Evolutionary Intelligence (2022) 15:759–783
https://doi.org/10.1007/s12065-020-00562-x

RESEARCH PAPER

A new metaheuristic algorithm based on water wave optimization for data clustering

Arvinder Kaur[1] · Yugal Kumar[1]

Received: 17 October 2020 / Revised: 20 December 2020 / Accepted: 30 December 2020 / Published online: 27 January 2021

Abstract
Data clustering is an important activity in the field of data analytics. It can be described as unsupervised learning for grouping the similar objects into clusters. The similarity between objects is computed through distance measure. Further, clustering has proven its significance for solving wide range of real-world optimization problems. This work presents water wave optimization (WWO) based metaheuristic algorithm for clustering task. It is seen that WWO algorithm is an effective algorithm for solving constrained and unconstrained optimization problems. But, sometimes WWO cannot obtain promising solution for complex optimization problems due to absence of global best information component and converged on premature solution. To address the absentia of global best information and premature convergence, some improvements are inculcated in WWO algorithm to make it more promising and efficient. These improvements are described in terms of modified search mechanism and decay operator. The absentia of global best information component is handled through updated search mechanism. While, the premature convergence is addressed through a decay operator. The performance of WWO algorithm is evaluated using thirteen benchmark clustering datasets using accuracy and F-score parameters. The simulation results are compared with several state of art existing clustering algorithms and it is observed proposed WWO clustering algorithm achieves a higher accuracy and F-score rates with most of clustering datasets as compared to existing clustering algorithms. It is also showed that the proposed WWO algorithm improves the accuracy and F-score rates an average of 4% and 7% respectively as compared to existing clustering algorithm. Further, statistical test is also conducted to validate the existence of proposed WWO algorithm and statistical results confirm the existence of WWO algorithm in clustering field.

Keywords Clustering · Data analysis · Meta-heuristic algorithms · Water wave optimization · Unsupervised learning

1 Introduction

Clustering is one of important data analysis method in the field of data mining. It is also known as unsupervised learning. The aim of unsupervised learning method is to explore the capabilities of the data without prior information of class label and due to this nature, these methods are widely adopted in machine learning field [1]. It discovers the unseen and unidentified pattern underlying in the given data, and further, partitions the data into different clusters based on either dissimilar characteristics or using a similarity measure. The underlying pattern and structure of data can be used for decision making, predicting future value and in diagnosis process. In past few decades, clustering proven its potentiality in various research fields such as web analysis [2], business [3], marketing [4], education [5], data science [6], medical diagnosis [7], image segmentation [8], text mining [9], bioinformatics [10], wireless sensor networks [11], text clustering [12] and financial analysis [13] etc. In practice, the clustering algorithms are divided into five sub classes. These are (1) partitional, (2) hierarchical, (3) density, (4) graph, and (5) optimization-based clustering algorithms. Each sub class having different mechanism to work with data, advantages and shortcomings [14]. The Partitional algorithms divide the data into multiple horizontal partitions and these partitions are optimal in nature using a similarity function [15]. However, these algorithms are sensitive to initial cluster centers, in turn, algorithm will be

✉ Yugal Kumar
yugalkumar.14@gmail.com

Arvinder Kaur
er.arvinderdhillon@gmail.com

[1] Department of Computer Science and Information Technology, Jaypee University of Information Technology, Waknaghat, Himachal Pradesh, India

converged on premature solution and also faced difficulty to partition the overlap data [16]. The hierarchical clustering algorithms explore the data through tree structure, but these algorithms cannot have prior information regarding number of clusters. But these algorithms are computationally extensive than partitional clustering algorithms [17]. The clusters with arbitrary shapes are determined through density-based clustering algorithms and these algorithms are more efficient to find outliers in data. Still, these algorithms are less adequate with high dimensional data and clusters with varying densities [18]. The graph-based clustering algorithms can be described as to divide the vertices into k-clusters by considering the edge structure of given graph and this arrangement considers many edges within each cluster and smaller number of edges in between clusters [19]. These algorithms are not suited well for dataset with large number of features, but works efficiently with minor features [20]. In present time, optimization-based clustering algorithms get wide attention from research community for solving clustering problems. These algorithms are more competitive than traditional algorithms and provide more attractive solution for clustering problems [21–24]. Few of these are Tabu Search (TS) [25], Simulated Annealing (SA) [26], Genetic Algorithm (GA) [27], Artificial Bee Colony (ABC) [28–30], Arrhenius Artificial Bee Colony algorithm [31], Ant Colony Optimization (ACO) [32, 33], Particle Swarm Optimization (PSO) [34], Cuckoo Search (CS) [35], Cat Swarm Optimization (CSO) [36, 37], Firefly Algorithm (FA) [38, 39], Gravitational Search Algorithm (GSA) [40, 41], Black Hole Algorithm (BH) [42], Charge System Search Algorithm (CSS) [43, 44], Teacher Learning Based Optimization (TLBO) [45], Artificial Chemical Reaction Optimization (ACRO) [46] and Big Bang-Big Crunch algorithm (BB-BC) [47, 48]. Furthermore, these algorithms consist of several inbuilt mechanisms for refining promise solutions and also explore the solution through local search and global search. The search process can be characterized through self-sustaining, dispersed and inhabitant behavior. These features made the optimization algorithms more powerful than traditional algorithms and competent to solve diverse problems. It is observed that exploitation and exploration are key foundations of population and meta-heuristic algorithms. The exploitation can be interpreted as discovering the candidate solution near to current solution, whereas, exploration can be described as searching of new candidate solution distant from the current solution location. To achieve the good solution for optimization problems, exploitation and exploration processes should be balanced [27]. The several other factors also consider for balancing the aforementioned processes such as searching the problem space, conflicts search objectives, dominating factors of search process. Several studies have been reported for effective balancing of the search processes of meta-heuristic and population-based algorithms [44]. The other aspect of meta-heuristic algorithm is stagnant in local optima [45]. It can be described as no change in the fitness function/cluster allocation in successive iterations. It also occurs due to lack of population diversity during the execution of algorithms. The local optima can affect the final optimal solution and also lead to premature convergence of algorithms. Several researchers focused on well-known local optima problem of clustering algorithm and presented some intelligent and effective solutions and strategies to overcome this problem [48]. The initialization sensitivity can also be considered as one of important problem related to clustering algorithm [49]. As, all clustering algorithm especially partitional clustering choose the initial cluster centers in random order and these selected centers have great impact on the final optimal solution. The empirical studies also considered this issue of clustering algorithm and reported promising solution for the same [50, 51].

In recent time, an algorithm inspired through water waves is developed, called water wave optimization (WWO) algorithm [52]. WWO attracts the attention of research community due to ease of implementation and simplicity. The key features of WWO algorithm are its diversity and adaptation mechanisms which makes the algorithm more suitable for diverse applications. In turn, wide range of optimization problems have been solved using WWO such as scheduling problems [52, 53], radio cognitive system [54], global optimization [55], multi objective optimization [56], parameter optimization of neural network [57], reactive power dispatch problem [58], feature selection [59], congestion control and quality of service in Wireless sensor networks [60], etc. However, WWO algorithm achieves at par results for most of optimization problems, but sometimes its performance is affected due to absentia of the global best information and converged to premature solutions [61, 62].

1.1 Motivation and contribution of work

This research aim is to address the premature convergence and absentia of global best information issues of WWO algorithm. It is seen that due to aforementioned issues, sometimes WWO algorithm could not attain global optimal solution and converged on local best solution [61]. It is also observed that premature convergence issue occurs due to lack of balance between local and global searches. So, this study also investigates the balancing factor between local and global searches. Furthermore, for enhancing the balance between local and global searches and also to handle premature convergence, a decay operator is incorporated into WWO algorithm. The aim of decay operator is to maintain the balancing between local and global searches; and explore search space effectively for addressing premature convergence issue. The absence of global best information issue is resolved through Particle Swarm Optimization (PSO) based search mechanism. The aim of this search mechanism is to guide the search towards the global optimal solution. Finally, the capabilities of improved WWO are explored for solving data clustering problems. Several benchmarks clustering datasets are taken into consideration for evaluating the performance of WWO algorithm. The simulation results are compared with wide range of meta-heuristic clustering algorithms. The key points of this research work are summarized as:

- To incorporate a decay operator into WWO algorithm for effectively balancing the local and global searches as well as premature convergence.
- The global search mechanism of WWO is improved through PSO based search mechanism. The algorithm adopts decay operator to make a balance between exploration and exploitation.
- The capability of improved WWO algorithm is explored for solving data clustering.
- Thirteen benchmark clustering datasets are considered for evaluating the performance of WWO algorithm.
- The statistical analysis is also performed for validating the proposed WWO algorithm.

1.2 Organization of paper

The remaining section of paper is organized in the following manner. Section 2 discusses the related works in the direction of partitional clustering algorithms along with research gap. Section 3 describes the basic WWO algorithm. Section 4 gives the description of improvements and proposed WWO algorithm with flowchart. Section 5 demonstrates the experimental results of proposed WWO clustering algorithm. Finally, the work is concluded with future scope in Sect. 6.

2 Literature review

2.1 Related works

This section discusses the various recent related works for partitional clustering.

Singh [63] introduced harris hawk's optimization (HHO) algorithm for solving data clustering problems in efficient manner. Further, this work considers a chaotic sequence number for guiding the search pattern of HHO algorithm and also overcomes the dependency of HHO algorithm on random numbers. The performance of HHO algorithm has been evaluated using twelve benchmark datasets and compared with six state of art techniques. The efficacy of proposed HHO algorithm is validated using several performance measures and statistical test. These tests confirm the effectiveness of HHO algorithm for solving data clustering problems.

For improving the effectiveness and efficiency of clustering, Tsal et al. [64] developed a new meta-heuristic algorithm, called coral reef optimization with substrate layers (CRO-SL) for handling large amount of data. The substrate layers concept integrates the particle swarm optimization (PSO) and genetic-k-means algorithm (GKA) for refining the end results. The aim of CRO-SL algorithm is to achieve better cluster result for big data analytics. Moreover, the proposed CRO-SL algorithm is implemented on cloud platform for reducing the response time of data analytics. Several state of art clustering algorithms like k-mean, GKA, PSO, simple coral reef optimization (SCRO) are picked for performance comparison. Seven benchmark and two artificial datasets are taken for implementing the CRO-SL algorithm. The simulation results showed that proposed CRO-SL algorithm accelerates the clustering results as compared to other algorithm using cloud platform.

Kuwil et al. [65] designed a distance-based clustering algorithm, called critical distance clustering algorithm. The proposed algorithm considers a new objective function for determining the similarity between data. The objective function is devised using Euclidean distance and basic statistics operation of mathematics. Moreover, the proposed algorithm can be worked with quantitative data, not qualitative, and categorical data. The performance of proposed algorithm is examined over twenty-six experiments. The simulation results proved that proposed algorithm produces the competitive clustering results as compared to k-means, DBSCAN and MST based clustering. It is also claimed that outliers are successfully handled through proposed distance-based algorithm.

Singh et al. [46] considered the slow convergence and local optima issues of clustering algorithms and developed a new heuristic algorithm, called artificial chemical optimization (ACRO) clustering algorithm. The convergence and local optima issues are addressed through position-based operator and neighborhood operator respectively. The performance of ACRO algorithm is tested over five benchmark and two artificial datasets. The standard clustering algorithms are taken to compare the simulation results of ACRO. The results showed that ACRO algorithm obtains better data clustering results in terms of intra cluster distance and f-measure. Furthermore, Friedman statistical test is also applied to validate the performance of ACRO. The results of statistical test confirm the effectiveness of the proposed ACRO algorithm in clustering filed.

Baalamurugan and Bhanu [66] introduced an efficient stud krill herd clustering (ESKH-C) technique to address data clustering in cloud environment. The objective of ESKH-C technique is to compute the optimum locations of clusters centers. Further, stud selection and crossover operator (SSC) has been integrated into krill herd clustering algorithm to make it more efficient. The SSC operator is inspired through genetic reproduction process and aim of this operator is to improve the convergence rate. The seven experiments are conducted for measuring the efficacy of ESKH-C algorithm. The simulation results are compared with k-means, PSO, ant colony optimization (ACO) and bacterial foraging algorithm (BFO) algorithms. The simulation results demonstrated that ESKH-C algorithm effectively works with different clusters number, densities and multi-dimensional datasets.

