

## A Hybrid Black Hole Algorithm with Genetic Algorithm for Solving Data Clustering Problems

Ahmed I. Taloba<sup>a</sup>, Omar Sabah Mohammed<sup>b</sup> and Adel A. Sewisy<sup>c</sup>

A, b

Information Science Department, Faculty Computer and Information, Assiut University, Assiut, Egypt

<sup>c</sup>Computer Science Department, Faculty Computer and Information, Assiut University, Assiut, Egypt

**Article History:** Received: 11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

**Abstract:** Clustering is a process of randomly selecting k-cluster centers also grouping the data around those centers. Issues of data clustering have recently received research attention and as such, a nature-based optimization algorithm called Black Hole (BH) has said to be suggested as an arrangement to data clustering issues. The BH as a metaheuristic which is elicited from public duplicates the black hole event in the universe, whereas circling arrangement in the hunt space addresses a solo star. Even though primordial BH has shown enhanced execution taking place standard datasets, it doesn't have investigation capacities yet plays out a fine local search. In this paper, another crossover metaheuristic reliant on the mix of BH algorithm as well as genetic algorithm suggested. Genetic algorithm represents its first part of the algorithm which prospect the search space and provides the initial positions for the stars. Then, the BH algorithm utilizes the search space and finds the best solution until the termination condition is reached. The proposed hybrid approach was estimated on synchronized nine popular standard functions where the outcomes indicated that the process generated enhanced outcome with regard to robustness compared to BH and the benchmarking algorithms in the study. Furthermore, it also revealed a high convergence rate which used six real datasets sourced of the UCI machine learning laboratory, indicating fine conduct of the hybrid algorithm on data clustering problems. Conclusively, the investigation showed the suitability of the suggested hybrid algorithm designed for resolving data clustering issues.

**Keywords:** Data clustering; Genetic algorithm; Blackhole algorithm; Optimization.

### 1. Introduction

The ground of data science, data clustering refers to the grouping of similar data together; similar items are positioned in solitary group while diverse items are placed in various groups. Data clustering is an unverified learning methodology described through the grouping of objects in undefined foreordained clusters. It is practically different from classification which is a supervised type of learning that requires the placement of objects to predetermined classes [1]. Data clustering has found applicability in several areas, such as data mining, image analysis, machine learning, pattern recognition, statistical data analysis, information retrieval, over and above. Since, clustering techniques as it may be classified into partitional, progressive, density-based, grid-based, as well as model-based methods [2].

Along with these clustering methods, most commonly used method is the partitional clustering technique, amid K-means algorithm being representative of the partitional and focus-based clustering algorithms. Being that the cluster centers are classified, the k-means clustering algorithm prone to local optima [3]. However, loads of nature-inspired evolutionary algorithms were created in past several decades to determine engineering design streamlining issues. These algorithms impersonate the conduct of living components in natural world; hence, they are known as Swarm Intelligence (SI) algorithms. SI algorithms normally look in favor of worldwide optima while being related with rapid union [4].

Metaheuristic looking through advancement has as of late been vigorously examined as per application on several engineering fields, for example power optimization control [5], robotic [6], communications and networking [7], computer vision and manufacturing engineering[8-10], in addition to machine learning[11-13]. In spite of the portrayal of the information angle by various ideas and inspirations, one central trait underlines their goal and that is the establishment of a solution using a discerning searching process to facilitate in the solution space propelled by heuristic knowledge. The solution ought towards upgrade an agreed objective function otherwise a bunch of objective functions if there should be an occurrence of multi-optimization in as much as the coordinated constraints are nurtured. Nowadays, algorithms have pulled in research consideration because of speedy upgrade of hardware speed and enhanced attainability into giving an answer for many engineering problems [14-16]. This is refined by sticking to the heuristic examining conceptualization with a simple constraints and objective design.

A number of nature-propelled probing optimization algorithms have been created dependent on stimulation from a range of natural events, for example hunting behavior of grey wolves [17]; krill herds [18]; black holes [19]; egg-laying conduct of cuckoos [20]; hunting behavior of bats [21]; food-searching manners of bees [22]; as well as initiative method of jazz musicians [23]. The "black hole" (BH) algorithm was because of late created as

a meta-heuristic optimization to facilitate mimic black hole behavior of heaving neighboring stars [19]. The BH algorithm was unambiguously impelled by the physical science of BH also, its connection with the incorporating stars. It is understood with the intention of a group of star shows the quantity of output in a given iteration and every star is liable to BHs' pulling force in the direction of greatest solution. Consequently, a new arrangement of yield is created by moving the stars toward BH in the accompanying cycle. The star that is inside a foreordained distance to the BH will be consumed by the BH, making ready for another star to be made. It permits algorithm to start investigation in the searching space before devouring streamlining time by means of a part fully determined with solutions. The BH algorithm applicable in tackling data clustering issues since the performance assessment confirmed that it is better evaluated to other alike processes. Methodology is acquainted with how to additionally upgrade to permit disclosure of a robust phenomenon in the solution space at the same time as guarantee a valuable clustering process. In this point of view, bequest of [19] can be expanded as of the objective function that does not guarantee the most excellent likely precision still while price is at global optimum. Indigenous BH algorithm is prone to feeble exploration, therefore, requires numerous emphases to achieve an ideal resolution. The BH algorithm and its comprehensive versions have found application in resolving a number of enhancements and engineering issues [24-35].

Hybrid optimization method is a procedure which includes improvement of a added robust algorithm which illustrates greater adaptability beside troublesome issues by means of consolidating a metaheuristic algorithm with an additional streamlining outline [36]. Local search algorithms are able to discover improved solutions by iteratively scanning the solution space using a well-defined neighborhood mechanism [36]. Metaheuristics be comprised of definite iterative generation operations which discover a search space through proficiently joining diverse sub-heuristics. Global optimum areas were discovered utilizing some learning strategies in such algorithms [36, 37] populace inspired metaheuristics rely upon accepted ideas to discover the solution field. They influence the populace to track down definitive result dependent on top of their particular influential techniques [36]. The performance of the populace-based metaheuristics on discovering local optima is superior to further direction strategies (prejudiced effectively through local optima). Consequently, hybrid metaheuristic is typically well-organized and flourishing because they possess of the reward of both course and populace-based methods in a legitimate way [36]. For instance, mix of the global research ability of GA and PSO's mix has been arranged by [38] intended for finding global ideal all the more precisely. A few examinations have projected diverse hybrid optimization techniques based on the blend of the influential mechanisms of such strategy. Consequently, hybrid frameworks are frequently creative in conditions of solution quality and running time [39]. These examinations suggest the BH algorithm by means of GA to tackle the concern of entrapment of BH algorithm in the local minima. An undeniable level hybridization is projected for this investigation, where GA will implement the global search (investigation) whereas BH will play out the local exploration subject to the GA-if best-picked solution. Noteworthy commitment of this project is the improvement of the global accessibility of BH algorithm using GA; to put it another way, all stars' underlying qualities in the BH population are the GA-picked better approach.

