Neighborhood search based improved bat algorithm for data clustering

Arvinder Kaur[1] · Yugal Kumar[1]

Abstract

Clustering is an unsupervised data analytic technique that can determine the similarity between data objects and put the similar data objects into one cluster. The similarity among data objects is determined through some distance function. It is observed that clustering technique gains wide popularity due to its unsupervised and can be used in diverse research filed such as image segmentation, data analytics, outlier detection, and so on. This work focuses on the data clustering problems and proposes a new clustering algorithm based on the behavior of micro-bats. The proposed bat algorithm to determine the optimal cluster center for data clustering problems. It is also observed that several shortcomings are associated with bat algorithm such as slow convergence rate, local optima, and trade-off among search mechanisms. The slow convergence issue is addressed through an elitist mechanism. While an enhanced cooperative method is introduced for handling population initialization issues. In this work, a Q-learning based neighbourhood search mechanism is also developed to effectively overcome the local optima issue. Several benchmark non-healthcare and healthcare datasets are selected for evaluating the performance of the proposed bat algorithm. The simulation results are evaluated using intracluster distance, standard deviation, accuracy, and rand index parameters and compared with nineteen existing meta-heuristic algorithms. It is observed that the proposed bat algorithm obtains significant results with these datasets.

Keywords Clustering · Echolocation · Elitist Strategy · Neighbourhood Search · Meta-heuristic

1 Introduction

In the data analytic field, clustering is a well-known data analysis method for determining similar data objects and grouped these data objects into one cluster [1, 2]. A cluster consists of similar data objects and has dissimilarity with the object present in other clusters [3]. The clustering task can be described through partitional, hierarchical, grid based, density based, and model-based clustering [4–6]. The partitional clustering aims to divide the set of data objects into many distinct clusters based on dissimilarity measures. In hierarchical clustering corresponds to the tree structure of clusters and each node of the tree acts as a cluster. It consists of two approaches- agglomerative (bottom up) and divisive (top down). In the agglomerative approach, each data point belongs to a separate cluster initially. As the process goes on, data points merge within a cluster, if having similar capabilities. While in the divisive approach, all data points belong to one cluster and are repeatedly divided into smaller clusters based on dissimilarity criteria. The grid-based clustering quantizes the data space into a finite number of cells and generates a grid like structure. While, in density-based clustering, the clusters are designed on the basis of data compactness. The compactness is computed through the number of data points presented in a given radius and the density of data points can be used to construct the clusters. The model-based clustering considers the generation of clusters using probability distribution function and each component represents a cluster. The clustering techniques has been proved their potentiality in various field such as image segmentation, stock market, pattern recognition, outlier detection, feature extraction, and medical data analysis [7–9].

Recently, meta-heuristic algorithms are widely adopted in the field of data clustering for obtaining the optimum

✉ Yugal Kumar
yugalkumar.14@gmail.com

Arvinder Kaur
er.arvinderdhillon@gmail.com

1 Department of Computer Science and Engineering & Information Technology, Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh, India

clustering solutions [10]. These algorithms are inspired from swarm intelligence and insect behaviour like particle swarm optimization (PSO) [11, 12], artificial bee colony (ABC) [13–15] and ant colony optimization (ACO) [16, 17]; well-known physics laws like magnetic optimization algorithm (MOA) [1], charged system search (CSS) [18], black hole (BH) [19] and big bang-big crunch (BB-BC) [20, 21]; chemical processes like artificial chemical reaction optimization (ACRO) [22]; evolutionary algorithms [23] like genetic algorithm (GA), genetic programing (GP) and biogeography based algorithm (BBA) [24]; animal behaviour based algorithms like grey wolf optimization (GWO) [25], elephant heard optimization [26], cat swarm optimization [27], lion optimizer [28]; and population based algorithms like sine cosine algorithm [29], stochastic fractal search algorithm [30], thermal exchange optimization algorithm [31]. These algorithms are differ to each other in terms of local and global search mechanisms. The different search mechanisms are adopted for computing local and global optimum solutions. Few algorithms having strong local search ability, for example CSO, BB-BC, SCA, BH etc., while rest of have strong global search ability like PSO, ABC, BBA, GWO etc. [10] However, it is observed that for getting optimal solution, local and global search abilities should be balanced [12]. Further, Abraham et al. [32] stated that data clustering can minimize the dissimilarity measures between the data points within a cluster and dissimilarity can be maximized between data points of other clusters. Several clustering algorithms are designed such k-means, c-means, tabu search, simulated annealing etc. But, it is observed that these algorithms are sensitive to initial solutions, thus easily trapped in local optima. For example, Choudhury et al., [33] designed an entropy based method to determine the initial solutions for the k-means algorithm. The aim of this method is to overcome the dependency of k-means on initial solutions. Moreover, Torrente and Romo [34] also considered the initialization issue of k-means and developed a new initialization method based on the concept of bootstrap and data depth for computing the optimal initial solutions. It is also noticed that traditional clustering algorithms faced difficulty with complex and large datasets. This issue of data clustering is effectively addressed through metaheuristic algorithms. For example, Ahmadi et al. [35] designed a clustering algorithm based on the grey wolf optimization method to tackle the data clustering problems, especially with large datasets. Several modifications like local search and balancing factor are incorporated into grey wolf optimization to effectively handle the data clustering problem. Ghany et al. [36] developed a hybrid clustering algorithm based on whale optimization algorithm (WOA) and tabu search (TS) for solving data clustering. The reason for hybridization is to overcome the local optima and also to improve the quality of clustering solutions. The results stated that hybridization of WOA and TS successfully handles the aforementioned issues.

Sorensen presented the critical evaluation of various well known metaheuristic algorithms [37]. It is stated that researchers focus on the actual mechanism behind the underlying concept rather than to develop the new metaheuristic algorithm and also concentrate promising research direction in the field of metaheuristic algorithms. To keep in mind rather than design a new metaheuristic algorithm, this work considers the existing metaheuristic algorithm i.e. Bat algorithm for solving data clustering problems. Recently, the bat algorithm become popular in the research community and provides optimal solution for various optimization problems [38–40]. Bat algorithm is developed by Yang et al. [38] based on the behavior of micro-bats into an algorithm, especially the echolocation feature of micro-bats. The micro-bats use the echolocation feature to detect prey (food) and avoid obstacles. For detection of prey, microbats emits a short pulse. The aim of short pulse is to produce echo and in turn, micro-bats recognize the shape and size of prey. It is seen that several performance issues are associated with the bat algorithm [40–42]. These issues are outlined as convergence rate, local optima, population initialization, and trade-off factor among local, and global searches. In turn, the bat algorithm converges on near to optimal solution instead of the optimal solution. The issues related to the performance of the bat algorithm are summarized as

- **Population Initialization:** The initial population provides a significant impact on the success of clustering algorithms [42, 43]. If, initial population is not selected in effective manner, then premature convergence problem can occur. As, meta-heuristic algorithms select the initial population using random function.
- **Local optima:** It is noticed that sometime, the population of bat algorithm is not updated in effective manner [39, 44]. In turn, the objective function returns same value in successive iteration. Finally, algorithm converges with same solution, but the solution is not optimal one. This situation is called local optima and it occurs due to lack of appropriate mechanism to update population of micro-bats.
- **Convergence Rate:** The convergence rate of an algorithm depends on the optimization process and exploration of the search space [45, 46]. The convergence rate can also affect due to lack of coordination between exploration (local search) and exploitation (global search) processes.

The contribution of work is given as:

1. To develop an enhanced cooperative co-evolution method to handle the population initialization issue.

2. An elitist strategy is developed for improving the convergence rate.
3. To incorporate limit operator to check the local optima situation in algorithm.
4. To develop the neighbourhood search mechanism for exploring optimal candidate solution in exploration process.
5. The proposed bat algorithm is applied to solve clustering problems.

2 Related works

The recent works reported on partitional clustering algorithm are summarized in this section. Since past few decades, numbers of clustering algorithms are developed for obtaining the optimum results for partitional clustering. Few of them are discussed below.

To determine best initial population and automatic cluster numbers, Rahman and Islam [47] designed a hybrid algorithm based on K-means (KM) and genetic algorithm (GA). Genetic algorithm was applied to determine optimized initial cluster centres for KM. C-means algorithm was adopted to obtain optimum clustering results. The performances of proposed algorithms were assessed on twenty datasets. The results were compared using well-known clustering techniques. It was claimed that fuzzy c-means with GA gives better clustering results.

Liu et al. [48] presented a clone selection algorithm for addressing automatic clustering. In automatic clustering, number of clusters can be detected in auto manner. Hence, in this work, authors introduce a genetic operator for detecting the number of clusters. The well-known twenty-three datasets are selected for measuring the performance of the clone selection algorithm. The results are compared with ILS, ACDE, VGA and DCPSO algorithms. Authors claimed that proposed algorithm provides better results without prior knowledge for number of clusters.

A two-step artificial bee colony algorithm was reported for obtaining the optimal clustering results [49]. Prior to implement, three improvements are inculcated in ABC algorithm to make it more robust and efficient. These improvements are summarized as initial cluster centre locations, updated search mechanism, and equations and abandoned food source. The initial cluster centre locations are determined through one step KM method. A PSO based search mechanism is used for exploring the promising search space. Hooke and Jeeves concept are considered for evaluating abandoned food source locations. The performance of proposed two-step ABC algorithm is tested on both artificial and benchmark data sets and compared with well-known clustering algorithms. It was observed from the results that

the proposed algorithm significantly improves the performance of conventional ABC algorithm.

Cao et al. [50] developed a new initialization method based on neighbourhood rough set model. The intra cluster and inter cluster similarities of an object were represented in terms of cohesion and coupling degrees. Furthermore, it is integrated with KM algorithm for improving clustering results. The efficacy of proposed algorithm is tested over three datasets and compared with other two initialization algorithms. The proposed initialization method provides superior results than traditional methods.

Han et al. [51] adopted a new diversity mechanism in gravitational search algorithm to handle clustering problems. The collective response of birds can be used to design diversity mechanism and implemented through three simple steps- (i) initialization, (ii) identification (nearest neighbours) and (iii) orientation alteration. The candidate population is generated into initialization step as a first step of algorithm. The second step corresponds to evaluate the nearest neighbours through a neighbourhood strategy. Third step can change the current location of candidate solution based on nearest neighbour. Thirteen datasets are chosen for evaluating the performance of algorithm and simulation results are compared with well-known clustering algorithms. Authors claimed that proposed algorithm achieves superior clustering results.

Senthilnath et al. [52] introduced two-phase fire fly algorithm (FA) for clustering task. This algorithm simulates the flashing pattern and social insect behaviours of fire flies. First phase of algorithm measures the variation of light intensity. Second phase towards the movement of fireflies. The efficiency of fire fly algorithm is assessed on thirteen standard datasets and compared with ABC and PSO. The simulation results favour the existence of FA algorithm in clustering filed.

To handle the initialization issue of K-mean algorithm, Erisoglu et al. [53] developed a new initialization method. This method is based on the bi-dimensional mapping of features. Initially, two features are chosen, the first feature is an attribute with maximum value of variation coefficient, called main axis. The second feature is determined using correlation values between main axis (first variable) and rest of attributes. Hence, the second feature is an attribute with minimum correlation. The several benchmark datasets are used to evaluate the performance of the proposed algorithm. From simulation results proved that proposed method significantly better than KM algorithm.

Kumar and Sahoo [54] hybridized the MCSS algorithm with PSO. The personal best mechanism of PSO algorithm was added into magnetic charge system search algorithm. Further, neighbourhood strategy was also introduced to avoid local optima situation. The ten datasets are selected for evaluating the performance of MCSS–PSO algorithm

and results are compared with wide range of clustering algorithms. Authors claimed that better quality results are achieved by MCSS-PSO algorithm.

Zhou et al. [55] introduced simplex method-based SSO algorithm for solving clustering task. In this work, simplex method is incorporated into SSO algorithm to enhance local search ability and improved convergence speed. The eleven datasets are considered for evaluating the simulation results of proposed algorithm and compared with well-known clustering algorithms. The proposed SSO algorithm perform well in terms of accuracy, robustness, and convergence speed.

Boushaki et al. [56] designed a new quantum chaotic CS algorithm for clustering task. To extend the global search ability of quantum chaotic cuckoo search algorithm, a non-homogeneous update mechanism was employed. Chaotic maps are incorporated in this algorithm to improve convergence speed. The performance of algorithm was compared with different variants CS algorithms and hybrid variants of clustering algorithms. Authors claimed that proposed CS algorithm provides more compact clusters than other algorithms.

A combination of GA and message-based similarity (MBS) measure was presented for effective cluster analysis by Chand et al. [57]. The MBS measure consists of two types of messages-responsibility and availability. The messages are exchanged among data points and cluster centres. The responsibility can be measured as evidence regarding cluster centres, while the availability corresponds to appropriateness of data point with respect to clusters. Further, GAMBS consists of variable-length real-valued chromosome representation and evolutionary operator. The artificial and real-life datasets are adopted for measuring the performance of GAMBS algorithm. The simulation results showed that the algorithm obtains significant clustering results.

Hatamlou [23] developed a new clustering algorithm inspired from black hole (BH) phenomenon. Like other clustering algorithms, BH algorithm starts with initial population selection and objective function evaluation. The performance of proposed algorithm is tested on six benchmark datasets and it is stated that black hole clustering algorithm provides better clustering results.

Zhang et al. [58] presented an ABC algorithm for data clustering. In ABC, onlooker bees and employed bees are responsible for global search, while scout bees are responsible for local search. Further, Deb's rule is incorporated to redirect search in solution space. The performance was tested on three real-life datasets and compared with other clustering algorithms. Results revealed that the proposed algorithm provides good quality results.

Taherdangkoo et al. [59] reported a new blind naked mole rat's algorithm in clustering field. This algorithm considers food search capability and colony protection characteristics of mole rats. The algorithm starts by initializing the population of mole rats and searches the entire space for optimal solution in random fashion. In next iterations, employed mole rats start movement to target food source and their neighbours. The performance of proposed algorithm was tested on six standard datasets and compared with other well-known clustering algorithms. Results revealed that blind naked mole rat's algorithm provides higher accuracy with faster convergence speed.

Hatamlou [60] considered the slow convergence rate of binary search algorithm and designed a new algorithm for cluster analysis. This algorithm chooses initial cluster points from differ locations. Further, the search direction is based on the successive objective function values. If current objective function is better than previous objective function, then search proceeds in same direction, otherwise in opposite direction. The six benchmark datasets are chosen for evaluating the efficacy of proposed algorithm. The results are compared with KM, GA, SA, TS, ACO, HBMO, and PSO algorithms. The proposed algorithm provides superior clustering results.

Bijari et al. [61] presented a memory-enriched BB-BC algorithm for clustering. It works in two phases- BB and BC phase. The BB phase corresponds for generation of random points near to initial seed points. While, BC phase corresponds for optimizing these generated points. The BB-BC algorithm is memory less algorithm. So, a memory concept is integrated into BB-BC algorithm for memorizing the best location and also maintaining the exploration and exploitation tasks. The performance of algorithm was tested on six data sets and compared with well-known algorithms like GA, PSO, GWO, and original BB–BC. Results stated that the clustering results are improved significantly.

Abualigah et al. [62] combined krill herd (KH) optimization algorithm with harmony search (HS) to overcome local optima problem in clustering. A global exploration operator and reproduction procedure was integrated in krill herd algorithm. The seven standard datasets are selected for measuring the performance of proposed algorithm and results are compared with GA, PSO, HS, KHA, H-GA, and H-PSO algorithms. Authors claimed that proposed combination (KH + HS) achieves more accurate clustering results.

Pakrashi and Chaudhuri [63] hybridized Kalman filtering algorithm with KM algorithm. In this work, authors consider the slow convergence rate of KM algorithm and it can be improved with the help of Kalman filtering algorithm. Further, a conditional restart mechanism was also incorporated in K-Means algorithm to handle local optima situation. The seven benchmark datasets are taken for evaluating the performance of proposed algorithm and results are compared with HKA, KGA, GAC, ABCC, and PSO algorithms. It is noticed that Kalman filtering algorithm successfully overcome the deficiency of KM algorithm.

Kang et al. [64] hybridized KM and mussels wandering optimization algorithm, called K-MWO. The proposed algorithm comprises of local search abilities of KM, while, global search is accomplished through mussels wandering optimization algorithm. The performance is tested on nine datasets and results are compared with K-M and K-PSO algorithms. Authors claimed that K-MWO is an effective clustering algorithm.

To solve clustering search space problems, Wang et al. [65] presented a hybrid version of flower pollination algorithm (FPA) and bee pollinator algorithm (BPA). The discard pollen operator of ABC is used for enhancing global search ability of flower pollination algorithm. Further, the local search mechanism is improved through elite mutation and crossover operator. Several artificial and benchmark datasets are selected for measuring the performance of proposed algorithm. The simulation results are compared with KM, FPA, CS, PSO, ABC, DE, algorithms. Results proved that combination of FPA and BPA provides more optimal results than others.

Hatamlou and Hatamlou [66] designed a two-stage clustering approach to overcome the drawbacks of particle swarm optimization like local optima and slow convergence speed. In first stage, PSO algorithm is adopted for generating the initial candidate solution. In second stage, HS algorithm is considered for improving the quality of solution. Seven datasets are chosen for measuring the performance of the proposed algorithm and results are compared with KM, PSO, GSA, BB-BC methods. It is seen that proposed algorithm determines good quality clusters.

A hybrid version of ABC algorithm with genetic algorithm is also presented for enhancing the information exchange mechanism among bees by Yan et al. [67]. It is applied for solving data clustering problems. The information exchange mechanism is enhanced with the help of crossover operator. The six standard datasets are adopted for evaluating the simulation results of proposed ABC algorithm and results are compared with other ABC, CABC, PSO, CPSO and GA clustering algorithms. The proposed ABC algorithm having better clustering results than others.