Sharma and Chhabra [67] considered the lack of equivalence between exploration and exploitation processes of clustering algorithms and developed a new clustering algorithm inspired through PSO and polygamous crossover, called PSOPC. The seven standard clustering datasets are taken to implement the PSOPC algorithm and simulation results are compared with PSO, genetic algorithm (GA), differential evolution (DE), firefly algorithm (FA) and grey wolf optimization (GWO). The cluster distance, cluster quality and convergence rate measures are adopted to evaluate the performance of PSOPC algorithm. It is stated that PSOPC algorithm outperforms in context of aforementioned measures.

Abdulwahab et al. [68] examined the exploration issue of clustering algorithms and designed an effective clustering algorithm, called Levy Flight Black Hole (LBH). The LBH algorithm is the combination of levy flight concept and black hole optimization algorithm. Authors stated that black hole (BH) algorithm provides superior results with benchmark datasets, but having lack of exploration capabilities for few datasets. In turn, BH algorithm cannot explore the search space away from the current black hole. Hence, this shortcoming of BH algorithm is handled through levy flight concept and aim of this concept is to increase the step size for the movement of star so that large search space can be explored for solution search. The performance of LBH algorithm is evaluated using six standard dataset and compared with several clustering algorithms. The results indicated that LBH algorithm displays robust clustering results.

Mustafa et al. [69] taken into consideration the equivalency issue of exploration and exploitation processes of clustering algorithm and developed an adaptive memetic differential evolution (ADME) optimization algorithm for data clustering. The ADME algorithm contains the advantages of memetic algorithm and adaptive differential evolution (DE) algorithm. In ADME algorithm, adaptive differential evolution mutation strategy is employed in memetic algorithm for better compromise between local and global searches. The experimental results specified that ADME algorithm can obtain more accurate clustering results than ME and DE algorithms. Furthermore, the statistical test also validates the proposed ADME algorithm in clustering filed.

Tarkhaneh and Moser [70] developed an improved differential evolution (IDE) algorithm for data clustering. IDE algorithm integrates Archimedean spiral, Mantegna levy distribution, and neighborhood search (NS) for effective cluster analysis, called adaptive differential evolution with neighborhood search (ADENS). Twelve experiments are conducted to investigate the performance of ADENS algorithm. The results showed that ADENS algorithm archives superior clustering results as compared to other algorithms. Further, Wilcoxon and Friedman statistical tests are also considered to validate the ADENS algorithm. The statistical results support the existence the ADENS algorithm for cluster analysis.

The random selection of initial cluster centers incurs the premature convergence sometimes, especially in clustering algorithms. Agbaje et al. [50] investigated the aforementioned issue of clustering algorithm by combining firefly algorithm (FA) and PSO algorithm, called FAPSO. In proposed FAPSO algorithm, initially FA algorithm is implemented to start the initial search and further, the PSO algorithm is employed for obtaining the optimal solution. The robustness of the proposed algorithm is assessed over twelve benchmark datasets and compared with four standard benchmark clustering algorithms. The simulation results illustrated that FAPSO having advantage over other clustering methods in terms of DB index and CS index.

Zhou et al. [49] investigated the dependency of K-means algorithm on initial solution and found that algorithm could be stuck in local optima if initial solution is not explored. In this work, authors propose an efficient algorithm inspired through symbiotic organism search (SOS) for data clustering. Ten experiments are conducted to examine the efficacy of SOS algorithm and simulation results are compared with DE, cuckoo search (CS), flower pollination algorithm (FPA), PSO, artificial bee colony (ABC), multi-verse optimizer, and k-means. Authors claimed that SOS algorithm generates more stable clustering results.

Aljarah et al. [71] inspected the trap in local optima and premature convergence issues of grey wolf optimizer (GWO) clustering algorithm. Authors stated that the aforementioned problems are occurred due to large number of variables. So, to address local optima and premature convergence issues, a tabu search (TS) method is incorporated in GWO algorithm, called TSGWO. The performance of proposed TSGWO is evaluated using thirteen benchmark clustering datasets and compared with several existing clustering algorithms. The simulation results showed that TSGWO achieves better convergence rate as compared to same class of algorithm and successfully overcomes the local optima issue.

Zhu et al. [72] considered the shortcomings of bat algorithm such as stagnation in local minima and accuracy, and developed an improved bat algorithm for effective cluster analysis. To alleviate the shortcomings of bat algorithm, two improvements are incorporated for improving global and local optimum abilities. The global optimum ability is improved using Gaussian based convergence factor and five different convergence factors. Whereas, local optimum ability is enhanced using whale-based optimization and sine position updating mechanism. The seven clustering datasets are adopted for evaluating the performance of improved bat algorithm. The simulation results showed that proposed improvements enhance the accuracy of bat algorithm in significant manner.

Kushwaha et al. [51] investigated the choice of initial cluster issue of k-means algorithm and propose electromagnetic clustering algorithm (ELMC) to address this problem. The ELMC is an improved version of electromagnetic filed optimization algorithm. Furthermore, the diversity of ELMC algorithm is maintained through concept of attraction–repulsion. A series of experiments are conducted to measure the efficiency of proposed ELMC algorithm and simulation results are compared with ACO, KFCM, KFC, PCM, and standard k-means clustering algorithms. It is observed that ELMC algorithm achieves more stable clustering results than other algorithms.

Senthilnath et al. [73] adopted flower pollination algorithm (FPA) for addressing the data clustering problem. The FPA algorithm is inspired through pollination process of flower. The objective of FPA algorithm is to compute the optimal cluster centers. The performance of FPA is evaluated using three clustering databases and compared with GA, PSO, CS, spider monkey optimization (SMO), GWO, DE, harmony search and bat algorithm. The simulation results demonstrate that FPA algorithm having minimum classification error as compared to rest of algorithms. Furthermore, statistical test is also employed to validate the effectiveness of FPA clustering algorithm. The statistical test proved that FPA is an effective algorithm to deal data clustering problem.

Mageshkumar et al. [74] developed a hybrid meta-heuristic algorithm for improving the efficacy of data clustering. The proposed hybrid algorithm integrates the important capabilities of ant lion optimization (ALO) algorithm with ant colony optimization (ACO) algorithm, called ACO-AOL. Furthermore, local minima problem overridden through Cauchy mutation operator. The four experiments are conducted to evaluate the performance of ACO-AOL algorithm and compared with K-means and ACO clustering algorithm. The simulation results showed that ACO-AOL algorithm obtains superior clustering results.

Kaur et al. [75] examined the local optima and slow convergence issues of K-means algorithm, especially for larger datasets and develop a new clustering algorithm based on chaos optimization and flower pollination algorithm, called chaotic FPA (CFPA). The sixteen datasets are considered to tested the performance of CFPA algorithm. The simulation results are compared with FPA, CS algorithm, BH algorithm, BA, PSO, FA and ABC clustering algorithms. The efficacy CFPA is measured using cluster integrity, execution time, number of iterations to converge (NIC) and stability performance measures. The simulation results showed that CFPA algorithm achieves better clustering results than other algorithms in terms of cluster integrity and execution time.

Xie et al. [76] investigated the sensitivity towards initial clusters and local optima problems of clustering algorithms and provide the solution by developing two variants of FPA algorithm, called inward intensified exploration FPA (IIEFPA) and compound intensified exploration FPA (CIEFPA). Further, the matrix-based search parameters and dispersing mechanisms are incorporated for improving the global and local searches of proposed variants of FPA. The fifteen datasets are adopted to assess the effectiveness of FPA variants. The simulation results showed that proposed FPA variants exhibit superior clustering results as minimum distance and higher accuracy rate than other algorithms.

Huang et al. [77] developed a memetic clustering algorithm based on PSO and GSA, called the memetic particle gravitation optimization (MPGO) algorithm. The aim of this algorithm is to perform efficient search and faster convergence. The important aspects of MPGO algorithm are highlighted as hybrid operation and enhanced diversity mechanism. The performance of MPGO is evaluated on six benchmark clustering datasets and compared with K-means, PSO, GSA], BH algorithm and WOA algorithm. The simulation results stated that MPGO outperforms than other algorithms in terms of better fitness function and more accurate clustering rate.

Dinkar and Deep [78] addressed the local optima and slow convergence issues of clustering algorithm and develop an improved ant lion optimization (ALO) algorithm, called OB-C-ALO, for performing data clustering in efficient manner. To make the algorithm more competitive, two amendments are integrated into ALO algorithm. These amendments are i) employ of Cauchy mutation operator for handling problem of local minima, and ii) utilizes opposition-based learning for addressing slow convergence rate. The six experiments are conducted to evaluate the performance of OB-C-ALO clustering algorithm. The simulation results are compared with AO and C-ALO clustering algorithm. It is noticed that OB-C-ALO algorithm obtains more promising result in terms of distance than others.

To improve the global search mechanism, Abualigah et al. [79] proposed hybrid algorithm (H-KHA) for solving data clustering and text clustering problem. In this work, krill herd (KH) algorithm is hybridized with harmony search (HS). The distance factor of HS has been adopted for improving the global search mechanism in KH. The performance has been evaluated on both seven data clustering datasets and six text document datasets. It is noticed from the results that proposed algorithm achieves state-of art results in terms of accuracy and convergence rate. The statistical analysis shows highest rank for H-KHA using F-measure as compared to other clustering and optimization algorithms in comparison. The work can be extended by using other powerful local search approaches, handle different clustering problem and evaluation of H-KHA on benchmark function datasets.

A hybrid proposal distribution method for pattern recognition was developed by Zeng et al. [80]. The aim is to estimate segment test and control lines accurately from gold immunochromatographic strip (GICS) images. A new dynamic state-space model has been adopted to describe relation between contour points on upper and lower boundaries using transition equation for test, and control lines. With uniformity measure and class variance, new observation equation has been developed. Further, deep-belief-network-based particle filter (DBN-PF) has been adopted to fins initial recognition and regions of high likelihood. The experimentation has been done using artificial dataset and GICS image. From the results, it is evaluated that proposed approach has resulted in superior performance for GICS images and significant improvement has been shown in terms of several indices. Further, this can be used with other approaches to combat the problems of particle filters.

Zeng et al. [81] proposed a dynamic-neighborhood-based switching PSO (DNSPSO) algorithm for improving exploration ability. The proposed algorithm has adopted dynamic neighborhood strategy to adjust personal and global best positions to overcome premature convergence problem. A new learning strategy has been developed to select acceleration coefficients, and update velocity in order to search the complete search space. Further, differential evolution method has been utilized for improving the diversity of PSO. The performance has been evaluated using fourteen benchmark functions. The experimental results have demonstrated that proposed algorithm gives 100% successful rate and overall ranks second for success performance on all benchmark functions. The DNSPSO can be applied to other research areas such as self-organizing RBF neural network, moving horizon estimation etc.

Abualigah [82] provided a comprehensive survey of group search optimizer (GSO). The various variants of GSO, results and its applications have been discussed from the year 2009 to 2020. From the set of candidate solution, GSO algorithm is able to find best solution. It helps to determine maximum or minimum objective function to solve the optimization problem. From the survey, it has been noticed that GSO is competent and gives promising results as compared to similar optimization algorithms. The study concludes that significant improvement can be done in the performance by enhancing GSO algorithms with different modifications, improvements, hybridizations or as per the need of the problem. It can be employed for solving multi-objective problems, and unsolved optimization problems by GSO. It can be modified for real-world problems.

A comprehensive review of multi-verse optimizer algorithm (MOA) from March-2015 till April- 2019 was presented by Abualigah [83]. The various features, advantages, disadvantages and its applications have been discussed. The study has also discussed various variants of MOA including-binary, hybridized, modified, multiobjective and chaotic. The performance of MOA has been evaluated for unimodal and multimodal functions; and compared to other related approaches. It has been depicted that MOA gives high exploration ability along with adjustable convergence rates. The study presents various future directions. It can be modified to solve real-world optimization problems and unsolved optimization problems. Further, it can be hybridized with other methods like hill climbing, differential evolution for significant improvement of results.

