated using respectively Eqn. (4), Eqn. (5), and Eqn. (7) In honey phase, it could be suggested that HBA conducts the search in a location close to x_{prey}.

4 Clustering honey badger algorithm

In this section, Clustering Honey Badger Algorithm (CHBA) is introduced. Smell intensity is a significant parameter that affects the performance of the algorithm as it is a constituent of digging process. In preliminary tests, HBA is applied to the clustering problems and the results are saved. It is observed that digging phase, in particular, is caught by local minimum traps and are not able to improve the results. Hence, it is considered that the smell intensity causes HBA to be caught in local minimums.
Smell intensity modelling is a critical component that directly affects the performance of CHBA. The inverse square law (Eqn. (2)) used in the original HBA assumes that smell intensity decreases linearly with distance. While this approach is physically valid for phenomena such as sound and light propagation, it is not realistic for scent dispersion in nature. In real-world conditions, the diffusion of odour molecules is significantly influenced by environmental factors such as wind speed, atmospheric turbulence, humidity levels, and temperature gradients. These factors lead to nonlinear and stochastic decay patterns in smell intensity. The proposed exponential decay method (Eqn. (9)) has been developed to model these realistic environmental conditions more accurately.

$$Ie_i = Se^{-r_8 d_i}, r_8 \; is \; a \; random \; number \; [0,1] \tag{9}$$

$$S = (x_i - x_{i+1}) \tag{10}$$

$$d_i = x_{prey} - x_i \tag{11}$$

Where Ie_i is the smell intensity. x_{prey} is the location of the prey, x_i is the location of the honey badger. S is the concentration strength. d_i represents the distance between the prey and the honey badger. The random parameter $r_8 \in [0,1]$ simulates the stochastic effects of environmental factors. Different values of this parameter influence the algorithm's exploration–exploitation balance as follows:
-$r_8 \rightarrow 0$: Smell intensity decreases slowly, resulting in a wider exploration area.
-$r_8 \rightarrow 1$: Smell intensity decreases rapidly, leading to a narrower exploitation area.
This exponential model enables the algorithm to avoid local minimum traps, particularly during the digging phase, and effectively addresses the problem of premature convergence. .
The smell intensity graphic that the two methods generate depending on the location is given in Figure 1.

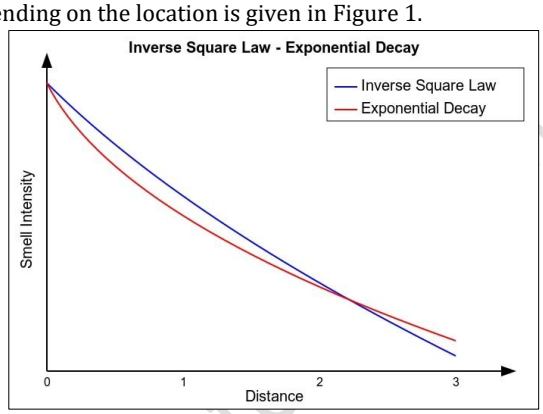

Figure 1. Comparison of the inverse square law and the exponential decay method

The density factor (α) that enables the transition between exploration and exploitation is calculated by the equation below (Eqn. (12)).

$$\alpha = C \times \exp\left(\frac{-t}{t_{max}}\right), \quad t_{max} = maximum \; number \; of \; iteration \tag{12}$$

Where t is iteration, t_{max} is maximum iteration, C is a constant and greater than 1. ($C = 2$).
The equation of digging phase is given below (Eqn. (13) and Eqn. (14)).

$$x_{new} = x_{prey} + F \times \beta \times Ie \times x_{prey} + F \times r_9 \times \alpha \times d_i \times |\cos(2\pi r_{10}) \times [1 - \cos(2\pi r_{11})]| \tag{13}$$

$$F = \begin{cases} 1, & r_{12} \le 0.5 \\ -1, & r_{12} > 0.5 \end{cases}, r_{12} \; is \; a \; random \; number \; [0,1] \tag{14}$$

Where x_{new} is the new location of the honey badger. x_{prey} is the location of the prey. In other words, it is the global best location. β represents the ability of honey badgers to find food. ($\beta > 1, default = 6$). r_9, r_{10} and r_{11} are random numbers that ranges between 0 and 1. F is a flag changing the direction of the search.
The equation of honey phase is below (Eqn. (15)).

$$x_{new} = x_{prey} + F \times r_{13} \times \alpha \times d_i, \quad r_{13} \; is \; a \; random \; number \; [0,1] \tag{15}$$

Where, x_{new} is the new location of the honey badger.
The selection of key parameters in CHBA is based on both theoretical foundations and extensive preliminary experiments:
β parameter (β=6): This parameter represents the foraging ability of honey badgers and is adopted from the original HBA study [40]. Preliminary tests evaluated β values in the range [4, 8], and it was observed that β=6 provides the most balanced exploration–exploitation performance for clustering problems.
C parameter (C=2): This constant is used in the calculation of the density factor. Ensuring that C>1 guarantees that the algorithm performs exploration in the early iterations and switches to exploitation in later stages. The value C=2 allows

the α factor to decrease appropriately throughout the iterations.

r_8 parameter ($r_8 \in [0,1]$): This random parameter in the exponential decay method is regenerated at each iteration. This stochastic approach increases diversity in the algorithm's search behavior and prevents local minimum traps caused by deterministic dynamics.