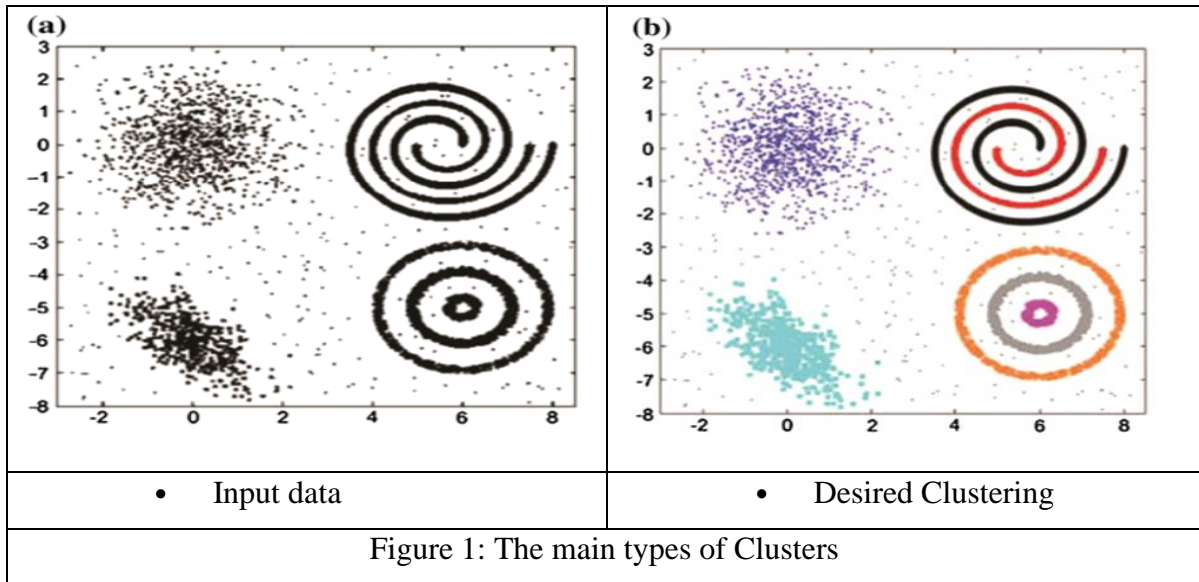
The enduring segment of this work is orchestrated in associated way: Segment 2 examined a portion of the recently projected research on data clustering. At that point, the BH algorithm and planned hybrid genetic-blackhole algorithm were accessible in Segment 3 and 4, respectively. Segment 5 laid out exploratory results got. Segment 6 concluded work in brief.

## **Overview of Data Clustering and Bh Algorithm**

### **2.1 The Problem of Data Clustering**

Definitely, clustering is a crucial independent arrangement procedure which is depicted by the allocating of a lot of models or vectors into the multi-dimensional space in bunches or assortment. Diverse closeness metrics between data objects are utilized to accomplish clustering; such data comparability/divergence in the database is viewed as utilizing distance estimation [40]. The activity is fueled by the possibility of data grouping using a specific amount of clusters by methods for distance belittlement among objects of each cluster itself. Cluster study is characterized as reworking of an assemblage of prototypes presented either as a vector of evaluations otherwise as a point in a multi-dimensional space. The point of examination is to get hold of cluster which is described by means of quality of comparability [41, 42].

Clusters are often used designed for different purposes, for model image processing, data statistical analysis, and medical imaging analysis. They are additionally utilized in other examination fields like science and engineering. Also, this is inseparable from statistical data analysis and recognized as an essential assignment of investigative data mining in several branches like ML, pattern recognition, image analysis, information retrieval, and bioinformatics. Differences among clusters in stipulations of shapes, sizes, and densities are represented in Figure 1.



Cluster detection might be confounded by the presence of boisterous data; in such cases, the ideal cluster is typically portrayed as negligible just as single set of points. Being that man have been proficient in two and three-dimensional clustering, automatic algorithms are usually used for high-dimensional data clustering. This detail, combined with indeterminate quantity of clusters set for data, constantly created several of clustering frameworks within distribution. Concerning acknowledgment, data analysis correlated with analytical modeling, where the training data is given while conduct of obscure test data anticipated (an undertaking called learning).

Distance measurement is needed to properly evaluate the similitude of data objects. The definition of the problem is as follows: known records of data with every documentation apportioned to one of clusters; clustering can be performed utilizing various rules as objective functions for optimization processes. Among the commonest characteristic is limiting the amount of squared Euclidean distance among each record and focal point of the consequent cluster as presented in equation 1.

$d = \sqrt{(x - c)^2 + (y - c)^2}$	(1)
------------------------------------	-----

Where  $d$  is the Euclidean distance among a data record and cluster center. Furthermore qualities of data records as well as numbers of clusters, individually.

### 2.2 Related Works

A range of studies have projected the work of metaheuristics for clustering problems. This section reviewed metaheuristic-based clustering algorithms but confined to strategies associated with the proposed algorithm in this study.

Van & Engelbrecht [43] made the first proposal for a data clustering perspective using two resources: first, is the PSO in which optimal centroids established moreover used as seeds in K-means algorithm. For 2nd approach, the PSO is used in rectifying the K-means-formed clusters. The two approaches were conformed acceptable.

Next, Shelokar et al. [44] conversed the Ant Colony Optimization (ACO) policy that is illustrated by utilization of conveyed specialists that emulates the way unfathomable ants find the briefest way to their food resource from their nest. The evaluation showed process as a viable heuristic for near-optimal cluster demonstration.

Another study by [45] focused on the capacity of Quantum-based PSO (QPPSO) for data clustering. This study presented the following 2 conclusions: i) it performed better than K-means along with the conventional PSO clustering algorithms; ii) combination of QPSO also K-means may enhance the QPSO clustering algorithm. Paterlini & Krink (2006) evaluated the performances of GA, PSO, and DE used for clustering complication and showed DE to continuously perform better.

A hybridized approach of amalgamation the PSO method, Nelder–Mead simplex inquiry, and K-means algorithm has been projected by [46]. The performance of K-NM-PSO was collated to those of PSO, NM-PSO, K-PSO and K-means clustering and found to be powerful and suited for data clustering.

A study by [47] hybridized PSO and k-harmonic means to achieve PSOKHM for improving the algorithm’s global search ability. Following this, [48] demonstrated the use of artificial bee colony (ABC) as a fresh approach, utilizing Deb’s rules rather than the greedy solution commonly seen in the ABC algorithm to resolve infeasible solutions. The evaluation of the algorithm showed encouraging and potential outcomes with regards to effectiveness and efficiency. Similarly, Karaboga & Ozturk [49] highlighted a fresh clustering approach using an ABC algorithm that simulates honey bee swarm’s tendencies for food foraging. Comparing the performance with other classification techniques, the algorithm was shown to be preferable than the others.