Zeng et al. [84] developed a framework for diagnosis of Alzheimer's disease (AD) and mild cognitive impairment (MCI). The model consists of pre-processing magnetic resonance (MRI) images, feature extraction, principal component analysis, and support vector machine (SVM). The proposed model has adopted switching delayed with PSO(SDPSO) in order to optimize SVM parameters. The proposed model has been evaluated over MRI scans taken from Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset for classification od AD and MCI. The results demonstrate that the developed framework gives the classification accuracies of 69.2308% for stable MCI (sMCI) vs progressive MCI (pMCI), 81.25% for Normal Control (NC) vs AD, 76.9231% for NC vs sMCI, 85.7143% for NC vs pMCI, 71.2329% for sMCI vs AD and 57.1429% for pMCI vs AD. The proposed method can be investigated with deep learning methods and also positron emission tomography (PET) image can be considered for AD diagnosis problems.

2.2 Research gaps

This subsection focuses on the research gaps in the existing studies. In related works section, twenty recently published research articles are discussed to determine the research gaps and all these articles are published in journals of repute. Through literature review, it is observed that data clustering problem attracts the wide attention from research community and large number of algorithms are reported for solving the data clustering problem in efficient manner. Further, Table 1 demonstrates the pros and cons of the metaheuristic algorithms as well as clustering problem. The several points are noted regarding clustering algorithms and data clustering problem and can be summarized as (1) recently, meta-heuristic algorithms take over the traditional algorithm for solving data clustering problems, (2) meta-heuristic

Table 1 Summarization of related works

Author Name	Problem Identified	Methods/Findings	Data Sets	Future Work
Singh [63]	Dependency on random numbers Improving the performance To handle large data	Proposed harris hawk's optimization (HHO) algorithm. Adopted a chaotic sequence of number for guiding the search pattern of HHO algorithm.	Jain, Flame, Compound, D31, R15, Aggregation, Path-based, Spiral, Wine, Glass, Iris, Yeast	For real-world application Multi-objective version
Tsai et al. [64]	To handle large data Better cluster result	Coral reef optimization with substrate layers (CRO-SL) was developed for handling large amount of data. The substrate layers concept integrates the particle swarm optimization (PSO) and genetic-k-means algorithm (GKA) for refining the end results	Iris, Wine, Breast Cancer Wisconsin, HTRU2, Spambase, User Locations Finland, Abalone, c20d6n2000, c20d6n200000	Dynamic parameter setting for convergence Hybridize with other metaheuristics algorithms Speed up of Graphics processing unit Using for other real-world applications like image processing, feature selection, WSN
Sharma and Chhabra [67]	Trap in local optima Premature convergence Balance between exploration and exploitation	Proposed a real encoded hybrid algorithm (PSOPC) using PSO for global search and polygamous approach for crossover to refine the exploration and exploitation strategy. Parameters like inertia weight, crossover probability and alpha values in arithmetic crossover are also tuned dynamically to refine the optimization process.	Wine, Glass, Haberman, Bhupa, Cancer, CMC, Iris	Can be extended for automatic clustering Multi objective version can be proposed
Abdulwahab et al. [68]	Trap in local optima Premature Convergence	Proposed a novel data clustering method ''Levy Flight Black Hole (LBH)''. Levy flight has been combined to Black hole algorithm. Movement of each star depends on the step size generated by the Levy distribution. It results in exploration of more area by the star from current black hole and vice-versa.	Iris, Wine, CMC, Glass, Cancer, Vowel	Can be extended to other applications like text document clustering etc.
Mustafa et al. [69]	Balance between exploration and exploitation	Proposed an adaptive memetic differential evolution optimization algorithm (AMADE). Superior mutation strategy has been used by employing adaptive DE with memetic Algorithm.	Wisconsin Breast Cancer, Glass, Vowel, Wine, Iris and Contraceptive Method Choice (CMC)	Can be enhanced by using other objective function for categorical and mixed data. Can be enhanced for multi objective problems.

Table 1 (continued)

Author Name	Problem Identified	Methods/Findings	Data Sets	Future Work
Tarkhaneh and Moser [70]	Convergence Speed Balance between exploration and exploitation	Proposed Adaptive Differential Evolution with Neighborhood Search (ADENS). Adopted new mutation strategy by combining Archimedian Spiral (AS) with Mantegna Levy flight for robust solutions. Self-adaptive strategy is applied for tuning control parameters.Initialization methodology is applied	Iris, Contraceptive method choice (denoted as CMC), Wisconsin breast cancer (WBC), Diabetes, Wine, Yeast, Vehicle, Letters, Liver, Ionosphere, Heart, and Cars	Can be enhanced through quantum or chaotic theory. Adopt more diversified approach for solutions. Considered other applications like text and image clustering
Agbaje et al. [50]	Premature convergence	Combined firefly algorithm (FA) and PSO algorithm, called FAPSO.FA algorithm is employed for initial search. PSO algorithm is employed for obtaining the optimal solution.	Breast, Compound, Flame, Glass, Iris, Jain, Path based, Spiral, Statlog, Thyroid, Two moons, Wine, Yeast	Can be enhanced using Levy flight by reducing foraging time. Local search approaches can be employed for solving real world clustering problems. Can be employed in other engineering optimization problems.
Zhou et al. [49]	Local optima Sensitive to initial selection	SOS algorithm is applied to solve the data clustering problem. New equations have been given for mutualism and commensalism phases.Adopted parasite vector in the parasitism phase	Artificial Dataset one, Artificial Dataset Two, Iris, Wine, Seeds, Statlog, Breast Cancer Wisconsin (Original), CMC, Haberman's Survival, Balance Scale	Can be extended for dynamically determine optimal number of clusters. Can be enhanced to solve complex clustering problems by combining with other approaches. Can be employed for other engineering applications.
Aljarah et al. [71]	Trapped in local optima Premature convergence Balance between exploration and exploitation	Proposed hybrid approach to enhance the efficiency and balance between exploratory and exploitative behaviors of the GWO algorithm. Tabu search has been employed as an operator in GWO	Iris, blood, breast cancer, glass, seeds, wine, Australian, diabetes, Haberman, heart, liver, planning index, tic-tac-toe	Analyzing spatial data and other synthetic datasets. Can be enhanced using Parallel approach for reducing run time.
Zhu et al. [72]	Trapped in local optima Improving accuracy rate	Proposed an improved bat algorithm To improve global search, a Gaussian-like convergence factor has been added. Five more different convergence factors have been proposed for global optimization. Hunting mechanism from whale optimization algorithm (WOA) and the sine position updating strategy are adopted for enhancing local optimization.	Iris, Wine, Bupa, Seeds, Heartstatlog, WDBC, and Wisconsin breast cancer	Can be enhanced for handling unstable search.

Table 1 (continued)

Author Name	Problem Identified	Methods/Findings	Data Sets	Future Work
Kushwaha et al. [51]	Problem of initial choice of clusters Local optima	Introduced an enhanced variant of electromagnetic field optimization (EFO) that is electromagnetic clustering algorithm (ELMC) for clustering. Utilizes the attraction–repulsion concept of the EFO algorithm to maintain the diversity of the population, making it less vulnerable towards the initial choice of centroids.	Iris, Vowel, Crude oil, Thyroid, IONO (Ionosphere database), GAS, Human Activity Recognition, and Contraceptive Method Choice (CMC)	Can be extended in the fields of image and text.
Senthilnath et al. [73]	Prior information for optimal cluster centers	Flower Pollination Algorithm (FPA) based approach is developed for data clustering. FPA is used to minimize the objective function to obtain the optimal solutions for the locations of cluster centers.	Image segmentation, vehicle, glass, Crop Type and Synthetic datasets for test samples	–
Kumar et al. [74]	Balance between exploration and exploitation Trapping in local minima Reduce the intra cluster distance	Proposed new hybrid ACO-ALO algorithm. Employed Cauchy's mutation operator to avoid local minima.	Zoo, iris, glass, wine	Can be hybrid using neural based algorithm for better parameter selection.
Kaur et al. [75]	Local optima Slow convergence for large datasets	Proposed a hybridized algorithm using Chaos optimization and flower pollination over K-means.	Iris, Wine, Breast Cancer, Glass, Balance, Dermatolgy, Haberman, Ecoli, Heart, Spambase, Ilpd, Leaf, Libras, Qualitative Bankruptcy Synthetic	Can be employed using other metaheuristics like krill herd algorithm, spider monkey agent ETC. Can be enhanced for constraint handling problems.
Xie et al. [76]	Initialization sensitivity Local optima trap	Proposed two variants of the Firefly Algorithm (FA)- (i)inward intensified exploration FA (IIEFA) and, (ii) compound intensified exploration FA (CIEFA). Matrix-based search parameters and dispersing mechanisms are incorporated to enhance exploration and exploitation. Adopted minimum Redundancy Maximum Relevance (mRMR)-based feature selection method to reduce feature dimensionality.	ALL-IDB2 database, a skin lesion data set, Sonar, Thyroid, Ozone, Iris, Wisconsin breast cancer diagnostic data set (Wbc1), Wine, Wisconsin breast cancer original data set (Wbc2, Balance, and E. coli.	Can be enhanced using other objective functions. Can be used for other optimization problems like discriminative feature selection, evolving deep neural network generation etc.

Table 1 (continued)

Author Name	Problem Identified	Methods/Findings	Data Sets	Future Work
Huang et al. [77]	Convergence rate Accuracy	Proposed a hybrid memetic clustering algorithm based on PSO and GSA names as memetic particle gravitation optimization (MPGO). Hybrid operation and diversity enhancement are important features of MPGO. PSO helps in exchange of individuals and GSA uses enhancement operator to enhance diversity of system.	Iris, Wine, Breast cancer, Car evaluation, Statlog, and Yeast; 52 benchmark function, six images for image segmentation	Computed tomography (CT) image enhancement using MPGO. Improve computing performance by dimension or pattern reduction. Can be enhanced using levy flight for improving diversity. For automatic object tracking and, text recognition MPGO can be implemented on Raspberry Pi.
Dinkar and Deep [78]	Local optima Slow convergence	Proposed opposition-based ALO using Cauchy distribution (OB-C- ALO). Adopted random walk based on Cauchy distribution for addressing local optima problem. Employed opposition-based learning model along with acceleration coefficient.	Iris, Glass, Wine, CMC, LD, WBC and 21 benchmark test functions	–
Abualigah [79]	Exploration ability Premature convergence	Proposed a novel hybrid of KH algorithm with harmony search (HS) algorithm as H-KHA. Adopted distance factor from HS to improve the global search mechanism in KH.	CMC, Iris, Vowel, Seeds, Cancer, Glass, WineText documents are: Classic4, Classic4, Reuters21578, 20Newsgroup, Reuters21578, 20Newsgroup	Can be investigated on test functions. Can be used for other clustering problems. Can be hybrid with other local search methods.
Zeng et al. [81]	Exploration ability Premature Convergence	Proposed a dynamic-neighborhood-based switching PSO (DNSPSO) algorithm for improving exploration ability. The proposed algorithm has adopted dynamic neighborhood strategy to adjust personal and global best positions to overcome premature convergence problem. Differential evolution method has been utilized for improving the diversity of PSO.	Fourteen benchmark functions	Can be employed in areas such as moving horizon estimation, self-organizing RBF neural network etc.

algorithms provides more effective and promising solution for data clustering than traditional algorithm due to inbuild local and global search mechanisms, (3) hybridization of different algorithms improve the performance as compared to stand alone algorithms, and (4) both simulation and statistical results are considered to validate the performance of clustering algorithms. Apart from these benefits, several issues are also associated with meta-heuristic algorithms such as (1) balance factor between local and global search, (2) stagnation in local minima, (3) sensitive to initial clusters selection, and (4) premature convergence due to lack of diversity in population. These issues affect the performance of meta-heuristic algorithm especially for solving data clustering problems. Lot of works have been reported on local minima and effective balancing between local and global search mechanisms. However, this work addresses the premature convergence and global best information issues of clustering algorithm. The premature convergence problem will be alleviated through a decay operator, whereas, global best information will be incorporated through an improved solution mechanism. This work introduces a water wave optimization (WWO) algorithm for solving data clustering problem in effective manner and the capability of WWO algorithm is improved through decay operator and improved solution search mechanism.

