Artificial Intelligence Review (2024) 57:287
https://doi.org/10.1007/s10462-024-10920-1

Exploring meta-heuristics for partitional clustering: methods, metrics, datasets, and challenges

Arvinder Kaur[1] · **Yugal Kumar**[2] · **Jagpreet Sidhu**[2]

Accepted: 22 August 2024

Abstract
Partitional clustering is a type of clustering that can organize the data into non-overlapping groups or clusters. This technique has diverse applications across the different various domains like image processing, pattern recognition, data mining, rule-based systems, customer segmentation, image segmentation, and anomaly detection, etc. Hence, this survey aims to identify the key concepts and approaches in partitional clustering. Further, it also highlights its widespread applicability including major advantages and challenges. Partitional clustering faces challenges like selecting the optimal number of clusters, local optima, sensitivity to initial centroids, etc. Therefore, this survey describes the clustering problems as partitional clustering, dynamic clustering, automatic clustering, and fuzzy clustering. The objective of this survey is to identify the meta-heuristic algorithms for the aforementioned clustering. Further, the meta-heuristic algorithms are also categorised into simple meta-heuristic algorithms, improved meta-heuristic algorithms, and hybrid meta-heuristic algorithms. Hence, this work also focuses on the adoption of new meta-heuristic algorithms, improving existing methods and novel techniques that enhance clustering performance and robustness, making partitional clustering a critical tool for data analysis and machine learning. This survey also highlights the different objective functions and benchmark datasets adopted for measuring the effectiveness of clustering algorithms. Before the literature survey, several research questions are formulated to ensure the effectiveness and efficiency of the survey such as what are the various meta-heuristic techniques available for clustering problems? How to handle automatic data clustering? What are the main reasons for hybridizing clustering algorithms? The survey identifies shortcomings associated with existing algorithms and clustering problems and highlights the active area of research in the clustering field to overcome these limitations and improve performance.

Keywords Automatic clustering · Descriptive analysis · Hybrid approaches · Meta-heuristic algorithms · Partitional clustering

Extended author information available on the last page of the article

Published online: 12 September 2024

1 Introduction

The process of exploring and analysing large data for new, valid, and profitable patterns is termed knowledge discovery. However, due to rapid increments in data generation and storage, it is becoming more and more difficult to retrieve information by traditional analysis methods. Data mining is a task that can be employed to retrieve valuable information and patterns from this large data. Data mining techniques are being used to scour databases so that new and convenient patterns can be effortlessly discovered. Data mining tasks are classified as predictive tasks and descriptive tasks (Tan et al. 2016). Predictive tasks determine the value of a particular attribute based on other attributes. Descriptive tasks derive patterns (correlations, trends, clusters) that summarize underlying relationships. Hence, clustering is a descriptive task that can group the objects based on some similarity measure. Broadly, clustering can be characterized as Partitional and hierarchical. Partitional clustering is grouping objects into non-overlapping clusters based on inter-cluster distances. Hierarchical clustering is a tree clustering either by an agglomerative (Bottom-up) approach or by Divisive (Top-down) approach. Several other clustering methods are reported in the literature (i) graph clustering, (ii) spectral clustering, (iii) model-based clustering, (iv) spectral clustering, (v) density-based clustering, etc. Graph clustering is based on a collection of vertices and edges (Schaeffer 2007). Graph clustering includes grouping of vertices based on edges within a cluster and relatively fewer among other clusters. Spectral clustering is a subset of graph clustering methods that utilize spectral analysis to cluster data points based on their graph representation (Kannan et al. 2004). This clustering method leverages graph theory and spectral analysis (eigenvalue decomposition) to cluster data points based on their similarity or affinity. Spectral clustering is an efficient technique to handle various heuristic problems. Model-based clustering uses the concept of finite mixture models (Schaeffer 2007). Model-based clustering is a statistical clustering approach and it is assumed that the data can be generated from a mixture of underlying probability distributions. In this clustering technique, data can be viewed as a combination of different probability distributions each corresponding to a cluster. In model-based clustering, the goal is to find the best-fitting model of the data by estimating the parameters of the underlying probability distributions. Density-based clustering techniques are designed to find clusters of arbitrary shapes. DBSCAN is a popular density-based clustering example (Hahsler and Bolaños 2016). The DBSCAN counts eps-neighbourhood and identifies core, border, and noise points on user-specified thresholds to estimate density around each data point.