Data clustering utilizing firefly algorithm was discussed by [50] in regards to the performance of FA on administered clustering issues with results underlining its ampleness and power. A new data clustering algorithm was highlighted by [51] utilizing a hybrid (HABC) which is a GA crossover administrator towards ABC to improve the honey bees’ data trade. HABC algorithm was shown to perform better and based on the report, [52]

subsequently presented a new PSO approach which was examined on two factitious datasets and five real datasets and the outcome showed its applicability to clustering problems with a disclosed or an undisclosed number of clusters alike.

Furthermore, Wan et al. [51] opted to test the Bacterial Foraging Optimization (BFO) on established benchmark datasets and subsequent comparison with 3 clustering technique. The evaluation indicated the effectiveness of BFO and the potential application for handling datasets of various cluster sizes, densities, and multiple dimensions. A brand new data clustering standpoint using Cuckoo search with levy flight was anticipated by [53]. The heavy-tailed trait of Levy flight permits powerful treatment of the yield area. It was also noted to perform better compared to GA and PSO.

Similarly, Hatamlou's new clustering algorithm [19] based on the BH phenomenon displayed superior performance compared to other conventional heuristic algorithms for various benchmark datasets.

A latest data clustering algorithm draw out from cuckoo query encroachment was produced by [54]. The conducting of new algorithm was assessed utilizing 4 datasets as well as benchmarked with K-means, PSO, GSA, the big bang–big crunch algorithm (BB-BC), in addition to the BH algorithm. Result of the analyses communicated force of the new technique in getting the finest qualities intended for practically all datasets.

A novel hybrid algorithm was accessible based on top of the cluster center initialization algorithm (CCIA), bees' algorithm, as well as level of dissimilarity evolution by [55] (collectively acknowledged as CCIA-BADE-K). Suggested algorithm and a subsequent comparison made with some alternatives successfully confirmed its better performance. A study by [56] combined DE with K-means to produce a hybrid clustering algorithm whose experimental evaluation manifested a greater performance of the hybrid forms with K-means algorithm analogized to non-hybrid types.

Basically, a hybrid algorithm employed for data clustering (HCSDE) in light of DE and CS was planned by [57]. The performance of this algorithm was shown to be better than those of CS, DE, PSO, GSA, BB-BC, and BH. Hybridized KHM among better PSO and standard CSA was obtainable by [58] to achieve sophisticated solutions so as to attain model solutions. The hybrid algorithm was found to be efficient but requires longer processing time. Moreover, it also converges due to CSA's flaws to the local optima. Nevertheless, KHM has also been hybridized with more alternative heuristic methods, with ICAKHM being deemed a novel finding. It is designed from the amalgamation of K-harmonic which means algorithm and altered form of absolutist aggressive algorithm (ICA) [59]. Such ICA form opted for the genetic mechanist of crossover and mutation for avoiding early-onset convergence. Its performance assessment indicated the appropriateness of the algorithm [60]. Recently, [61] proposes a most recent algorithm designed for data clustering based on K-means and improved CS. The method adjusts two CS parameters to fine-tune the solution vectors while extending the abilities of CS further. Its efficacy was tested using three microarray datasets and the outcomes showed that it performed better than the others.

The general population-based clustering approach based on concerning PSO and Lloyd's K-means algorithm was presented by [62]. Comparative studies on the algorithm showed its capability to generate better and steadier solutions contrasted its five-individual-based counterparts much of the time. [63] Presented a clustering algorithm for overcoming the K-Means local optimum problem via hybridization with Crow Search Optimization (CSA) which is novel broader populace metaheuristic accession algorithm that mimics the crows' intelligent behavior. Moreover, the Elephant Herding Optimization was created by [64] endeavors by methods for reducing the intra-cluster distance and cost function.

### 2.3 Black Hole (BH) Algorithm

BH algorithm was anticipated based on the black hole hypothesis as well as on the primary idea of a region of space facilitating where no nearby object can escape from its gravitational pull from a broad volume of mass. Whatever falls into incidence would be disposed from universe, including light. There are two main position of BH algorithm: i) star movement, and ii) star re- instatement crossing into D-dimensional hyper sphere in the region of BH (that is incident horizon). Its function is as follows: First, + 1 star, (where populace size) are discretionarily introduced in search space. Subsequent to evaluating their fitness, the most excellent worth is alluded to black hole. Black hole is motionless, which means no development till different stars accomplish superior goal. Along these lines, the quantity of people searching for the ideal worth equivalents to. Then, in the next generation dependent on the condition beneath each star moves towards the BH:

$$\text{rand} \times \text{distance\_to\_BH} \quad (2)$$

Where rand = a incidental number inside an interval [0,1].

BH algorithm likewise demonstrates any star which exists in the BHs' where the wiped out will be event horizon. Event horizon's radius (R) is depicted as follows:

$$R = \frac{1}{\text{fitness\_BH} - \text{fitness\_star}_i} \quad (3)$$

Where and are the fitness values of BH and ith star is the No. of stars.

While the distance between BH and a star is < R, candidate breakdowns, plus another applicant is created and self-assertively spread in search space. BH is usually connected in the midst of a plain design along with simplicity

of execution, just like parameter-free algorithm. Its intermingling occurs in all runs to the global optimum, whereas other heuristics possibly will experience capture in local optima [19, 65].

Despite the fact that BH has shown fascinating execution as a clustering technique, it experiences powerless adjusting of its investigation and utilization capacities. Star may change its course if some other star tracks down a superior arrangement contrasted with the answer for the current BH, consequently turning into the new BH. Moreover, the anticipation of the event horizon made as the stars may perhaps show moderately quick combination for search space to be involved by means of the black hole because of the short of investigation abilities. Nonetheless, it denies heightening of investigation otherwise gathering information with respect to the recently called solution. It's just a start again strategy exposed to each star separately [66]. Consequently, this study hybridized BH algorithm with GA in favor of effective data clustering.

- Hybrid GA-BH algorithm (GBH)

This segment the construction of the anticipated algorithm named Genetic-Black Hole (GBH) algorithm clarified. GBH consolidates GA and BH algorithms; by means of civilizing the stars' creating stage utilizing GA, fundamental design of BH algorithm was changed. In the customary form of BH, initial tread is the instatement or generation of the stars. Moreover, at the point when a star has a fitness value not exactly, by then, it should be taken out, just as another star should be delivered; in any case, in the extended type of BH, both normalized allotment conditions in the two stages were superseded by GA to redesign the assessment capacity of BH algorithm. In broad, primary construction of GBH is given in Figure XX.