3 Water wave optimization

This section presents the basic Water Wave Optimization (WWO). WWO is a metaheuristic algorithm inspired by water wave theory for solving global optimization problems [52]. WWO considers solution space similar to a seabed area where each solution represents a “wave” using height (h) and wavelength (λ). The fitness of each wave is measured using seabed depth and shorter distance to still water level represents the higher fitness. The population of WWO algorithm is described in terms of waves and each wave is represented through h_{max} and λ equal to 0.5. In WWO, three operations are defined for attaining global optimum such as propagation, refraction, and breaking at each iteration. In propagation operation, a new wave (X’) is created using displacement at each dimension (d) for each wave (X) and added to the original wave as given in Eq. 1.

$$X\prime = X + \mathrm{rand}\,(-1,1) \times \lambda \times L_d \tag{1}$$

where rand is random function for generating random numbers in specified range and L_d is the length for dth dimension of search space. The fitness (f(X′)) of new wave (X’) is higher than the fitness (f(X)) of old wave (X), then replace old wave (X) by new wave (X’) and reset height to h_{max}; Otherwise, wave height is decreased by one.

As deep-water waves have low wave heights and long wavelengths. Similarly, shallow water waves have low wave heights and short wavelengths. So, wavelength decreases if the wave moves from deep water to shallow water. The wavelength (λ) of each wave is calculated using Eq. 2.

$$\lambda = \lambda \times \alpha^{\frac{-(f(X)-fmin+\varepsilon)}{(fmax-fmin+\varepsilon)}} \tag{2}$$

where fmax is maximum and fmin is minimum fitness value within the current population, α is wavelength reduction coefficient parameter, and ε is a small constant used to avoid division-by-zero, f(X) is the fitness of wave X. This helps the propagation of higher fitness waves within smaller ranges and smaller wavelengths.

The refraction operator is adopted when waves height decreases to zero. The new wave (X’) is calculated using a Gaussian function described in terms of mean and standard deviation.

$$X\prime = \mathrm{Gaussian}(\mu, \sigma) \tag{3}$$

In Eq. 3, μ can be described as Mean, whereas σ can be defined as standard deviation and these are computed using Eqs. 4, 5.

$$\mu = \frac{Xbestd + Xd}{2} \tag{4}$$

$$\sigma = \frac{Xbestd - Xd}{2} \tag{5}$$

The mean (μ) is computed using present wave (X) and best wave (X_{bestd}). The standard deviation (σ) can be described as difference between the best wave (X_{bestd}) and present wave (X). Further, the wave height is reset to h_{max}, and wavelength is set using Eq. 6.

$$\lambda\prime = \frac{f(X)}{f(X\prime)} \tag{6}$$

In Eq. 6, λ′ is the wavelength of next wave, f(X′) is the fitness of the new wave (X’), f(X) is the fitness of the old wave (X) and λ is the previous wavelength. In WWO, breaking operator breaks the wave (X), when it attains the better location than the current best solution (X_{best}). The solitary wave (X’) is computed using Eq. 7.

$$X\prime = X + \mathrm{Gaussian}(0,1) \times \beta \times Ld \tag{7}$$

where β denotes the breaking coefficient, Gaussian(0, 1) generates the random number in the range of 0 and 1. If the wave X’ is better than X, then it replaces X. The pseudocode of WWO is mentioned in Algorithm 1.

Algorithm 1: Pseudocode of WWO algorithm

```
Step 1:   Initialize the population P of n waves randomly
Step 2:   While stop criterion is not satisfied do
Step 3:   For each X ε P do
Step 4:    Propagate X to new X′ using equation 1
Step 5:    If f(X′)>f(X) then
Step 6:      If f(X′)>f(X*) then
Step 7:         Break X′ using equation 7
Step 8:         Update X* with X′
Step 9:      Replace X with X′
Step 10:  Else
Step 11:   Decrease X.h by 1
Step 12:   If X.h=0 then
Step 13:  Refract X to new X′ using equation 5 and 6
Step 14:  Update wavelength using equation 2
Step 15:  Return X*
```

4 Proposed WWO clustering algorithm

This section presents a water wave optimization-based clustering algorithm for data clustering problems and also highlights the modification incorporated in WWO algorithm. Subsection 4.1 presents the proposed modifications in WWO, while, the steps of WWO clustering algorithm and flowchart is mentioned in subsection 4.2.

4.1 Proposed modifications

WWO consists of three phases i.e., propagation, refraction, and breaking. WWO algorithm having a strong local search mechanism, but the global search mechanism is weak [62]. Through literature, it is revealed that the PSO algorithm consists of a strong global search mechanism [85]. Hence, to improve the global search mechanism of WWO and also determine the optimal solution, an updated solution search equation inspired through PSO is designed for the propagation phase of the WWO algorithm. It is also observed that the WWO algorithm is also suffered from premature convergence problems due to the refraction phase [61]. In the refraction phase, wave heights decrease continuously, suddenly tends to zero and algorithm converges without obtaining the optimal solution. So, to overcome the premature convergence problem, a decay operator is proposed in the refraction phase of the WWO algorithm. The proposed improvements in the WWO algorithm for addressing weak global search mechanism and premature convergence are discussed below.

4.1.1 Decay operator

The issue associated with the WWO algorithm is premature convergence. In the refraction phase, the height of the wave is continuously decreased and suddenly, it becomes zero. In turn, the algorithm converges without attaining the global best solution. Further, it is observed that WWO algorithm is not explored the entire search space effectively due to small step size and in turn, local search and global search become imbalance. Hence, to overcome the aforementioned problems, a decay operator is integrated with the location updated equation of wave in the refraction phase and also to increase the step size. The updated location search equation is given as:

$$X\prime = N\left(\frac{X^{*} + X}{2}, \frac{\left|X^{*} - X\right|}{2}\right) \times \left[(1 - \rho) + \Delta X\right] \quad (8)$$

In Eq. 8, ρ denotes as decay operator and it can be defined as $\rho \in [0, 1]$ and ΔX represents the difference between two consecutive waves, X represents the current wave and X^{*} denotes local best wave.

4.1.2 Global best information component

To improve the global search mechanism of the WWO algorithm, an updated search equation is proposed to determine the optimal solution. The aim of this equation is also to guide the direction of search towards an optimum solution and exploit the entire search space. The solution search equation of WWO algorithm is given as Eq. 1.

where in Eq. 1, $X\prime$ describes the new wave, X denotes the current wave, rand function generates a random number in the interval of − 1 to 1, L(d) denotes the length of search space and λ describes wavelength. Through Eq. 1, it is observed that the new wave is generated using old wave, random function, and wavelength. It is noticed that the new wave is generated without any guidance of global best and local best information of wave. In turn, the global exploration capability of WWO is affected. Hence to improve the global search mechanism, an updated search equation is developed based on the PSO concept which is mentioned in Eq. 9.

$$X\prime = X + C_{best} + \mathrm{rand}(-1, 1) \times \lambda \times L(d) \tag{9}$$

4.1.3 Improved WWO cluster formulation

This research work applies improved WWO for solving data clustering problems. In data clustering, a dataset consists of n number of data objects such as $(X = X_1, X_2, X_3 \ldots X_n)$ and d number of dimensions i.e., $(X = X_{1,1}, X_{1,2}, X_{1,3} \ldots X_{1,d})$. Each data object is repersented using $X_{i,j}$, where X_{ij} represents ith object with jth dimension. The vector representation of data is given as $X_{i,j} = X_{i,1}, X_{i,2} \ldots X_{i,3}..X_{i,d}$. The objective of data clustering is divided the dataset into K distinct clusters such as $(C = C_1, C_2, C_3, \ldots, C_K$. In this work, wave is considered as cluster $C_k \in (k = 1, 2, \ldots, K)$. The aim is to arrange the data objects $X \in \left(X_1, X_2, X_3 \ldots X_n\right)$ into cluster $C \in \left(C_1, C_2, C_3 \ldots C_K\right)$ with minimum distance. The distance between the data objects and clusters is computed using Eq. 10.

$$D\left(X_i, C_k\right) = \sqrt{\sum_{j=1}^{d} \left(X_{ij}, C_{kj}\right)^2} \tag{10}$$

where $D\left(X_i, C_k\right)$ denotes distance between ith data object and kth cluster, $X_i$and C_k denote ith data object of dataset (X) and kth cluster (C). After computing the distance between data objects and each cluster center, the data objects are assigned to clusters using minimum distance. The accuracy of clusters is evaluated using assigned data objects. Further, the computational steps of WWO algorithm are used to determine the updated cluster positions. Again, the distance between the data objects and new cluster centers are computed and assigned the data objects to nearest clusters. The above-mentioned process is repeated until either optimal cluster centers are not obtained or termination condition occurs. Hence, the WWO algorithm is used to obtain optimal cluster center such that higher accuracy can be achieved.

4.2 Computational steps of improved WWO algorithm

The algorithmic steps of the proposed WWO algorithm are mentioned in Algorithm 2, while flowchart of proposed WWO algorithm for cluster analysis is shown in Fig. 1.

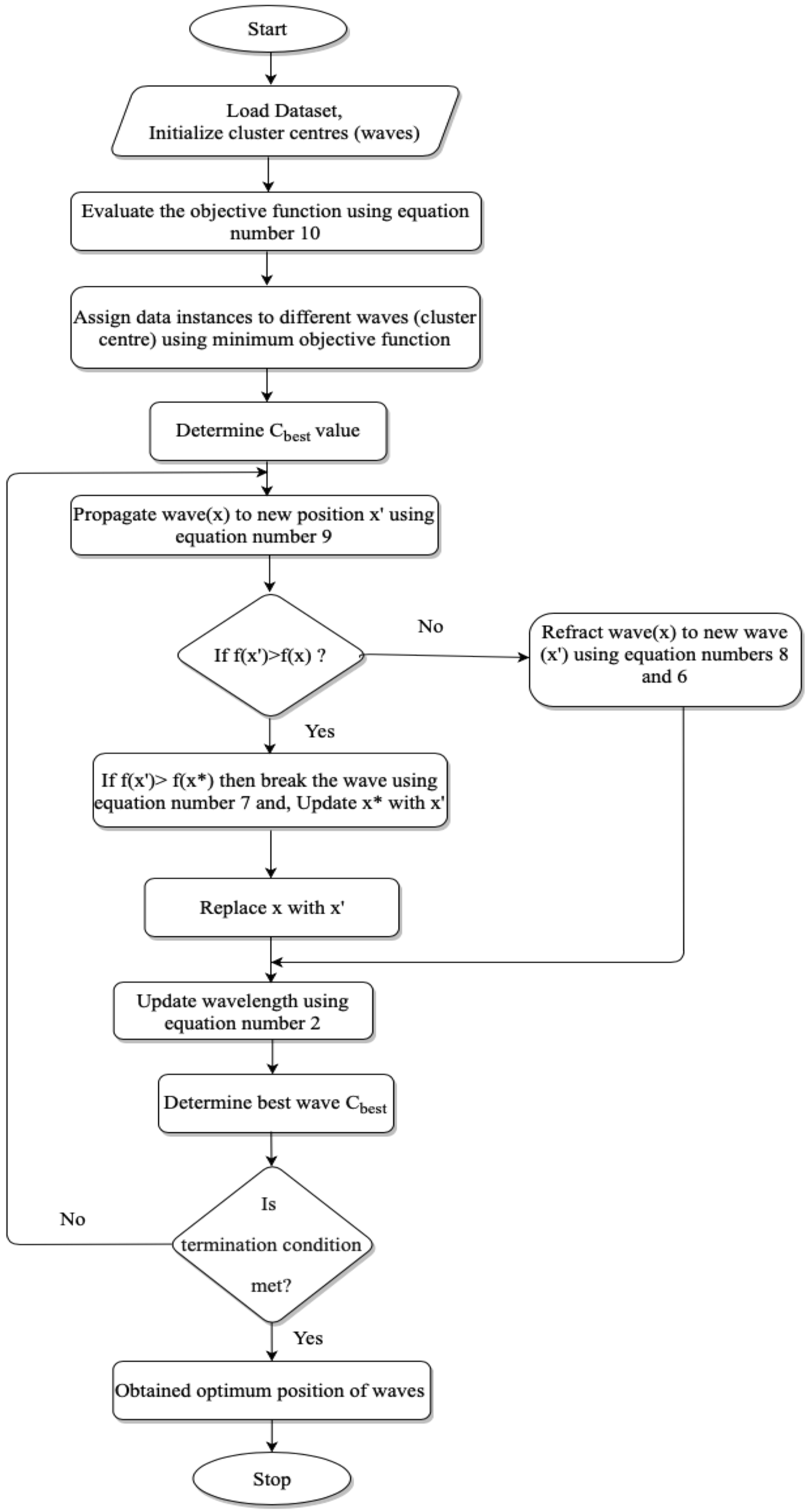
Table 2 Descriptions of different clustering datasets