Maximum number of iterations (t_{max}=500): This value is chosen to provide sufficient time for exploration and convergence. In clustering problems, 500 iterations offer an adequate time window for the algorithm to reach optimal solutions.

The pseudocode of CHBA is presented in Algorithm 1.

Algorithm 1. Pseudo code of CHBA

Set parameters t_{max}, N, β, C.
Initialize population with random positions.
Evaluate the fitness of each honey badger position $x_i$using objective function and assign to $f_i, i \in [1,2, \dots, N]$.
Save best position x_{prey} and assign fitness to f_{prey}
while $t \leq t_{max}$ **do**
  Update the decreasing factor α using Eqn. (12)
  for $i = 1$ *to* N **do**
    Calculate the intensity I_i using Eqn. (9)
    if $r \leq 0.5$ **then**
      Update the position x_{new} using Eqn. (13)
    else
      Update the position x_{new} using Eqn. (15)
    end if
    Evaluate new position and assign to f_{new}
    if $f_{new} \leq f_i$ **then**
      $x_i = x_{new}, f_i = f_{new}$
    end if
    if $f_{prey} \leq f_{prey}$ **then**
      $x_{prey} = x_{new}, f_{prey} = f_{new}$
    end if
  end for
end while
Return x_{prey}

5 Performance evaluation criteria

5.1 Accuracy evaluation

Accuracy is related to the comparison of the label that the algorithms assign to an object to the real label of that object. Accuracy is defined as the ratio of the number of successful assignments to the total number of assignments in the dataset. It is one of the most popular external measurements. For an algorithm to be acknowledged successful, the accuracy parameter is expected to be high. Accuracy is calculated by the Eqn. (16).

$$Accuracy = \frac{num.\,of\ correct\ data\ objects\ identified}{total\ number\ of\ data\ objects} \quad (16)$$

5.2 F-score evaluation

F-score is one of the commonly used external measurements that is used to compare the success of the algorithms. Having a high F-score indicates that there is a good clustering. F-score is calculated by the harmonic mean of precision and recall. F-score is calculated by Eqn. (17).

$$F - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (17)$$

5.3 Precision evaluation

F-score Precision is an important external metric used in evaluating clustering performance. Precision measures the ratio of data objects within a cluster that actually belong to that cluster. In other words, it indicates how many of the examples assigned to a cluster by the algorithm are correctly classified. Precision is calculated for a specific cluster using the following formula (Eqn. (18)):

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (18)$$

Where True Positive represents the number of examples correctly assigned to that cluster, and False Positive represents the number of examples incorrectly assigned to that cluster. A high precision value indicates that the algorithm is reliable in cluster assignment.

5.4 Sensitivity evaluation

F-score Sensitivity, also known as recall, is another crucial external metric for clustering performance evaluation. Sensitivity measures the algorithm's ability to correctly identify and assign data objects that truly belong to a specific cluster. It represents the proportion of actual cluster members that are successfully detected and assigned to the correct cluster by the algorithm. Sensitivity is calculated for a specific cluster using the following formula (Eqn. (19)):

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (19)$$

Where True Positive represents the number of examples correctly assigned to the cluster, and False Negative represents the number of examples that actually belong to the cluster but were incorrectly assigned to other clusters. A high sensitivity value indicates that the algorithm has good detection capability and successfully captures most of the examples that belong to each cluster.

5.5 Friedman test

To validate the statistical significance of performance differences between algorithms, the Friedman test is employed. The Friedman test is a non-parametric statistical test used for comparing multiple algorithms across multiple datasets. It is particularly suitable for clustering performance evaluation as it does not assume normal distribution of the data and can handle ties in rankings.

The test statistic follows a chi-square distribution with (k-1) degrees of freedom, where k is the number of algorithms. A p-value less than 0.05 indicates statistically significant differences between algorithms, allowing us to reject the null hypothesis and conclude that the observed performance differences are not due to random chance.

The Friedman test results are reported alongside the performance comparisons in Table 7, Table 8, Table 9, Table 10 and

Table 11 providing statistical validation for the superiority of the proposed CHBA algorithm.