However, in the literature, it is found that Partitional clustering is a prominent one among all clustering methods for data analysis. Partitional clustering is a widely used approach in data analysis, machine learning, and data mining. It divides a dataset into non-overlapping groups, such that each data point belongs to exactly one cluster. This clustering technique aims to minimize within-cluster variance and maximize inter-cluster variance, resulting in clusters that are as distinct and cohesive as possible. While Partitional clustering methods such as k-means and k-mediods are popular due to their simplicity and efficiency, these algorithms have some limitations including sensitivity to initial conditions, potential convergence to local optima, and challenges in determining the optimal number of clusters. To handle these limitations and enhance clustering performance, meta-heuristic algorithms have been proposed as alternatives or enhancements to traditional methods. Meta-heuristic algorithms offer a flexible and adaptive approach to Partitional clustering. These algorithms consist of intelligent search strategies to explore the solution space and

optimize clustering assignments. The metaheuristics are optimization algorithms that help in finding the solutions to the complex problems. Thus, metaheuristic algorithms provide a powerful approach to optimizing the different aspects during the clustering process. This helps to improve the cluster quality and can efficiently handle complex clustering problems. Different metaheuristics approaches have been developed and used for optimizing the clustering process. The clustering process using a metaheuristic consists of various steps. The clustering problem is defined by initializing the number of clusters and objective function. Initialize the population and randomly generate the initial set of solutions. The objective function further evaluates the quality of each solution and the fitness values of each solution define the satisfying criteria of the clustering objective. The Metaheuristic approach is used for iterating thru the candidate solution and improving the fitness value and quality of clusters. The best solutions when found are updated in the current population. When the convergence criteria are met, the best solutions are returned as cluster centroid. Further, the quality of clusters can be evaluated using different performance measures or metrics such as compactness, separation, or clustering stability. Further, metaheuristic algorithms also help in improving the quality of clustering by modifying the cluster centres iteratively concerning the fitness requirements such as minimum intra-cluster distance. These algorithms are also capable of handling non-convex clusters through the exploration of intricate search spaces and the determination of non-linear cluster boundaries. However, it also observed that metaheuristic algorithms also have some limitations such as being stuck in local optima, convergence rate, unbalanced search mechanism, population diversity, and initialization issues (Yao et al. 2018; Bahrololoum et al. 2015; Bijari et al. 2018; Chang et al. 2016). Hence, the objective of this survey is to identify the different metaheuristic algorithms available in the literature for Partitional clustering, shortcomings associated with these algorithms, alleviation of the shortcomings, objective functions, and benchmark datasets for clustering. Before proceeding, several research questions are designed to find the accurate outcome for this survey. These research questions are highlighted below. Further, metaheuristic algorithms also help in improving the quality of clustering through modifying the cluster centres iteratively with respect to the fitness requirements such as minimum intra-cluster distance. These algorithms also capable to handle the non-convex clusters through the exploration of intricate search spaces and the determination of non-linear cluster boundaries. However, it also observed that metaheuristic algorithms also have some limitations such as stuck in local optima, convergence rate, unbalanced search mechanism, population diversity, and initialization issues (Yao et al. 2018; Bahrololoum et al. 2015; Bijari et al. 2018; Chang et al. 2016). The visualization in Fig. 1a–d illustrates the examination of meta-heuristics in data clustering using VOS Viewer (Abbasi and Choukolaei 2023). This analysis involved exploring various key terms within research articles from 2015 to 2024 from Science Direct, leveraging meta-heuristics in data clustering. VOS Viewer is a specialized software tool designed for constructing and visualizing bibliometric networks. Widely embraced in academic circles, VOS Viewer facilitates the analysis and visualization of relationships among scientific publications, authors, keywords, and other entities within a specific research domain (Emrouznejad et al. 2023). These visualizations assist researchers in discerning patterns, clusters, and trends within the literature, providing valuable insights into the structure and dynamics of the field under investigation.

The primary aim of this survey is to identify different metaheuristic algorithms presented in the literature for Partitional clustering, along with their associated shortcomings, methods for mitigating these shortcomings, objective functions, and benchmark datasets for clustering. To achieve this objective, several research questions have been formulated to ensure the accuracy of the survey findings. These research questions are outlined below.