Sr. no.	Datasets	Clusters (K)	Instances	Dimension
1	Iris	3	150	4
2	Wine	3	178	13
3	Vowel	6	871	3
4	Balance	3	625	4
5	Glass	7	214	9
6	CMC	3	1473	9
7	Thyroid	3	215	5
8	Dermatology	6	358	34
9	BC	2	683	9
10	WDBC	2	569	30
11	LD	2	345	6
12	Heart	2	270	13
13	Diabetes	2	768	8

Table 3 Parameter setting for proposed WWO

Parameter	Value
h_{max}	12
α	1.2
k_{max}	12
β	Linearly decreases from 0.25 to 0.01
n	10

Fig. 1 Flowchart of proposed WWO algorithm for clustering



Algorithm 2: Pseudocode of Improved WWO algorithm	
Step 1:	Initialize the population of wave (C) such as $C_j \in (j = 1, 2, ..., n)$
Step 2:	Evaluate the objective function value using equation 10.
Step 3:	Assign the data instance to different waves using minimum objective function value and determine the best wave (C_{best}).
Step 4:	While (stopping condition is not met), do the following
Step 5:	For each wave $x \in C$
Step 6:	Propagate the wave (x) to a new position x' using equation 9.
Step 7:	If $f(x') > f(x)$ then
Step 8:	If $f(x') > f(x^*)$
Step 9:	Break the wave x' using equation 7.
Step 10:	Update the x^* with x'
Step 11:	Replace x with x'.
Step 12:	Else, Refract the wave (x) to new x' using equations 8 and 6.
Step 13:	Update the wavelength using equation 2.
Step 14:	Determine the best wave (C_{best})
Step 15:	End while
Step 16:	Compute the optimum position of waves

5 Experimental results and discussion

This section presents the experimental results of the proposed WWO clustering algorithm. The performance of the proposed algorithm is assessed over thirteen benchmark clustering datasets. These datasets were extracted from the UCI repository. Table 2 presents descriptions of these datasets. The proposed algorithm has been implemented on window operating system in MATLAB 2010a on corei5 processor with 4 GB. Table 3 gives the parameter setting for proposed WWO.

Table 4 Demonstrate experimental results of the proposed WWO clustering algorithm in contrast to existing clustering algorithms based on the accuracy performance metric

Dataset	Algorithm					
	FCM	PSO	K-means	GA	WWO	Proposed WWO
Iris	**99.33**	84.13	78.53	78.34	85.23	93.21
Wine	70.22	67.94	67.61	65.73	66.82	**74.63**
Vowel	78.68	84.91	73.45	84.7	84.66	**89.47**
Balance	86.64	**89.76**	84.99	78.62	85.93	88.78
Glass	60.35	58.02	62.45	57.27	64.64	**68.81**
LD	96.19	91.05	69.15	92.32	92.78	**96.51**
BC	55.36	91.4	90.96	92.52	92.47	**94.32**
CMC	39.71	41.97	35.66	43.05	42.54	**48.23**
Thyroid	86.05	85.43	83.32	87.32	88.63	**90.23**
Heart	58.89	84.18	66.27	86.48	85.91	**88.64**
Dermatology	56.82	85.61	59.53	72.43	88.21	**90.55**
WDBC	85.24	80.22	61.45	82.56	86.78	**91.67**
Diabetes	89.25	90.15	67.33	88.35	91.29	**94.02**

Bold indicates best values obtained by the algorithms for particular dataset

Whereas the parameter values for the compared algorithms are taken the same as reported in corresponding literature Further, accuracy and f-score parameters are employed to evaluate the performance of the proposed algorithm. The experimental results are compared with various existing meta-heuristic clustering algorithms. Results are taken as an average of thirty independent runs.

5.1 Performance measures

This subsection describes performance measures to evaluate the proposed WWO clustering algorithm. Accuracy and f-score are taken as performance measures.

Accuracy It determines the correctness of an algorithm as compared to true class labels. Accuracy can be described as the true label of an object "i" to cluster "c" is matched with cluster label using the map function. Clustering results are accurate when a high value of accuracy is obtained.

$$\text{Accuracy} = \sum_{i=1}^{n} \delta(\text{Truelabel}, \text{map}(c))/n \tag{11}$$

F-Score It is computed as the harmonic mean of precision and recall for testing the accuracy of the algorithm. Precision is calculated using the number of true positive results divided by the number of all true positive results. The recall represents the number of true positives divided by the number of all relevant results.

Table 5 Demonstrate experimental results of the proposed WWO clustering algorithm in contrast to existing hybrid clustering algorithms based on the accuracy performance metric

Dataset	Hybrid Algorithm				
	Fuzzy-PSO	KFCM	Fuzzy-MOC	PSO-GA	Proposed WWO
Iris	67.33	83.33	92.666	87.56	**93.21**
Wine	70.25	71.91	73.03	69.34	**74.63**
Vowel	76.27	63.31	79.82	86.28	**89.47**
Balance	84.12	71.12	81.44	86.42	**88.78**
Glass	60.29	50	**70.79**	64.09	68.81
LD	**96.69**	74.96	96.49	93.05	96.51
BC	55.36	55.36	53.62	**94.59**	94.32
CMC	43.11	38.97	45.96	44.52	**48.23**
Thyroid	58.14	56.28	64.19	88.85	**90.23**
Heart	58.89	76.3	61.48	85.61	**88.64**
Dermatology	27.65	25.42	35.2	88.35	**90.55**
WDBC	85.24	65.73	87.35	85.32	**91.67**
Diabetes	90.34	91.22	92.67	91.25	**94.02**

Bold indicates best values obtained by the algorithms for particular dataset

Table 6 Illustrates the experimental results of the proposed WWO clustering algorithm in contrast to other existing clustering algorithms based on the F-Score measure

Dataset	Algorithm					
	FCM	PSO	K-means	GA	WWO	Proposed WWO
LD	0.516	0.493	0.467	0.482	0.532	**0.585**
BC	0.958	0.814	0.829	0.819	0.894	**0.963**
CMC	0.357	0.331	0.334	0.324	0.371	**0.509**
Thyroid	0.480	0.778	0.731	0.763	0.748	**0.812**
Heart	0.422	0.521	0.319	0.401	0.726	**0.821**
Dermatology	0.293	0.255	0.213	0.223	0.283	**0.317**
WDBC	0.131	0.661	0.903	0.589	0.669	**0.893**
Diabetes	0.221	0.325	0.13	0.224	0.334	**0.481**
Iris	0.778	0.782	0.778	0.776	0.784	**0.793**
Wine	0.520	0.518	0.521	0.515	0.522	**0.531**
Vowel	0.649	**0.659**	0.652	0.647	0.651	0.655
Balance	0.734	**0.746**	0.724	0.716	0.721	0.744
Glass	0.548	0.573	0.563	0.561	0.576	**0.611**

Bold indicates best values obtained by the algorithms for particular dataset

$$\mathrm{F-Score} = 2\frac{\mathrm{precision*recall}}{\mathrm{precision+recall}} \tag{12}$$

5.2 Results and discussion

The results of proposed WWO clustering algorithm are compared with ten benchmark algorithms. Out of nine algorithms, five are popular benchmark clustering algorithms, these are K-Means, FCM, PSO, WWO and GA. Rest of are popular hybrid variant algorithm of aforementioned algorithms, these hybrid algorithms are Fuzzy-PSO, KFCM, Fuzzy-MOC, and PSO-GA. The results are presented in the form of average accuracy and f-measure.

Table 4 presents the experimental results of the proposed WWO clustering algorithm and other existing clustering algorithms like FCM, PSO, K-Means, GA and WWO using accuracy parameter. It is observed that the proposed WWO clustering algorithm obtains significant results with most of datasets except iris and balance datasets. The proposed algorithm achieves higher accuracy results in comparison in contrast to other existing clustering algorithms. For iris dataset, FCM algorithm achieves higher accuracy i.e. 99.33%, but the proposed WWO algorithm obtains 93.21% accuracy rate i.e. second highest accuracy rate among all other algorithms. The PSO algorithm obtains better accuracy rate i.e. 89.26% for balance dataset, while proposed WWO clustering algorithm obtains 88.78% accuracy rate. It is observed that performance of PSO algorithm is slightly better than proposed WWO algorithm using balance dataset. Table 5 provides the comparative results of proposed WWO clustering algorithm and some popular hybrid algorithms such as Fuzzy-PSO, KFCM, Fuzzy-MOC and PSO-GA using accuracy parameter. The significant results are obtained by the proposed algorithm using most of datasets. It's also revealed that proposed WWO clustering algorithm is not obtained better results (highest accuracy rate) for glass, LD, and BC datasets. It is observed that Fuzzy-MOC algorithm obtains better accuracy rate (70.79) for glass dataset as compared to WWO clustering algorithm (68.81). For LD dataset, Fuzzy-PSO algorithm archives highest accuracy (96.69), while WWO clustering algorithm obtains 96.51% accuracy rate. For BC dataset, PSO-GA obtains best accuracy rate (94.59), whereas, the accuracy rate of proposed WWO clustering algorithm is 94.32. But it is seen that performance of proposed WWO algorithm having almost similar to Fuzzy-PSO and PSO-GA in case of LD and BC datasets.