Furthermore, one more intensification of the BH algorithm was projected in this paper. Main steps of GBH are:

Step1:

Read the datasets as well as access the parameters which are the integer of chromosomes (), quantity of stars (),

Step2:

Define the objective function.

Step3: engender the population for GA casually through standardized allocation, where is the amount of chromosomes through the homogeneous allocation equation, as follows:

(4)

Step4:

Evaluate each chromosome/solution using.

Step5:

Selection

It is an administrator which chooses 2 parents for the generation of new-fangled litter. During selection, probability of choosing a string among stumpy robustness esteem be elevated than those among high fitness esteem. Parents are indiscriminately chosen in GA.

Step6:

Crossover

This is an operator for the generation of latest litter as of chosen parents. GA has numerous crossover operators.

Step7: Mutation

This is an operator which changes components in offspring produced by the crossover operator. In the local search algorithms, a mutation operator is able to view as a migration from a current solution to its neighborhood solution.

Step8: Elitist strategy (Replacing):

This engages an unsystematic elimination of a string as of surviving populace as well as restoring from the previous population same with the best string.

Step9: Repairing:

It includes making improbable individuals from populace plausible. This method for most part targets isolating the attainable people from improbable ones in the populace through fixing improbable people, accordingly, guaranteeing conjunction with improbable individuals till they turn out to be feasible.

Step10:

If the particular quantity of generations is pleased, subsequently, go to Step 11, or else, go back to Step 3.

Step11:

Pick the pinnacle solutions, after that embark populace of the stars based on them.

Step12:

Analyze the most excellent star, and set it as

Step13:

Proceed the entire stars in the direction of the black hole vi eq. X

Step14:

For the fitness function examine each star in the new position the fitness function.

Step15:

If the latest fitness value is finer, then,

Step16:

Calculate the event horizon by Eq.X

Step17:

If the latest appropriateness is finer than the previous, after that, keep the new position, otherwise, keep the previous position.

Step18:

The pseudo code of recommended algorithm is revealed in Algorithm 1, whereas flowchart is given in Figure X.

**Algorithm 1: Genetic Black Hole Algorithm**

**Input** Dataset or Test Function, *GAMaxItr*, *StarSize*, *Upper*, *Lower*, *PopSize*, rate of elitism, rate of mutation, *BHMaxItr*

**Output** Solution  $X_{BH}$

**Procedure**

1. Define Objective Function  $f(x_i)$
2. Initialize all the solutions  $x_i$  in the population via eq. 4
3. Evaluate the fitness value of each star  $X$  in the population via  $f$
4. **While**  $GA_{itr} < GAMaxItr$
5.     Perform the selection process
6.     Apply the crossover operator
7.     Apply the mutation operator
8.     Update the solutions in the population for the next generation
9. **End While**
10. Initialize the Stars based on the best  $S$  solutions from Population
11. Set the best solution in the Stars as Black Hole  $x_{BH}$
12.     **While**  $itr \leq MaxItr$
13.         **For** each star  $X_i$  in the population
14.             Update the position of each star  $X_i$
15.             Check the boundaries of each star  $X_i$
16.             Evaluate the fitness value of the star  $X_i$
17.             Set the best star in the population as Black Hole  $x_{BH}$
18.         **End For**
19.         Calculate the event horizon ( $R$ ) via eq. 3
20.         **For** each star  $X_i$  in the population
21.             **If**  $X_i$  crosses the event horizon ( $R$ ) **Then**
22.                 Remove the star  $X_i$
23.                 Generate a new star via eq. 4
24.             **End If**
25.         **End For**
26.         Set the best star in the population as Black Hole  $X_{BH}$
27.     **Loop**
28.     **Return**  $X_{BH}$

• Results and Discussion

In this study, the assessment was done on a PC (Core i7, 3.6 GHz, 16 GB of RAM, 64-bit Windows 10 Operating System) by means of Matlab 2017a. Proposed algorithm was evaluated for data clustering on 6 datasets - Iris, Wine, Glass, Cancer, Contraceptive Method Choice (CMC), and Vowel. Table 1 presents the respective attributes of the datasets. These data sets were extracted from the UCI machine learning laboratory.

Table 1: Main uniqueness of the test datasets

Datasets	No. of classes	No. of features	No. of instances	Size of classes
Iris	3	4	150	50,50,50
Wine	3	13	178	59,71,48
CMC	3	9	1473	629,334,510
Cancer	2	9	683	444,178
Glass	6	9	214	70,17,76,13,9,29
Vowel	6	3	871	72,89,172,151,207,180

- Iris dataset

Consists of 150 arbitrary models of flowers with 4 properties from iris. The samples were grouped hooked on 3 groups of 50 instances, with every group representing figure of iris plant (Setosa, Versicolor and Virginica).

- Wine dataset

Wine dataset presents nature of three sorts of wine dependent on physicochemical properties (filled within an indistinguishable area in Italy however sourced from three cultivars, individually). All of the three kinds of wine were associated with 178 events, with 13 numeric credits tending to measures of 13 sections evoked in them.

- CMC dataset

CMC dataset was made as a subset of Indonesia's 1987 National Contraceptive Prevalence Survey by TjenSien Lim. Example range was comprised of wedded ladies who were either not pregnant or not mindful of their pregnancy during the inquiry time frame. The dataset marked the concern of forecasting the current contraceptive method choice (i.e. no use, long-term method, or short-term methods) according to the demographic and socioeconomic characteristics of women.

- Cancer dataset

Cancer data set is a portrayal of Wisconsin breast cancer database, comprising of 683 occasions with 9 parts. It comprised: Clump Thickness, Cell Size Uniformity, Cell Shape Uniformity, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nuclei, and Mitoses. Every one of the occurrences was perhaps of one set, either kind or dangerous.

- Glass dataset

Glass dataset comprised of 214 items having 9 attributes, which are magnesium, aluminum, barium, silicon, refractive index, sodium, potassium, calcium, as well as iron. The data sampling was based on 6 category of glass - non-float method building windows, tableware, float-processed vehicle windows; float processed building windows, containers, as well as headlamps.

- Vowel dataset

Vowel data set is encompassed of 871 Indian Telugu vowel sounds with 3 ascribes which relate towards 1st, 2nd and 3rd vowel occurrence, in addition to six covering sets.