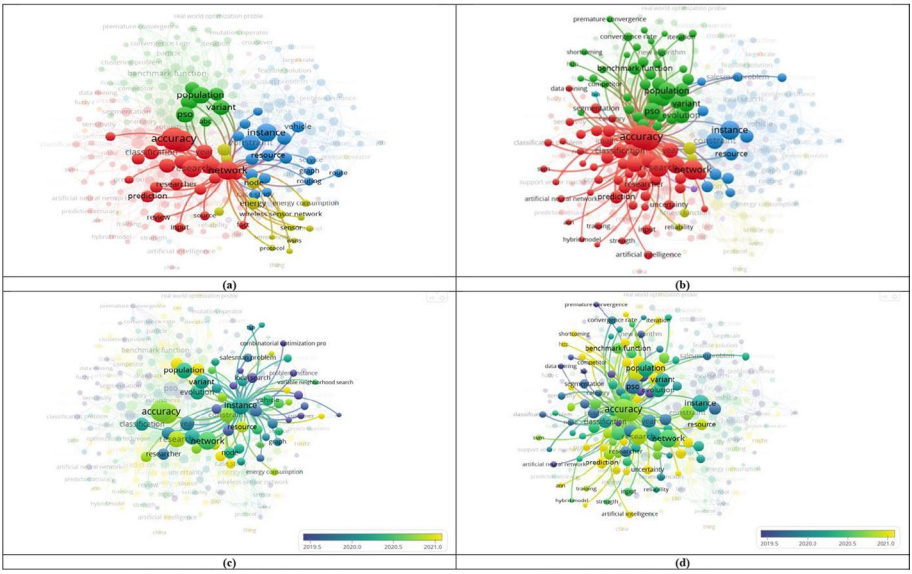

Fig. 1 **a**–**d** Network analysis based on meta-heuristics in data clustering keywords

1.1 Research questions (RQ)

The primary survey objective is to find answers to the following Research Questions (RQ):

RQ 1 What are the various meta-heuristic techniques available for clustering problems?

RQ 2 How to handle automatic data clustering?

RQ 3 How to handle high dimensional data (problems) with clustering?

RQ 4 What are the main reasons for hybridizing the clustering algorithms?

RQ 5 What are different objective functions (distance function), different performance measures, and benchmark datasets adopted to evaluate the performance of Partitional clustering algorithms?

1.2 Purpose of this survey

The purpose of this survey paper is to provide a comprehensive review of the field of partitional clustering. This study aims to identify the recent advancement in the context of meta-heuristic algorithms, exploring the structure of the meta-heuristic algorithms and, the strengths and weaknesses of the algorithms for handling the partitional clustering problems. This survey also synthesizes the knowledge from both classical and contemporary approaches for partitional clustering, including optimization-based methods (meta-heuristic algorithms), improved algorithms, hybrid algorithms, and adaptive control parameters.

It also highlights the various distance functions adopted as similarity measures for clustering tasks and considers the benchmark datasets that can be adopted for evaluating the efficacy of the clustering algorithms. By examining the strengths, limitations, and potential areas for improvement of these methods, this paper seeks to offer insights into the evolution of partitional clustering and guide future research directions. The goal of this survey is to serve as a valuable resource for researchers for selecting and designing effective meta-heuristic algorithms for complex clustering tasks and for understanding the current state of partitional clustering. To analysis this rich literature, several research questions are designed. The paper is divided into six sections. Section second summarizes the methodology adopted for the survey. The different techniques adopted for cluster analysis are discussed in section three. Section four presents the diverse clustering objective functions, performance metrics, and datasets considered for clustering problems. Section five discusses the various open issues and challenges related to clustering. Section six concludes the entire article, including the research questions devised in section two.

2 Methodology for the survey

This section including research questions, source of information, and inclusion and exclusion criteria of research articles for an effective and efficient survey. Figure 2 illustrates the process of collecting research articles for this survey.

2.1 Source of information

The following databases are explored for the domain of data clustering.

- Google scholar (www.scholar.google.co.in)

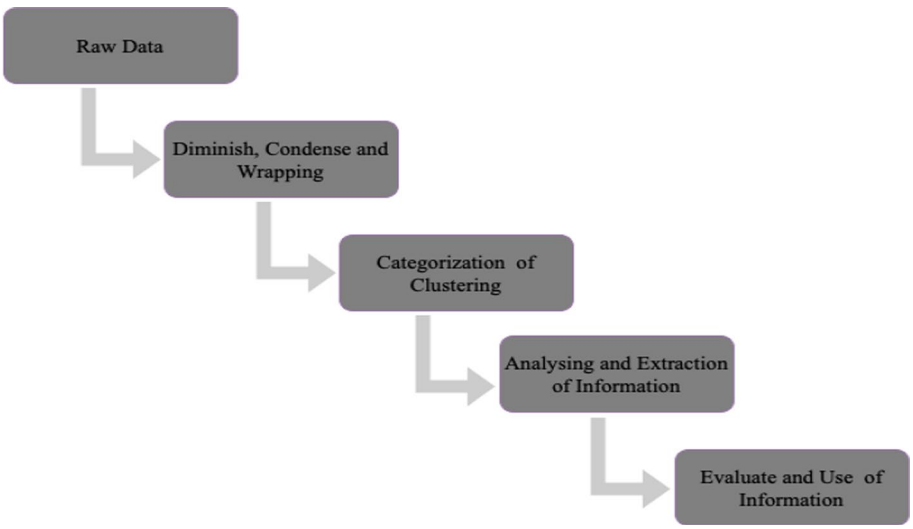

Fig. 2 Research articles collection process

- IEEE (www.ieeexplore.ieee.org)
- Springer (www.springerlink.com)
- Science Direct (www.sciencedirect.com)
- ACM digital library (dl.acm.org)
- Semantics scholar (www.semanticscholar.org)
- Elsevier (www.elsevier.co.in) and others