6 Results and discussion

In this section, the results of the proposed CHBA are compared to the results of six popular MHAs. The algorithms picked for comparison are Grey Wolf Optimizer (GWO) [41], Artificial Rabbits Optimization (ARO) [42], Arithmetic Optimization Algorithm (AOA) [43], Marine Predators Algorithm (MPA) [44], Whale Optimization Algorithm (WOA) [45], Honey Badger algorithm (HBA) [40]. The reason why these algorithms are picked is that they are commonly used in the literature and their validity is proved. MHAs are highly sensitive to initial parameters. Therefore, adjusting these parameters is a delicate process. The parameter adjustments of the algorithms in their original articles are made in a detailed way. The values of initial parameters of the algorithms used in this study are taken from their original articles. The values of initial parameters of the algorithms are given in Table 1.

Table 1. Parameter values of CHBA and competing algorithms

Algorithms	Parameters	T_{max}/N
GWO	$a = 2$	
ARO	$-$	
AOA	$\alpha = 5, \mu = 0.5$	
MPA	$U = 0 \text{ or } 1, p = 0.5, FADs = 0.2$ $R = uniform\ random\ vector\ [0,1],$	500/50
WOA	$l = -1 \text{ or } 1, r = random\ vector\ [0,1],$ $a = linear\ reduction\ [2,0]$	
HBA	$\beta = 6, C = 2, r_1, \dots, r_7 = [0,1]\ random$	
CHBA	$\beta = 6, C = 2, r_8, \dots, r_{13} = [0,1]\ random$	

The proposed CHBA and competitor algorithms are applied to six clustering dataset. Clustering datasets are received from UCI data base. Metrics about the datasets are given in Table 2. The experiments are carried out on a computer that has 64 GB RAM and WINDOWS operating system with CORE I9 processor. Algorithms are coded in the language of Python. The iteration number of all the algorithms are set to be 500 and 30 independent running results are saved.

Table 2. Characteristics of seven benchmark clustering datasets from UCI repository

Datasets	Clusters	Instances	Features
Cancer	2	683	9
Iris	3	150	4
CMC	3	1473	9
Wine	3	178	13
Vowel	6	871	3
Glass	6	214	9
Ecoli	8	336	7

In Table 3, the results of all algorithms based on accuracy performance metric are given. The proposed algorithm is the most successful one in the datasets of Cancer, Iris, CMC, Wine, and Vowel. In Glass dataset, ARO is the most competitive one. The performance of HBA is weaker than AOA and MPA in Iris dataset while it is weaker than ARO in CMC, Wine, and Vowel datasets. The results of CHBA indicates that the method suggested for improving HBA is successful. In datasets, ARO has consistent results. This case could be explained through the fact that it does not have an initial parameter. Not having initial parameter could ease the process of adaptation to problems.

Table 3. Accuracy performance comparison of CHBA against six meta-heuristic algorithms across seven datasets

Dataset	Algorithms						
	GWO	ARO	AOA	MPA	WOA	HBA	CHBA
Cancer	92.56	91.4	90.96	92.52	92.47	94.32	94.86
Iris	89.78	90.42	93.16	92.47	89.91	92.08	93.94
CMC	40.23	46.13	38.46	42.75	41.22	44.56	46.33
Wine	71.83	73.30	68.44	70.74	69.46	72.89	73.36
Vowel	88.72	88.97	79.81	78.23	77.45	88.63	89.21
Glass	62.43	67.88	59.84	63.60	61.78	66.58	59.16
Ecoli	76.19	83.04	71.43	66.07	82.74	82.44	84.52

In Table 4, F-scored based results of all the algorithms are given. The proposed algorithm is the most successful one in Cancer, Iris, CMC, Wine, and Vowel datasets. In Glass dataset, Aro is the optimizer with the best results. Even though CHBA is not the most successful algorithm in Glass dataset, it has better F-score than HBA in all the datasets. Besides, CHBA, along with MPA, is the second-best optimizer in Glass dataset with 0.592 F-score. In clustering problems, it should be kept in mind that F-score is a better measurement than accuracy [1].

Table 4. F-score performance comparison of CHBA against six meta-heuristic algorithms across seven datasets

Dataset	Algorithms						
	GWO	ARO	AOA	MPA	WOA	HBA	CHBA
Cancer	0.948	0.926	0.873	0.842	0.914	0.945	0.952
Iris	0.779	0.785	0.780	0.782	0.779	0.790	0.792
CMC	0.491	0.490	0.456	0.464	0.456	0.486	0.495
Wine	0.523	0.525	0.517	0.518	0.523	0.527	0.529
Vowel	0.652	0.649	0.650	0.649	0.650	0.649	0.654
Glass	0.580	0.604	0.586	0.592	0.589	0.590	0.592
Ecoli	0.723	0.741	0.566	0.590	0.598	0.676	0.778

Table 5 presents the results of all algorithms based on the precision performance metric. The proposed algorithm achieved the most successful results in Cancer, Iris, CMC, Wine, Vowel, and Ecoli datasets. In the Glass dataset, ARO was the algorithm with the best performance. CHBA's precision values were higher than HBA across all datasets. This indicates that CHBA's reliability in cluster assignment is better compared to HBA. Since the precision metric measures how many of the examples assigned to a cluster actually belong to that cluster, it can be concluded that CHBA has a lower rate of incorrect cluster assignments. The ARO algorithm generally exhibited the second-best performance except for the Glass dataset, which can be explained by its parameter-free structure's ability to adapt to problems.