To perform efficient clustering, Kwedlo [68] combined differential evolution (DE) algorithm with KM. KM algorithm is used to tune candidate solutions generated through mutation and crossover operators of DE. Additionally, a recording procedure is also introduced to handle redundant solutions. The performance of proposed algorithm was compared with five other well-known clustering algorithms. It was noticed that DE-KM algorithm gives state of art clustering results.

Yin et al. [69] presented a hybridized version of improved GSA with KHM for solving clustering problems. This work considers the convergence rate of KHM and diversity mechanism of GSA to develop new algorithm. The performance of proposed algorithm is tested on seven benchmark datasets and compared with other well-known clustering algorithms. Authors claimed that combination of KHM-GSA achieves better convergence.

A hybrid version of ant algorithm is presented for handling clustering problems [70]. KHM algorithm is used for hybridizing the Ant algorithm, called KHM-Ant. The proposed algorithm contains the merit of both algorithms such as initialization characteristic of KHM and local optima characteristic of ant. The five benchmark datasets are considered for measuring the performance of KHM-Ant algorithm. The simulation results are compared with KHM and ACA. Authors claimed that more optimal results are achieved by KHM-Ant algorithm.

Xiao et al. [71] developed a quantum-inspired GA (QGA) for partitional clustering. In this work, Q-bit based representation and rotation operation of quantum gates are applied for achieving better search mechanisms. Several standards and simulated datasets are selected for evaluating the performance of QGA algorithm. The QGA is able for finding optimal clusters without prior knowledge of number of clusters centres.

Aljarah et al. [72] hybridized grey wolf optimizer (GWO) with tabu search (TS) for cluster analysis. TS is incorporated as an operator in GWO for searching neighbourhood. It helps in balancing exploration and exploitation of GWO. The proposed GWOTS is tested on thirteen real datasets and results have been compared with other popular metaheuristics. The experiment results show that GWOTS is superior in terms of convergence behaviour and optimality of results.

A PSO based clustering algorithm is presented in [73]. The concept of cooperative evaluation is incorporated into PSO for improving convergence rate and diversity. The cooperative co-evolution method worked as decomposer and PSO algorithm as optimizer. The standard and simulated datasets are selected for measuring the performance of PSO and compared with SRPSO, ACO, ABC, DE, and KM algorithms. The concept of cooperative evaluation improves the performance of PSO in significant manner.

To solve clustering problems effectively, an improved CSO algorithm is reported in [74]. Several modifications are incorporated in CSO algorithm to make it effective. These modifications are described in terms of search equations. Further, a local search method is also developed for handling local optima problem. The performance is evaluated on five datasets and compared with several known clustering algorithms. Simulation results showed that improved CSO obtains effective clustering results.

A class room teaching based meta-heuristic algorithm is also presented for handling clustering problems [75]. The properties of K-means algorithm were also investigated for effective clustering results [76]. Six benchmark datasets are used to evaluate the performance of aforementioned

algorithm. The performance is measured in terms of overlapping, number of clusters, dimensionality and cluster size. An intelligent system for spam detection was presented in [77]. More relevant features were identified using evolutionary random weight networks. Table 1 gives the summary of the various studies of literature.

3 Bat algorithm

Bat algorithm is based on the echolocation behaviour of microbats especially prey detection and obstacle avoidance [38]. In search of prey, microbat emits short pulse and considers the echo of nearby objects to determine its shape and size. The loudness, emission rate and random variable of bat algorithm are initialized using Eqs. (1–2).

$$A_i^{t+1} = \alpha\left(A_i^t\right) \tag{1}$$

$$r_i^{t+1} = \left[1 - \exp\left(-\alpha\right)\right] \tag{2}$$

Where, A_i^t is loudness, r_i^t is pulse emission rates and (α) is a user specified variable having values between [0–1]. The frequency and velocity of bats are computed using Eqs. (3–4).

$$f_i^t = f_{min}^t + \left(f_{max}^t - f_{min}^t\right) \text{rand}() \tag{3}$$

$$v_i^t = v_i^{t-1} + \left(x_i^t - x_*\right)f_i^t \tag{4}$$

Where, f_{min}^t and f_{max}^t denotes lowest and highest frequencies at time stamp (t) and rand () is a random function having values between [0–1], v_i^{t-1} is initial velocity, x_i^t is current position and x_* is current best position. The positions of bats are updated using Eqs. (5–6)

$$x_i^{t+1} = X_{new} + v_i^t \tag{5}$$

$$X_{new} = x_i^t + \text{randi}[-1, 1]A_i^t \tag{6}$$

Where, X_{new} is new position and x_i^{t+1} is final updated position.

4 Improved Bat algorithm for cluster analysis

This section presents the proposed improvements in bat algorithm. These improvements are (i) enhanced co-operative co-evaluation for population initialization, (ii) elitist strategy for better convergence rate as well as tradeoff between local and global solutions, and (iii) neighborhood strategy to avoid local optima and explore good candidate solutions.

4.1 Enhanced co-operative co-evolution method

It is observed that efficiency of clustering algorithm also depends on initial cluster points [42, 44]. Several initialization methods are reported to address initial cluster selection problem [47–50]. For improving the performance of bat algorithm, an enhanced co-operative co-evolution framework is introduced to select initial cluster centers. The co-operative co-evolution method works on divide and conquers paradigm. This paradigm divides the problem into sub problems and sub problems can be solved individually and final solution is obtained by combing each sub problem solution. Hence, in this work, a co-operative co-evolution method with centroid selection mechanism is proposed. This method considers number of partitions, size of partition, and selection criteria for population initialization.

4.1.1 Population partitions and size description

This subsection gives description about number of partitions and their size to implement the co-operative co-evolution method. First task is to divide the population (data instances) into several predefined partitions. The partitions are equal to number of clusters (K) for a given dataset as mentioned in Eq. 7.

$$p_n \propto K \tag{7}$$

Where, p_n denotes number of partitions and K denotes the number of clusters. The size of subpopulation is obtained through Eq. 8.

$$p_s = T/K \tag{8}$$

Where, T is total population size i.e., data instances in a dataset, K is number of clusters and p_s denotes size of subpopulations. The number of subpopulations is computed through Eq. 9.

$$\begin{cases} p_{s1} = 1 \text{ to } \lceil p_s \rceil \\ p_{s2} = \lceil UL(p_{s1}) + 1 \rceil \text{ to } \lceil p_{s1} + p_s \rceil \\ \quad \dots\dots\dots\dots\dots\dots\dots\dots \qquad n = \{1, 2, 3 \dots \dots K\} \\ p_{s(n-1)} = \lceil UL(p_{s(n-2)}) + 1 \rceil \text{ to} \dots + \lceil p_{s2} \rceil + \lceil p_{s1} \rceil + \lceil p_s \rceil \\ p_{sn} = \lceil UL(p_{s(n-1)}) + 1 \rceil + \sum_{n=0}^{K-1} \lceil p_s \rceil \end{cases} \tag{9}$$

In Eq. 9, p_{s1}, p_{s2}, …p_{sn} denotes subpopulations. UL represents the upper limit in a sub population. Further, from each subpopulation an appropriate centroid is selected using Eq. 10.

$$C_{kn} = \min\left(p_{sn}\right) + \left(\max\left(p_{sn}\right) - \min\left(p_{sn}\right)\right) * \text{rand}(0, 1); \text{Where n} = 1 \text{ to K} \tag{10}$$

Where, C_k denotes k^{th} cluster center, $\min(p_n)$ and $\max(p_n)$ are minimum and maximum values of each nth sub population (p_{sn}) and rand () is a random number.

Table 1 Summarization of the recent state of art works in the direction of partitional clustering

Author Name	Source of Inspiration	Shortcomings	Amendments	Future Work
Rahman & Islam [47]	- K- means and Genetic Algorithm	-Initial Population-Automatic cluster numbers	-Proposed a hybrid algorithm based on K-means (KM) and genetic algorithm (GA). -Optimized initial cluster centres identified using GA for KM. -Optimum clustering results obtained using C-means.	-Reduction of time complexity for GenClust.
Liu et al. [48]	-Gene Transposon	- Automatic Clustering	-Proposed an improved variant of clonal selection algorithm to determine number of clusters. -A novel operation called antibody gene transposon is introduced to the framework of clonal selection algorithm which can realize to find satisfied number of cluster automatically	- Employ more improved cluster validity indices to form antibody affinity making it suitable for non-convex data sets.
Kumar & Sahoo [49]	-Artificial Bee Colony algorithm	-Issues of ABC addressed such as food sources initial position, solution search equation and abandoned food location	-One step KM method has been used for finding initial cluster centre locations. -Search space exploration performed using PSO based search mechanism. -Further, Hooke and Jeeves concept are employed to evaluate abandoned food source locations.	-Other initialization methods can be explored for generating initial optimal solutions for clustering algorithms in future.
Cao et al. [50]	-Neighbourhood rough set model	-Initialization issue of K-means	-Degree of cohesion and coupling between neighbourhoods of objects have been defined based on the neighbourhood rough set model -Integrated with K-means	-Extreme machine learning concept will be investigated for improving clustering results.
Han et al. [51]	-Gravitational Search Algorithm	-Local Optima-Premature convergence	-Modified Gravitational Search Algorithm BFGSA was proposed using bird flock mechanism. -Three step mechanism has been employed. First step initializes the candidate solution. Second step evaluates nearest neighbours using neighbourhood strategy. Third step can based on nearest neighbour change the current location of candidate solution based on nearest neighbour. -The collective response mechanism helps BFGSA explores wider range of search space and escapes suboptimal solutions.	-Some evolutionary operators will be considered to overcome local optima. -Design new approaches for generating diverse population.

Table 1 (continued)

Author Name	Source of Inspiration	Shortcomings	Amendments	Future Work
Senthilnath et al. [52]	-Fire fly algorithm	- Applied for clustering task	-Simulates flashing pattern and social insect behaviours of fire flies in two phases. -First phase of algorithm measures variation of light intensity. -Second phase towards the movement of fireflies.	-Design new heuristics to recognize flashing pattern and light intensity.
Erisoglu et al. [53]	-Bi-dimensional mapping of features	-Initialization issue of K-means	-Two features are chosen. -First feature is an attribute with maximum value of variation coefficient, called main axis. -Second feature is determined using minimum correlation values between main axis and rest of attributes.	- Focus on bi-directional clustering
Kumar & Sahoo [54]	-MCSS and PSO	-Local Optima-Solution search	-Hybridized MCSS algorithm with PSO. -Personal best mechanism of PSO algorithm added into magnetic charge system search algorithm. -Neighbourhood strategy introduced to resolve local optima situation.	
Zhou et al. [55]	-Social spider optimization (SSO)	-Initial cluster centers -Local optima -Convergence Rate	-Incorporated simplex method into SSO algorithm -It enhances local search and improves convergence rate.	-Extend fitness function to explicitly optimize intra-cluster distances. -Perform testing with higher-dimension problems and more patterns. -Work to be extended for dynamically determining optimal number of clusters.
Boushaki et al. [56]	-Quantum Chaotic Cuckoo Search algorithm	-Initial population -Search ability -Convergence speed	-Designed quantum chaotic CS algorithm for clustering task.-Initial population is created using ergodicity and non-repetition properties of chaotic map. -Also, employed nonhomogeneous update mechanism for global search ability improvement. -Incorporated boundary strategy for enhancing search procedure.	-Apply proposed approach on specific datasets. -Also applying quantum theory to other recent meta-heuristic algorithms.

Table 1 (continued)

Author Name	Source of Inspiration	Shortcomings	Amendments	Future Work
Chang et al.[57]	-GA, message-based similarity (MBS)	-Initial cluster numbers -Clustering performance	-Developed GAMS algorithm for clustering problem with unknown cluster number. -Decisions are made using domain specific knowledge by GAMS. -It can find optimal number of clusters as well as cluster centres automatically. - New similarity measure uses both distance between data point with the nearest centre and with neighbouring data points.	-New method will be designed to determine initial solutions that can identify the clusters in automatic manner.
Hatamlou [19]	Black Hole Optimization	-Quality clusters	-Proposed BH for solving clustering problem. - It is simple, easy to implement and free from parameter tuning.	-Can be applied to other areas of applications. -For more effectiveness, BH can be combined with other approaches.
Zhang et al. [58]	-Artificial Bee colony	-Quality clusters	-Developed an artificial bee colony algorithm to solve clustering problems. -Onlooker bees and employed bees in ABC are responsible for global search, whereas scout bees are responsible for local search.	-Can be combined with effective local search strategy and hybrid using other metaheuristics. -Proposed algorithm can be applied to solve other optimization problem.
Taherdangkoo et al. [59]	-Blind, naked mole-rats (BNMR) algorithm	-Initial cluster centers -Local optima -Convergence Rate	-Reported a new blind naked mole rat's algorithm in clustering field. -Considers food search capability and colony protection characteristics of mole rats. -The algorithm starts by initializing the population of mole rats and searches the entire space for optimal solution in random fashion. - In next iterations, employed mole rats start movement to target food source and their neighbours.	-Increasing the speed and accuracy of proposed algorithm by combining it with other approaches. -Self adaptive parameter tuning

Table 1 (continued)

Author Name	Source of Inspiration	Shortcomings	Amendments	Future Work
Hatamlou [60]	-Binary search algorithm	-Quality cluster -Convergence rate	-Proposed novel binary search algorithm for data clustering. -It finds high quality clusters and converges to the same solution in different runs. -In the proposed algorithm a set of initial centroids are chosen from different parts of the test dataset and then optimal locations for the centroids are found by thoroughly exploring around of the initial centroids.	- Integrate the neighbourhood structure to determine optimal population.
Bijari et al. [61]	-Memory-enriched Big Bang-Big Crunch algorithm	-Slow convergence speed -Local optima -Exploration and exploitation trade-off	-Reported a memory-enriched BB-BC algorithm for clustering. -Works in two phases that is BB and BC phase. -BB phase corresponds for generation of random points near to initial seed points. -BC phase corresponds for optimizing these generated points. - A memory concept is integrated into BB-BC algorithm to memorize best location and balancing exploration and exploitation tasks.	-BB-BC can be hybrid with k-means to overcome limitations of K-means. -Can be used for multi-objective optimization -Application of ME-BB–BC in technical settings such as power dispatch systems
Abualigah et al. [62]	-Krill herd and harmony search	-Local Optima-Global search ability	-Proposed hybrid algorithm by combining krill herd (KH) optimization algorithm with harmony search (HS). -A global exploration operator and reproduction procedure integrated in krill herd algorithm.	-Investigating on benchmark function datasets
Pakrashi & Chaudhuri [63]	-K-Means and Heuristic Kalman algorithm	-Convergence rate -Local Optima	-Hybridized Kalman filtering algorithm with K-Means algorithm (HKA-K). -Kalman filtering helps in improving convergence rate -Incorporated a conditional restart mechanism in K-Means algorithm to handle local optima situation.	-Investigate HKA by hybridizing with other methods. -Improving performance of HKA can be investigated. -A new and effective multi-solution framework can be developed for this proposed method instead of conditional restart method

Table 1 (continued)

Author Name	Source of Inspiration	Shortcomings	Amendments	Future Work
Kang et al. [64]	-KM and mussels wandering optimization algorithm	-Improve local search and global search abilities	-Hybridized KM with mussels wandering optimization algorithm, called K-MWO. -The proposed algorithm encompasses local search abilities of KM and global search is achieved through mussels wandering optimization algorithm.	-Study new updating rules of the weight information in the proposed clustering ensemble framework. -Exploring the reason why proposed method cannot perform well in some particular dataset. -Develop a more influential clustering ensemble by combining with other approaches.
Wang et al. [65]	-Flower Pollination Algorithm and Bee Pollinator	-Diversity of population -Local Optima -Convergence Speed	-Hybridized flower pollination algorithm (FPA) and bee pollinator algorithm (BPA). -Global search ability of FPA is enhanced using discard pollen operator of ABC. -Elite mutation and crossover operator are used for improving local search mechanism.	-Extend fitness function to explicitly optimize intra-cluster distances. -Investigation of proposed algorithm on higher dimensional problems and large number of patterns. - Determine optimal number of clusters dynamically in proposed algorithm.
Hatamlou & Hatamlou [66]	-Particle Swarm Optimization (PSO) and Heuristic Search (HS)	-Local Optima -Convergence Speed	-Designed a two-stage clustering approach by hybridizing PSO and HS. -First stage adopts PSO algorithm for generating initial candidate solution. -Second stage employs HS algorithm for improving quality of solution.	-Wrapping methods can be adopted for generating initial solutions.
Yan et al.[67]	-Artificial Bee Colony and GA	-Convergence Speed of ABC	-Proposed a Hybrid Artificial Bee Colony (HABC) algorithm to improve optimization ability of canonical ABC. -Information exchange between bees is done through crossover operator of Genetic Algorithm (GA).	-Finding features of functions and improving HABC for which it does not work well.
Kwedlo [68]	-Differential Evolution and K-means	-Effective Clustering by combining evolutionary algorithm with local optimization method	-Combined differential evolution algorithm (DE) with K-means for clustering. -K-means algorithm fine-tunes each candidate solution attained by mutation and crossover operators of DE. -Proposed a reordering procedure to tackle redundant representation problem.	-Replacement of K-means by better local search method. -Reducing the runtime of DE-KM through parallelization

Table 1 (continued)

Author Name	Source of Inspiration	Shortcomings	Amendments	Future Work
Yin et al. [69]	-Gravitational Search Algorithm and KHM	-Local Optima of KHM -Slow convergence Speed of IGSA	-Proposed a hybridized version of improved GSA with KHM. -Convergence rate of KHM and diversity mechanism of GSA are considered.	-For effective clustering Integrate better local search algorithms into KHM.
Jiang et al. [70]	-Ant clustering algorithm, K-harmonic means	-Initialization-Local Optima	-Propose a new clustering algorithm using Ant clustering algorithm with K-harmonic means clustering (ACAKHM). -Proposed algorithm utilizes merits of KHM and ACA. -KHM is less sensitive to initialization and ACA can avoid trapping in local optimal solution.	-Improving Ant clustering algorithm to reduce runtime of ACAKHM algorithm. - Study KHM algorithm with other combinatorial optimization techniques.
Xiao et al. [71]	-Quantum-inspired GA, K-means	-Initial cluster centroid	-Proposes a quantum-inspired genetic algorithm for k-means clustering (KMQGA). -A Q-bit based representation is employed for exploration and exploitation. -Variable length of a Q-bit in KMQGA is considered during evolution. -KMQGA obtains optimal number of clusters and provide optimal cluster centroids.	-Investigate exploring search space with small number of individuals
Aljarah et al. [72]	-Grey Wolf Optimizer, Tabu search	-Local Optima -Premature convergence	-Hybridized GWO with TS -Incorporated TS as an operator in GWO. -TS helps balancing exploration and exploitation of GWO.	-Investigation of GWOTS on synthetic datasets and arbitrary shapes datasets. -Run time reduction of GWOTS using parallel computation.
Jiang & Wang [73]	-PSO, bare-bone particle swarm optimization (BPSO), Cooperative Co-evolution (CC)	-Convergence rate -Diversity of population	-Developed CC framework to improve the performance of PSO on clustering high-dimensional datasets. -To solve each sub-problem cooperatively, BPSO is employed. -A new centroid-based encoding schema is designed for each particle. -Chern off bounds is applied for deciding population size.	-Develop CC framework for clustering problems where numbers of clusters is not known prior. -Detection of natural clusters and clustering in sub-space with CC framework.