2.2 Inclusion and search criteria

The objective is to find various meta-heuristic algorithms for effective handling of clustering problems. Figure 3 describes the process of inclusion and exclusion of research articles. The meta-heuristic algorithms considered meet the following criteria:

(i) Related to meta-heuristic algorithms.
(ii) Includes data on high dimensional clustering, data clustering, dynamic, and automatic clustering.
(iii) Related to single objective and multi-objective clustering.
(iv) Work published in between 2015 to 2024.
(v) Published in SCI and SCOPUS-listed journals.

Initial search considered all relevant work with key words: (Data clustering) < OR > (Meta heuristic algorithms) < OR > (Single objective Clustering) < OR > (Multi-objective clustering) < OR > (High dimensional clustering) < OR > (Data clustering) < OR > (dynamic and automatic clustering) < OR > (Graph clustering).The above query generated literature rather than a title or abstract.

2.3 Exclusion criteria

An exclusion criterion is also adopted for the exclusion of non-relevant research papers. Research articles from journals of high repute are only considered (SCI and free Scopus). The exclusion criterion includes research published in books, national and international conferences, magazines, newsletters and educational courses, symposium workshops, and journals of less repute.

2.4 Extraction of articles

Initially, 956 articles are collected from various research databases. A huge amount of research articles were found due to the keyword “clustering”. The next step is to exclude non-relevant as per the criteria. It resulted in 455 research articles. Further, research articles published in the journal of repute are considered by manually removing articles from non-repute journals, books, and magazines. It resulted in the exclusion of 182 more research articles. During the study, 189 research articles didn’t fit well in the predefined search criteria. Finally, 130 research articles are analysed during the survey. Table 1 illustrates the data. Further, a team of four researchers is formed to manually select articles on predefined search criteria. Initially, two researchers select the articles, and the selected articles are further crosschecked by the third and fourth researchers. In case of a conflict,