Table 5. Precision performance comparison of CHBA against six meta-heuristic algorithms across seven datasets

Dataset	Algorithms						
	GWO	ARO	AOA	MPA	WOA	HBA	CHBA
Cancer	0.813	0.899	0.834	0.884	0.852	0.878	0.920
Iris	0.949	0.966	0.927	0.944	0.962	0.942	0.972
CMC	0.396	0.414	0.342	0.381	0.363	0.403	0.428
Wine	0.703	0.708	0.632	0.652	0.650	0.703	0.739
Vowel	0.744	0.778	0.699	0.733	0.684	0.773	0.788
Glass	0.532	0.619	0.528	0.555	0.551	0.602	0.559
Ecoli	0.704	0.682	0.614	0.598	0.651	0.624	0.742

Table 6 presents the results of all algorithms based on the sensitivity performance metric. The proposed CHBA algorithm achieved the highest sensitivity values in Cancer, Iris, CMC, Wine, Vowel, and Ecoli datasets. In the Glass dataset, the ARO algorithm provided the best results. Since the sensitivity metric indicates how successfully the algorithm can detect examples belonging to a specific cluster, CHBA's high sensitivity values reveal that the algorithm has strong detection capability. CHBA obtained higher sensitivity values than HBA across all datasets, demonstrating that the exponential decay method improves the algorithm's exploration capability. Particularly in imbalanced datasets (CMC, Glass, Ecoli), CHBA's sensitivity performance shows that the algorithm can successfully detect minority classes as well.

Table 6. Sensitivity performance comparison of CHBA against six meta-heuristic algorithms across seven datasets

Dataset	Algorithms						
	GWO	ARO	AOA	MPA	WOA	HBA	CHBA
Cancer	0.855	0.892	0.908	0.866	0.843	0.937	0.944
Iris	0.964	0.944	0.928	0.917	0.929	0.929	0.990
CMC	0.417	0.427	0.396	0.397	0.385	0.429	0.479
Wine	0.722	0.745	0.705	0.711	0.720	0.709	0.774
Vowel	0.815	0.842	0.784	0.745	0.737	0.836	0.865
Glass	0.607	0.611	0.554	0.617	0.547	0.625	0.606
Ecoli	0.811	0.743	0.601	0.553	0.776	0.743	0.816

Intra-cluster distance represents the total distance between each data point within a cluster and the central point of that cluster.

Table 12 presents intra-cluster distances that the proposed algorithm and the competitor algorithms have out of clustering datasets. Intra-cluster distances' being small as the data are close to the center of the cluster is a desired case. In the best value metric, CHBA is the most successful one in Cancer, Iris, Wine, Vowel and Ecoli datasets. In Glass dataset, ARO is the algorithm with the most successful result. In the mean value metric, there seems to be a similar case. In the intra-cluster distance metric, CHBA outperforms HBA in all datasets. This could be explained through the fact that CHBA does not get caught in local minimum traps. Moreover, ARO outperforms HBA. This could be explained through the fact that ARO has a structure that is able to adapt to problems, as explained earlier. While comparing algorithms, it is not enough to compare only the results of the problems. The results should be statistically meaningful. Therefore, the accuracy, F-score, and intra-cluster distance metrics of the proposed algorithm and the competitor algorithms are evaluated according to the average rank values of the Friedman test. Table 7 indicates the rank values which are calculated by considering the accuracy criterion of all the algorithms. The last row of the table shows the average success rank of the algorithms in all datasets. CHBA is the algorithm with the best clustering performance in the all the datasets. ARO and HBA have similar performances. AOA and WOA are the least successful optimizers.

Table 7. Friedman test statistical ranking results for accuracy metric across seven datasets with p-values

Dataset	Algorithms							
	GWO	ARO	AOA	MPA	WOA	HBA	CHBA	p-value
Cancer	3	6	7	4	5	2	1	1.484E-5
Iris	7	5	2	3	6	4	1	3.393E-6
CMC	6	2	7	4	5	3	1	2.387E-5
Wine	4	2	7	5	6	3	1	1.211E-3
Vowel	3	2	5	6	7	4	1	1.259E-3
Glass	4	1	6	3	5	2	7	4.726E-5
Ecoli	5	2	6	7	3	4	1	3.419E-3
Avg.	4.57	2.86	5.71	4.57	5.29	3.14	1.86	

Table 8 shows the average rank values which are calculated by considering the F-score of all the algorithms. In the last row of the table, there is the average success ranks of all the algorithms for all the datasets. ARO is the second with 3.33 score and HBA is the third best optimizer with 3.50 score. AOA is the least successful algorithm according to F-score.