Table 1 (continued)

Author Name	Source of Inspiration	Shortcomings	Amendments	Future Work
Kumar & Singh [74]	-Cat Swarm optimization (CSO)	-Unbalance exploration and exploitation processes. -Lack of diversity -Local Optima -Slow convergence	-Proposed an Improved CSO -New search equation is proposed for steering search towards global optimal solution. -Incorporated local search method for quality solutions and overcome problem of local optima.	-Multi objective variant of CSO algorithm will be developed for improving clustering quality.
Kumar & Singh [75]	-Teaching learning based optimization (TLBO)	-Unbalance between local and global search -Premature Convergence	-Proposed chaotic version of TLBO algorithm using different chaotic mechanisms. -Incorporated a local search method for improving quality of solution and better trade-off among local and global search.	-Instead of local search, global search mechanism will be incorporated and investigated its performance.
Fränti & Sieranoja [76]	-K-means	-Initialization -Clustering accuracy	-Introduced clustering basic benchmark. -Performance of k-means is studied using this benchmark. -Performance is measured on factors-overlap of clusters, dimensionality, number of clusters, and cluster sizes	
Faris et al. [77]	-GA, Random Weight Network (RWN)	-Automatic detection of features	-Proposed a hybrid e-mail spam detection system by combining RWN and GA. -Automatic Feature Selection (FS) and e-mail classification is performed in two stages. -Wrapper approach is used in FS stage, whereas RWN is employed in classification stage.	-Imbalanced classification tasks to be investigated

The working of co-operative co-evolution method is illustrated using iris dataset. This dataset contains one hundred fifty data instances and four attributes. These data instances are divided into three classes. Hence number of clusters (K) considered for iris dataset is three. The co-operative co-evolution method contains three Eqs. (7–10) to determine the initial population for clustering algorithm in terms of cluster centers.

a) First step is to divide the population into subpopulation. Equation 7 is used to compute the subpopulation.

$p_n \propto K$, For iris dataset, $K = 3$;
$$p_n = 3$$

b) In second step, size of subpopulations is computed. The size of subpopulation is computed through Eq. 8.

$p_s = T/K$, where $T = 150$ and $K = 3$;
$$p_s = 150/3 = 50.$$

The size of subpopulation (p_s)is 50. Further, the number of subpopulations is determined using Eq. 9.

$p_{s1} = 1$ to $\lceil p_s \rceil$; $p_s = 50$;
$p_{s1} = 1$ to $50, LB(p_{s1}) = 1$ and $UB(p_{s1}) = 50$;
$p_{s2} = \lceil UB(p_{s1}) + 1 \rceil$ to $\lceil p_{s1} + p_s \rceil$; $UB(p_{s1}) = 50, p_s = 50$;
$p_{s2} = 51$ to $100; LB(p_{s2}) = 51, UB(p_{s2}) = 100$;

$$p_{sn} = \lceil UB(p_{s(n-1)}) + 1 \rceil + \sum_{i=0}^{n-1} \lceil p_s \rceil; n = 3$$

$$p_{s3} = \lceil UB(p_{s2}) + 1 \rceil + \sum_{i=0}^{2} \lceil p_s \rceil; \ p_{s0} = p_s$$

$p_{s3} = 101$ to 150;

iii) In third step, Eq. 10 is used for computing initial cluster centers for clustering algorithm.

- When $n = 1$　$c_{k1} = \min(1:50) + (\max(1:50) - \min(1:50)) * \text{rand}()$

  {5.5221 4.0109 1.7333 0.5074}

- W　　　　　h　　　　e　　　　n
  $n = 2$　$c_{k2} = \min(51:100) + (\max(51:100) - \min(51:100)) * \text{rand}()$
  {6.8022 3.2681 4.9022 1.7246}

- When $n = 3$　$c_{k3} = \min(101:150) + (\max(101:150) - \min(101:150)) * \text{rand}()$
  {5.2810 2.4032 4.8048 1.5397}

4.2 Elitist strategy

This subsection discusses another important aspect of clustering that is convergence speed. The convergence rate depends on the searching pattern of algorithm. To improve convergence speed, an improved elitist strategy is incorporated in bat algorithm. According to elitist strategy, best positions move from previous iteration to next iteration. In this work, elitist strategy is implemented in two phases- Evaluation Phase and Updating Phase.

4.2.1 Evaluation phase

In this phase, personal best and global best positions are computed using Eqs. (11–12). The comparison operator is used to calculate global best position i. e. X_{Gbest} and personal best position X_{Pbest}.

$$X_{Pbest} = \min(\text{fitness value}) \tag{11}$$

$$X_{Gbest} = \min(\text{distance value}) \tag{12}$$

The personal best (X_{Pbest}) is obtained using the fitness function as described in Eq. 13.

$$F(C_K) = \sum_{K \in 1}^{K} \frac{SSE(C_K)}{\sum_{K=1}^{K} SSE(C_K)} \tag{13}$$

Where, SSE denotes the sum of square error and C_K represents the Kth centroid object. After evaluation of fitness function minimum value is selected as personal best. In next step, the global best (X_{Gbest}) is evaluated using Eq. 13, which is minimum value of distance function or objective function.

4.2.2 Updating phase

In this phase, the personal best and global best positions are compared with previous iteration values. If, current values are better than previous values, than positions are updated using Eqs. (14–15). Otherwise, previous values are considered.

$$X_{Pbest} = \begin{cases} X_{Pbest}^{t-1} = X_{pbest}^{t} & \text{fit}(t) <= \text{fit}(t-1) \\ X_{Pbest}^{t} = X_{pbest}^{t-1} & \text{fit}(t) >= \text{fit}(t-1) \end{cases} \tag{14}$$

$$X_{Gbest} = \begin{cases} X_{Gbest}^{t-1} = X_{Gbest}^{t} & s(t) \le s(t-1) \\ X_{Gbest}^{t} = X_{Gbest}^{t-1} & s(t) \ge s(t-1) \end{cases} \tag{15}$$

To achieve optimum trade-off among search mechanisms, the basic frequency, velocity and search equations of bat algorithm are modified using Eqs. (16–19).

$$f_i^t = \frac{\min\left(X_{Gbest}^t\right) + \left(\max\left(X_{Gbest}^t\right) - \min\left(X_{Gbest}^t\right)\right)\beta}{\max\left(X_{Pbest}^t\right)} \quad (16)$$

$$v_i^t = v_i^{t-1} + \left(X_{Gbest}^t - X_{Pbest}^t\right)f_i^t \quad (17)$$

$$X_{new} = X_{Gbest}^t + randi[-1, 1] \quad (18)$$

$$X_i^t = \begin{cases} X_{Gbest} & \text{if } rand > r_i \\ X_{new} + v_i^t & \text{otherwise} \end{cases} \quad (19)$$

where, f_i^t represent frequency of ith bat, v_i^t denotes the velocity of ith bat, and X_i^t represents the position of ith bat; $\min\left(X_{Gbest}^t\right)$ and $\max\left(X_{Gbest}^t\right)$ denote minimum and maximum values of sum function associated with X_{Gbest}^t position. $\max\left(X_{Pbest}^t\right)$ represents the maximum value of fitness function associated with personal best position and β denotes a random value between [0, 1].

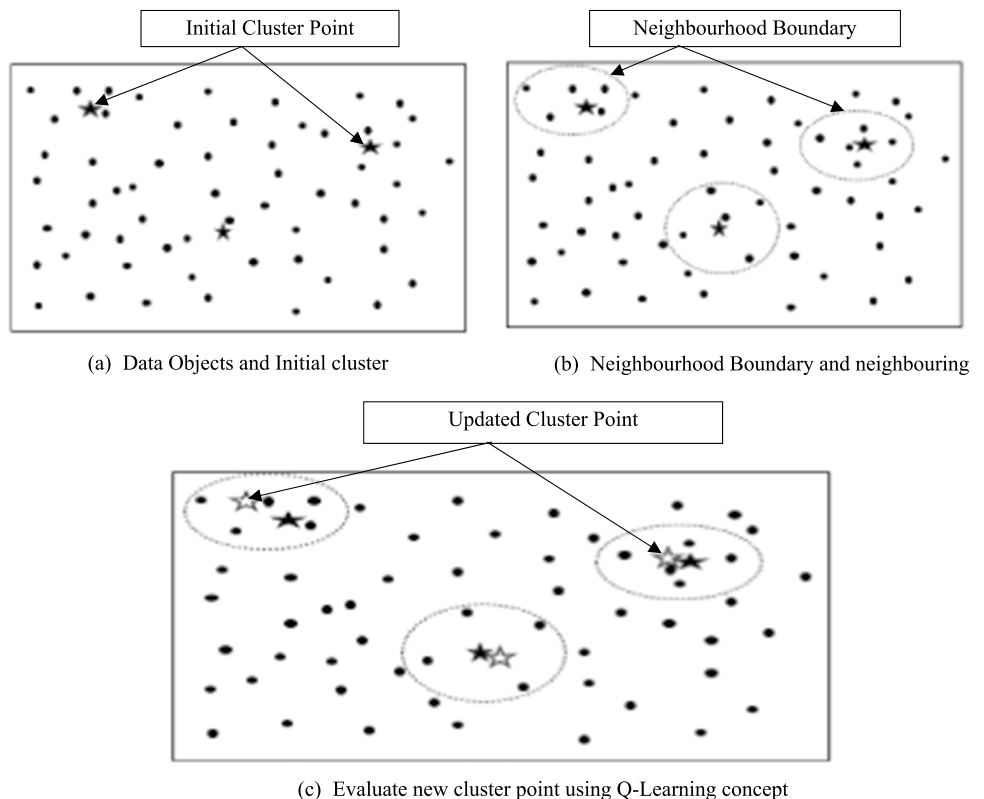## 4.3 Q-learning based neighborhood search mechanism

This subsection describes the Q-learning based neighborhood search for handling the local optima problem of bat algorithm. The performance of clustering algorithms can be degraded due to local optima issue [78]. The various strategies are developed for avoiding the local optima issue in literature [79, 80]. This work also presents a Q-learning based neighborhood search mechanism for effectively handling the local optima issue of clustering algorithms. The proposed Q-learning based concept works into two steps- Identification Step and Evaluation Step. The first step corresponds to determine the neighborhood boundary and neighboring data objects. Whereas, second step corresponds for evaluating the updated position of initial cluster points through Q-learning concept. Fig. 1(a-c) illustrates the Q-learning based neighborhood search mechanism.

4.3.1 Identification step

This step corresponds to determine neighboring data points of initial cluster centers as shown in Fig. 1(a). The Euclidean distance measure is considered for evaluating the neighboring data points. In this work, neighboring data objects are set to 5. Hence, five data objects with minimum Euclidean distance are selected as neighboring data points of given cluster center as shown in Fig. 1(b). Let X_i represents i^{th} cluster center and $X_{i,\ neigh}$ represents set of neighboring data points of ith cluster center. $X_{i,\ neigh}$ is described as $X_{i,\ neigh} = \{X_{i,\ 1}, X_{i,\ 2}, \ldots\ldots X_{i,\ 5}\}$ where neigh = 1 to 5.

Fig. 1 (a-c) Illustrate Q-learning based Neighbourhood Search Mechanism

(a) Data Objects and Initial cluster

(b) Neighbourhood Boundary and neighbouring

(c) Evaluate new cluster point using Q-Learning concept

4.3.2 Evaluation phase

This step corresponds to determine the updated position of initial cluster points as shown in Fig. 1(c). In this step, Q-learning [81] is used instead of arithmetic mean to compute the position of neighborhood data items. The Q-learning algorithm follows simple procedure until good quality solution is received. It consists of initializing the Q-table, choosing an action, performing the action, measuring the rewards and updating the Q-table. The $Q[s, a]$ of neighbouring data points is updated using Eq. 20:

$$Q\varepsilon(s, a) = Q(s, a) + \alpha \left[R(s, a) + \gamma \max Q'(s' + a') - Q(s, a) \right] \tag{20}$$

where $Q''(s, a)$ represents the new Q-value for state (s) and action (a), $Q(s, a)$ gives the current value, α is the learning rate, $R(s, a)$ represents the reward for taking action for a state, γ is the discount rate, and $\max Q'(s' + a')$ represents the maximum expected future reward.

Algorithm 1: Pseudo code of Improved Bat algorithm for cluster analysis
Step 1: upload the data sets and specify number of clusters $K_i \in (i = 1, 2, \ldots, n)$
Step 2: Choose initial cluster locations using enhanced cooperative co-evolution strategy Eqs. (7–10)
Step 3: Initialize pulse rates (r_i) loudness (A_i) and initial velocity (v_i) Eqs. (1–2)
Step 4: Evaluate the objective function using Eq. (13) and assign objects to clusters with minimize objective values
Step 5: Apply neighbourhood strategy to find X_{neigh} position using Eq. (20)
Step 6: Apply elitist strategy, calculate X_{Pbest} and X_{Gbest} Eqs. (11–12)
Step 7: While (t < max number of iterations), do
Step 8: Calculate pulse frequency (f_i^t) and velocity (v_i^t) using Eqs. (16–17)
Step 9: if (rand $> r_i$)
Accept X_n position as solution
Else
Update the position using Eqs. (18–19)
Increase r_i and reduce A_i
Accept the new solution
Step 10: Recalculate the objective function and assign items with minimum objective function values.
Step 11: Update X_{Pbest} and X_{Gbest} using Eqs. (14–15)
Step 12: Apply neighbourhood strategy
Step 13: Memorize the result and store in memory pool t=t+1;
Step 14: Check termination Condition
Step 15: Obtain Global optimal solution

The algorithmic steps of improved bat algorithm for cluster analysis are given in algorithm 1. The working of proposed approach is divided into three broad steps- (i) initialization, (ii) evaluation and assignment, and (3) updation. Fig. 2 illustrates the flow chart of aforementioned approach.

- **Initialization Step**: This step corresponds to initialization of algorithmic parameters of improved bat algorithm. Further, details of datasets are also described such as dimension, number of objects etc. A cooperative co-

evolution (described in subsection 4.1) method is used to select initial centroid. The other parameters like loudness, emission rate and random variable are specified.

- **Evaluation and Assignment Step:** The work of this step is to evaluate objective function and allocate objects to nearest clusters. The Euclidean distance can be acted as objective function and allocates objects to nearest clusters using minimized Euclidean distance. Moreover, a Q-learning based neighborhood search mechanism is incorporated to overcome local optima situation. A limit operator is applied to determine the local optima situation. If, the candidate solution is not improved in predefined limit operator range, then it is assumed that algorithm traps in local optima and neighborhood mechanism can be invoked.

- **Updation Step:** This step corresponds for updating the positions of bats through search mechanism. The emission rate of bat algorithm is compared with random function. If random function provides less value than emission rate, then neighborhood value is accepted. Otherwise, a new value is calculated using parameters variations i.e., loudness and emission rate. If, a termination criterion is met, then algorithm stops its working and final solution is obtained. Otherwise, it repeats phase 2–3.

4.4 Time complexity

The working of proposed algorithm is summarized using three steps- initialization, evaluation and assignment, and updation. The complexity of each step is computed and combined to get the final complexity of the algorithm.

Step 1: The initialization step starts with the selection of initial centroid positions. These positions are evaluated using Eqs. (7–10). In turn, dataset divides into M number of distinct parts and M can be interpreted as number of clusters (K) i.e. M = K. Now, the population of bat algorithm is described in terms of number of bats. Hence, the time required by initialization phase of Improved Bat algorithm is given as O(number of bats × number of attributes). Here, number of bats indicates the number of clusters (K) and number of attributes indicates the dimension (d) i.e. the length of cluster centre O(K × d).