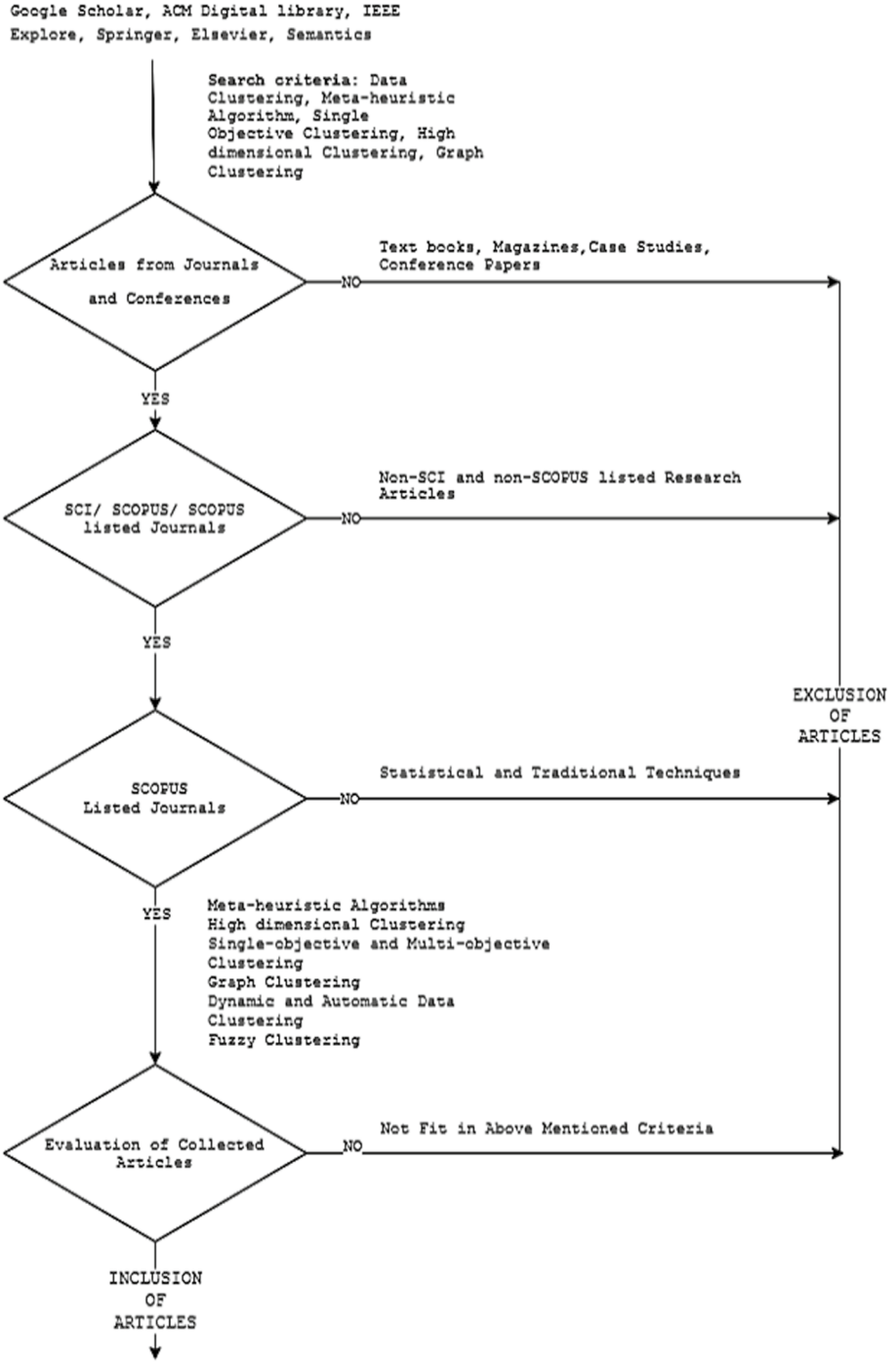

Fig. 3 Process of inclusion and exclusion of research articles for review

Table 1 Journal composition after selection

Sr. No.	Journal name	Publisher	No. of papers
1	Pertanika J. Sci. & technology	University of Putra	1
2	Pattern recognition letters	Elsevier	2
3	Expert system with applications	Elsevier	3
4	Knowledge based system	Elsevier	5
5	Neural computing and applications	Springer	6
6	European journal of operational research	Elsevier	1
7	Journal of big data	Springer	2
8	Transactions on pattern analysis and machine intelligence	IEEE	1
9	Vietnam journal of computer science	Springer	1
10	Neurocomputing	Elsevier	3
11	Information sciences	Elsevier	2
12	Journal of clinical monitoring and computing	Springer	1
13	International journal of machine learning and cybernetics	Springer	2
14	Pattern recognition	Elsevier	4
15	Applied soft computing	Elsevier	13
16	Transactions on knowledge and data engineering	IEEE	2
17	System journal	IEEE	1
18	IEEE access	IEEE	2
19	Transactions on evolutionary computation	IEEE	3
20	Internet of things journal	IEEE	1
21	Transactions on cybernetics	IEEE	2
22	Transactions on neural networks and learning systems	IEEE	1
23	Soft computing	Springer	8
24	Signal processing	Elsevier	1
25	Fuzzy sets and systems	Elsevier	1
26	Engineering applications of artificial intelligence	Elsevier	4
27	Information sciences	Elsevier	1
28	Swarm and evolutionary computation	Elsevier	1
29	Journal of intelligent systems	De Gruyter	1
30	Progress in artificial intelligence	Springer	1
31	Technology in cancer research & treatment	SAGE	1
32	Journal of information and communication technology	University of Utara	1
33	AI communications	IOS Press	1
34	Applied intelligence	Springer	2
35	Ain shams engineering journal	Elsevier	1
36	Alexandria engineering journal	Elsevier	1
37	Arabian journal for science and engineering	Springer	2
38	Computers & industrial engineering	Elsevier	1
39	Evolutionary intelligence	Springer	6
40	Hacettepe journal of mathematics and statistics	Hacettepe University	1
41	IEEE transactions on cybernetics	IEEE	1
42	IEEE transactions on emerging topics in computing	IEEE	1
43	IEEE transactions on evolutionary computation	IEEE	1

Table 1 (continued)