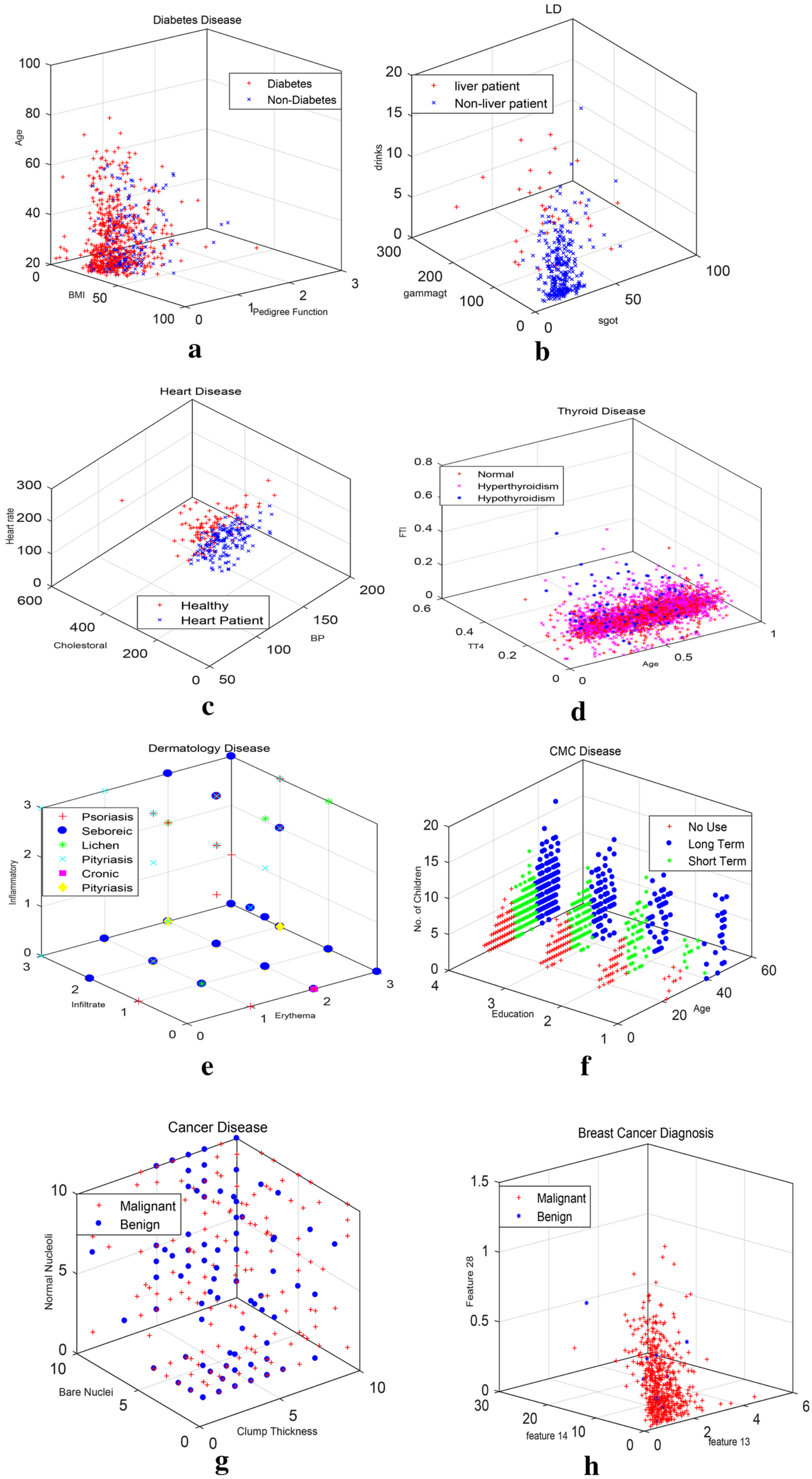
F-score is also an important performance measure to assess the performance of the clustering algorithm. It considers precision and recall for evaluating the performance of the proposed algorithm and provides more accurate results as compared to accuracy as a measure. Table 6 illustrates experimental results based on the F-Score measure. The proposed clustering algorithm obtains a higher F-Score rate in contrast to other existing clustering algorithms except on vowel and balance datasets. For vowel and balance datasets, PSO algorithm achieves higher f-score rate (0.659 and 0.746) as compared to other clustering algorithms. It is also seen that proposed algorithm obtains 0.655 and 0.744 f-score rate for vowel and balance datasets. On comparing the results of PSO and proposed WWO algorithms, it can be stated that performance of PSO is not far better with respect to WWO clustering algorithm, both algorithms having similar performance in case of vowel and balance datasets. Table 7 presents the simulation results of proposed WWO clustering algorithm with respect to some popular hybrid clustering algorithms such as Fuzzy-PSO, KFCM, Fuzzy-MOC and PSO-GA using f-score parameter. Its claimed that proposed WWO algorithm attains higher accuracy rate for LD, BC, CMC, Heart, Dermatology, WDBC, Diabetes, Iris,

Table 7 Illustrates the experimental results of the proposed WWO clustering algorithm in contrast to other existing hybrid clustering algorithms based on the F-Score measure

Dataset	Hybrid algorithm				
	Fuzzy-PSO	KFCM	Fuzzy-MOC	PSO-GA	Proposed WWO
LD	0.484	0.438	0.587	0.498	**0.585**
BC	0.42	0.253	0.946	0.821	**0.963**
CMC	0.329	0.351	0.502	0.335	**0.509**
Thyroid	0.23	0.199	0.698	**0.837**	0.812
Heart	0.422	0.76	0.61	0.525	**0.821**
Dermatology	0.168	0.258	0.222	0.261	**0.317**
WDBC	0.131	0.633	0.871	0.656	**0.893**
Diabetes	0.325	0.225	0.33	0.350	**0.481**
Iris	**0.795**	0.783	0.791	0.790	0.793
Wine	0.523	0.521	0.527	0.524	**0.531**
Vowel	0.651	0.646	0.650	0.653	**0.655**
Balance	0.746	0.727	**0.749**	0.735	0.744
Glass	0.568	0.493	0.601	0.604	**0.611**

Bold indicates best values obtained by the algorithms for particular dataset

Fig. 2 a–h Illustrate the clustering of different clustering datasets (LD, BC, CMC, Thyroid, Heart, Dermatology, WDBC, and Diabetes) proposed WWO based clustering algorithm



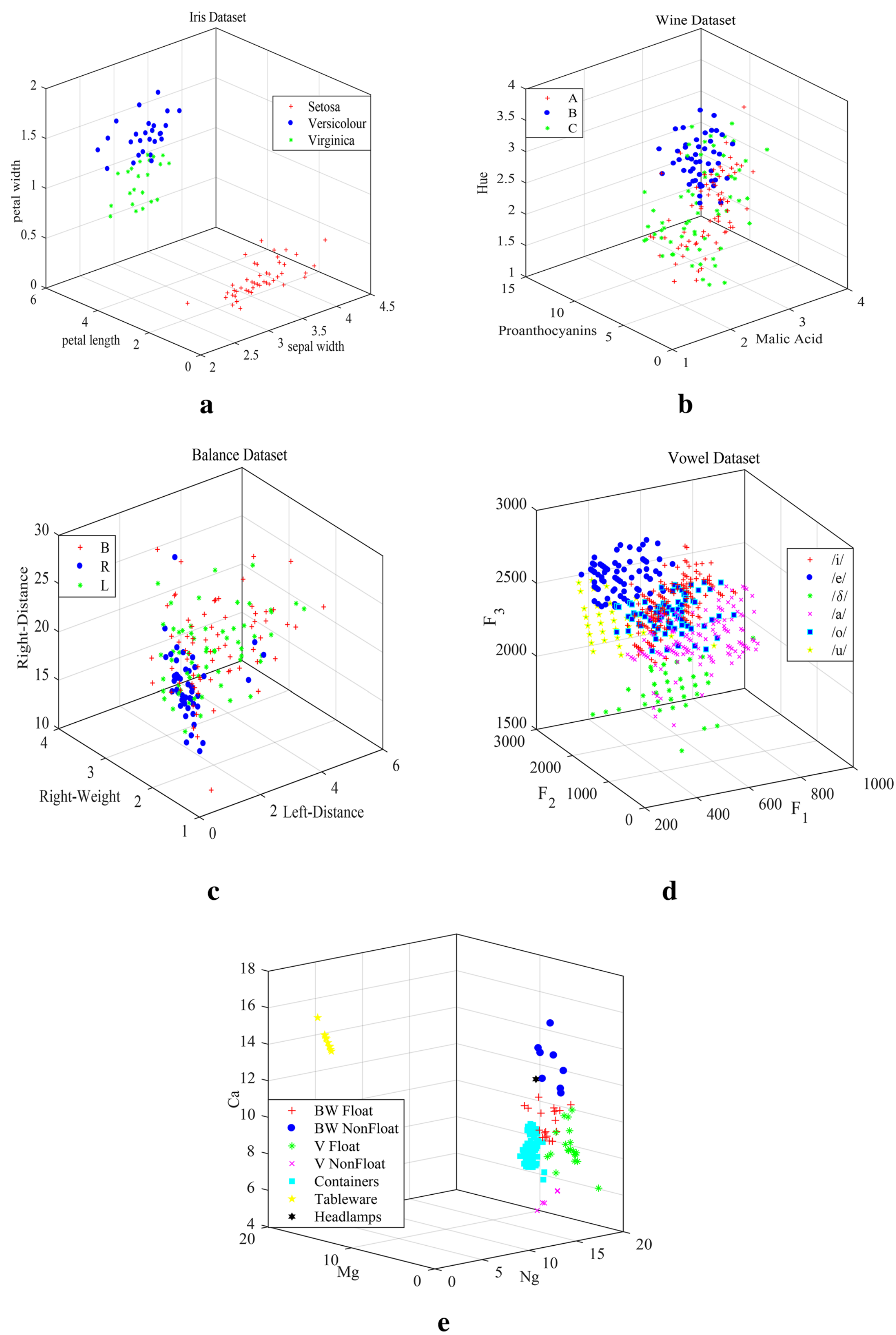


Fig. 3 a–e Demonstrate clustering of data objects using WWO based clustering algorithm for Iris, Wine, Vowel, Balance, and Glass datasets

Table 8 Average ranking of proposed WWO and benchmark clustering algorithm using an accuracy parameter

FCM	PSO	K-means	GA	WWO	Proposed WWO
3.92	3.69	5.23	4.08	2.92	1.15

Table 9 Results of the Friedman test on benchmark clustering algorithms using an accuracy parameter

Method	Statistical Value	*p*-Value	Degree of freedom	Critical Value	Hypothesis
Friedman Test	25.625455	1.62E−06	5	11.070504	Rejected

Wine, Vowel, and Glass datasets than other clustering algorithms. For rest of datasets, the proposed WWO algorithm does not obtains higher accurate results, but still competitive as compared to most of clustering algorithms. The PSO-GA algorithm achieves higher f-measure rate (0.837) than other clustering algorithms for thyroid dataset, but the proposed algorithm obtains second highest f-measure rate (0.812). For iris and balance datasets, Fuzzy-PSO and Fuzzy-MOC clustering algorithms obtains better f-measure rates (0.795 and 0.749) respectively, while proposed WWO algorithm achieves 0.793 and 0.744 f-measure rates. Hence, it is stated that proposed algorithm is also a competitive algorithm for such datasets. It can be stated that the proposed WWO clustering algorithm is one of the robust, viable, and efficient technique for analyzing benchmark clustering datasets.

Figure 2a–h illustrates the categorization of healthcare data using the proposed WWO clustering algorithm. Figure 2a considers the diabetes dataset and proposed clustering algorithm groups data into 2 clusters (i) diabetes and (ii) non-diabetes clusters. Figures 2b, c illustrates liver and heart diseases dataset in various cluster groups. It is observed that the proposed algorithm is capable to determine the clusters of healthy and non-healthy patients. Figure 2d demonstrates the thyroid disease dataset. The proposed clustering algorithm significantly divides the thyroid disease dataset into

Table 11 Average ranking of proposed WWO and hybrid variants of clustering algorithm using an accuracy parameter

Fuzzy-PSO	KFCM	Fuzzy-MOC	PSO-GA	Proposed WWO
3.75	4.42	2.92	2.67	1.25

three clusters i.e. normal, hyperthyroidism, and hypothyroidism. Figure 2e illustrates dermatology disease and it is observed that the proposed clustering algorithm divides it into 6 different clusters (1) psoriasis, (2) saboreic, (3) lichen, (4) pityriasis, (5) chronic and (6) pityriasis. It is being observed from results the proposed clustering algorithm well separated the dermatology data into 6 clusters. Figure 2f shows the CMC disease data. It is observed that the proposed clustering algorithm determines all three clusters in CMC dataset i.e. (1) no use, (2) long term, and (3) short term. Figures 2g, h illustrate the cancer and breast cancer datasets. It is observed that the proposed WWO clustering algorithm effectively analyze the both datasets. Hence, it can be stated that the proposed WWO clustering algorithm is an efficient for analyzing healthcare dataset.

Figure 3a–e demonstrates the clustering of data objects based on the WWO based clustering algorithm on non-healthcare datasets. Figure 3a considers the iris dataset and proposed WWO algorithm groups data into 3 clusters, (1) setosa (2) versicolour and (3) virginica. Figure 3b, c depict the clustering of wine and balance datasets. The proposed algorithm categorizes the wine dataset into three clusters i.e. (1) A, (2) B and (3) C. Data objects of balance dataset are also divided into three clusters which are B, L, and R. Figure 3d demonstrates the clustering results of vowel dataset. The proposed WWO algorithm divides the data objects into seven clusters i.e. (1)/i/, (2)/e/, (3)/δ/, (4)/a/, (5)/o/, and (6)/u/. It is also observed that the proposed algorithm separates data objects of the vowel dataset effectively. Figure 3e illustrates clustering of glass dataset. The proposed algorithm divides data objects into seven different clusters. It is also observed that one cluster is linearly separable from the other six clusters. Whereas, the rest of the six are non-linearly separable. It is stated that the proposed algorithm effectively performs the clustering of data objects into different

Table 10 Results of post hoc test on benchmark clustering algorithm using accuracy parameter

Techniques	FCM	PSO	K-means	GA	WWO	Proposed WWO
FCM	NA	=	+	=	=	+
PSO	=	NA	+	=	=	+
K-means	+	+	NA	+	+	+
GA	=	=	+	NA	=	+
WWO	=	=	+	=`	NA	+
Proposed WWO	+	+	+	+	+	NA

Table 12 Results of the Friedman test on hybrid clustering algorithms using an accuracy parameter

Method	Statistical Value	*p*-Value	Degree of freedom	Critical Value	Hypothesis
Friedman Test	31.76834	2.13E−06	4	9.487729	Rejected

Table 13 Results of post hoc test on hybrid variants of clustering algorithm using accuracy parameter

Techniques	Fuzzy-PSO	KFCM	Fuzzy-MOC	PSO-GA	Proposed WWO
Fuzzy-PSO	NA	=	+	+	+
KFCM	=	NA	+	+	+
Fuzzy-MOC	+	+	NA	=	+
PSO-GA	+	+	=	NA	+
Proposed WWO	+	+	+	+	NA

Table 14 Average ranking of proposed WWO and benchmark clustering algorithm using an F-score parameter

FCM	PSO	K-means	GA	WWO	Proposed WWO
4.12	3.38	4.35	5.15	2.77	1.23

clusters for non-healthcare datasets. Hence, it can be said that the proposed WWO based clustering algorithm is an effective algorithm for the clustering of data objects.