Step 2: The Evaluation and Assignment Phase comprises of objective function, allocate the data respective clusters and fitness function.

a. To compute the objective function and allocate data to respective cluster require $O(n^2 \times K \times d)$ time.
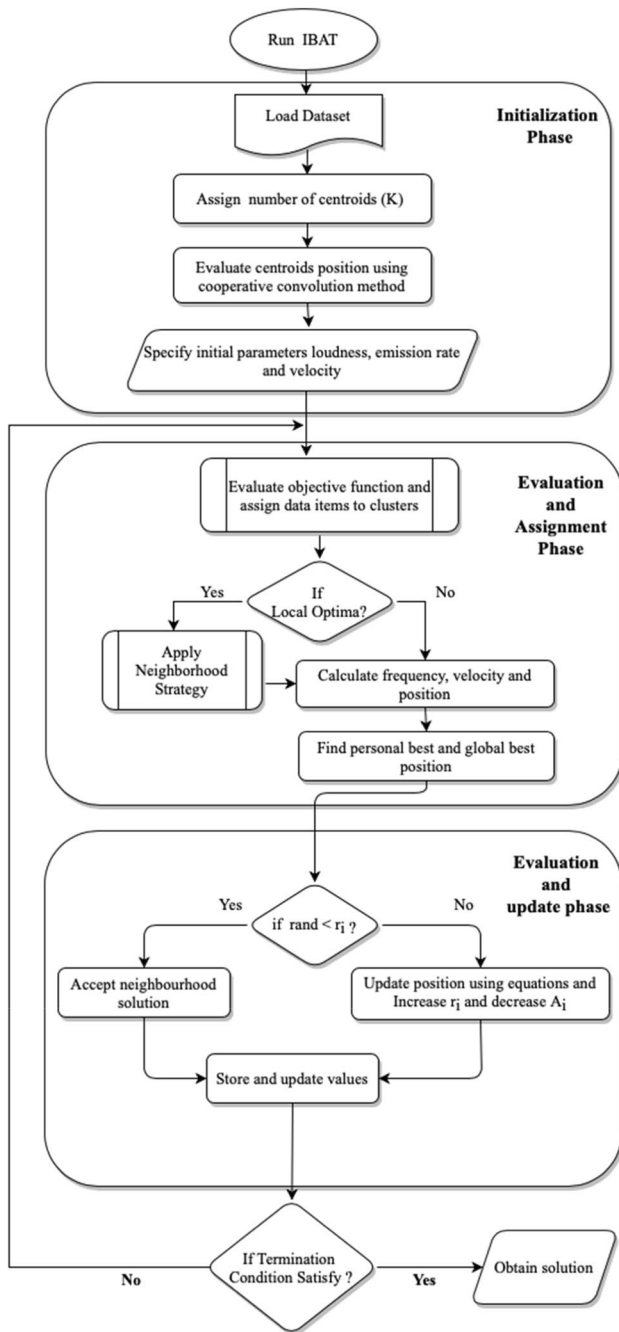
Fig. 2 Flow chart of improved bat algorithm

   b. In worst case scenario, the neighbourhood strategy can require $O(n \times K \times d)$ time.

   c. The fitness function evaluation requires $O(n \times K \times d)$ time.

Hence, the total complexity of Step 2 can be given as $O((n^2 \times K \times d) + (n \times K \times d) + (n \times K \times d)) = O(n^2 \times K \times d)$.

Step 3: In evaluation and update phase, the position of individual bat i.e., cluster centre is updated. This updating requires $O(K \times d)$ time. Further, the above process is repeated up to termination condition i.e., max_iter. Hence, the total time required in this phase is given as $O(K \times d \times max\_iter)$.

The overall complexity of the Improved Bat algorithm is the combination of all above steps and can be given as $O(n^2 \times K \times d \times max\_iter)$. Whereas, the space complexity of the proposed algorithm is O(number of attributes $\times$ number of bats) i. e. $(d \times K)$.

5 Experimental setup and results

This section presents the simulation results of proposed BAT algorithm for solving clustering problems. Several non-healthcare and healthcare datasets are chosen for evaluating the performance of proposed algorithm. Table 2 summarizes the description of these datasets. These datasets are (i) Iris, (ii) Glass, (iii) Wine, (iv) Ionosphere, (v) Control, (vi) Vowel, (vii) Balance, (viii) Crude Oil, (ix) CMC, (x) Liver Disorder (LD), (xi) Wisconsin Breast Cancer (WBC), (xii) Thyroid. These data sets are downloaded from UCI repository and freely available to assess the efficiency of newly proposed clustering algorithm. The user defined parameters of proposed bat algorithm are described as population $= K \times d$, $\alpha = 0.5$, loudness $(A_i) \in (0.1, 0.9)$, initial velocity $= 0.1$, number of partitions $= K$, $max\_iteration = 200$. The average intra cluster distance (IntraCD), standard deviation (SD), accuracy and rand index parameters are used for evaluating the simulation results of proposed bat algorithm. The simulation results of proposed BAT algorithm are compared with several existing meta-heuristic algorithms. The different state of art meta-heuristic algorithms are selected for comparing the simulation results of proposed BAT algorithm. In this work, a total nineteen (19) meta-heuristic algorithms are chosen for comparison purpose. All nineteen meta-heuristic algorithms are divided into three groups (1) standard and well known clustering algorithm, (2) hybrid meta-heuristic clustering algorithm, and (3) recently developed clustering algorithm. In first group, eight standard meta-heuristic algorithms are selected for comparing the clustering results of proposed BAT algorithm which are highly cited in clustering literature. These algorithms are (i) Particle swarm optimization (PSO) [11], (ii) ant colony optimization (ACO) [70], (iii) artificial bee colony (ABC) [49], (iv) differential evolution (DE) [68], (v) genetic algorithm (GA) [78], (vi) big bang- big crunch (BB-BC) [61], (vii) Bat [42], and (viii) K-means [33]. The second group consists of

Table 2 Summary of non-healthcare and healthcare datasets

Sr. No.	Data sets	K	D	N	Description
Non-Healthcare datasets					
1	Iris	3	4	150	Fisher's iris data
2	Glass	6	9	214	Glass identification data
3	Wine	3	13	178	Wine data
4	Ionosphere	2	34	351	Radar observations
5	Control	6	60	600	Synthetic control chart
6	Vowel	6	3	871	Indian Telugu vowel
7	Balance	3	4	625	Balance
8	Crude Oil	3	5	56	Crude Oil data
Healthcare datasets					
9	CMC	3	9	1473	Contraceptive method choice
10	LD	2	6	345	Liver Disorder Data
11	WBC	2	9	683	Wisconsin Breast Cancer
12	Thyroid	3	5	215	Thyroid

six popular hybrid meta-heuristics algorithms that provide state of art clustering results in corresponding literature and also focus on local optima and convergence issues especially of clustering problems. These algorithms are (i) MEBBC [61], (ii) H-KHA [62], (iii) IKH [80], (iv) ICMPKHM [82], (v) PSO-BB-BC [83], and (vi) CBPSO [73]. Third group consists of recently reported meta-heuristic algorithm for solving clustering task. The reason behind the selection of these algorithms is to compare the performance of the proposed IBAT algorithm with recent clustering algorithms. These algorithms are (i) VS [84], (ii) MBOA [85], (iii) WOA [86], (iv) ICSO [74], (v) Chaotic TLBO [75]. The reason behind the selection of these hybrid and recent clustering algorithms is to make fair comparison of the performance of IBAT. The parameters settings of aforementioned algorithms are taken same as reported in corresponding literature. Here, population size of all algorithms is $K \times d$. The maximum iteration is set to 200. The results are evaluated using intra cluster distance, accuracy, rand index and rank measures.

5.1 Experiment 1: benchmark clustering datasets (non-healthcare datasets)

This subsection presents the simulation results of proposed BAT algorithm for non-healthcare datasets. The simulation results of proposed BAT algorithm are compared with three categories of clustering algorithms such as standard/well-known clustering algorithms, hybrid clustering algorithms and recently reported clustering algorithms.

5.1.1 Comparison of simulation results of proposed BAT and standard/well-known clustering algorithms

This subsection discusses the simulation results of proposed BAT and standard/well-known clustering algorithms using non-healthcare datasets. The results are evaluated using intra cluster distance (intra), standard deviation (SD), accuracy, random index and rank measures. Table 3 presents the performance comparison of proposed BAT algorithm and well-known clustering algorithms like K-means, PSO, ACO, ABC, DE, GA, BB-BC, BAT. The results are evaluated in terms of average intra cluster distances (intra), standard deviation (SD) and rank measures. The results show that proposed BAT algorithm gives minimum intra cluster distance values for Iris (9.16E+01), Glass (1.96E+02), Ionosphere (8.01E+02), Vowel (1.49E+05) and balance (5.01E+04) datasets. While DE achieves minimum intra value (1.58E+04) for wine dataset, BB-BC outperforms for Control dataset with intra value as 2.38E+04 and for crude oil dataset, ACO gives the minimum intra value (2.47E+02). Standard deviation is used to assess the efficiency of algorithms. It represents the dispersion of data objects within a cluster. The minimum value of standard deviation parameter tells that data objects are tightly bound with clusters. From simulation results it is also noted that most of aspects the value of standard deviation parameter is minimum for proposed BAT algorithm than rest of algorithms. The results of accuracy parameter of proposed BAT algorithm and other cluster algorithms are illustrated in Table 4. The results are measured in terms of average case values. The proposed BAT algorithm gives the higher accuracy values for iris (93.00), glass (69.17), wine (76.01), ionosphere (71.94), control (75.30), and crude oil (76.64). While in case of vowel dataset, GA have higher accuracy value (84.7) and PSO have higher accuracy value (89.76) for balance dataset than proposed algorithm. Hence, it is concluded that proposed BAT algorithm provides more accurate results in clustering field. The simulation results of proposed BAT algorithm and other well-known clustering algorithms for non-healthcare datasets using rand index measure are given in Table 5. From the results it is observed that proposed BAT gives better results for datasets iris (0.72), glass (0.427), control (0.799) and crude oil (0.074). While it gives promising results for other datasets when compared to well-known clustering algorithms for non-healthcare datasets. Hence, it is stated that proposed BAT algorithm is one of competent algorithm for cluster analysis.

using intra cluster distance parameter

Table 3 Simulation results of proposed BAT and standard clustering algorithms using intra cluster distance (intra) and standard deviation (SD) measures

Datasets	Measure		Standard/ Well-known Clustering Algorithms								
		K-means	PSO	ACO	ABC	DE	GA	BB-BC	BAT	IBAT
Iris	Intra	9.20E+01	9.86E+01	1.01E+02	1.08E+02	1.21E+02	1.25E+02	9.68E+01	1.15E+02	9.16E+01
	SD	2.67E+01	4.67E-01	1.31E+00	3.63E+00	5.23E+00	1.46E+01	2.22E+00	3.76E+01	2.12E+01
	Rank	2	4	5	6	8	9	3	7	1
Glass	Intra	3.79E+02	2.76E+02	2.19E+02	3.29E+02	3.62E+02	2.82E+02	6.64E+02	3.75E+02	1.96E+02
	SD	7.05E+01	8.59E+00	3.36E+00	1.14E+01	1.21E+01	4.14E+00	6.89E+01	4.29E+00	1.98E+00
	Rank	8	3	2	5	6	4	9	7	1
Wine	Intra	1.81E+04	1.64E+04	1.62E+04	1.69E+04	1.58E+04	1.65E+04	1.67E+04	1.71E+04	1.61E+04
	SD	9.06E+02	8.55E+01	3.69E+01	4.74E+02	5.60E+01	7.84E+01	2.88E+00	5.66E+01	3.54E+01
	Rank	9	4	3	7	1	5	6	8	2
Ionosphere	Intra	2.42E+03	1.00E+03	8.16E+02	1.11E+03	1.13E+03	1.00E+03	1.07E+03	1.33E+03	8.01E+02
	SD	4.55E+02	3.34E+02	4.48E+02	2.61E+02	3.17E+02	4.13E+02	2.99E+02	2.34E+02	1.53E+02
	Rank	9	3	2	6	7	4	5	8	1
Control	Intra	1.01E+06	4.18E+04	2.39E+04	5.12E+04	5.23E+04	4.62E+04	2.38E+04	2.68E+04	2.39E+04
	SD	5.05E+03	1.02E+03	1.71E+02	1.32E+03	9.16E+02	1.58E+03	1.09E+02	1.78E+02	1.24E+02
	Rank	9	5	3	7	8	6	1	4	2
Vowel	Intra	1.60E+05	1.58E+05	1.89E+05	1.70E+05	1.81E+05	1.59E+05	1.94E+05	1.96E+05	1.49E+05
	SD	4.52E+03	2.88E+03	2.58E+03	4.64E+03	2.86E+03	3.11E+03	2.44E+04	3.98E+03	1.15E+03
	Rank	4	2	6	5	7	3	8	9	1
Balance	Intra	1.20E+05	6.20E+04	5.94E+04	6.61E+04	6.78E+04	6.91E+04	5.96E+04	6.02E+04	5.01E+04
	SD	9.28E+03	4.01E+03	7.56E+02	6.79E+02	5.25E+03	5.62E+03	3.72E+02	8.26E+02	3.59E+02
	Rank	9	8	2	5	6	7	3	4	1
Crude oil	Intra	2.91E+02	2.86E+02	2.47E+02	2.81E+02	3.69E+02	2.81E+02	2.77E+02	2.89E+02	2.51E+02
	SD	2.63E+01	1.14E+01	7.11E+00	1.09E+01	2.33E+01	8.14E+00	1.17E+02	1.76E+01	1.06E+02
	Rank	9	6	1	5	3	7	4	8	2
Average Rank		7.4	4.4	3	5.8	5.8	5.6	4.9	6.9	1.4

Table 4 Simulation results of proposed BAT and standard clustering algorithms using accuracy measure

Dataset	Measure	Standard/Well-known Clustering Algorithms								
		K-means	PSO	ACO	ABC	DE	GA	BB-BC	BAT	IBAT
Iris	Accuracy	82.33	84.13	72.87	89.03	88.37	78.34	83.25	90.5	93
	Rank	7	5	9	3	4	8	6	2	1
Glass	Accuracy	51.87	53.73	37.36	48.43	48.48	48.97	55.53	48.76	69.17
	Rank	4	3	9	8	7	5	2	6	1
Wine	Accuracy	67.53	67.94	59.21	70.34	71.1	65.73	66.43	65.48	76.01
	Rank	5	4	9	3	2	7	6	8	1
Ionosphere	Accuracy	71.23	64.84	60.7	64.37	63.02	60.1	62.6	60.08	71.94
	Rank	2	3	7	4	5	8	6	9	1
Control	Accuracy	59.72	41.22	39.53	35.56	39.26	46	39.41	66.84	75.3
	Rank	3	5	6	9	8	4	7	2	1
Vowel	Accuracy	51.16	84.04	51.69	56.31	53.41	84.7	84.32	57.21	67.11
	Rank	9	3	8	6	7	1	2	5	4
Balance	Accuracy	84.99	89.76	74.28	76.67	74.96	78.01	79.69	86.75	88.92
	Rank	4	1	9	7	8	6	5	3	2
Crude oil	Accuracy	65.45	76.46	59.12	56.78	66.52	63.19	63.57	63.17	76.64
	Rank	4	2	8	9	3	6	5	7	1
Average Rank		4.8	3.3	8.1	6.1	5.5	5.6	4.9	5.3	1.5

Table 5 Simulation results of proposed BAT and standard clustering algorithms using rand index measure

Dataset	Measure	Standard/ Well-known Clustering Algorithms								
		K-means	PSO	ACO	ABC	DE	GA	BB-BC	BAT	IBAT
Iris	Rand Index	0.016	0.561	0.61	0.73	0.026	0.52	0.701	0.692	0.72
	Rank	9	6	5	2	8	7	3	4	1
Glass	Rand Index	0.425	0.375	0.404	0.258	0.426	0.401	0.392	0.306	0.427
	Rank	5	7	3	9	2	4	6	8	1
Wine	Rand Index	0.841	0.768	0.363	0.372	0.721	0.241	0.305	0.321	0.374
	Rank	1	2	6	5	3	9	8	7	4
Ionosphere	Rand Index	0.318	0.309	0.311	0.327	0.323	0.306	0.3	0.308	0.319
	Rank	4	6	5	1	2	8	9	7	3
Control	Rand Index	0.798	0.782	0.758	0.766	0.723	0.729	0.746	0.763	0.799
	Rank	2	3	6	4	9	8	7	5	1
Vowel	Rand Index	0.841	0.831	0.838	0.856	0.843	0.829	0.844	0.842	0.846
	Rank	6	8	7	1	4	9	3	5	2
Balance	Rand Index	0.546	0.582	0.537	0.543	0.541	0.521	0.525	0.528	0.574
	Rank	3	1	6	4	5	9	8	7	2
Crude oil	Rand Index	0.05	0.069	0.073	0.072	0.036	0.068	0.071	0.07	0.074
	Rank	8	7	2	3	9	6	4	5	1
Average Rank		4.8	5	5	3.6	5.3	7.5	6	6	1.9

The convergence behaviour of proposed BAT, BAT, BB-BC, GA, DE, ABC, ACO, PSO and K-means clustering algorithm is shown in Fig. 3(a-h). In this graphical illustration, X-axis labels the number of iteration and Y-axis labels the intra-cluster distance. It is observed that proposed BAT algorithm converges on minimum values accept the balance and control dataset. Although in most of aspect, the proposed algorithm provides better convergence rate. Hence, it is stated that the proposed BAT outperforms than other well-known clustering algorithms.