Sr. No.	Journal name	Publisher	No. of papers
44	Indian academy of sciences	Springer	1
45	Intelligent decision technologies	IOS press	1
46	Journal of ambient intelligence and humanized compute	Springer	1
47	Knowledge and information systems	Springer	1
48	Pattern analysis & applications	Springer	1
49	Cluster computing	Springer	1
50	Computers & electrical engineering	Elsevier	1
51	IEEE transactions on fuzzy systems	IEEE	1
52	IEEE transactions on knowledge and data engineering	IEEE	1
53	Procedia computer science	Elsevier	1
54	Proceedings of 11th international conference on bioinformatics and computational biology	EPiC series in computing	1
55	International journal of intelligent engineering and systems	Intelligent networks and systems society	1
56	Applied artificial intelligence	Taylor & Francis	2
57	Multimedia tools and applications	Springer	2
58	Natural computing	Springer	1
59	Engineering with computers	Springer	1
60	Journal of intelligent & fuzzy systems	IOS press	1
61	International journal of machine learning and cybernetics	Springer	1
62	Journal of intelligent manufacturing	Springer	1
63	The journal of supercomputing	Springer	1
64	Expert systems with applications	Elsevier	4
65	Iran journal of computer science	Springer	1
66	Knowledge-based system	Elsevier	1
67	Scientific reports	Springer	1
68	Symmetry	MDPI	1
69	The institution of engineering and technology	Wiley	1
70	Plos one	Plos one	1

a collective decision has been taken by the team. This process has been repeated in every phase of study selection. Table 1 and Fig. 3 illustrate journals considered for the survey.

Figure 3 provides a comprehensive visualization of the distribution of research articles across various journals within the surveyed literature. The figure presents a tabular representation with three columns: Sr. No., Journal Name, Publisher, and No. of Papers. Each row in the table corresponds to a specific journal and includes details such as the journal name, publisher, and the number of papers published within the surveyed literature. This detailed breakdown allows for a clear understanding of the publication landscape and the relative contribution of each journal to the body of research on clustering algorithms. From prestigious publishers like Elsevier and Springer to specialized journals such as IEEE Transactions, the table encompasses a wide array of publication outlets. It highlights the diversity of sources from which researchers draw when exploring clustering algorithms, reflecting the interdisciplinary nature of the field. By presenting this information in a

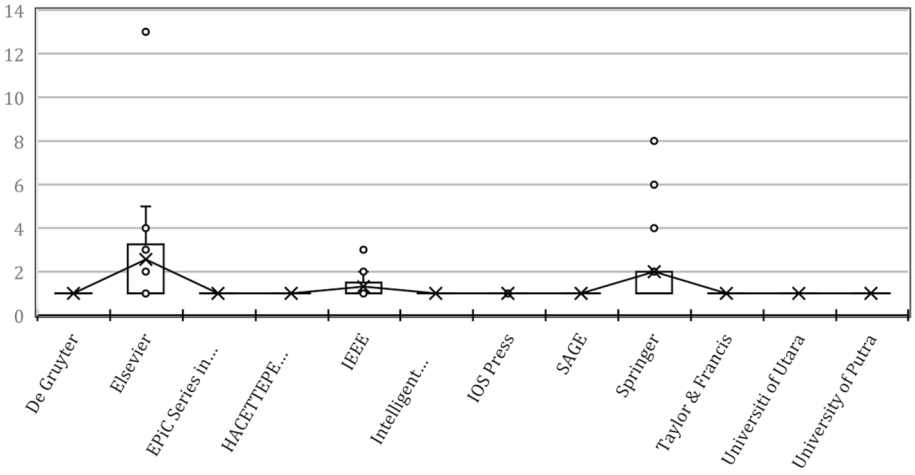

Fig. 4 Box and whiskers diagram representation of article composition during the survey

structured and easily digestible format, Fig. 4 offers valuable insights into the dissemination of knowledge within the clustering research community, aiding researchers in identifying key journals and publishers within the domain.

2.5 Data classification process

Finally, articles are classified into five and explored thoroughly to find key points for comparative study. Articles are reanalysed and evaluated on parameters (i) Algorithm/methodology used (ii) Type of clustering (iii) Data sets used (iv) Performance metrics and (v) Authors.

3 Literature survey

The literature survey is divided into five subsections.

This section analyses various meta-heuristic algorithms reported for clustering problems. Further, clustering problems are divided into Partitional clustering, dynamic and automatic clustering, and fuzzy clustering.