5.3 Statistical results

This subsection discusses the statistical analysis to validate the performance of proposed WWO based clustering algorithm for various benchmark datasets. Table 8 shows the average ranking of the proposed WWO based clustering algorithm and other benchmark clustering algorithms in terms of accuracy parameter. It is seen that proposed WWO achieves first rank that is 1.15 among the other clustering algorithms. While K-means achieves the lowest among all the clustering algorithms in comparison with value 5.23.

Table 9 presents the statistics of Friedman test using accuracy parameter on benchmark clustering algorithms. The statistical value and critical value of Friedman Test is 25.62545 and 11.070504 respectively with degree of freedom as 5, whereas *p* value is 1.62E−06. Hence, it can be stated that the null hypothesis (H_0) is rejected at the confidence level of 0.05. A significant difference occurs between

Table 15 Results of the Friedman test on benchmark clustering algorithms using F-score parameter

Method	Statistical Value	*p*-Value	Degree of freedom	Critical Value	Hypothesis
Friedman Test	35.462555	1.62E−06	5	11.070504	Rejected

the performance of the proposed WWO based clustering algorithm and the rest of the clustering algorithms.

Further, post hoc test is conducted to find out from where the differences actually came. Table 10 shows the results of the post hoc test on benchmark clustering algorithm using accuracy parameter. + symbol represents significantly different, = represents similar and NA is not applicable. It is seen from the results that FCM, PSO, GA and WWO yields similar results with respect to each other, while there is significant difference in results for K-means and proposed WWO with respect to FCM, PSO, GA and WWO. Thus, it is stated that proposed WWO outperforms the state-of-art clustering algorithms. The clustering algorithms are divided into five groups on the basis of post hoc test results. These groups are (FCM, PSO, GA and WWO), (K-means), (FCM, PSO, GA), (FCM, PSO, WWO) and (Proposed WWO). The algorithms lie within the group having similar performance. It is stated that proposed WWO algorithm is statistically different than other algorithms.

Table 11 shows the average ranking of the proposed WWO based and hybrid variants of clustering algorithms using accuracy parameter. It is seen that proposed WWO achieves first rank that is 1.25 among the other clustering algorithms. While KFCM achieves the lowest among hybrid variants clustering algorithms with value 4.42.

Table 12 presents the statistics of Friedman test using accuracy parameter on hybrid variants of clustering algorithms. The statistical value and critical value of Friedman Test is 31.76834 and 9.487729 respectively with degree of freedom as 4, whereas p-value is 2.13E−06. Hence, it can be stated that the null hypothesis (H_0) is rejected at the confidence level of 0.05.

A significant difference occurs between the performance of proposed WWO based clustering algorithm and the hybrid variants of clustering algorithms. Further, post hoc test is conducted. Table 13 shows the results of the post hoc test on hybrid clustering algorithm using accuracy parameter. It is seen that Fuzzy-PSO resulted in similar performance to KFCM for accuracy but Fuzzy-MOC, PSO-GA and proposed WWO give significantly different results. Fuzzy-MOC with respect to PSO-GA gives the similar results. It is observed that the proposed WWO clustering algorithms

Table 16 Results of post hoc test on benchmark clustering algorithm using F-score parameter

Techniques	FCM	PSO	K-means	GA	WWO	Proposed WWO
FCM	NA	=	=	+	+	+
PSO	=	NA	=	+	=	+
K-means	=	=	NA	=	+	+
GA	+	+	=	NA	+	+
WWO	+	=	+	+	NA	+
Proposed WWO	+	+	+	+	+	NA

Table 17 Average ranking of proposed WWO and hybrid variants of clustering algorithm using F-score parameter

Fuzzy-PSO	KFCM	Fuzzy-MOC	PSO-GA	Proposed WWO
3.85	4.23	2.54	3	1.46

Table 18 Results of the Friedman test on hybrid clustering algorithms using F-score parameter

Method	Statistical Value	*p*-Value	Degree of freedom	Critical Value	Hypothesis
Friedman Test	25.930502	3.27E−05	4	9.487729	Rejected

Table 19 Results of post hoc test on hybrid variants of clustering algorithm using F-score parameter

Techniques	Fuzzy-PSO	KFCM	Fuzzy-MOC	PSO-GA	Proposed WWO
Fuzzy-PSO	NA	=	+	=	+
KFCM	=	NA	+	+	+
Fuzzy-MOC	+	+	NA	=	+
PSO-GA	=	+	=	NA	+
Proposed WWO	+	+	+	+	NA

performs effectively with respect to other hybrid variants of clustering algorithm. Further, the algorithms are grouped into three different categories as per the post hoc test results. These categories are (Fuzzy-PSO, KFCM), (Fuzzy-MOC, PSO-GA) and (Proposed WWO). The algorithms placed in same category exhibit similar performance.

The results of Friedman statistical test using the F-score parameter are reported in Tables 14, 15 and 16. Table 14 depicts the average ranking of the proposed WWO based clustering algorithm and the other clustering algorithms. It is seen that the proposed algorithm obtains the first rank i.e., 1.23. It is also noted that the GA algorithm achieves the lowest rank i.e., 5.15 among other algorithms in comparison. Table 15 presents the statistics of the Friedman test. The statistical value of the Friedman test (0.05, 9) is 34.868132 and the critical value is 34.868132 with degree of freedom as 5, whereas the p-value is 1.62E−06. Hence, it can be stated that the null hypothesis (H_0) is rejected at the confidence level of 0.05. It is stated that the performance of proposed WWO algorithm significantly differs from clustering algorithms in comparison.

Further, post hoc test is conducted using F-score parameter. Table 16 shows the results of post hoc test on benchmark clustering algorithm. It is observed that PSO, FCM, K-means and GA yield similar results. For most of the cases, GA, WWO and proposed WWO gives significantly different results (denoted by + symbol) for F-score parameter. It is observed that proposed WWO algorithms performs significantly better as compared to benchmark clustering algorithms. As per the results of post hoc test, the algorithms are allocated to divided into six groups such as (FCM, PSO, K-means), (FCM, PSO, K-means, WWO), (FCM, PSO, K-means, GA), (K-means, GA, PSO, WWO) and (Proposed WWO). The proposed WWO algorithm is not grouped with other algorithms. Hence, it can be concluded that proposed WWO algorithm exhibits dissimilar performance.

Table 17 shows the average ranking of the proposed WWO and hybrid variants of clustering algorithm. It is seen that the proposed algorithm obtains the first rank i.e. 1.46. It is also noted that the KFCM achieves the lowest rank i.e. 4.23 in contrast to other algorithms. Table 18 presents the statistics of the Friedman test. The statistical value of the Friedman test is 25.930502and the critical value is 9.487729 with degree of freedom as 4, whereas the p-value is 3.27E−05. The null hypothesis (H_0) is therefore rejected at the confidence level of 0.05.

Similarly, post hoc test is conducted using F-score parameter for hybrid variants of clustering algorithm. Table 19 shows the results of the post hoc test on hybrid variants of clustering algorithm. It is observed that Fuzzy-PSO and KFC, Fuzzy-PSO and PSO-GA, Fuzzy-MOC and PSO-GA yield similar results. It is seen that Fuzzy-PSO and proposed WWO perform significantly differently in contrast to Fuzzy-PSO. Also, Fuzzy-MOC, PSO-GA and proposed WWO give significance difference compared to KFCM. With respect to PSO-GA, KFCM and proposed WWO gives significance difference among the results and algorithms group into five different categories such as (Fuzzy-PSO, KFCM, PSO-GA), (Fuzzy-PSO, KFCM), (Fuzzy-MOC, PSO-GA), (Fuzzy-PSO, Fuzzy-MOC, PSO-GA) and (Proposed WWO). Hence, the test reveals that proposed WWO performs significantly better in contrast to other hybrid variants of clustering algorithm for F-score parameter.

6 Conclusion

In this work, the WWO based clustering algorithm is proposed for solving data clustering problems. Prior to adoption, two improvements are incorporated in WWO algorithm for generating more promising and efficient clustering results. These improvements are described in terms of an updated global search mechanism and decay operator. The objective of decay operator is to address the premature convergence issue of WWO algorithm. Further, an updated global search mechanism based on the global best concept of PSO algorithm is incorporated in WWO to improve the accuracy rate as well as guide the global optimum solution. The performance of proposed WWO algorithm is assessed over thirteen benchmark datasets on the basis accuracy and F-score parameter. The simulation results of the proposed WWO based clustering algorithm are compared with the state of art clustering algorithms from the literature. It is observed that the proposed clustering algorithm achieves higher accuracy and F-score rate in contrast to other existing clustering algorithms reported in the literature. It is concluded that the proposed WWO clustering algorithm obtains better clustering results for most of the clustering datasets. Hence, it can be stated that the proposed WWO based clustering algorithm is a promising and efficient clustering algorithm for analyzing the data. In future, the proposed WWO algorithm can be enhanced using neighborhood method. Moreover, the applicability of WWO algorithm will be evaluated feature selection, parameter optimization and multi objective optimization problems.

References

1. Jain AK (2008) Data clustering: 50 years beyond k-means. In: Joint European conference on machine learning and knowledge discovery in databases. Springer, Berlin, Heidelberg, pp 3–4
2. Gong S, Hu W, Li H, Qu Y (2018) Property clustering in linked data: an empirical study and its application to entity browsing. Int J Semant Web Inf Syst (IJSWIS) 14(1):31–70
3. Chou CH, Hsieh SC, Qiu CJ (2017) Hybrid genetic algorithm and fuzzy clustering for bankruptcy prediction. Appl Soft Comput 56:298–316
4. Holý V, Sokol O, Černý M (2017) Clustering retail products based on customer behaviour. Appl Soft Comput 60:752–762
5. Navarro ÁAM, Ger PM (2018) Comparison of clustering algorithms for learning analytics with educational datasets. IJIMAI 5(2):9–16
6. Hyde R, Angelov P, MacKenzie AR (2017) Fully online clustering of evolving data streams into arbitrarily shaped clusters. Inf Sci 382:96–114
7. Wang L, Zhou X, Xing Y, Yang M, Zhang C (2017) Clustering ecg heartbeat using improved semi-supervised affinity propagation. IET Softw 11(5):207–213
8. Mekhmoukh A, Mokrani K (2015) Improved fuzzy C-means based particle swarm optimization (PSO) initialization and outlier rejection with level set methods for MR brain image segmentation. Comput Methods Prog Biomed 122(2):266–281
9. Abualigah LM, Khader AT, Al-Betar MA, Alomari OA (2017) Text feature selection with a robust weight scheme and dynamic dimension reduction to text document clustering. Expert Syst Appl 84:24–36
10. Triguero I, del Río S, López V, Bacardit J, Benítez JM, Herrera F (2015) ROSEFW-RF: the winner algorithm for the ECBDL'14 big data competition: an extremely imbalanced big data bioinformatics problem. Knowl-Based Syst 87:69–79
11. Zhu J, Lung CH, Srivastava V (2015) A hybrid clustering technique using quantitative and qualitative data for wireless sensor networks. Ad Hoc Netw 25:38–53
12. Abualigah LMQ (2019) Feature selection and enhanced krill herd algorithm for text document clustering. Springer, Berlin, pp 1–165
13. Marinakis Y, Marinaki M, Doumpos M, Zopounidis C (2009) Ant colony and particle swarm optimization for financial classification problems. Expert Syst Appl 36(7):10604–10611
14. Saraswathi S, Sheela MI (2014) A comparative study of various clustering algorithms in data mining. Int J Comput Sci Mob Comput 11(11):422–428
15. Hartigan JA, Wong MA (1979) Algorithm AS 136: a k-means clustering algorithm. J R Stat Soc Ser C Appl Stat 28(1):100–108
16. Celebi ME, Kingravi HA, Vela PA (2013) A comparative study of efficient initialization methods for the k-means clustering algorithm. Expert Syst Appl 40(1):200–210
17. Han J, Pei J, Kamber M (2011) Data mining: concepts and techniques. Elsevier, Amsterdam
18. Moreira A, Santos MY, Carneiro S (2005) Density-based clustering algorithms–DBSCAN and SNN. University of Minho-Portugal, pp 1–18
19. Schaeffer SE (2007) Graph clustering. Comput Sci Rev 1(1):27–64
20. Hufnagl B, Lohninger H (2020) A graph-based clustering method with special focus on hyperspectral imaging. Anal Chim Acta 1097:37–48
21. Nanda SJ, Panda G (2014) A survey on nature inspired metaheuristic algorithms for partitional clustering. Swarm Evol Comput 16:1–18
22. Nayyar A, Le DN, Nguyen NG (eds) (2018) Advances in swarm intelligence for optimizing problems in computer science. CRC Press, Boca Raton
23. Nayyar A, Nguyen NG (2018) Introduction to swarm intelligence. Adv Swarm Intell Optim Probl Comput Sci:53–78
24. Nayyar A, Garg S, Gupta D, Khanna A (2018) Evolutionary computation: theory and algorithms. In: Advances in swarm intelligence for optimizing problems in computer science. Chapman and Hall/CRC, pp 1–26