Table 8. Friedman test statistical ranking results for F-score metric across seven datasets with p-values

Dataset	Algorithms							
	GWO	ARO	AOA	MPA	WOA	HBA	CHBA	p-value
Cancer	2	4	6	7	5	3	1	1.384E-4
Iris	6.5	3	5	4	6.5	2	1	2.184E-3
CMC	2	3	6.5	5	6.5	4	1	4.411E-4
Wine	4.5	3	7	6	4.5	2	1	2.152E-4
Vowel	2	6	3.5	6	3.5	6	1	2.419E-3
Glass	7	1	6	2.5	5	4	2.5	3.628E-5
Ecoli	3	2	7	6	5	4	1	2.028E-3
Avg.	3.86	3.14	5.86	5.21	5.14	3.57	1.21	

Table 9 shows the average rank values calculated according to the precision metric. Based on the average success ranking across all datasets, CHBA achieved the best clustering performance with an average rank of 1.29. The ARO algorithm ranked second with an average rank of 2.00, while HBA was the third most successful algorithm with an average rank of 3.71. The AOA algorithm was the least successful algorithm in terms of the precision metric with an average rank of 6.57. CHBA's achievement of statistically significant p-values ($p < 0.05$) across all datasets confirms that its superiority in precision performance is not due to random chance. Particularly, the very low p-values in Cancer (p=2.187E-5) and Iris (p=1.415E-4) datasets reveal that CHBA has strong statistical significance in terms of precision.

Table 9. Friedman test statistical ranking results for precision metric across seven datasets with p-values

Dataset	Algorithms							
	GWO	ARO	AOA	MPA	WOA	HBA	CHBA	p-value
Cancer	7	2	6	3	5	4	1	2.187E-5
Iris	4	2	7	5	3	6	1	1.415E-4
CMC	4	2	7	5	6	3	1	3.148E-3
Wine	3	2	7	5	6	3	1	3.746E-4
Vowel	4	2	6	5	7	3	1	8.423E-5
Glass	6	1	7	4	5	2	3	2.581E-3
Ecoli	2	3	6	7	4	5	1	1.074E-3
Avg.	4.29	2.00	6.57	4.86	5.14	3.71	1.29	

Table 10 presents the average rank values calculated according to the sensitivity metric. In the average success ranking across all datasets, CHBA demonstrates the superior sensitivity performance with an average rank of 1.57. The ARO algorithm ranks second with an average rank of 3.00, while HBA ranks third with an average rank of 3.14. WOA and AOA algorithms exhibited the lowest performance in terms of sensitivity with an average rank of 5.57. According to the Friedman test results, the p-values obtained across all datasets are less than 0.05, indicating that the sensitivity differences between algorithms are statistically significant. The p-values in CMC dataset (3.296E-5) and Ecoli dataset (2.619E-4) are particularly low, demonstrating that CHBA's superiority in sensitivity performance is based on strong statistical foundations.

Table 10. Friedman test statistical ranking results for sensitivity metric across seven datasets with p-values

Dataset	Algorithms							
	GWO	ARO	AOA	MPA	WOA	HBA	CHBA	p-value
Cancer	6	4	3	5	7	2	1	3.846E-4
Iris	2	3	6	7	4	4	1	4.271E-4
CMC	4	3	6	5	7	2	1	3.296E-5
Wine	3	2	7	5	4	6	1	5.753E-4
Vowel	4	2	5	6	7	3	1	3.461E-3
Glass	4	3	6	2	7	1	5	4.914E-3
Ecoli	2	4	6	7	3	4	1	2.619E-4
Avg.	3.57	3.00	5.57	5.29	5.57	3.14	1.57	

Table 11 indicates the average rank values which are calculated by considering the intra-cluster distances of CHBA and competitor algorithms. CHBA is the best optimizer according to both the best value metric and the average value metric. According to the best value metric, ARO is the second and GWO is the third best algorithm. For the average value metric, ARO is the second and the HBA is the third best optimizer. The Friedman test results show a statistically significant p-value of 2.713E-4 ($p < 0.005$), confirming that the observed differences in intra-cluster distance performance between algorithms are not due to random chance and that CHBA's superiority in minimizing intra-cluster distances is statistically validated.

Table 11. Friedman test statistical ranking results for intra-cluster distance metric across seven datasets with p-values

Algorithms	Ranking based on average	Ranking based on best
GWO	4.43	4.14
ARO	2	2.29
AOA	5.43	5
MPA	5	4.57
WOA	5.57	5.57
HBA	4.43	5
CHBA	1.14	1.29
p-value: 2.713E-4		

While these statistical results statistically confirm the superiority of CHBA over competing algorithms, analysis of the underlying reasons for this performance difference is also important.