3.1 Meta-heuristic algorithms for partitional clustering

Meta-heuristic algorithms are higher-level procedures and heuristics for optimization problems. These algorithms are optimization algorithms inspired by natural phenomena such as biological evolution and swarm behaviour. These algorithms aim to find the optimal and near-to-optimal solution for Partitional problems. Further, several assumptions are taken into consideration for solving optimization tasks. These algorithms have been applied to clustering tasks to improve the quality of the clustering process and overcome challenges such as determining the optimal number of clusters, handling complex data distributions, and dealing with outliers. In this section, we explore improved meta-heuristic clustering

algorithms that have been developed to enhance clustering performance, focusing on novel strategies and recent advancements. Meta-heuristic clustering algorithms, such as Genetic Algorithms (GAs), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO), use population-based search strategies to optimize clustering objectives. In Partitional clustering, these algorithms aim to find a set of cluster assignments that maximize intra-cluster similarity while minimizing inter-cluster similarity. Moreover, the data are partitioned into a fixed number of clusters using some distance measures. It is also noticed that the number of clusters is fixed and known in advance. In Partitional clustering, Euclidean distance is applied to determine the optimal set of clusters in most cases. Partitional clustering is also known as non-overlapping clustering because the data belongs to only one cluster. The popular example of Partitional clustering is K-mean and it is also known as hard clustering. Table 2, illustrates Partitional clustering literature during the survey. Table 2, illustrates Partitional clustering literature in terms of meta-heuristic algorithms that can be applied for improving the efficacy of the clustering problems.

3.1.1 Meta-heuristic algorithms for dynamic and automatic partitional clustering

Dynamic and automatic clustering is a sub-branch of Partitional clustering that focuses on grouping data points into meaningful clusters in scenarios where the data itself is changing over time, or new data is constantly being added. This presents a challenge because static clustering techniques, which rely on fixed data sets, might not be suitable for data that evolves. Dynamic clustering techniques aim to adapt to changes in the data set by adjusting cluster structures and numbers as new data is introduced or as data distribution changes. Automatic clustering involves algorithms that automatically determine the optimal number of clusters and other parameters required to generate the clusters. When combined, dynamic and automatic clustering can provide an effective approach for evolving data sets without requiring extensive manual intervention. Recently, meta-heuristic algorithms are optimization algorithms that can be used effectively in dynamic clustering because they provide flexible and efficient methods for exploring the search space. These algorithms are particularly useful in solving complex optimization problems and can adapt to changing environments. These meta-heuristic algorithms can be applied to dynamic and automatic clustering by defining an appropriate objective function, such as minimizing intra-cluster distance or maximizing inter-cluster distance. As the data changes over time, these algorithms can adapt the clusters accordingly, ensuring that the clustering remains relevant and meaningful. This clustering includes very large data, data streams, incomplete data, noisy data, unbalanced data, and structured data. In dynamic and automatic clustering, it is important to evaluate the model performance regularly, ensuring that the clusters remain meaningful as the data evolves. The choice of the specific algorithm will depend on the characteristics of the data set, including its size, dimensionality, and the rate at which it changes over time. This subsection highlights the recent work reported on dynamic and automatic Partitional clustering. Table 3, illustrates various dynamic and automatic clustering algorithms considered during the survey.

3.1.2 Meta-heuristic algorithms for fuzzy clustering (generalization of the partitional clustering)

Fuzzy clustering is also known as soft clustering. It is a generalization of the Partitional clustering method. In this clustering, each data can belong to more than one cluster. Fuzzy

Table 2 Illustrates partitional clustering

Author name	Algorithm used	Objective	Results
Bijari et al. (2018)	Memory-enriched big bang-big crunch algorithm	Proposed memory-enriched big bang big crunch (ME-BB–BC) to enhance the exploitation of the algorithm	Improved shortcomings of the k-means method
Dos Santos and Zarate (2015)	Tax map clustering mechanism	Distinct measures have been reported using TaxMap clustering to determine similar patterns in categorical data	Provides stable and satisfactory results
Ferrari and Castro (2015)	Meta-learning systems	Proposed new ways to obtain meta-knowledge for clustering tasks	Better clustering results generated
Gebru et al. (2016)	FWD-EM algorithm and WD-EM algorithm	Employed weighted-data Gaussian mixture model for heterogeneous and multimodal datasets	Optimized results generated
Gutierrez-Rodríguez et al. (2015)	Pattern-based clustering algorithm	Proposed algorithm to obtain patterns in numerical datasets without priori discretization	Provides better results than pattern-based clustering algorithms
Hahsler and Bolanos (2016)	DBSTREAM clustering algorithm	Proposed an algorithm to explore density level between micro-clusters with the help of a shared density graph	Improved quality of clusters generated
Jing et al. (2015)	Ensemble clustering with stratified sampling	Proposed ensemble clustering method to perform clustering in high dimensional data	Achieved better clustering results
Kaur and Datta (2015)	SUBSCALE Algorithm	Proposed SUBSCALE to find sub-space cluster	k Database scans for k- k-dimensional datasets to find the subspace clusters
Senthilnath et al. (2019)	Flower Pollination Algorithm	Proposed flower pollination algorithm for data clustering problems	FPA efficiently clusters the data and performs better than the state-of-the-art methods
Kumar et al. (2016b)	Differential search algorithm	Employed differential search algorithm to find the near-optimal solution for partitional clustering problems	The proposed algorithm provides state of art results for solving Partitional clustering problems
Kumar and Sahoo (2014)	CSS algorithm	Presented an algorithm based on charged particles to find solutions to partitional clustering problems	CSS algorithm provides enhanced and more precise results as compared to other algorithms