5.1.2 Comparison of simulation results of proposed BAT and existing hybrid clustering algorithms

This subsection discusses about the proposed BAT simulation results for benchmark clustering datasets and compared to existing hybridized clustering algorithms. Furthermore, the performance of proposed algorithm is compared with six hybridized clustering algorithms. Table 6 demonstrates the simulation results of H-KHA, MEBBC, IKH, ICMPKHM, PSO-BB-BC, CBPSO and proposed BAT algorithm using average intra cluster distance (intra) and standard deviation (SD). It is observed that proposed BAT algorithm obtains minimum intra cluster distance for iris (9.16E+01), glass (1.96E+02), wine (1.61E+04), ionosphere (8.01E+02), control (2.39E+04), balance (5.01E+04) and crude oil (2.51E+02). While in vowel dataset, ICMPKHM has minimum intra cluster value (1.47E+05) than proposed algorithm. As well as, the values of standard deviation are minimum for proposed BAT algorithms for the dataset's iris (2.12E+01), wine (3.54E+01). It gives the favourable

results in most of cases when compared to other hybridized clustering algorithms. The results of accuracy parameter of proposed BAT algorithm and other hybridized clustering algorithms are demonstrated in Table 7. It is noticed that proposed BAT algorithm provides more accurate results for iris (93.00), wine (76.01), ionosphere (71.94), control (75.30), vowel (67.11), and crude oil (76.64). For glass and balance datasets, PSO-BB-BC performs better with accuracy value as 69.52 and 89.21, after that proposed BAT gives higher accuracy rate as 69.17 and 88.92 respectively. Moreover, rand index is also computed to prove its effectiveness in clustering field. Table 8 demonstrates the simulation results of proposed BAT algorithm and other hybridized clustering algorithms using rand index parameter for benchmark clustering datasets. The proposed BAT algorithm obtains better results for wine (0.374), ionosphere (0.319), glass (0.427), control (0.799), balance (0.574) and crude oil (0.074) datasets for rand index measure as compared to hybridized variants of clustering algorithms. While, K-KHA achieves better rand index (0.734) for iris, PSO-BB-BC (0.852) for vowel dataset. From the results it can be stated that proposed BAT algorithm is competent with other hybridized variants of clustering techniques over benchmark clustering datasets.

5.1.3 Comparison of simulation results of proposed BAT and recently reported clustering algorithms

The performance of proposed BAT algorithm is also compared with recent clustering algorithms. Table 9 demonstrates the simulation results of VS, MBOA, WOA, ICSO, Chaotic TLBO and proposed BAT algorithm using average
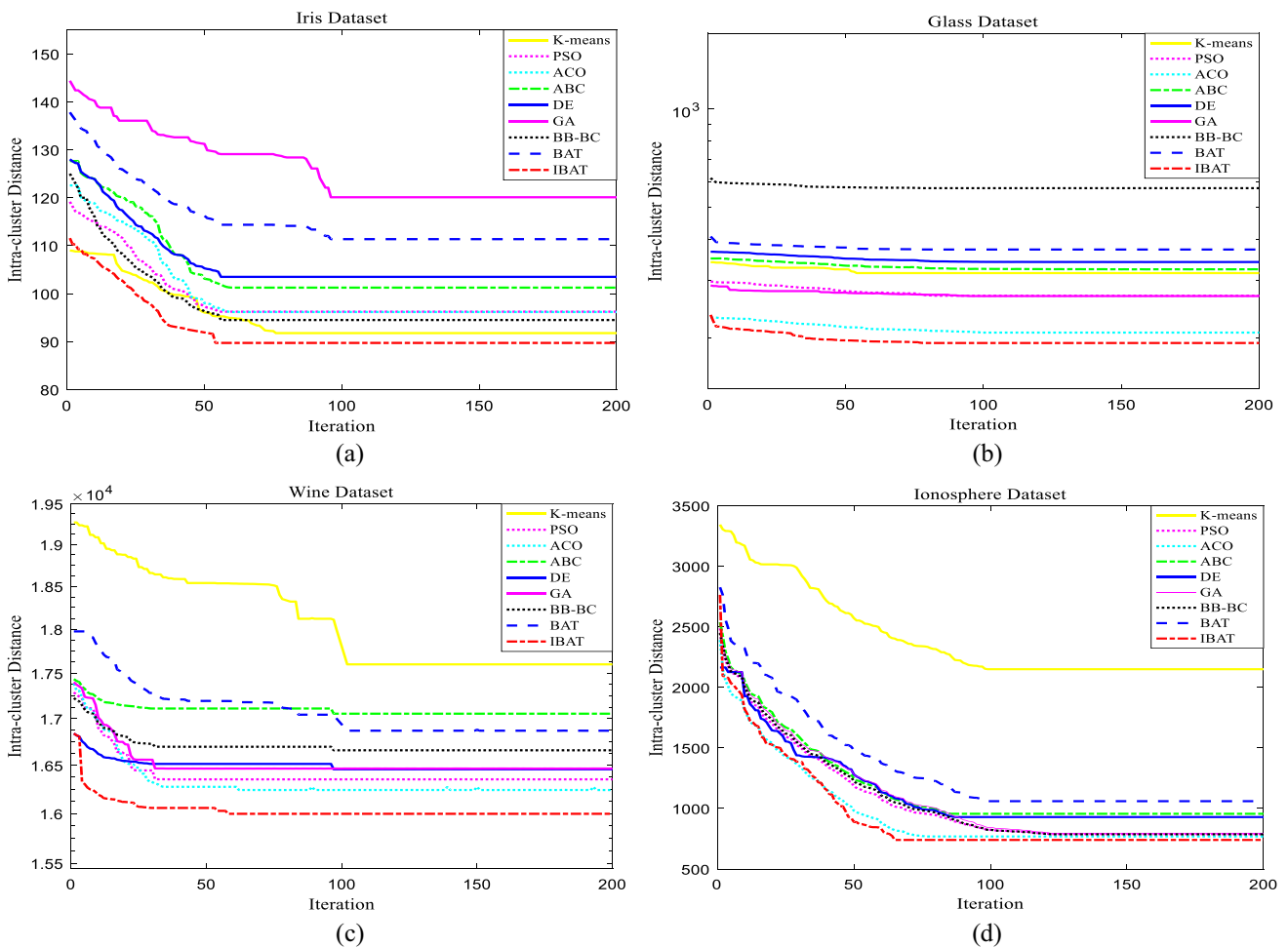
Fig. 3 (**a**-**h**) Convergence behaviour of IBAT, BAT, BB-BC, GA, DE, ABC, ACO, PSO and K-means algorithm

intra cluster distance parameter and standard deviation. It is observed that proposed BAT algorithm obtains minimum intra cluster distance for iris (9.16E+01), glass (1.96E+02), wine (1.61E+04), ionosphere (8.01E+02), control (2.39E+04), balance (5.01E+04) and crude oil (2.51E+02). The values of standard deviation are minimum for proposed BAT algorithms for the dataset's iris (2.12E+01), glass (1.98E+00), ionosphere (1.53E+01), control (3.62E+01), vowel (1.15E+02) and crude oil (1.06E+02). Whereas for the datasets wine (3.54E+01) and balance (3.59E+02), proposed algorithm is the second most after Chaotic TLBO. From results, it is concluded that proposed BAT is competent and outperforms in most of cases when compared to other hybridized clustering algorithms. Table 10 demonstrates the results of accuracy parameter of proposed BAT algorithm and recent clustering algorithms for benchmark clustering datasets. It is noticed that proposed BAT algorithm provides more accurate results for wine (76.01), ionosphere (71.94), control (75.30), vowel (67.11), balance (88.92) and crude oil (76.64). Except for iris and glass dataset in which Chaotic TLBO performs better with accuracy

value as 91.19 and 69.52, after that proposed BAT gives accuracy value as 93.00 and 69.17. Additionally, rand index is also calculated to verify its efficacy in clustering field. Table 11 illustrates the simulation results of proposed BAT algorithm and recent clustering algorithms using rand index parameter for benchmark clustering datasets. The proposed BAT algorithm obtains better results for most of datasets that is iris (0.72), glass (0.427), wine (0.374), control (0.799), vowel (0.846), and balance (0.534) for rand index measure as compared to recent clustering algorithms. Whereas, for ionosphere, ICSO achieves better rand index (0.319) and MBOA achieves higher rand index (0.078) for crude oil dataset. The results show that proposed BAT outperforms and is proficient as compared to recent clustering algorithms.

5.2 Experiment 2: healthcare datasets

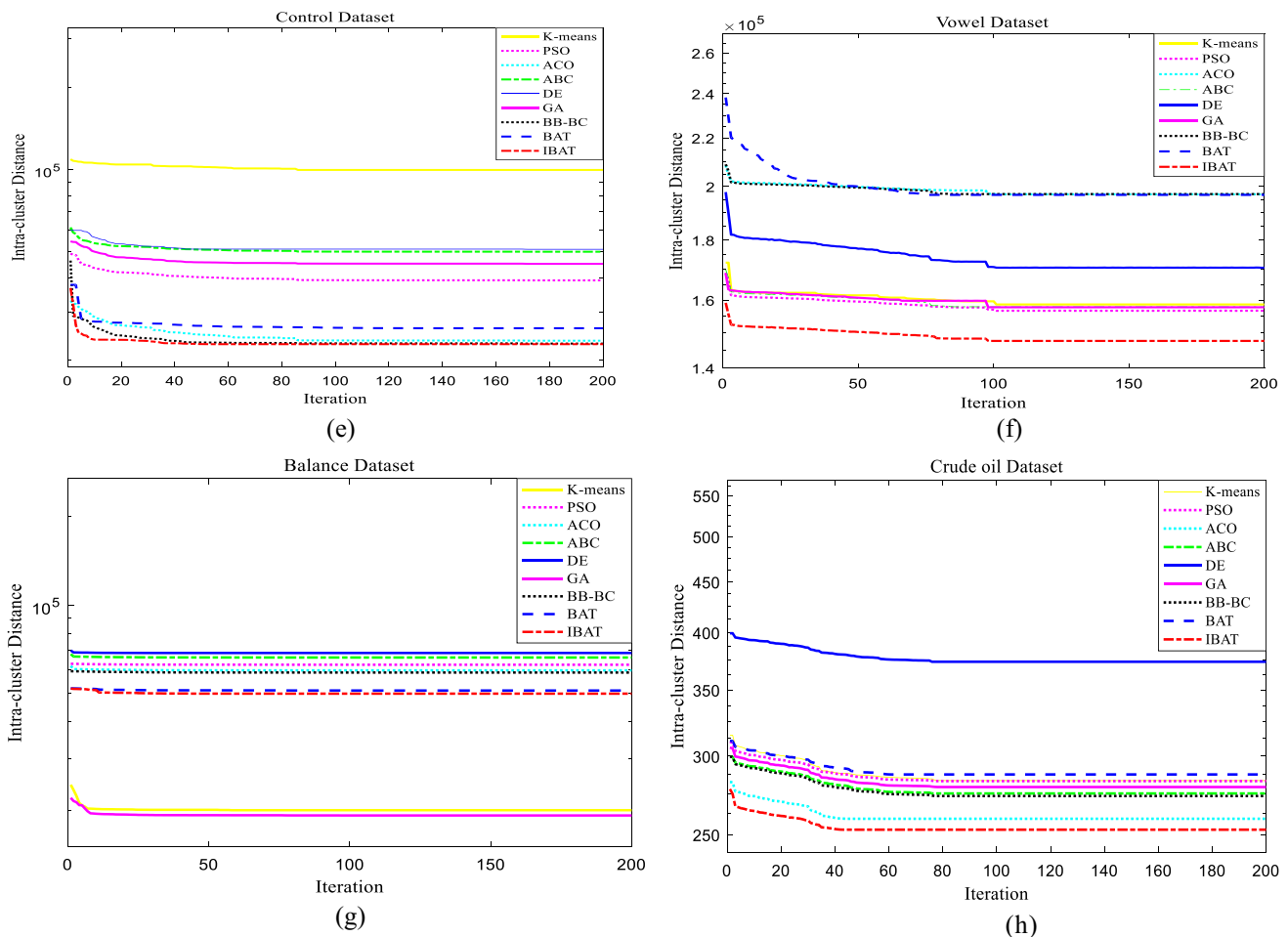This subsection presents the simulation results of proposed BAT clustering algorithm for healthcare datasets.

Fig. 3 (continued)

5.2.1 Comparison of simulation results of proposed BAT and standard/well-known clustering algorithms

The performance comparison of proposed BAT algorithm with K-means, PSO, ACO, ABC, DE, GA, BB-BC, and BAT algorithms are presented in Table 12. The results are evaluated in terms of average intra cluster distance (intra), standard deviation (SD) and rank. Four healthcare datasets are considered to test and compare the performance of proposed BAT with well-known clustering algorithms. Simulation results showed that proposed BAT algorithm obtains minimum intra cluster distance values for all the considered healthcare datasets that is CMC (5.52E+03), LD (2.31E+02), WBC (2.89E+03), and thyroid (2.51E+02) as compared to well-known clustering algorithms. The standard deviation is minimum value computed for assessing the efficiency of algorithms. It represents dispersion of data objects within a cluster. It is also analysed that in most of aspects standard deviation parameter is minimum for proposed BAT algorithm than rest of algorithms. The results of

accuracy parameter of proposed BAT algorithm and other well-known clustering algorithms are illustrated in Table 13. Results show that proposed BAT algorithm gives higher accuracy for CMC with value 48.21, WBC as 96.61 and Thyroid as 71.98. In case of LD dataset, PSO gives better results with accuracy value as 54.05. Even then the proposed algorithm with accuracy value as 54.02 outperforms than rest of the well-known clustering algorithms in case of LD dataset. So, it is concluded that proposed BAT algorithm gives more accurate results in clustering field for considered healthcare datasets. Table 14 presents the simulation results of proposed BAT algorithm and other well-known clustering algorithms using rand index measure for healthcare datasets. It is seen from the results that proposed BAT algorithm obtains better results as compared to other clustering algorithms for CMC, LD, WBC and thyroid datasets with values 0.28, 0.492, 0.276 and 0.383 respectively. Hence, proposed BAT algorithm is considered to be one of proficient algorithm for cluster analysis.

Table 6 Simulation results of proposed BAT and hybrid clustering algorithms using intra cluster distance (intra) and standard deviation (SD) measures

Datasets	Measure	Hybrid Clustering Algorithms						
		H-KHA	MEBBC	IKH	ICMPKHM	PSO-BB-BC	CBPSO	IBAT
Iris	Intra	9.65E+01	9.68E+01	9.71E+01	9.58E+01	9.60E+01	9.69E+01	9.16E+01
	SD	4.45E+00	2.10E+00	9.80E+00	1.06E+00	4.16E+02	3.18E+00	2.12E+01
	Rank	4	5	7	2	3	6	1
Glass	Intra	2.16E+02	2.27E+02	2.23E+02	1.99E+02	2.19E+02	2.13E+02	1.96E+02
	SD	1.07+01	1.26E+01	2.66E+00	1.27E-01	2.01E+00	2.09E+00	1.98E+00
	Rank	4	7	6	2	5	3	1
Wine	Intra	1.66E+04	1.68E+04	1.65E+04	1.67E+04	1.63E+04	1.64E+04	1.61E+04
	SD	231	6.90E+01	7.07E+01	1.08E+02	1.56E+02	9.41E+01	3.54E+01
	Rank	5	7	4	6	2	3	1
Ionosphere	Intra	8.95E+02	2.04E+03	9.34E+02	2.18E+03	2.45E+03	8.08E+02	8.01E+02
	SD	4.03E+02	3.67E+02	3.93E+02	3.69E+02	3.62+02	1.35E+01	1.53E+02
	Rank	3	4	7	5	6	2	1
Control	Intra	2.51E+04	3.01E+04	2.49E+04	2.92E+04	5.04E+04	2.54E+04	2.39E+04
	SD	1.28E+02	1.34E+03	1.68E+02	1.99E+02	7.08E+02	2.88E+02	1.24E+02
	Rank	3	6	2	5	7	4	1
Vowel	Intra	3.52E+05	1.49E+05	1.56E+05	1.47E+05	1.52E+05	1.50E+05	1.51E+05
	SD	1.25E+03	4.52E+02	1.74E+03	1.23E+03	7.69E+02	7.76E+02	1.15E+03
	Rank	7	2	6	1	5	3	4
Balance	Intra	6.78E+04	5.83E+04	6.01E+04	6.26E+04	6.19E+04	6.20E+04	5.01E+04
	SD	9.32E+02	3.74E+03	8.12E+03	3.32E+03	7.32E+02	2.80E+02	3.59E+02
	Rank	7	2	3	6	4	5	1
Crude oil	Intra	2.82E+02	2.76E+02	2.80E+02	2.81E+02	2.88E+02	2.78E+02	2.51E+02
	SD	1.92E+01	1.32E+01	2.00E+01	1.81E+01	7.78E+00	1.07E-01	1.06E+02
	Rank	6	2	4	5	7	3	1
Average Rank		4.9	4.4	4.9	4	4.9	3.6	1.4

Table 7 Simulation results of proposed BAT and hybrid clustering algorithms using accuracy measure

Dataset	Measure	Hybrid Clustering Algorithms						
		H-KHA	MEBBC	IKH	ICMPKHM	PSO-BB-BC	CBPSO	IBAT
Iris	Accuracy	89.24	90.02	89.87	92.44	90.52	90.79	93
	Rank	7	5	6	2	4	3	1
Glass	Accuracy	58.89	58.73	68.39	69.06	69.52	51.92	69.17
	Rank	5	6	4	3	1	7	2
Wine	Accuracy	75	72.63	72.37	72.88	73.58	71.31	76.01
	Rank	2	5	6	4	3	7	1
Ionosphere	Accuracy	67.92	63.08	58.93	60.23	68.19	69.11	71.94
	Rank	4	5	7	6	3	2	1
Control	Accuracy	74.58	72.32	69.56	71.87	72.09	67.06	75.3
	Rank	2	3	6	5	4	7	1
Vowel	Accuracy	66.98	59.21	61.76	59.65	60.18	51.72	67.11
	Rank	2	6	3	5	4	8	1
Balance	Accuracy	75.42	68.42	76.42	77.42	89.21	76.62	88.92
	Rank	6	7	5	3	1	4	2
Crude oil	Accuracy	65.77	68.21	69.32	66.87	55.02	74.57	76.64
	Rank	6	4	3	5	7	2	1
Average Rank		4.3	5.1	5	4.1	3.4	5	1.3

Table 8 Simulation results of proposed BAT and hybrid clustering algorithms using rand index measure

Dataset	Measure	Hybrid Clustering Algorithms						
		H-KHA	MEBBC	IKH	ICMPKHM	PSO-BB-BC	CBPSO	IBAT
Iris	Rand Index	0.734	0.712	0.719	0.716	0.726	0.729	0.720
	Rank	1	7	5	6	3	2	4
Glass	Rand Index	0.366	0.253	0.306	0.262	0.419	0.255	0.427
	Rank	3	7	4	5	2	6	1
Wine	Rand Index	0.369	0.368	0.371	0.373	0.370	0.372	0.374
	Rank	6	7	4	2	5	3	1
Ionosphere	Rand Index	0.313	0.312	0.314	0.311	0.310	0.308	0.319
	Rank	2	4	3	5	6	7	1
Control	Rand Index	0.749	0.783	0.729	0.745	0.739	0.741	0.799
	Rank	3	2	7	4	6	5	1
Vowel	Rand Index	0.839	0.813	0.848	0.824	0.852	0.828	0.846
	Rank	4	7	2	6	1	5	2
Balance	Rand Index	0.53	0.549	0.553	0.529	0.565	0.536	0.574
	Rank	7	4	3	6	2	5	1
Crude oil	Rand Index	0.069	0.071	0.047	0.038	0.061	0.073	0.074
	Rank	4	3	6	7	5	2	1
Average Rank		3.8	5.1	4.3	5.1	3.8	4.4	1.8

using intra cluster distance parameter for healthcare datasets

Figure 4(a-d) show the convergence behavior of IBAT and well-known clustering algorithms (BAT, BB-BC, GA, DE, ABC, ACO, PSO and K-means). In the graphical illustration X-axis labels number of iterations and Y-axis labels the intra-cluster distance. It is observed that IBAT algorithm converges on minimum values for all the considered healthcare datasets. The proposed algorithm provides better convergence rate in most of the cases. It is stated that the IBAT outperforms than other clustering algorithms for the considered healthcare datasets.