The superior performance of CHBA can be attributed to several key technical improvements:

- Enhanced Smell Intensity Modelling

The exponential decay method (Equation 9) provides a more realistic representation of environmental factors compared to HBA's inverse square law (Equation 2). While inverse square law assumes linear intensity reduction with distance, exponential decay captures the stochastic nature of real-world scent propagation affected by wind turbulence, atmospheric conditions, and humidity variations. This results in more diverse search patterns and prevents premature convergence.

- Improved Exploration-Exploitation Balance

The stochastic parameter $r_8 \in [0,1]$ in the exponential model dynamically adjusts the search radius. When $r_8 \to 0$, the algorithm maintains wider exploration areas, while $r_8 \to 1$ focuses on intensive exploitation. This adaptive mechanism allows CHBA to automatically balance between global and local search based on the problem landscape.

- Prevention of Local Minimum Traps

Traditional HBA's deterministic intensity calculation often leads to repetitive search patterns around the same regions. CHBA's stochastic exponential approach generates varying intensity values even for identical distances, creating escape mechanisms from local optima and enabling discovery of global solutions.

- Enhanced Convergence Characteristic

The exponential model's mathematical properties ensure smoother convergence compared to the abrupt changes in HBA's inverse square method, particularly in complex clustering landscapes with irregular cluster boundaries and overlapping data distributions.

Table 12. Intra-cluster distance performance comparison showing best and mean values for CHBA and competing algorithms

Dataset		**GWO**	**ARO**	**AOA**	**MPA**	**WOA**	**HBA**	**CHBA**
Cancer	Best	3108.462	2964.876	3026.186	2989.758	3076.472	2974.876	**2964.647**
	Mean	3245.362	2970.128	3032.758	2995.461	3107.563	2994.643	**2967.874**
Iris	Best	97.0283	96.698	96.847	97.142	97.462	97.175	**96.642**
	Mean	97.285	96.9238	97.044	97.492	97.599	97.253	**96.763**
CMC	Best	5617.846	5538.462	5746.374	**5532.855**	5673.492	5549.184	5534.749
	Mean	5849.463	5672.184	5973.171	5698.163	5712.841	5687.942	**5662.219**
Wine	Best	16330.182	16332.516	16583.467	16469.263	16486.591	16364.758	**16304.402**
	Mean	16366.843	16361.728	16676.481	16491.946	16520.137	16402.403	**16334.184**
Vowel	Best	149846.3	149782.2	150461.5	150467.3	151742.9	150763.5	**149489.6**
	Mean	150023.6	150018.5	150703.2	150602.5	152011.2	151236.1	**149869.7**

Glass	Best	219.481	**217.415**	219.472	224.486	220.648	221.794	218.284
	Mean	221.472	**218.179**	222.99	227.634	224.948	223.576	219.973
Ecoli	Best	27.493	25.212	28.882	28.163	25.146	28.882	**24.181**
	Mean	28.351	26.371	30.214	29.826	26.212	29.915	**25.816**

7 Conclusion

This study introduces a novel meta-heuristic approach called the Clustering Honey Badger Algorithm (CHBA), designed to solve clustering problems. CHBA replaces the inverse square law used in the classical Honey Badger Algorithm (HBA) with an exponential decay model that more realistically accounts for environmental factors in scent propagation. This modification helps the algorithm overcome issues of premature convergence and entrapment in local optima.

CHBA was evaluated on seven clustering datasets and compared with six well-known meta-heuristic algorithms, including GWO, ARO, AOA, MPA, WOA, and the original HBA. Performance was assessed using five external metrics: accuracy, F-score, precision, sensitivity, and intra-cluster distance. The results show that CHBA achieved the highest performance in most datasets across all metrics. Furthermore, statistical analyses using the Friedman test yielded p-values below 0.005 for all metrics, confirming that CHBA's superiority is statistically significant rather than due to chance.

Despite its strengths, CHBA has certain limitations. The algorithm has only been tested on single-objective clustering problems and static datasets. Additionally, no adaptive mechanism has been integrated to automatically tune its parameters during execution.

Future studies may focus on adapting CHBA to multi-objective clustering scenarios, extending its applicability to dynamic or streaming datasets, and developing self-adaptive versions with automatic parameter control. Further improvements may also involve enhancing components such as the honey phase or hybridizing CHBA with local search strategies to improve convergence speed and overall performance.

In conclusion, CHBA stands out as a competitive alternative in the field of meta-heuristic clustering, particularly due to its high performance and statistically validated superiority on complex and imbalanced datasets. Author contribution statement

In this study, Author 1 focused on forming the idea, conducting experimental studies, evaluating the results, contributing to the literature review, spelling, and checking the article's content.