25. Sung CS, Jin HW (2000) A tabu-search-based heuristic for clustering. Pattern Recogn 33(5):849–858
26. Selim SZ, Alsultan K (1991) A simulated annealing algorithm for the clustering problem. Pattern Recogn 24(10):1003–1008
27. Maulik U, Bandyopadhyay S (2000) Genetic algorithm-based clustering technique. Pattern Recogn 33(9):1455–1465
28. Karaboga D, Ozturk C (2011) A novel clustering approach: artificial Bee Colony (ABC) algorithm. Appl Soft Comput 11(1):652–657
29. Sahoo G, Kumar Y (2017) A two-step artificial bee colony algorithm for clustering. Neural Comput Appl 28(3):537–551
30. Nayyar A, Puri V, Suseendran G (2019) Artificial bee Colony optimization—population-based meta-heuristic swarm intelligence technique. Data management, analytics and innovation. Springer, Singapore, pp 513–525
31. Kumar S, Nayyar A, Kumari R (2019) Arrhenius artificial bee colony algorithm. International conference on innovative computing and communications. Springer, Singapore, pp 187–195
32. Shelokar PS, Jayaraman VK, Kulkarni BD (2004) An ant colony approach for clustering. Anal Chim Acta 509(2):187–195
33. Nayyar A, Singh R (2016) Ant colony optimization—computational swarm intelligence technique. In: 2016 3rd International conference on computing for sustainable global development (INDIACom), IEEE, pp 1493–1499
34. Niknam T, Amiri B (2010) An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis. Appl Soft Comput 10(1):183–197
35. 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
36. 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
37. Kumar Y, Sahoo G (2015) A hybrid data clustering approach based on improved cat swarm optimization and K-harmonic mean algorithm. AI Commun 28(4):751–764
38. Senthilnath J, Omkar SN, Mani V (2011) Clustering using firefly algorithm: performance study. Swarm Evol Comput 1(3):164–171
39. Durbhaka GK, Selvaraj B, Nayyar A (2019) Firefly swarm: metaheuristic swarm intelligence technique for mathematical optimization. Data Management, Analytics and Innovation. Springer, Singapore, pp 457–466
40. Han X, Quan L, Xiong X, Almeter M, Xiang J, Lan Y (2017) A novel data clustering algorithm based on modified gravitational search algorithm. Eng Appl Artif Intell 61:1–7
41. Kumar Y, Sahoo G (2014) A review on gravitational search algorithm and its applications to data clustering & classification. Int J Intell Syst Appl 6(6):79
42. Hatamlou A (2013) Black hole: a new heuristic optimization approach for data clustering. Inf Sci 222:175–184
43. Kumar Y, Sahoo G (2014) A charged system search approach for data clustering. Prog Artif Intell 2(2–3):153–166
44. 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
45. Kumar Y, Singh PK (2019) A chaotic teaching learning based optimization algorithm for clustering problems. Appl Intell 49(3):1036–1062
46. Singh H, Kumar Y, Kumar S (2019) A new meta-heuristic algorithm based on chemical reactions for partitional clustering problems. Evol Intel 12(2):241–252
47. Hatamlou A, Abdullah S, Hatamlou M (2011) Data clustering using big bang–big crunch algorithm. In: International conference on innovative computing technology. Springer, Berlin, Heidelberg, pp 383–388
48. Singh H, Kumar Y (2019) Hybrid big bang-big crunch algorithm for cluster analysis. In: International conference on futuristic trends in networks and computing technologies. Springer, Singapore, pp 648–661
49. Zhou Y, Wu H, Luo Q, Abdel-Baset M (2019) Automatic data clustering using nature-inspired symbiotic organism search algorithm. Knowl-Based Syst 163:546–557
50. Agbaje MB, Ezugwu AE, Els R (2019) Automatic data clustering using hybrid firefly particle swarm optimization algorithm. IEEE Access 7:184963–184984
51. Kushwaha N, Pant M, Sharma S (2019) Electromagnetic optimization-based clustering algorithm. Expert Syst:e12491
52. Zhao F, Zhang L, Liu H, Zhang Y, Ma W, Zhang C, Song H (2019) An improved water wave optimization algorithm with the single wave mechanism for the no-wait flow-shop scheduling problem. Eng Optim 51(10):1727–1742
53. Singh G, Rattan M, Gill SS, Mittal N (2019) Hybridization of water wave optimization and sequential quadratic programming for cognitive radio system. Soft Comput 23(17):7991–8011
54. Zhao F, Liu H, Zhang Y, Ma W, Zhang C (2018) A discrete water wave optimization algorithm for no-wait flow shop scheduling problem. Expert Syst Appl 91:347–363
55. Zhang J, Zhou Y, Luo Q (2018) An improved sine cosine water wave optimization algorithm for global optimization. J Intell Fuzzy Syst 34(4):2129–2141
56. Shao Z, Pi D, Shao W (2019) A novel multi-objective discrete water wave optimization for solving multi-objective blocking flow-shop scheduling problem. Knowl-Based Syst 165:110–131
57. Liu A, Li P, Sun W, Deng X, Li W, Zhao Y, Liu B (2019) Prediction of mechanical properties of micro-alloyed steels via neural networks learned by water wave optimization. Neural Comput Appl:1–16
58. Zhou Y, Zhang J, Yang X, Ling Y (2018) Optimal reactive power dispatch using water wave optimization algorithm. Oper Res:1–17
59. Ibrahim AM, Tawhid MA, Ward RK (2020) A binary water wave optimization for feature selection. Int J Approximate Reasoning 120:74–91
60. Manshahia MS (2017) Water wave optimization algorithm-based congestion control and quality of service improvement in wireless sensor networks. Trans Netw Commun 5(4):31–31
61. Hematabadi AA, Foroud AA (2019) Optimizing the multi-objective bidding strategy using min–max technique and modified water wave optimization method. Neural Comput Appl 31(9):5207–5225
62. Soltanian A, Derakhshan F, Soleimanpour-Moghadam M (2018) MWWO: modified water wave optimization. In: 2018 3rd conference on swarm intelligence and evolutionary computation (CSIEC). IEEE, pp 1–5
63. Singh T (2020) A chaotic sequence-guided Harris hawks optimizer for data clustering. Neural Comput Appl
64. Tsai CW, Chang WY, Wang YC, Chen H (2019) A high-performance parallel coral reef optimization for data clustering. Soft Comput 23(19):9327–9340
65. Kuwil FH, Shaar F, Topcu AE, Murtagh F (2019) A new data clustering algorithm based on critical distance methodology. Expert Syst Appl 129:296–310
66. Baalamurugan KM, Bhanu SV (2019) An efficient clustering scheme for cloud computing problems using metaheuristic algorithms. Cluster Comput 22(5):12917–12927
67. Sharma M, Chhabra JK (2019) An efficient hybrid PSO polygamous crossover-based clustering algorithm. Evol Intell:1–19
68. Abdulwahab HA, Noraziah A, Alsewari AA, Salih SQ (2019) An enhanced version of black hole algorithm via levy flight

for optimization and data clustering problems. IEEE Access 7:142085–142096
69. Mustafa HM, Ayob M, Nazri MZA, Kendall G (2019) An improved adaptive memetic differential evolution optimization algorithm for data clustering problems. PLoS ONE 14(5):e0216906
70. Tarkhaneh O, Moser I (2019) An improved differential evolution algorithm using Archimedean spiral and neighborhood search-based mutation approach for cluster analysis. Fut Gener Comput Syst 101:921–939
71. Aljarah I, Mafarja M, Heidari AA, Faris H, Mirjalili S (2020) Clustering analysis using a novel locality-informed grey wolf-inspired clustering approach. Knowl Inf Syst 62(2):507–539
72. Zhu LF, Wang JS, Wang HY, Guo SS, Guo MW, Xie W (2020) Data clustering method based on improved bat algorithm with six convergence factors and local search operators. IEEE Access 8:80536–80560
73. Senthilnath J, Kulkarni S, Suresh S, Yang XS, Benediktsson JA (2019) FPA clust: evaluation of the flower pollination algorithm for data clustering. Evol Intell:1–11
74. Mageshkumar C, Karthik S, Arunachalam VP (2019) Hybrid metaheuristic algorithm for improving the efficiency of data clustering. Cluster Comput 22(1):435–442
75. Kaur A, Pal SK, Singh AP (2019) Hybridization of chaos and flower pollination algorithm over k-means for data clustering. Appl Soft Comput:105523
76. Xie H, Zhang L, Lim CP, Yu Y, Liu C, Liu H, Walters J (2019) Improving K-means clustering with enhanced Firefly Algorithms. Appl Soft Comput 84:105763
77. Huang KW, Wu ZX, Peng HW, Tsai MC, Hung YC, Lu YC (2019) Memetic particle gravitation optimization algorithm for solving clustering problems. IEEE Access 7:80950–80968
78. Dinkar SK, Deep K (2019) Opposition-based antlion optimizer using Cauchy distribution and its application to data clustering problem. Neural Comput Appl:1–29
79. Abualigah LM, Khader AT, Hanandeh ES, Gandomi AH (2017) A novel hybridization strategy for krill herd algorithm applied to clustering techniques. Appl Soft Comput 60:423–435
80. Zeng N, Wang Z, Zhang H, Kim KE, Li Y, Liu X (2019) An improved particle filter with a novel hybrid proposal distribution for quantitative analysis of gold immunochromatographic strips. IEEE Trans Nanotechnol 18:819–829
81. Zeng N, Wang Z, Liu W, Zhang H, Hone K, Liu X (2020) A dynamic neighborhood-based switching particle swarm optimization algorithm. IEEE Trans Cybern
82. Abualigah L (2020) Group search optimizer: a nature-inspired meta-heuristic optimization algorithm with its results, variants, and applications. Neural Comput Appl:1–24
83. Abualigah L (2020) Multi-verse optimizer algorithm: a comprehensive survey of its results, variants, and applications. Neural Comput Appl:1–21
84. Zeng N, Qiu H, Wang Z, Liu W, Zhang H, Li Y (2018) A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease. Neurocomputing 320:195–202
85. Zhu G, Kwong S (2010) Gbest-guided artificial bee colony algorithm for numerical function optimization. Appl Math Comput 217(7):3166–3173