5.2.2 Comparison of simulation results of proposed BAT and existing hybrid clustering algorithms

This subsection presents the performance of proposed algorithm compared with six hybridized clustering algorithms. Table 15 demonstrates the simulation results of H-KHA, MEBBC, IKH, ICMPKHM, PSO-BB-BC, CBPSO and proposed BAT algorithm using average intra cluster distance (intra), standard deviation (SD) and rank measures. From the results, it is observed that proposed BAT algorithm attains minimum intra cluster distance in all the considered healthcare datasets with intra cluster distance value for CMC as 5.52E+03, LD as 2.31E+02, WBC as 2.89E+03 and thyroid as 2.51E+02 as compared to other hybridized clustering algorithms. In vowel dataset H-KHA has minimum intra cluster value than proposed algorithm. The proposed BAT algorithm also gives minimum value of standard deviation for most of the cases in comparison to hybridized algorithm

on considered healthcare datasets. Table 16 demonstrates the results of accuracy parameter as average case of proposed BAT algorithm and other hybridized clustering algorithms for healthcare datasets. It is perceived that proposed BAT algorithm provides more accurate results as compared to other hybridized clustering algorithms. The proposed BAT gives higher accuracy values for CMC as 48.21, LD as 54.02. While there is marginal difference for accuracy values of proposed BAT over WBC (IBAT = 96.61, CBPSO = 96.89) and thyroid (IBAT = 71.98, CBPSO = 72.21) as compared to CBPSO but then also the values are higher than CBPSO and other hybridized clustering algorithms. It is stated that proposed BAT provides more accurate results compared with other hybridized clustering algorithms. Rand index is also computed for healthcare datasets to prove its effectiveness in clustering. Table 17 shows the simulation results of proposed BAT algorithm and other hybridized clustering algorithms for healthcare datasets using rand index measure. It is noticed that proposed BAT algorithm gives better results for rand index on CMC with value 0.280, WBC with value 0.257 and thyroid with 0.383. While proposed BAT gives identical rand index value compared with MEBBC for LD (0.496). Thus, it is indicated that proposed BAT obtains better results than other hybridized variants of clustering algorithms for considered healthcare datasets.

5.2.3 Comparison of simulation results of proposed BAT and recently reported clustering algorithms

This subsection presents the performance comparison of proposed algorithm with recent clustering algorithms.

Table 9 Simulation results of proposed BAT and recent clustering algorithms using intra clustering distance (intra) and standard (SD) measures

Datasets	Measure	Recent Clustering Algorithms					
		VS	MBOA	WOA	ICSO	Chaotic TLBO	IBAT
Iris	Intra	9.89E+01	9.83E+01	9.84E+01	9.57E+01	9.69E+01	9.16E+01
	SD	1.21E+00	1.24E+00	1.06E+00	1.92E+00	2.88E+00	2.12E+01
	Rank	6	4	5	2	3	1
Glass	Intra	2.34E+02	2.31E+02	2.18E+02	2.26E+02	2.38E+02	1.96E+02
	SD	1.01E+01	1.27E+01	1.37E+01	1.10E+01	9.09E+00	1.98E+00
	Rank	5	4	2	3	6	1
Wine	Intra	1.83E+04	1.91E+04	1.84E+04	1.69E+04	1.68E+04	1.61E+04
	SD	6.90E+01	7.07E+01	4.83E+01	3.26E+01	2.94E+01	3.54E+01
	Rank	4	6	5	3	2	1
Ionosphere	Intra	2.25E+03	2.93E+03	2.45E+03	1.45E+03	1.11E+03	8.01E+02
	SD	3.67E+01	4.39E+01	3.69E+01	36.2	2.35E+01	1.53E+02
	Rank	4	6	5	3	2	1
Control	Intra	3.15E+04	3.04E+04	2.99E+04	2.51E+04	2.55E+04	2.39E+04
	SD	1.17E+02	9.68E+02	8.47E+01	7.08E+01	4.83E+01	1.24E+02
	Rank	6	5	4	2	3	1
Vowel	Intra	1.59E+05	1.61E+05	1.57E+05	1.49E+05	1.52E+05	1.51E+05
	SD	4.52E+02	2.94E+02	2.54E+02	1.77E+02	1.78E+02	1.15E+03
	Rank	5	6	4	1	3	2
Balance	Intra	5.74E+04	5.96E+04	6.06E+04	5.39E+04	5.36E+04	5.01E+04
	SD	2.74E+02	1.81E+02	1.83E+02	2.17E+02	1.29E+02	3.59E+02
	Rank	4	5	6	3	2	1
Crude oil	Intra	2.81E+02	2.86E+02	2.83E+02	2.64E+02	2.59E+02	2.51E+02
	SD	1.54E+01	2.08E+01	1.62E+01	1.48E+01	1.37E+01	1.06E+02
	Rank	4	6	5	3	2	1
Average Rank		4.8	5.3	4.5	2.5	2.9	1.1

Table 10 Simulation results of proposed BAT and recent clustering algorithms using accuracy measure

Dataset	Measure	Recent Clustering Algorithms					
		VS	MBOA	WOA	ICSO	Chaotic TLBO	IBAT
Iris	Accuracy	94.14	95.43	94.56	91.35	91.19	93.00
	Rank	4	6	5	2	1	3
Glass	Accuracy	58.89	58.73	68.39	69.06	69.52	69.17
	Rank	5	6	4	3	1	2
Wine	Accuracy	69.74	70.31	68.46	73.24	72.53	76.01
	Rank	5	4	6	2	3	1
Ionosphere	Accuracy	63.26	64.58	61.12	68.73	69.41	71.94
	Rank	5	4	6	3	2	1
Control	Accuracy	57.81	60.32	62.04	70.25	72.83	75.3
	Rank	6	5	4	3	2	1
Vowel	Accuracy	64.96	56.92	58.74	65.28	64.91	67.11
	Rank	3	6	5	2	4	1
Balance	Accuracy	71.34	71.95	69.27	78.61	81.04	88.92
	Rank	5	4	6	3	2	1
Crude oil	Accuracy	62.34	65.21	63.49	74.58	73.46	76.64
	Rank	6	4	5	2	3	1
Average Rank		4.9	4.9	5.1	2.5	2.3	1.4

Table 11 Simulation results of proposed BAT and recent clustering algorithms using rand index measure

Dataset	Measure	Recent Clustering Algorithms					
		VS	MBOA	WOA	ICSO	Chaotic TLBO	IBAT
Iris	Rand Index	0.681	0.694	0.676	0.709	0.712	0.72
	Rank	5	4	6	3	2	1
Glass	Rand Index	0.368	0.382	0.374	0.411	0.406	0.427
	Rank	6	4	5	3	2	1
Wine	Rand Index	0.319	0.332	0.341	0.359	0.363	0.374
	Rank	6	5	4	3	2	1
Ionosphere	Rand Index	0.314	0.311	0.309	0.321	0.315	0.319
	Rank	4	5	6	1	3	2
Control	Rand Index	0.736	0.748	0.746	0.781	0.773	0.799
	Rank	6	5	4	2	3	1
Vowel	Rand Index	0.826	0.831	0.823	0.836	0.835	0.846
	Rank	5	4	6	2	3	2
Balance	Rand Index	0.519	0.514	0.52	0.524	0.531	0.574
	Rank	5	6	4	3	2	1
Crude oil	Rand Index	0.07	0.078	0.069	0.073	0.071	0.074
	Rank	5	1	6	3	4	2
Average Rank		5.3	4.3	5.1	2.5	2.6	1.4

Table 12 Simulation results of proposed BAT and standard clustering algorithms using intra cluster distance (intra) and standard deviation (SD) measures

Datasets	Measure	Standard/Well-known Clustering Algorithms								
		K-means	PSO	ACO	ABC	DE	GA	BB-BC	BAT	IBAT
CMC	Intra	5.59E+03	5.85E+03	5.83E+03	5.94E+03	5.95E+03	5.76E+03	5.71E+03	5.79E+03	5.52E+03
	SD	6.76E+00	4.89E+01	1.23E+02	1.31E+02	8.69E+01	5.04E+01	2.86E+01	3.67E+01	5.39E+00
	Rank	2	7	6	8	9	4	3	5	1
LD	Intra	1.17E+04	2.39E+02	2.41E+02	9.85E+03	1.15E+04	5.44E+02	2.32E+02	2.36E+02	2.31E+02
	SD	6.68E+02	2.88E+01	1.64E+01	8.20E+02	2.07E+03	4.18E+01	2.41E+01	1.52E+01	1.67E+01
	Rank	9	4	5	7	8	6	2	3	1
WBC	Intra	1.93E+04	4.26E+03	3.37E+03	3.50E+03	3.73E+03	3.00E+03	2.96E+03	3.06E+03	2.89E+03
	SD	5.14E-12	2.08E+02	4.17E+01	2.12E+02	1.84E+02	2.25E+02	5.57E+02	1.98E+02	1.33E+02
	Rank	9	8	5	6	7	4	2	3	1
Thyroid	Intra	2.39E+03	1.11E+04	1.99E+03	1.98E+03	2.96E+02	1.22E+04	1.94E+03	3.85E+02	2.51E+02
	SD	2.46E+02	2.71E+01	3.09E+01	2.23E+02	2.06E+01	3.26E+01	1.95E+02	2.28E+01	1.32E+01
	Rank	7	8	6	5	3	9	4	2	1
Average Rank		6.8	6.8	5.5	6.5	6.8	5.8	2.8	3.3	1

Table 18 demonstrates the simulation results of VS, MBOA, WOA, ICSO, Chaotic TLBO and proposed BAT algorithm using average intra cluster distance (intra), standard deviation (SD) and rank measures for healthcare datasets. From the results, it is witnessed that proposed BAT algorithm attains minimum intra cluster distance in healthcare datasets with intra cluster distance value for LD as 2.31E+02, WBC as 2.89E+03 and thyroid as 2.51E+02 as compared to recent clustering algorithms. For CMC dataset, it is seen that MBOA gives the minimum intra cluster distance value as 5.21E+03 than proposed algorithm. Also, the proposed BAT

algorithm also gives minimum value of standard deviation for almost all the considered healthcare datasets except thyroid in which ICSO give minimum standard deviation value as 1.16E+01 and then comes proposed BAT with value 1.32E+01 as compared to recent clustering algorithms. The results of accuracy parameter as average case of proposed BAT algorithm and recent clustering algorithms for healthcare datasets are illustrated in Table 19. It is noticed that proposed BAT algorithm provides more accurate results as compared to recent clustering algorithms. The proposed BAT gives higher accuracy values for CMC as 48.21, LD

Table 13 Simulation results of proposed BAT and standard clustering algorithms using accuracy measure

Dataset	Measure	Standard/Well-known Clustering Algorithms								
		K-means	PSO	ACO	ABC	DE	GA	BB-BC	BAT	IBAT
CMC	Accuracy	39.69	44.1	36.89	40.06	39.58	43.3	44.67	42.62	48.21
	Rank	7	3	9	6	8	5	2	4	1
LD	Accuracy	52.16	54.05	52.89	49.89	52.01	49.28	50.2	53.07	54.02
	Rank	5	1	4	8	6	9	7	3	2
WBC	Accuracy	96.5	93.62	77.92	95.01	94.68	55.73	57.04	89.21	96.61
	Rank	2	5	7	3	4	9	8	6	1
Thyroid	Accuracy	63.76	68.93	64.87	64.39	65.76	63.2	63.86	63.82	71.98
	Rank	8	2	4	5	3	9	6	7	1
Average Rank		5.5	2.8	6	5.5	5.3	8	5.8	5	1.3

Table 14 Simulation results of proposed BAT and standard clustering algorithms using rand index measure

Dataset	Measure	Standard/Well-known Clustering Algorithms								
		K-means	PSO	ACO	ABC	DE	GA	BB-BC	BAT	IBAT
CMC	Accuracy	0.262	0.234	0.231	0.238	0.224	0.218	0.221	0.236	0.280
	Rank	2	4	6	3	7	9	8	5	1
LD	Accuracy	0.468	0.466	0.475	0.482	0.471	0.465	0.476	0.478	0.492
	Rank	7	9	5	2	6	8	4	3	1
WBC	Accuracy	0.246	0.253	0.256	0.255	0.251	0.249	0.258	0.254	0.276
	Rank	9	6	3	4	7	8	2	5	1
Thyroid	Accuracy	0.31	0.379	0.361	0.359	0.32	0.329	0.365	0.356	0.383
	Rank	9	2	4	5	8	7	3	6	1
Average Rank		6.8	5.3	4.5	3.5	7	8	4.3	4.8	1

as 54.02, WBC as 96.61 and Thyroid as 71.98. From the results, it is indicated that IBAT provides more accurate results as compared to recent clustering algorithms. To prove the effectiveness of proposed algorithm in clustering for healthcare datasets, Rand index is computed. The simulation results of proposed BAT algorithm and recent clustering algorithms using rand index measure for healthcare datasets are shown in Table 20. It is seen that proposed BAT algorithm gives better results for rand index on WBC with value 0.276 and thyroid with 0.383 as compared to recent clustering algorithms. While for CMC dataset ICSO gives the better rand index value 0.283 and then comes proposed BAT with value 0.280 and chaotic TLBO obtains better rand index rate as 0.498. Thus, it is signified that proposed BAT is competent and obtains better results for most of the considered healthcare datasets.