8 Ethics committee approval and conflict of interest statement

There is no need for an ethics committee approval in the prepared article. There is no conflict of interest with any person/institution in the prepared article.

9 References

[1] Kaur A, Kumar Y. "A new metaheuristic algorithm based on water wave optimization for data clustering". *Evolutionary Intelligence*, 15(1), 759-785, 2022.

[2] Tarkhaneh O, Moser I. "An improved differential evolution algorithm using Archimedean spiral and neighborhood search-based mutation approach for cluster analysis". *Future Generation Computer Systems*, 101, 921-939, 2019.

[3] Gong S, Hu W, Li H, Qu Y. "Property Clustering in Linked Data: An Empirical Study and Its Application to Entity Browsing". *International Journal on Semantic Web and Information Systems*, 14(1), 31-70, 2018

[4] Chou CH, Hsieh SC, Qiu CJ. "Hybrid genetic algorithm and fuzzy clustering for bankruptcy prediction". *Applied Soft Computing*, 56, 298-316, 2017.

[5] Hyde R, Angelov P, MacKenzie AR. "Fully online clustering of evolving data streams into arbitrarily shaped clusters". Information Sciences, 382-383, 96-114, 2017.

[6] Wang L, Zhou X, Xing Y, Yang M, Zhang C. "Clustering ECG heartbeat using improved semi-supervised affinity propagation". *IET Software*, 11(5), 207-213, 2017.

[7] Yao H, Duan Q, Li D, Wang J. "An improved K-means clustering algorithm for fish image segmentation". *Mathematical and Computer Modelling*, 58(3-4), 790-798, 2013.

[8] Zhu J, Lung CH, Srivastava V. "A hybrid clustering technique using quantitative and qualitative data for wireless sensor networks". *Ad Hoc Networks*, 25, 38-53, 2015.

[9] Marinakis Y, Marinaki M, Doumpos M, Zopounidis C. "Ant colony and particle swarm optimization for financial classification problems". *Expert Systems with Applications*, 36(7), (10604-10611), 2009.

[10] Mishra RK, Saini K, Bagri S. "Text document clustering on the basis of inter passage approach by using K-means". *International Conference on Computing, Communication and Automation (ICCCA)*, Greater Noida, India,15-16 May 2015

[11] Jain AK. "Data clustering: 50 years beyond K-means". *Pattern Recognition Letters*, 31(8), 651-666, 2010.

[12] Celebi ME, Kingravi HA, Vela PA. "A comparative study of efficient initialization methods for the k-means clustering algorithm". *Expert Systems with Applications*, 40(1), (200-210), 2013.

[13] Jothi R, Mohanty SK, Ojha A. "DK-means: a deterministic K-means clustering algorithm for gene expression analysis". *Pattern Analysis and Applications*, 22(2), (649-667), 2019.

[14] Sahoo RC, Kumar T, Tanwar P, Pruthi J, Singh S. "An efficient meta-heuristic algorithm based on water flow optimizer for data clustering". *Journal of Supercomputing*, 80(8), (10301-10326), 2024.

[15] Jia H, Rao H, Wen C, Mirjalili S. "Crayfish optimization algorithm" *Artificial Intelligence Review*, 56, 1919-1979, 2023.

[16] Kumar Y, Sahoo G. "A two-step artificial bee colony algorithm for clustering". *Neural Computing and Applications*, 28(3), 537-551, 2017.

[17] Kumar Y, Singh PK. "A chaotic teaching learning based optimization algorithm for clustering problems". *Applied Intelligence*, 49(3), 1036-1062, 2019.

[18] Singh H, Kumar Y, Kumar S. "A new meta-heuristic algorithm based on chemical reactions for partitional clustering problems". *Evolutionary Intelligence*, 12(2), 241-252, 2019.

[19] Bouyer A, Hatamlou A. "An efficient hybrid clustering method based on improved cuckoo optimization and modified particle swarm optimization algorithms". *Applied Soft Computing*, 67, 172-182, 2018.

[20] Kumar Y, Singh PK. "Improved cat swarm optimization algorithm for solving global optimization problems and its