Algorithmic functioning be evaluated as well as evaluated using 2 attributes:

- Sum of intra-cluster distances as an internal quality measure: This is amount of the distances among each data plus the center of corresponding cluster. It was determined utilizing equation 1. Normally, a minor intra-cluster distances sum addresses a superior clustering worth. In this study an evaluation criterion for fitness was served by the sum of intra-cluster distances

- Error Rate (ER) as external quality measure: proportion of mislaid data objects as shown in the equation below:

	(8)
--	-----

The performance of anticipated algorithm was measured up to with those of quite a few heuristic techniques, the same as K-means [41], PSO [67], ABC [68], BAT [69], GSA [70], BB-BC [71], CS [72], GWO [73] and BH [19]. Contrarily, GBH was analysed against the new hybrid and customized meta-heuristics revealed in literature, including enhanced krill herd algorithm [74], hybrid clustering method using imitation bee colony and Mantegna levy allocation [75], latest quantum chaotic cuckoo explore algorithm [72], Hd-ABC history-driven counterfeit bee colony [76]. ICAKHM is viewed as a original technique whose plan depend on hybridization of K-harmonic means algorithm in addition to an altered version of the imperialist spirited algorithm (ICA) [77] and grey wolf optimizer with levy flight steps [73].

Table 2 and Table 3 illustrates sum of intra-cluster distances and error rate use typical meta-heuristics clustering algorithm, hybrids meta-heuristics algorithms, as well as the altered metaheuristics to ensure a better benchmarking with GBH.

A précis of intra-cluster distance also error rate be shown at Table 2. Each independent algorithm was simulated for 30 independent implementations before calculating the values designed for the top, customary, worst, standard deviation plus error rate. Finest-attained values through algorithms were shown bold in each dataset. The experimental outcome showed GBH execute improved compared to BH and K-means. Besides, recommended algorithm has negligible customary deviation contrasted with different algorithms, implying that it accomplished the minimum value each time.

GBH outperformed the auxiliary algorithms on Iris dataset by achieving an intra-cluster distance value of 96.5403 and standard deviation of 0.00014 compared to the other algorithms. Regarding Wine dataset, the suggested GBH algorithm acquired the ideal estimation of “16,291.99” which was strikingly better analyzed than the other relative algorithms. Comparison of CMC dataset additionally demonstrated GBH to be much better contrasted with different algorithms, with the most exceedingly terrible solution attained at 5532.58940. On the other hand, this value is still superior to the greatest solutions originated by the other algorithms. On Cancer dataset, the projected GBH algorithm performed more than K-means, PSO and GSA algorithms, on the other hand BB-BC algorithm performed better contrasted to GBH as far as standard deviation.

GBH algorithm got an average of 210.97180 on Glass dataset though different algorithms failed to accomplish a solution. In the interim, GBH accomplished the top average solution in addition to standard deviation on Vowel dataset contrasted with different algorithms. Consequently, GBH presented improved solution quality as well as

more modest standard deviation compared to the supplementary algorithms. GBH can build up te ideal solutions as proven in most examined cases at the same time as the other algorithms possibly will be trapped in local optima.

Table 2 The result acquired through GBH in addition to standard algorithms on diverse data sets.

Data set	Criteria	Kmeans	PSO	ABC	BA	GSA	BB-BC	CS	GW O	BH	GBH
Iris	Best	97.32590	96.89428	-	97.433	96.68794	96.67648	97.98364	96.65826	96.65589	<b>96.5403</b>
	Average	106.5766	97.23280	96.65502	103.036	96.73105	96.76537	102.51332	99.12574	96.65681	<b>96.5622</b>
	Worst	123.9695	97.89733	-	108.870	96.82463	97.42865	106.76087	-	96.66306	<b>96.5873</b>
	standard	12.938	0.347168	2.213958	3.410	0.02761	0.20456	2.182.24	-	0.001.73	<b>0.00014</b>
	Error rate	13.42	12.58	10.00	10.78	10.04	10.05	09.80	10.74	10.02	9.40
Wine	Best	16,555.68	16,345.97	-	16,391.46	16,313.87	16,298.67	16,363.12	16,307.1	16,293.41	<b>16,291.99</b>
	Average	17,251.35	16,417.47	16,308.53	16,606.90	16,374.30	16,303.41	16,420.81	16,318.4	16,294.31	<b>16,292.99</b>
	Worst	18,294.85	16,562.32	-	17,160.39	16,428.86	16,310.11	16,525.72	-	16,300.22	<b>16,296.89</b>
	standard	874.148	85,497.4	5.096923001	237.740	34,671.22	2,661.98	45,540.86	-	16,512.70	<b>0.90340</b>
	Error rate	31.14	28.52	28.76	28.92	29.15	28.52	29.10	29.56	28.47	28.40
CMC	Best	5703.20000	5700.98500	-	5671.52600	5542.27631	5534.09483	5778.45388	-	5532.883.23	<b>5531.99988</b>
	Average	5705.37000	5820.96500	5.584,630.1	5802.14400	5581.94502	5574.75174	5962.09604	-	5533.631.22	<b>5532.29789</b>
	Worst	5704.57000	5923.24900	-	5966.19000	5658.76293	5644.70264	6205.93042	-	5534.777.38	<b>5532.58940</b>
	standard	1.033	46.95969	10.16857871	88.219	41.13648	39.43494	115.23954	-	0.59940	0.58878
	Error rate	54.48	54.49	57.68	56.00	55.67	54.52	57.18	-	54.39	54.35
Cancer	Best	2988.43000	2973.50000	-	3021.483000	2965.76394	2964.38753	3089.77652	2964.390	2964.38878	<b>2961.95000</b>
	Average	2988.99000	3050.04000	-	3107.125000	2972.66312	2964.38798	3200.79638	2964.395	2964.39539	<b>2963.90000</b>
	Worst	2999.19000	3318.88000	-	3250.525000	2993.24458	2964.38902	3476.06894	-	2964.45074	2988.43000
	standard	315.14560	110.8013	-	77.110	8.91860	<b>0.00048</b>	102.96476	-	0.00921	0.0072
	Error rate	04.39	05.25	-	03.79	03.74	03.70	04.94	03.65	03.70	3.65
Glass	Best	215.73000	270.57000	-	232.00700	224.98410	223.89410	220.12580	265.8142	210.51549	<b>209.99689</b>
	Average	218.70000	275.71000	254.03500	241.91600	233.54329	231.23058	225.19820	302.0415	211.49860	<b>210.97180</b>
	Worst	227.35000	283.52000	-	247.08500	248.36721	243.20883	227.02230	-	213.95689	<b>211.56990</b>
	standard	2.456	4.557134	10.107	5.059	6.139.46	4.650.13	5.662.3	-	1.182.30	<b>0.99869</b>
	Error rate	38.44	30.58	38.67	40.56	41.39	41.37	41.89	40.90	36.51	<b>30.50</b>
	Best	149.398.66	148,976.01	-	155.163.59	151.317.56	149.038.51	149.417.31	-	148.985.61	<b>148.965.64</b>



<b>Vowel</b>	Average	151.987 .98	148,999 .82	153,218. 45	147.411. 21	152.931 .81	151.010 .03	150.186 .12	-	149.848. 18	<b>149.466 .52</b>
	Worst	162.455 .69	149,121 .18	-	160.783. 94	155.346 .69	153.090 .44	150.841 .40	-	153.058. 98	<b>149.484 .69</b>
	standard	3425.25 0	28.8134 692	162.2703	3001.824 5	2486.70 285	1859.32 353	1576.36 97	-	1306.95 375	<b>1297.64 781</b>
	Error rate	43.57	41.92	42.87	42.55	42.39	41.89	42.41	-	41.65	<b>41.36</b>

Table 3: The sum of intra-cluster distances and error rate attained by GBH with modified algorithms on various data sets.