5.3 Statistical test

This subsection describes the statistical test to determine the best performing algorithm among proposed IBAT and other existing clustering algorithms. The statistical tests are used to establish a new algorithm and also compute weather a significant difference is occurred among the performances of the new algorithm and existing algorithms [87–89]. In

this work, Friedman statistical test is considered for identifying the best performing algorithm among all. To perform the statistical test, two hypothesis (H_0 and H_1) are designed at the significance level 0.05. Hypothesis (H_0) corresponds to no significant difference among performances of new algorithm and rest of algorithms. Hypothesis (H_1) corresponds to significant difference among performances of new algorithm and rest of algorithms. If, significant difference is not occurred, then hypothesis (H0) is not rejected and it is said that the proposed algorithm (IBAT) similar performance like other algorithms. Otherwise, hypothesis (H_0) is rejected and hypothesis (H_1) is true and it indicates that there is a significant difference occurs between the performances of newly proposed algorithm and rest of algorithms. So, it can perform better than other existing algorithms and better performing algorithm. Hence, statistical tests are widely adopted for analyzing performance of newly proposed algorithm and statistical test results gives the clear idea about better performing algorithm. In this work, Friedman statistical test is applied for determining the best performing algorithm. In the first step of Friedman test, a rank is assigned to each algorithm with each dataset and furthermore, average ranking is computed for all algorithms using all dataset. The ranking of each algorithm with each dataset is reported in Table 21. The ranking of algorithms is computed using
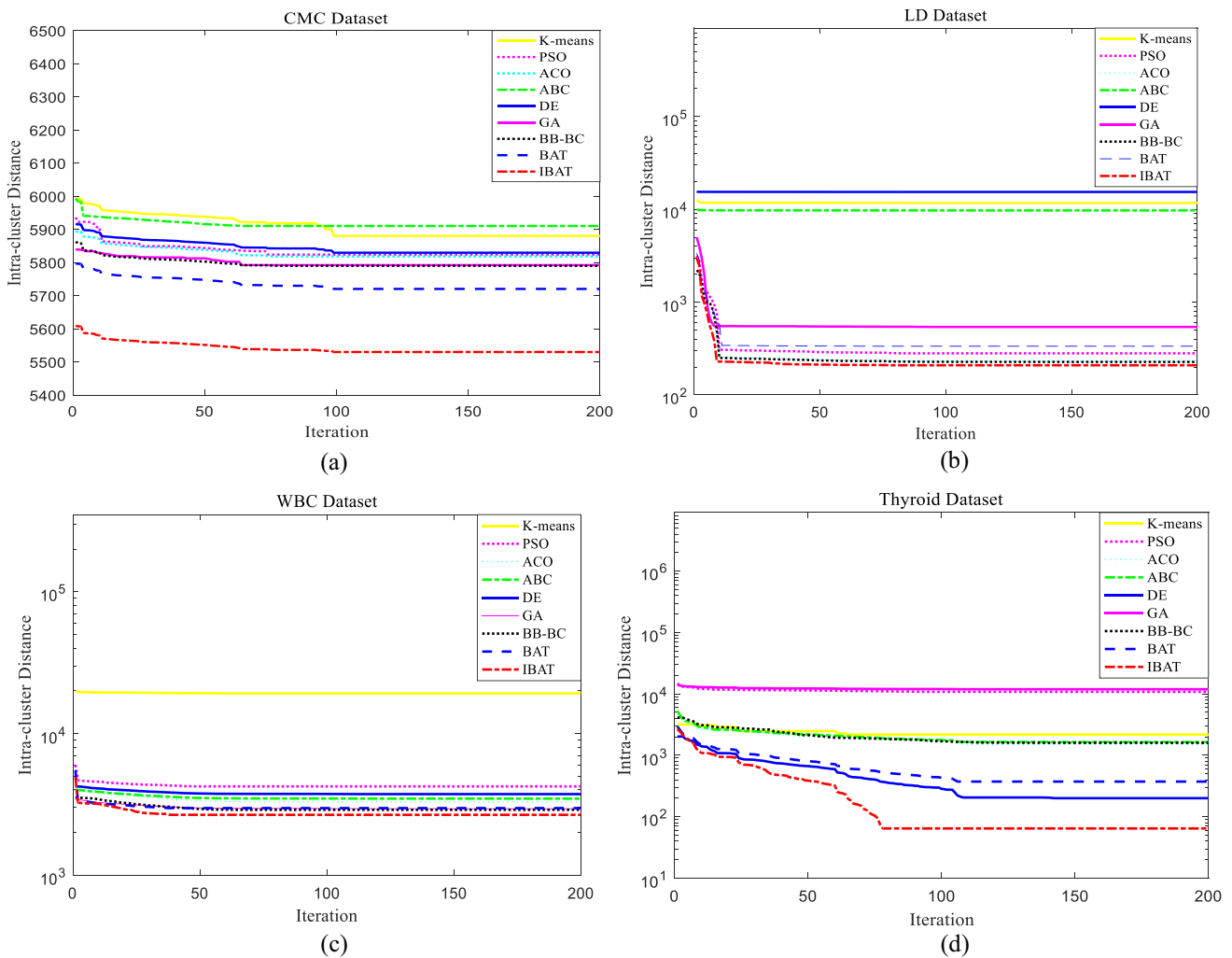
Fig. 4 (**a-d**) Convergence behaviour of IBAT, BAT, BB-BC, GA, DE, ABC, ACO, PSO and K-means algorithm

Table 15 Simulation results of proposed BAT and hybrid clustering algorithms using intra cluster distance (intra) and standard deviation (SD) measures

Datasets	Measure	Hybrid Clustering Algorithms						
		H-KHA	MEBBC	IKH	ICMPKHM	PSO-BB-BC	CBPSO	IBAT
CMC	Intra	5.60E+03	5.53E+03	5.69E+03	5.70E+03	5.57E+03	5.54E+03	5.52E+03
	SD	3.06+02	2.60E-01	7.98E-03	8.26E-01	4.98E-01	3.35E+00	5.39E+00
	Rank	5	2	6	7	4	3	1
LD	Intra	3.14E+03	1.36E+04	3.11E+03	3.09E+03	9.98E+03	1.00E+04	2.31E+02
	SD	2.00E+03	6.52E+02	2.01E+03	1.95E+02	4.39E+02	4.40E+02	1.67E+01
	Rank	4	7	3	2	5	6	1
WBC	Intra	2.99E+03	3.01E+03	2.94E+03	3.02E+03	2.86E+03	2.97E+03	2.89E+03
	SD	1.14E+01	8.15E+00	1.26E-03	3.79E-01	2.29E-02	1.77E+00	1.33E+02
	Rank	5	6	3	7	1	4	2
Thyroid	Intra	1.80E+03	2.56E+02	1.77E+03	1.39E+03	2.82E+02	1.86E+03	2.51E+02
	SD	1.31E+01	1.81E+01	1.39E+01	1.08E+00	7.78E+00	9.97E+00	1.32E+01
	Rank	5	2	4	3	7	6	1
Average Rank		4.8	4.3	4	4.8	4.3	4.8	1.3

Table 16 Simulation results of proposed BAT and hybrid clustering algorithms using accuracy measure

Dataset	Measure	Hybrid Clustering Algorithms						
		H-KHA	MEBBC	IKH	ICMPKHM	PSO-BB-BC	CBPSO	IBAT
CMC	Accuracy	47.45	46.58	46.63	46.69	47.61	39.58	48.21
	Rank	3	6	5	4	2	7	1
LD	Accuracy	51.91	49.86	52.96	52.15	52.17	53.65	54.02
	Rank	4	7	3	5	6	2	1
WBC	Accuracy	61.01	88.05	72.07	46.69	95.5	96.89	96.61
	Rank	6	4	5	7	3	1	2
Thyroid	Accuracy	65.4	65.22	66.91	66.82	56.83	72.21	71.98
	Rank	5	6	3	4	7	1	2
Average Rank		4.5	5.8	4	5	4.5	2.8	1.5

Table 17 Simulation results of proposed BAT and hybrid clustering algorithms using rand index measure

Dataset	Measure	Hybrid Clustering Algorithms						
		H-KHA	MEBBC	IKH	ICMPKHM	PSO-BB-BC	CBPSO	IBAT
CMC	Rand Index	0.261	0.252	0.258	0.249	0.264	0.244	0.280
	Rank	3	5	4	6	2	7	1
LD	Rand Index	0.470	0.496	0.493	0.472	0.465	0.467	0.492
	Rank	5	1	2	4	7	6	3
WBC	Rand Index	0.242	0.243	0.232	0.246	0.239	0.24	0.257
	Rank	4	3	7	2	5	6	1
Thyroid	Rand Index	0.332	0.35	0.371	0.329	0.36	0.382	0.383
	Rank	6	5	3	7	4	2	1
Average Rank		4.5	3.5	4	4.8	4.5	5.3	1.5

Table 18 Simulation results of proposed BAT and recent clustering algorithms using intra cluster distance (intra) and standard deviation (SD) measures

Datasets	Measure	Recent Clustering Algorithms					
		VS	MBOA	WOA	ICSO	Chaotic TLBO	IBAT
CMC	Intra	5.56E+03	5.21E+03	5.69E+03	5.32E+03	5.53E+03	5.52E+03
	SD	6.42E+01	5.60E+01	6.98E+00	7.45E+00	6.38E+00	5.39E+00
	Rank	6	1	5	2	4	3
LD	Intra	1.52E+03	1.32E+03	9.11E+02	4.09E+02	4.98E+02	2.31E+02
	SD	6.32E+01	6.52E+01	5.91E+01	2.95E+01	2.39E+01	1.67E+01
	Rank	6	5	4	2	3	1
WBC	Intra	4.09E+03	3.61E+03	3.76E+03	3.12E+03	2.99E+03	2.89E+03
	SD	4.33E+01	2.87E+01	2.26E+01	1.81E+01	2.35E+01	1.33E+02
	Rank	6	4	5	3	2	1
Thyroid	Intra	1.96E+03	2.16E+03	1.67E+03	9.90E+02	6.79E+02	2.51E+02
	SD	1.81E+01	1.66E+01	2.09E+01	1.16E+01	1.48E+01	1.32E+01
	Rank	6	5	4	3	2	1
Average Rank		6	3.8	4.5	2.5	2.8	1.5

Table 19 Simulation results of proposed BAT and recent clustering algorithms using accuracy measure

Dataset	Measure	Recent Clustering Algorithms					
		VS	MBOA	WOA	ICSO	Chaotic TLBO	IBAT
CMC	Accuracy	41.09	44.23	42.37	46.78	46.54	48.21
	Rank	6	4	5	2	3	1
LD	Accuracy	50.94	50.67	51.42	53.02	53.12	54.02
	Rank	5	6	4	3	2	1
WBC	Accuracy	79.21	85.25	83.43	91.87	92.08	96.61
	Rank	6	4	5	3	2	1
Thyroid	Accuracy	60.41	59.36	62.14	68.24	67.38	71.98
	Rank	5	6	4	2	3	1
Average Rank		5.5	5	4.5	2.5	2.5	1

Table 20 Simulation results of proposed BAT and recent clustering algorithms using rand index measure

Dataset	Measure	Recent Clustering Algorithms					
		VS	MBOA	WOA	ICSO	Chaotic TLBO	IBAT
CMC	Rand Index	0.267	0.272	0.259	0.283	0.276	0.280
	Rank	5	4	6	1	3	2
LD	Rand Index	0.471	0.468	0.462	0.467	0.498	0.492
	Rank	3	4	6	5	1	2
WBC	Rand Index	0.253	0.247	0.242	0.256	0.251	0.257
	Rank	3	5	6	2	4	1
Thyroid	Rand Index	0.348	0.355	0.364	0.369	0.373	0.383
	Rank	6	5	4	3	2	1
Average Rank		4.25	4.5	5.5	2.75	2.5	1.5

accuracy measure and it is also illustrated the average ranking of each technique. It is seen that proposed BAT (IBAT) algorithm obtains 2.1 rank which is highest rank among rest of algorithms and ACO algorithm obtains lower rank (18) among all algorithms. It also noticed that chaotic TLBO achieves second highest rank (5.1) among all algorithms. The statistical results of Friedman test are shown in Table 22. It is observed that the statistical value of Friedman test is 63.6531. The degree of freedom is 19 and the critical value is 301,435 at the significance level of 0.05. The p value computed for Friedman test is 1.01E-06. On the analysis of statistical results, it is concluded that p value is considerably less than critical value. Hence, the hypothesis (H_0) is rejected and a significant difference occurs between the performance of proposed BAT (IBAT) and other existing algorithms. These results certified that proposed algorithm (IBAT) performs better than other existing algorithms and also validates the performance of proposed algorithm as compared to existing clustering algorithms. Moreover, a posthoc test is also conducted to determine the possible grouping of the similar algorithms. The results of posthoc test is presented into Table 23. The symbol "+" indicates

the significant difference between the performances of algorithms, while symbol "-" indicates no significant difference between the performances of algorithms. On the analysis of posthoc test, it is stated that several algorithms exhibit the similar performance and can be clubbed into a single group. In turn, twelve groups are determined that having similar performance according to posthoc test. The description of these groups are listed as Group 1 K-means, PSO, ABC, DE, GA, BB-BC, BAT, MEBBC, IKH, ICMPKHM, CBPSO, VS, MBOA, and WOA algorithm. Group 2 consists of K-Means, K-means, PSO, GA, BB-BC, BAT, H-KHA, MEBBC, IKH, ICMPKHM, PSO-BB-BC, CBPSO, VS, MBOA, WOA, ICSO, and Chaotic TLBO algorithms. The algorithms in group 3 are ACO, ABC, DE, GA, and BAT. Group 4 contains K-means, ACO, ABC, DE, GA, BB-BC, BAT, MEBBC, IKH, VS, MBOA, and WOA algorithms. Group 5 consists of contains K-means, ACO, ABC, DE, GA, BB-BC, BAT, MEBBC, IKH, VS, MBOA, CBPSO, and WOA algorithms. Group 6 consists of K-means, PSO, ACO, ABC, DE, GA, BB-BC, BAT, MEBBC, IKH, VS, MBOA, CBPSO, and WOA algorithms. Group 7 consists of PSO, H-KHA, MEBBC, IKH, ICMPKHM, PSO-BB-BC,

Table 21 Average rank of each clustering algorithm using accuracy measure for non-healthcare datasets

Clustering Algorithms	Dataset								Average Rank
	Iris	Glass	Wine	Ionosphere	Control	Vowel	Balance	Crude	
K-means	18	15	16	2	13	20	5	11	12.5
PSO	16	13	15	8	16	3	1	2	9.3
ACO	20	20	20	16	17	19	16	18	18.3
ABC	14	19	11	10	20	16	11	19	15.0
DE	15	18	10	13	19	17	15	9	14.5
GA	19	16	18	18	15	1	9	15	13.9
BB-BC	17	12	17	14	18	2	7	13	12.5
BAT	10	17	19	19	10	14	4	16	13.6
H-KHA	13	8.5	2	7	2	5	14	10	7.7
MEBBC	11	10.5	6	12	4	12	20	7	10.3
IKH	12	6.5	8	20	8	9	13	6	10.3
ICMPKHM	5	4	5	17	6	11	10	8	8.3
PSO-BB-BC	9	1.5	3	6	5	10	2	20	7.1
CBPSO	8	14	9	4	9	18	12	4	9.8
VS	3	8.5	13	11	14	7	18	17	11.4
MBOA	1	10.5	12	9	12	15	17	12	11.1
WOA	2	6.5	14	15	11	13	19	14	11.8
ICSO	6	5	4	5	7	6	8	3	5.5
Chaotic TLBO	7	1.5	7	3	3	8	6	5	5.1
IBAT	3	3	1	1	1	4	3	1	2.1

Table 22 Statistical results of Friedman test

Method	Statistical Value	p Value	Degree of Freedom (DF)	Critical value	Hypothesis
Friedman Test	63.6531	1.01E-06	19	30.1435	Reject

CBPSO, VS, MBOA, WOA, ICSO, and Chaotic TLBO algorithms. Group 8 consists of K-means, PSO, ABC, DE, GA, BB-BC, BAT, HKA, MEBBC, IKH, VS, MBOA, CBPSO, and WOA algorithms. Group 9 contains K-Means, PSO, BB-BC, H-KHA, MEBBC, IKH, ICMPKHM, PSO-BB-BC, CBPSO, VS, MBOA, WOA, ICSO, and Chaotic TLBO algorithms. Group 10 consists of K-Means, PSO, DE, GA, BB-BC, H-KHA, MEBBC, IKH, ICMPKHM, PSO-BB-BC, CBPSO, VS, MBOA, WOA, ICSO, and Chaotic TLBO algorithms. Group 11 consists of PSO, H-KHA, ICMP-KHM, PSO-BB-BC, CBPSO, ICSO, and Chaotic TLBO algorithms. Group 12 contains only a single algorithm i.e. IBAT. Posthoc test reveals that algorithms within a group exhibits similar performance. It is also seen that proposed BAT (IBAT) algorithm is not presented in any groups except group 12 and it is single algorithm presented in group 12. Hence, it is concluded that proposed BAT algorithm provides significant different performance than other existing algorithms and it is best performing algorithm than rest of algorithms.

6 Conclusion

An improved variant of the bat algorithm is proposed for data clustering task. The proposed algorithm efficiently deals the initial population selection, convergence rate and local optima issues of bat algorithm. To resolve initial population selection problem an enhanced co-operative co-evolution method is developed and integrated into bat algorithm. A neighborhood search-based mechanism is designed to handle local optima condition. Furthermore, an elitist strategy is incorporated in bat algorithm for achieving optimal trade off among search mechanisms. The solution search equations of bat algorithm are also improved for determining optimum solution in last iterations. The performance of proposed BAT algorithm is tested using eight benchmark (non-healthcare) and four healthcare clustering datasets. The simulation results of IBAT algorithm are compared with nineteen meta-heuristic algorithms such as K-means, PSO, ACO, ABC, DE, GA, BB-BC, BAT, and hybridized H-KHA, MEBBC, IKH, ICMPKHM, PSO-BB-BC and CBPSO clustering algorithms; and also, with recent clustering algorithms like VS, MBOA, WOA, ICSO, and Chaotic TLBO. The proposed algorithm achieves better quality clustering results in terms of intra cluster distance, accuracy and rand index with most

Table 23 Results of posthoc test after rejection of hypothesis (H_0)

	KMEAN	PSO	ACO	ABC	DE	GA	BB-BC	BAT	HKHA	MEBBC
K-MEAN		-	+	-	-	-	-	-	+	-
PSO	-		+	+	+	-	-	-	-	-
ACO	+	+		-	-	-	+	-	+	+
ABC	-	+	-		-	-	-	-	+	-
DE	-	+	-	-		-	-	-	+	-
GA	-	-	-	-	-		-	-	+	-
BB-BC	-	-	-	-	-	-		-	+	-
BAT	-	-	-	-	-	-	-		+	-
H-KHA	+	-	+	+	+	+	+	+		-
MEBBC	-	-	+	-	-	-	-	-	-	
IKH	-	-	+	-	-	-	-	-	-	-
ICM-PKHM	-	-	+	+	+	+		+		-
PSO-BB-BC	+	-	+	+	+	+	+	+	-	-
CBP+O	-	-	+	+	-	-	-	-	-	-
VS	-	-	+	-	-	-	-	-	-	-
MBOA	-	-	+	-	-	-	-	-	-	-
WOA	-	-	+	-	-	-	-	-	-	-
ICSO	-	-	-	-	-	-	-	-	-	-
Chaotic TLBO	+	-	+	+	+	+	+	+	-	+
IBAT	+	+	+	+	+	+	+	+	+	+

	-IKH	-ICMPKHM	-PSOBB-BC	-CBPSO	-VS	-MBOA	-WOA	-ICSO	-CTLBO	-IBAT
K-MEAN	-	+	-	-	-	-	-	+	+	+
PSO	-	-	-	-	-	-	-	-	-	+
ACO	+	+	+	+	+	+	+	+	+	+
ABC	-	+	+	+	-	-	-	+	+	+
DE	-	+	+	-	-	-	-	+	+	+
GA	-	+	+	-	-	-	-	+	+	+
BB-BC	-	-	+	-	-	-	-	+	+	+
BAT	-	+	+	-	-	-	-	+	+	+
H-KHA	-	-	-	-	-	-	-	-	-	+
MEBBC	-	-	-	-	-	-	-	+	+	+
IKH	-	-	-	-	-	-	-	+	+	+
ICM-PKHM	-	-	-	-	-	-	-	-	-	+
PSO-BB-BC	-	-		-	-	-	-	-	-	+
CBP+O	-	-	-		-	-	-	-	-	+
VS	-	-	-	-		-	-	+	+	+
MBOA	-	-	-	-	-		-	+	+	+
WOA	-	-	-	-	-	-		+	+	+
ICSO	-	-	-	-	+	+	+		-	+

Table 23 (continued)

	-I K H PK HM	-I C M	-P S O BB-BC	-C B P S O	-V S	-M B O A	-W O A	-I C S O	-C T L B O	-I B A T
Chaotic TLBO	+	-	-	-	+	+	+	-		+
IBAT	+	+	+	+	+	+	+	+	+	

of datasets. Furthermore, Friedman statistical test is also applied to determine the best performing algorithm. The statistical results showed that the hypothesis (H_0) is rejected at confidence level of 0.05. In turn, significant difference is occurred between the performance of proposed IBAT and other clustering algorithms. Hence, it is stated that proposed IBAT algorithm is best performing algorithms than rest of clustering algorithms. Finally, it is also concluded that proposed IBAT algorithm is a robust and an effective algorithm for handling data clustering task.