application to clustering" *Applied Intelligence*, 48(9), (2681-2697), 2018.
[21] Premkumar M, Sinha G, Ramasamy MD, Sahu S, Subramanyam CB, Sowmya R, Abualigah L, Derebew B. "Augmented weighted K-means grey wolf optimizer: An enhanced metaheuristic algorithm for data clustering problems". *Scientific Reports*, 14(1), 1-33, 2024.
[22] Abualigah LM, Khader AT, Hanandeh ES, Gandomi AH. "A novel hybridization strategy for krill herd algorithm applied to clustering techniques". *Applied Soft Computing*, 60, 423-435, 2017.
[23] Boushaki SI, Kamel N, Bendjeghaba O. "A new quantum chaotic cuckoo search algorithm for data clustering". *Expert Systems with Applications*, 96, 358-372, 2018.
[24] Sag T, Ihsan A. "Particle Swarm Optimization with a new intensification strategy based on K-Means". *Pamukkale University Journal of Engineering Sciences*, 29(3), 264-273, 2023.
[25] Li Y, Li Y, Li G, Zhao D, Chen C. "Two-stage multi-objective OPF for AC/DC grids with VSC-HVDC: Incorporating decisions analysis into optimization process". *Energy*, 147, 286-296, 2018.
[26] Mustafa HMJ, Ayob M, Nazri MZA, Kendall G. "An improved adaptive memetic differential evolution optimization algorithms for data clustering problems". *PLoS One*, 14(5), e0216906, 2019.
[27] Toru E, Yılmaz G. "A multi-depot vehicle routing problem with time windows for daily planned maintenance and repair service planning". *Pamukkale University Journal of Engineering Sciences*, 29(8), 913-919, 2023.
[28] Özkış A, Karakoyun M. "A binary enhanced moth flame optimization algorithm for uncapacitated facility location problems". *Pamukkale University Journal of Engineering Sciences*, 29(7), 737-751, 2023.
[29] Verma H, Verma D, Tiwari PK, "A population based hybrid FCM-PSO algorithm for clustering analysis and segmentation of brain image". *Expert Systems with Applications*, 167, 114121, 2021.
[30] Al-Behadili HNK. "Improved Firefly Algorithm with Variable Neighborhood Search for Data Clustering". *Baghdad Science Journal*, 19(2), 409-421, 2022.
[31] Xia H, Liu L. "Basketball Big Data and Visual Management System under Metaheuristic Clustering". *Mobile Information Systems*, 2022(1), 1-14 2022.
[32] Singh H, and Kumar Y. "An Enhanced Version of Cat Swarm Optimization Algorithm for Cluster Analysis". *International Journal of Applied Metaheuristic Computing*, 13(1), 1-25, 2022.
[33] Kuo RJ, Zheng YR, Nguyen TPQ. "Metaheuristic-based possibilistic fuzzy k-modes algorithms for categorical data clustering". *Information Sciences*, 557, 1-15, 2021.
[34] Kushwaha N, Pant M, Sharma S. "Electromagnetic optimization-based clustering algorithm". *Expert Systems*, 39(7), e12491, 2022.
[35] Hashemi SE, Tavana M, Bakhshi M. "A New Particle Swarm Optimization Algorithm for Optimizing Big Data Clustering". *SN Computer Science*, 3(4), 1-16, 2022.
[36] Mohammadi M, Mobarakeh MI. "An integrated clustering algorithm based on firefly algorithm and self-organized neural network". *Progress in Artificial Intelligence*, 11(3), 207-217, 2022.
[37] Chaurasia S, Kumar K, Kumar N, "MOCRAW: A Meta-heuristic Optimized Cluster head selection based Routing Algorithm for WSNs" Ad Hoc Networks, 141, 103079, 2023.
[38] Zeng N, Wang Z, Liu W, Zhang H, Hone K, Liu X. "A Dynamic Neighborhood-Based Switching Particle Swarm Optimization Algorithm". *IEEE Transactions on Cybernetics*, 52(9), 9290-9301, 2020.
[39] Singh T. "A chaotic sequence-guided Harris hawks optimizer for data clustering". *Neural Computing and Applications*, 32(23), 17789-17803, 2020.
[40] Hashim FA, Houssein EH, Hussain K, Mabrouk MS, Al-Atabany W. "Honey Badger Algorithm: New metaheuristic algorithm for solving optimization problems" Mathematics and Computers in Simulation, 192, 84-110, 2022
[41] Mirjalili S, Mirjalili SM, Lewis A, "Grey Wolf Optimizer". *Advances in Engineering Software*, 69, 46-61, 2014.
[42] Wang L, Cao Q, Zhang Z, Mirjalili S, and Zhao W. "Artificial rabbits optimization: A new bio-inspired meta-heuristic algorithm for solving engineering optimization problems". *Engineering Applications of Artificial Intelligence*, 114, 105082, 2022.
[43] Abualigah L, Diabat A, Mirjalili S, Abd Elaziz M, Gandomi AH. "The Arithmetic Optimization Algorithm". *Computer Methods in Applied Mechanics and Engineering*, 376, 113609, 2021.
[44] Faramarzi A, Heidarinejad M, Mirjalili S, A. H. Gandomi AH. "Marine Predators Algorithm: A nature-inspired metaheuristic". *Expert Systems with Applications*, 152, 113377, 2020.
[45] Mirjalili S, Lewis A. "The Whale Optimization Algorithm". *Advances in Engineering Software*, 95, 51-67, 2016.