<b>Datasets</b>	<b>Criteria</b>	<b>IKH</b>	<b>ABCL</b>	<b>QCCS</b>	<b>Hd-ABC</b>	<b>ICAKHM</b>	<b>EGWO</b>	<b>GBH</b>
<b>Iris</b>	Best	96.65550	-	96.65548	-	96.63620	96.65230	<b>96.5403</b>
	Average	96.65550	96.65550	96.65623	94.47	96.66640	99.12530	<b>96.5622</b>
	Worst	96.65550	-	96.66771	-	96.69190	-	<b>96.5873</b>
	standard	9.8E - 06	1.351718	0.00266	-	0.01055	-	<b>0.00014</b>
	Error rate	9.78	10.45	09.43	0.0	11.23	9.76	9.40
<b>Wine</b>	Best	16,292.21	-	16,292.26	-	16,293.90	16,292.15	<b>16,291.99</b>
	Average	16,294.30	16,295.30	16,293.26	16280.96	16,295.60	16.292.43	<b>16,292.99</b>
	Worst	16,292.84	-	16,294.34	-	16,296.94	-	<b>16,296.89</b>
	standard	0.706742	1.09745	0.71534	-	1.002372	-	<b>0903.40</b>
	Error rate	28.90	29.80	28.70	4.68	28.73	28.71	28.40
<b>CMC</b>	Best	5693.720	-	5532.22476	-	5699.21830	-	<b>5531.99898</b>
	Average	5693.779	5533.7790	5532.71992	5692.75	5705.14850	-	<b>5532.29789</b>
	Worst	5693.735	-	5535.29050	-	5721.17790	-	<b>5532.58940</b>
	standard	0.007975	0.85343	0.134	-	1.268275	-	0.58878
	Error rate	55.90	57.12	57.11	2.81	54.47	-	54.35
<b>Cancer</b>	Best	2964.387	-	2964.38951	-	2962.42000	2964.11000	<b>2961.95000</b>
	Average	2964.393	-	2964.41463	-	3022.81000	2964.49000	<b>2963.90000</b>
	Worst	2964.389	-	2964.49945	-	3150.15000	-	2988.43000
	standard	0.001258	-	0.02761	-	0.396	-	0.0072
	Error rate	3.69	-	03.51	-	4.27	3.75	3.65
<b>Glass</b>	Best	210.2520	-	-	-	199.86000	214.42500	<b>209.99689</b>
	Average	222.8008	2.2009e+0	-	217.89	202.41000	242.43800	<b>210.97180</b>
	Worst	215.9355	-	-	-	209.77000	-	<b>211.56990</b>

	standar d	2.737919	4.636733	-	-	0.26	-	<b>0.99869</b>
	Error rate	33.90	32.56		34.45	32.61	33.60	30.50
<b>Vowel</b>	Best	148,967.24	-	-	-	149,201.63	-	<b>148,965.64</b>
	Average	158,600.52	149,600.5	-	-	161,431.04	-	<b>149,466.52</b>
	Worst	150,172.42	-	-	-	165,804.67	-	<b>149,484.69</b>
	standar d	1732.4516	1128.941	-	-	2746.041	-	<b>1297.64781</b>
	Error rate	41.56	41.90	-	-	41.98	-	41.36

Table 3 is evidence for GBH to acquire the best performance based on nominal intra-cluster distances as well as error rate while compared with other algorithms. It also executed improved on every 6 datasets compared to the further qualified algorithms. Evidently, noteworthy balance between utilization as well as investigation magnified performance of suggested GBH.

Standard deviation for anticipated GBH algorithm on the Iris dataset was 0.00014, which was altogether a lesser amount than other relative algorithms. Contrarily, top solution was 96.5403 while most terrible was 96.5873 for GBH albeit still improved than supplementary algorithms. Moreover, Wine dataset demonstrated the projected GBH algorithm got perfect estimation of “16,291.99”, has outperformed the other algorithms.

GBH additionally accomplished better performance on CMC dataset contrasted to the other algorithms; here, the most terrible solution accomplished by GBH was 5532.88940 which stayed enhanced than the top solutions attained by supplementary algorithms. In favor of Cancer dataset, the recommended GBH achieved 2961.95000 as the best solution and 2963.90000 as the average solution while the standard deviation was 0.00723. This was better looked at than the performance of ABCL, QCCS, HD-ABC, ICAKHM, and EGO.

At last, ICAKHM algorithm accomplished the best solution (199.86000) on the Glass dataset. Meanwhile, the Vowel dataset showed GBH to provide the best average solution of 149,466.52. GBH adequately yielded the best results on practically the entirety of the datasets when contrasted with the other relative algorithms; in this manner, ending up being successful in settling complex streamlining issues just by adding new operators.

### Conclusion

This paper reported the hybridization of GA with BH algorithm with the aim of improving clustering performance. The suggested approach was evaluated on six datasets and the outcome of the evaluations showed the suggested GBH to perform well inefficient data clustering. It also illustrated that the proposed GBH flee local optima as well as effectively explores the search space. Further work may well investigate other applications of GBH, such as text documents clustering.

### References

1. Abdeyazdan, M., Data clustering based on hybrid K-harmonic means and modifier imperialist competitive algorithm. *The Journal of Supercomputing*, 2014. 68(2): p. 574-598.
2. Aggarwal, C.C. and C.K. Reddy, *Data clustering: algorithms and applications*. 2013: CRC press.
3. Al Sudani, Z.A., S.Q. Salih, and Z.M. Yaseen, Development of Multivariate Adaptive Regression Spline Integrated with Differential Evolution Model for Streamflow Simulation. *Journal of Hydrology*, 2019.
4. Al-Musawi, A.A., et al., Shear strength of SFRCB without stirrups simulation: implementation of hybrid artificial intelligence model. *Engineering with Computers*, 2018: p. 1-11.
5. Alwanas, A.A.H., et al., Load-carrying capacity and mode failure simulation of beam-column joint connection: Application of self-tuning machine learning model. *Engineering Structures*, 2019. 194: p. 220-229.
6. Alzaidi, A.A., et al., Sine-Cosine Optimization-Based Bijective Substitution-Boxes Construction Using Enhanced Dynamics of Chaotic Map. *Complexity*, 2018. 2018.
7. Anderberg, M.R., *Cluster analysis for applications*. 1973, Office of the Assistant for Study Support Kirtland AFB N MEX.
8. Arora, S. and I. Chana. A survey of clustering techniques for big data analysis. in *Confluence The Next Generation Information Technology Summit (Confluence)*, 2014 5th International Conference-. 2014. IEEE.
9. Azizipanah-Abarghooee, R., et al., Short-term scheduling of thermal power systems using hybrid gradient based modified teaching-learning optimizer with black hole algorithm. *Electric Power Systems Research*, 2014. 108: p. 16-34.
10. Babrdel Bonab, M., et al., An effective hybrid of bees algorithm and differential evolution algorithm in data clustering. *Mathematical Problems in Engineering*, 2015. 2015.