Abbreviations ABC: Artificial Bee Colony; ACA: Ant Clustering Algorithm; ACDE: Automatic Clustering Differential Evolution; ACO: Ant Colony Optimization; BATC: Bat Algorithm based Clustering; BB-BC: Big Bang–Big Crunch; CABC: Cooperative Artificial Bee Colony; CCSSA: Chaotic Charge System Search Algorithm; CPSO: Cooperative Particle Swarm Optimization; CS: Cuckoo Search; CSO: Cat Swarm Optimization; CSS: Charge System Search; DCPSO: Dynamic Clustering Particle Swarm Optimization; DE: Differential Evolution; FA: Firefly algorithm; FPAC: Flower Pollination Algorithm based Clustering; GA: Genetic Algorithm; GAMS: Genetic Algorithm with Message-based Similarity; GTCSA: Gene Transposon based Clone Selection Algorithm; GWA: Grey Wolf Algorithm; GWO: Grey Wolf Optimizer; HABC: Hybrid Artificial Bee Colony; HBMO: Honey Bee Mating Optimization; KH: Krill Herd; KHM: K-harmonic Means; K-MWO: K-means and Mussels Wandering Optimization; HS: Harmony Search; IBAT: Improved Bat; ICSO: Improved Cat Swarm Optimization; ILS: Iterated Local Search; MCSS: Magnetic Charge System Search; MO: Magnetic Optimization; PSO: Particle Swarm Optimization; SA: Simulated Annealing; TLBO: Teaching learning Based Optimization; TS: Tabu Search; VGA: Variable-string-length Genetic Algorithm; MBOA: Modified Butterfly Optimization Algorithm; WOA: Whale Optimization Algorithm; ICSO: Improved cat swarm optimization; Chaotic TLBO: Chaotic Teaching Learning based optimization; VS: Vortex Search

References

1. Kushwaha N, Pant M, Kant S, Jain VK (2018) Magnetic optimization algorithm for data clustering. Pattern Recogn Lett 115:59–65
2. Kant S, Ansari IA (2016) An improved K means clustering with Atkinson index to classify liver patient dataset. International Journal of System Assurance Engineering and Management 7(1):222–228
3. Aggarwal CC, Reddy CK (2014) Data clustering. Algorithms and applications. Chapman & Hall/CRC Data mining and Knowledge Discovery series, Londra
4. Xu R, Wunsch D (2005) Survey of clustering algorithms. IEEE Trans Neural Netw 16(3):645–678
5. Chang D-X, Zhang X-D, Zheng C-W (2009) A genetic algorithm with gene rearrangement for K-means clustering. Pattern Recogn 42(7):1210–1222
6. Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In Kdd (Vol. 96, No. 34, pp. 226-231)
7. Scheunders P (1997) A genetic c-means clustering algorithm applied to color image quantization. Pattern Recogn 30(6):859–866
8. Gomez-Muñoz VM, Porta-Gándara MA (2002) Local wind patterns for modeling renewable energy systems by means of cluster analysis techniques. Renew Energy 2:171–182
9. Mitra S, Banka H (2006) Multi-objective evolutionary bi clustering of gene expression data. Pattern Recogn 39:2464–2477
10. Nanda SJ, Panda G (2014) A survey on nature inspired metaheuristic algorithms for partitional clustering. Swarm and Evolutionary computation 16:1–18
11. Cura T (2012) A particle swarm optimization approach to clustering. Expert Syst Appl 39(1):1582–1588
12. Jordehi AR (2015) Enhanced leader PSO (ELPSO): a new PSO variant for solving global optimisation problems. Appl Soft Comput 26:401–417
13. Karaboga, D. (2005) An idea based on honey bee swarm for numerical optimization, Erciyes University, Kayseri, Turkey, Technical Report-TR06
14. Karaboga D, Basturk B (2007) A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. J Glob Optim 39(3):459–471
15. Karaboga D, Akay B (2009) A comparative study of artificial bee colony algorithm. Appl Math Comput 214(1):108–132
16. Dorigo M, Birattari M, Stutzle T (2006) Artificial ants as a computational intelligence technique. IEEE Comput Intell Mag 1:28–39
17. Dorigo M, Maniezzo V, Colorni A (1996) Ant system: optimization by a colony of cooperating agents. IEEE Transactions on Systems, man, and cybernetics, Part B: Cybernetics 26(1):29–41
18. Kumar Y, Sahoo G (2014) A charged system search approach for data clustering. Progress in Artificial Intelligence 2(2–3):153–166
19. Hatamlou A (2013) Black hole: A new heuristic optimization approach for data clustering. Inf Sci 222:175–184
20. Erol OK, Eksin I (2006) A new optimization method: big bang–big crunch. Adv Eng Softw 37(2):106–111
21. Jordehi AR (2014) A chaotic-based big bang–big crunch algorithm for solving global optimisation problems. Neural Comput & Applic 25(6):1329–1335
22. Alatas B (2011) ACROA: artificial chemical reaction optimization algorithm for global optimization. Expert Syst Appl 38(10):13170–13180
23. Maulik U, Bandyopadhyay S (2000) Genetic algorithm-based clustering technique. Pattern Recogn 33(9):1455–1465

24. Ergezer M, Simon D, Du D Oppositional biogeography-based optimization. 2009 IEEE international conference on systems, man and cybernetics. IEEE, 2009

25. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69:46–61

26. Wang GG, Deb S, Coelho LDS (2015) Elephant herding optimization. In 2015 3rd International Symposium on Computational and Business Intelligence (ISCBI) (pp. 1–5). IEEE

27. Chu SC, Tsai PW, Pan JS (2006) Cat swarm optimization. In: Pacific Rim international conference on artificial intelligence. Springer, Berlin, Heidelberg, pp 854–858

28. Yazdani M, Jolai F (2016) Lion optimization algorithm (LOA): a nature-inspired metaheuristic algorithm. Journal of computational design and engineering 3(1):24–36

29. Mirjalili S (2016) SCA: a sine cosine algorithm for solving optimization problems. Knowl-Based Syst 96:120–133

30. Salimi H (2015) Stochastic fractal search: a powerful metaheuristic algorithm. Knowl-Based Syst 75:1–18

31. Kaveh A, Dadras A (2017) A novel meta-heuristic optimization algorithm: thermal exchange optimization. Adv Eng Softw 110:69–84

32. Abraham A, Das S, Roy S (2008) Swarm intelligence algorithms for data clustering. In: Soft computing for knowledge discovery and data mining. Springer, Boston, pp 279–313

33. Chowdhury K, Chaudhuri D, Pal AK (2021) An entropy-based initialization method of K-means clustering on the optimal number of clusters. Neural Comput & Applic 33(12):6965–6982

34. Torrente A, Romo J (2021) Initializing k-means clustering by bootstrap and data depth. Journal of Classification 38(2):232–256

35. Ahmadi R, Ekbatanifard G, Bayat P (2021) A Modified Grey Wolf Optimizer Based Data Clustering Algorithm. Appl Artif Intell 35(1):63–79

36. Ghany KKA, AbdelAziz AM, Soliman THA, Sewisy AAEM (2020) A hybrid modified step whale optimization algorithm with tabu search for data clustering. Journal of King Saud University-Computer and Information Sciences

37. Sörensen K (2015) Metaheuristics—the metaphor exposed. Int Trans Oper Res 22(1):3–18

38. Yang X-S A new metaheuristic bat-inspired algorithm. Nature inspired cooperative strategies for optimization (NICSO 2010). Springer, Berlin, Heidelberg, 2010. 65–74

39. Ashish T, Kapil S, Manju B (2018) Parallel bat algorithm-based clustering using mapreduce. In: Networking Communication and Data Knowledge Engineering. Springer, Singapore, pp 73–82

40. Fister I Jr, Fister D, Yang XS (2013) A hybrid Bat algorithm. ELEKTROTEHNIˇSKI VESTNIK 80(1–2):1–7

41. Yilmaz S, Kucuksille EU (2013) Improved bat algorithm (IBA) on continuous optimization problems. Lecture Notes on Software Engineering 1(3):279

42. Senthilnath J, Kulkarni S, Benediktsson JA, Yang XS (2016 Apr) A novel approach for multispectral satellite image classification based on the bat algorithm. IEEE Geosci Remote Sens Lett 13(4):599–603

43. Neelima S, Satyanarayana N, Murthy PK (2018) Minimizing Frequent Itemsets Using Hybrid ABCBAT Algorithm. In: Data Engineering and Intelligent Computing. Springer, Singapore, pp 91–97

44. Aboubi Y, Drias H, Kamel N (2016) BAT-CLARA: BAT-inspired algorithm for Clustering LARge Applications. IFAC-PapersOnLine. 49(12):243–248

45. Fister I, Fong S, Brest J (2014) A novel hybrid self-adaptive bat algorithm. Sci World J 2014:70973

46. Zhao D, He Y (2015) Chaotic binary bat algorithm for analog test point selection. Analog Integr Circ Sig Process 84(2):201–214

47. Rahman MA, Islam MZ (2014) A hybrid clustering technique combining a novel genetic algorithm with K-Means. Knowl-Based Syst 71:345–365

48. Liu R et al (2012) Gene transposon based clone selection algorithm for automatic clustering. Inf Sci 204:1–22

49. Kumar Y, Sahoo G (2017) A two-step artificial bee colony algorithm for clustering. Neural Comput & Applic 28(3):537–551

50. Cao F, Liang J, Jiang G (2009) An initialization method for the K-Means algorithm using neighborhood model. Computers & Mathematics with Applications 58(3):474–483

51. Han XH et al (2017) A novel data clustering algorithm based on modified gravitational search algorithm. Eng Appl Artif Intell 61:1–7

52. Senthilnath J, Omkar SN, Mani V (2011) Clustering using firefly algorithm: performance study. Swarm and Evolutionary Computation 1(3):164–171

53. Erisoglu M, Calis N, Sakallioglu S (2011) A new algorithm for initial cluster centers in k-means algorithm. Pattern Recogn Lett 32(14):1701–1705

54. Kumar Y, Sahoo G (2015) Hybridization of magnetic charge system search and particle swarm optimization for efficient data clustering using neighborhood search strategy. Soft Comput 19(12):3621–3645

55. Zhou Y et al (2017) A simplex method-based social spider optimization algorithm for clustering analysis. Eng Appl Artif Intell 64:67–82

56. Boushaki SI, Kamel N, Bendjeghaba O (2018) A new quantum chaotic cuckoo search algorithm for data clustering. Expert Syst Appl 96:358–372

57. Chang D et al (2012) A genetic clustering algorithm using a message-based similarity measure. Expert Syst Appl 39(2):2194–2202

58. Zhang C, Ouyang D, Ning J (2010) An artificial bee colony approach for clustering. Expert Syst Appl 37(7):4761–4767

59. Taherdangkoo M et al (2013) A robust clustering method based on blind, naked mole-rats (BNMR) algorithm. Swarm and Evolutionary Computation 10:1–11

60. Hatamlou A (2012) In search of optimal centroids on data clustering using a binary search algorithm. Pattern Recogn Lett 33(13):1756–1760

61. Bijari K et al (2018) Memory-enriched big bang–big crunch optimization algorithm for data clustering. Neural Comput & Applic 29(6):111–121

62. Abualigah LM et al (2017) A novel hybridization strategy for krill herd algorithm applied to clustering techniques. Appl Soft Comput 60:423–435

63. Pakrashi A, Chaudhuri BB (2016) A Kalman filtering induced heuristic optimization based partitional data clustering. Inf Sci 369:704–717

64. Kang Q et al (2016) A weight-incorporated similarity-based clustering ensemble method based on swarm intelligence. Knowl-Based Syst 104:156–164

65. Wang R et al (2016) Flower pollination algorithm with bee pollinator for cluster analysis. Inf Process Lett 116.1:1–14

66. Hatamlou A, Hatamlou M (2013) PSOHS: an efficient two-stage approach for data clustering. Memetic Computing 5(2):155–161

67. Yan X et al (2012) A new approach for data clustering using hybrid artificial bee colony algorithm. Neurocomputing 97:241–250

68. Kwedlo W (2011) A clustering method combining differential evolution with the K-means algorithm. Pattern Recogn Lett 32(12):1613–1621

69. Yin M et al (2011) A novel hybrid K-harmonic means and gravitational search algorithm approach for clustering. Expert Syst Appl 38(8):9319–9324

70. Jiang H et al (2010) Ant clustering algorithm with K-harmonic means clustering. Expert Syst Appl 37(12):8679–8684

71. Xiao J et al (2010) A quantum-inspired genetic algorithm for k-means clustering. Expert Syst Appl 37(7):4966–4973
72. Žalik KR (2008) An efficient k′-means clustering algorithm. Pattern Recogn Lett 29(9):1385–1391
73. Jiang B, Wang N (2014) Cooperative bare-bone particle swarm optimization for data clustering. Soft Comput 18(6):1079–1091
74. Kumar Y, Singh PK (2018) Improved cat swarm optimization algorithm for solving global optimization problems and its application to clustering. Appl Intell 48(9):2681–2697
75. Kumar Y, Singh PK (2019) A chaotic teaching learning based optimization algorithm for clustering problems. Appl Intell 49(3):1036–1062
76. Fränti P, Sieranoja S (2018) K-means properties on six clustering benchmark datasets. Appl Intell 48(12):4743–4759
77. Faris H, Ala'M AZ, Heidari AA, Aljarah I, Mafarja M, Hassonah MA, Fujita H (2019) An intelligent system for spam detection and identification of the most relevant features based on evolutionary random weight networks. Information Fusion 48:67–83
78. Maulik U, Bandyopadhyay S (2000) Genetic algorithm-based clustering technique. Pattern Recognition 33.9:1455–1465
79. Kumar Y, Sahoo G (2017) An Improved Cat Swarm Optimization Algorithm Based on Opposition-Based Learning and Cauchy Operator for Clustering. JIPS 13(4):1000–1013
80. Jensi R, Wiselin Jiji G (2016) 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
81. Watkins CJ, Dayan P (1992) Q-learning. Mach Learn 8(3–4):279–292
82. Bouyer A, Hatamlou A (2018) An efficient hybrid clustering method based on improved cuckoo optimization and modified particle swarm optimization algorithms. Appl Soft Comput 67:172–182
83. Hatamlou A (2017) A hybrid bio-inspired algorithm and its application. Appl Intell 47(4):1059–1067
84. Doğan B, Ölmez T (2015) A new metaheuristic for numerical function optimization: vortex search algorithm. Inf Sci 293:125–145
85. Mirjalili S, Lewis A (2016) The whale optimization algorithm. Adv Eng Softw 95:51–67
86. Wang G-G, Deb S, Cui Z (2019) Monarch butterfly optimization. Neural Comput & Applic 31(7):1995–2014
87. Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. The Journal of Machine Learning Research 7:1–30
88. Derrac J, García S, Molina D, Herrera F (2011) A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. Swarm and Evolutionary Computation 1(1):3–18
89. García S, Fernández A, Luengo J, Herrera F (2010) Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. Inf Sci 180(10):2044–2064

Arvinder Kaur received her B.Tech degree in Information Technology from Punjab Technical University, Jalandhar, Punjab, India in 2007 and M.Tech. degree in Computer Science and Engineering from Punjab Technical University, Jalandhar, Punjab, India in 2010. She is pursuing Ph.D. degree in Computer Science and Engineering from Jaypee University of Information Technology, Waknaghat, Himachal Pradesh, India. She is working on data clustering. She has published papers in international journals and conferences of repute.

Yugal Kumar is presently working as Assistant Professor (Senior Grade) in Department of Computer Science & Engineering at Jaypee University of Information Technology (JUIT), Waknaghat, Himachal Pradesh, India. He has more than 15 years of teaching and research experience at reputed colleges and universities of India. He has completed his Ph.D. in Computer Science & Engineering from Birla institute of Technology, Mesra, Ranchi. His primary area of research includes meta-heuristic algorithms, data clustering, swarm intelligence, pattern recognition, medical data international journals and conferences of repute. He is serving as editorial review board member of various journals including Soft Computing, Neurocomputing, Computer Methods and Programs in Biomedicine, PLOSE ONE, Journal of Advanced Computational Intelligence and Intelligent Informatics and Journal of Information Processing System.