11. Blum, C., A. Roli, and M. Sampels, Hybrid metaheuristics: an emerging approach to optimization. Vol. 114. 2008: Springer.
12. Boucekara, H., Optimal power flow using black-hole-based optimization approach. *Applied Soft Computing*, 2014. 24: p. 879-888.
13. Boucekara, H.R., Optimal design of electromagnetic devices using a black-hole-based optimization technique. *IEEE Transactions on Magnetics*, 2013. 49(12): p. 5709-5714.
14. Boushaki, S.I., N. Kamel, and O. Bendjeghaba, A new quantum chaotic cuckoo search algorithm for data clustering. *Expert Systems with Applications*, 2018. 96: p. 358-372.
15. Bouyer, A. and A. Hatamlou, An efficient hybrid clustering method based on improved cuckoo optimization and modified particle swarm optimization algorithms. *Applied Soft Computing*, 2018. 67: p. 172-182.
16. Bouyer, A., An optimized k-harmonic means algorithm combined with modified particle swarm optimization and Cuckoo Search algorithm. *Foundations of Computing and Decision Sciences*, 2016. 41(2): p. 99-121.
17. Bouyer, A., H. Ghafarzadeh, and O. Tarkhaneh, An efficient hybrid algorithm using cuckoo search and differential evolution for data clustering. *Indian Journal of Science and Technology*, 2015. 8(24).
18. Cura, T., A particle swarm optimization approach to clustering. *Expert Systems with Applications*, 2012. 39(1): p. 1582-1588.
19. Diaz, J.E., J. Handl, and D.-L. Xu, Integrating meta-heuristics, simulation and exact techniques for production planning of a failure-prone manufacturing system. *European Journal of Operational Research*, 2018. 266(3): p. 976-989.
20. Doraghinejad, M. and H. Nezamabadi-pour, Black hole: a new operator for gravitational search algorithm. *International Journal of Computational Intelligence Systems*, 2014. 7(5): p. 809-826.
21. Eskandarzadehalamdary, M., B. Masoumi, and O. Sojodishijani. A new hybrid algorithm based on black hole optimization and bisecting k-means for cluster analysis. in *Electrical Engineering (ICEE), 2014 22nd Iranian Conference on*. 2014. IEEE.
22. Gandomi, A.H. and A.H. Alavi, Krill herd: a new bio-inspired optimization algorithm. *Communications in nonlinear science and numerical simulation*, 2012. 17(12): p. 4831-4845.
23. Geem, Z.W., J.H. Kim, and G.V. Loganathan, A new heuristic optimization algorithm: harmony search. *simulation*, 2001. 76(2): p. 60-68.
24. Ghafarzadeh, H. and A. Bouyer, An Efficient Hybrid Clustering Method Using an Artificial Bee Colony Algorithm and Mantegna Lévy Distribution. *International Journal on Artificial Intelligence Tools*, 2016. 25(02): p. 1550034.
25. Gupta, G.P. and S. Jha, Integrated clustering and routing protocol for wireless sensor networks using Cuckoo and Harmony Search based metaheuristic techniques. *Engineering Applications of Artificial Intelligence*, 2018. 68: p. 101-109.
26. Hartigan, J.A., *Clustering algorithms*. 1975.
27. Hatamlou, A., Black hole: A new heuristic optimization approach for data clustering. *Information Sciences*, 2013. 222: p. 175-184.
28. Hatamlou, A., S. Abdullah, and H. Nezamabadi-Pour. Application of gravitational search algorithm on data clustering. in *International Conference on Rough Sets and Knowledge Technology*. 2011. Springer.
29. Hatamlou, A., S. Abdullah, and M. Hatamlou, Data clustering using big bang–big crunch algorithm, in *Innovative Computing Technology*. 2011, Springer. p. 383-388.
30. Jain, A.K., Data clustering: 50 years beyond K-means. *Pattern recognition letters*, 2010. 31(8): p. 651-666.
31. Jaiprakash, K.P. and S.J. Nanda, Elephant Herding Algorithm for Clustering, in *Recent Developments in Machine Learning and Data Analytics*. 2019, Springer. p. 317-325.
32. Janardhanan, M.N., et al., Metaheuristic algorithms for balancing robotic assembly lines with sequence-dependent robot setup times. *Applied Mathematical Modelling*, 2019. 65: p. 256-270.
33. Jensi, R. and G.W. Jiji, An improved krill herd algorithm with global exploration capability for solving numerical function optimization problems and its application to data clustering. *Applied Soft Computing*, 2016. 46: p. 230-245.
34. Jensi, R. and G.W. Jiji, MBA-LF: A NEW DATA CLUSTERING METHOD USING MODIFIED BAT ALGORITHM AND LEVY FLIGHT. *ICTACT Journal on Soft Computing*, 2015. 6(1).
35. Karaboga, D. and C. Ozturk, A novel clustering approach: Artificial Bee Colony (ABC) algorithm. *Applied soft computing*, 2011. 11(1): p. 652-657.
36. Kennedy, J. and R. Eberhart, Particle swarm optimization in *Proceedings of IEEE International Conference on Neural Networks*. Piscataway December, 1995.
37. Kumar, G.N., et al., Multi-Objective Optimal Power Flow Using Metaheuristic Optimization Algorithms With Unified Power Flow Controller to Enhance the Power System Performance, in *Advancements in Applied Metaheuristic Computing*. 2018, IGI Global. p. 1-33.
38. Kumar, S., D. Datta, and S.K. Singh, Black hole algorithm and its applications, in *Computational intelligence applications in modeling and control*. 2015, Springer. p. 147-170.

39. Kumar, V., J.K. Chhabra, and D. Kumar, Grey wolf algorithm-based clustering technique. *Journal of Intelligent Systems*, 2017. 26(1): p. 153-168.
40. Lakshmi, K., N.K. Visalakshi, and S. Shanthi, Data clustering using K-Means based on Crow Search Algorithm. *Sādhanā*, 2018. 43(11): p. 190.
41. Lei, X., X. Huang, and A. Zhang. Improved artificial bee colony algorithm and its application in data clustering. in *Bio-Inspired Computing: Theories and Applications (BIC-TA)*, 2010 IEEE Fifth International Conference on. 2010. IEEE.
42. Lenin, K., B.R. Reddy, and M.S. Kalavathi, Dwindling of active power loss by enhanced black hole algorithm. *Int. J. Res. Electron. Comm. Tech.*, 2014. 1(4): p. 11-15.
43. Li, H., et al., Genetic algorithm search space splicing particle swarm optimization as general-purpose optimizer. *Chemometrics and Intelligent Laboratory Systems*, 2013. 128: p. 153-159.
44. Malik, A., et al., The Implementation of a Hybrid Model for Hilly Sub-Watershed Prioritization Using Morphometric Variables: Case Study in India. *Water*, 2019. 11(6): p. 1138.
45. Mangat, V., Survey on particle swarm optimization based clustering analysis, in *Swarm and evolutionary computation*. 2012, Springer. p. 301-309.
46. Marinakis, Y., M. Marinaki, and N. Matsatsinis, A stochastic nature inspired metaheuristic for clustering analysis. *International Journal of Business Intelligence and Data Mining*, 2008. 3(1): p. 30-44.
47. Mirjalili, S., S.M. Mirjalili, and A. Lewis, Grey wolf optimizer. *Advances in engineering software*, 2014. 69: p. 46-61.
48. Nemati, M. and H. Momeni, Black holes algorithm with fuzzy Hawking radiation. *International Journal of Scientific & Technology Research*, 2014. 3(6): p. 85-88.
49. Niu, B., et al., A population-based clustering technique using particle swarm optimization and k-means. *Natural computing*, 2017. 16(1): p. 45-59.
50. Osman, I.H. and G. Laporte, *Metaheuristics: A bibliography*. 1996, Springer.
51. Pandey, A.C., D.S. Rajpoot, and M. Saraswat. Data clustering using hybrid improved cuckoo search method. in *Contemporary Computing (IC3)*, 2016 Ninth International Conference on. 2016. IEEE.
52. Pashaei, E. and N. Aydin, Binary black hole algorithm for feature selection and classification on biological data. *Applied Soft Computing*, 2017. 56: p. 94-106.
53. Pashaei, E., M. Ozen, and N. Aydin. An application of black hole algorithm and decision tree for medical problem. in *2015 IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE)*. 2015. IEEE.
54. Piotrowski, A.P., J.J. Napiorkowski, and P.M. Rowinski, How novel is the “novel” black hole optimization approach? *Information Sciences*, 2014. 267: p. 191-200.
55. Premalatha, K. and R. Balamurugan. A nature inspired swarm based stellar-mass black hole for engineering optimization. in *2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. 2015. IEEE.
56. Raidl, G.R. and J. Puchinger, Combining (integer) linear programming techniques and metaheuristics for combinatorial optimization, in *Hybrid metaheuristics*. 2008, Springer. p. 31-62.
57. Rodrigues, D., et al. Black hole algorithm for non-technical losses characterization. in *2015 IEEE 6th Latin American Symposium on Circuits & Systems (LASCAS)*. 2015. Ieee.
58. Runkler, T.A., Ant colony optimization of clustering models. *International Journal of Intelligent Systems*, 2005. 20(12): p. 1233-1251.
59. Saida, I.B., K. Nadjet, and B. Omar, A new algorithm for data clustering based on cuckoo search optimization, in *Genetic and Evolutionary Computing*. 2014, Springer. p. 55-64.
60. Salih, S.Q. and A.A. Alsewari, Solving large-scale problems using multi-swarm particle swarm approach. *International Journal of Engineering & Technology*, 2018. 7(3): p. 1725-1729.
61. Salih, S.Q., A.A. Alsewari, and Z.M. Yaseen. Pressure Vessel Design Simulation: Implementing of Multi-Swarm Particle Swarm Optimization. in *Proceedings of the 2019 8th International Conference on Software and Computer Applications*. 2019. ACM.
62. Salih, S.Q., et al. Novel Multi-swarm Approach for Balancing Exploration and Exploitation in Particle Swarm Optimization. in *International Conference of Reliable Information and Communication Technology*. 2018. Springer.
63. Sarstedt, M. and E. Mooi, Cluster analysis, in *A concise guide to market research*. 2019, Springer. p. 301-354.
64. Senthilnath, J., et al. Clustering using levy flight cuckoo search. in *Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012)*. 2013. Springer.
65. Senthilnath, J., S. Omkar, and V. Mani, Clustering using firefly algorithm: performance study. *Swarm and Evolutionary Computation*, 2011. 1(3): p. 164-171.
66. Shelokar, P., V.K. Jayaraman, and B.D. Kulkarni, An ant colony approach for clustering. *Analytica Chimica Acta*, 2004. 509(2): p. 187-195.

67. Sun, J., W. Xu, and B. Ye, Quantum-behaved particle swarm optimization clustering algorithm. *ADVANCED DATA MINING AND APPLICATIONS, PROCEEDINGS*, 2006. 4093: p. 340-347.
68. Tvrdik, J. and I. Křivý, Hybrid differential evolution algorithm for optimal clustering. *Applied Soft Computing*, 2015. 35: p. 502-512.
69. Wan, M., et al., Data clustering using bacterial foraging optimization. *Journal of Intelligent Information Systems*, 2012. 38(2): p. 321-341.
70. Xiao, X., et al. Gene clustering using self-organizing maps and particle swarm optimization. in *Parallel and Distributed Processing Symposium*, 2003. Proceedings. International. 2003. IEEE.
71. Yaghoobi, S., S. Hemayat, and H. Mojallali. Image gray-level enhancement using Black Hole algorithm. in *2015 2nd International Conference on Pattern Recognition and Image Analysis (IPRIA)*. 2015. IEEE.
72. Yang, F., T. Sun, and C. Zhang, An efficient hybrid data clustering method based on K-harmonic means and Particle Swarm Optimization. *Expert Systems with Applications*, 2009. 36(6): p. 9847-9852.
73. Yang, X.-S. and S. Deb. Cuckoo search via Lévy flights. in *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*. 2009. IEEE.
74. Yang, X.-S. Engineering optimizations via nature-inspired virtual bee algorithms. in *International Work-Conference on the Interplay Between Natural and Artificial Computation*. 2005. Springer.
75. Yang, X.-S., A new metaheuristic bat-inspired algorithm, in *Nature inspired cooperative strategies for optimization (NICSO 2010)*. 2010, Springer. p. 65-74.
76. Yang, X.-S., *Nature-inspired metaheuristic algorithms*. 2010: Luniver press.
77. Zabihi, F. and B. Nasiri, A Novel History-driven Artificial Bee Colony Algorithm for Data Clustering. *Applied Soft Computing*, 2018. 71: p. 226-241.