Table 2 (continued)

Author name	Algorithm used	Objective	Results
Kushwaha et al. (2018)	Magnetic Optimization Algorithm	Employed electro-magnetic force based algorithm to determine optimal centroid for clusters	Better clustering results in terms of accuracy
Li et al. (2017)	Density-based clustering mechanism	Identified dense regions as clusters in multi-target detection	The investigation is considered to be valid and efficient for multi-target detection in a cluttered environment
Montgomery et al. (2016)	Near-infrared spectroscopy algorithm	Employed K-means algorithm with Gaussian mixture model to obtain precise partitioning of the dataset	Helps to monitor cerebral auto-regulation function
Pohl et al. (2016)	Online sub-event detection mechanism	Proposed a framework to investigate the problem of identifying the real-time sub-events in social media data (i.e., Twitter, Flickr and YouTube) during emergencies	Provides better support to an emergency responder
Safarinejadian and Hasanpour (2016)	MABDEM algorithm	The proposed algorithm is applied to determine density distribution	The proposed algorithm effectively measures the density regions
Santi et al. (2016)	Variable neighborhood search algorithm	Introduced mathematical programming model based on p median to cluster data and employed variable neighborhood search	The proposed model provides effective clustering results
Tang et al. (2016)	Intrusive tumour growth-inspired optimization algorithm	Presented a new meta-heuristic algorithm for data clustering inspired by tumor growth	The proposed algorithm is more robust, efficient, and effective than traditional algorithms
Vo et al. (2016)	VQ_fk_nps algorithm	Proposed an approach to handle incomplete educational data clustering	Provides optimized clusters with arbitrary shapes in data space
Zhang et al. (2016b)	Novel random-walk-based graph-based clustering	Presented a random walk-based clustering method to find attractor vertices	Prove its superiority over graph clustering methods
Zhang et al. (2016c)	Multivariate optimization algorithm	Applied multi-variant optimization to determine optimal solutions for clustering problems	The proposed algorithm gives more stable results than other algorithms

Table 2 (continued)

Author name	Algorithm used	Objective	Results
Nazari et al. (2019)	k-Mea ns and fuzzy-c means	Proposed a cluster-level weighting framework that ensembles the clusters based on the reliability of cluster	Provides better results for cluster ensemble
Das and Das (2018a)	Student learning ability	Proposed a new population-based class topper Optimization (CTO) algorithm that is inspired by the knowledge aptitude of students in a class and led to achieving the global best solution	Helps in solving real-life complex optimization problems
Quio and Zohu et al. (2019)	k-Means and water cycle algorithm	Proposed an algorithm using a simple water cycle algorithm and perculation operator for cluster analysis. The rainfall process is discarded in the proposed algorithm	The proposed WCA performs significantly better in terms of speed, stability, and quality
Deb et al. (2018)	C-Elephant search algorithm	Proposed a new data clustering algorithm C-ESA	SBD is used for measuring the distance between each time series data, computation time, and clustering accuracy
Alswaitti et al. (2018)	Gravitational clustering	Proposed a data clustering algorithm based on the universal gravity rule to balance between the exploitation and exploration	Clusters formed are more homogeneous than those formed using standard methods
Pimentel and Carvalho (2019)	Meta-models	Proposed a new approach based on meta-models for the performance prediction of clustering algorithms	Meta-features have helped to improve the recommendation
Narayana and Vasumathi (2018)	K-medoids	Proposed a similarity-based K-medoids clustering technique to process large datasets	Better results in terms of clustering accuracy, error rate, execution time (s), adjusted Rand index (ARI), and convergence time (s)
Zhu and Ma (2018)	New clustering validity index	For finding the optimal number of clusters and clustering partition	Results in optimal clusters

Table 2 (continued)

Author name	Algorithm used	Objective	Results
Salem et al. (2018)	k-Means	Introduced a new categorical method based on partitions known by Manhattan frequency k-Means (MFk-M)	MFk-M outperforms the k-modes while clustering categorical data and the k-means while clustering quantitative datasets
Xie et al. (2019)	K-means & fire fly algorithm	Proposed IIEFA and CIEFA to solve the problem of initialization and getting trapped in local optima	Superior in both performance and distance measures for clustering
Kuwil et al. (2019)	Similarity between clusters	Proposed a novel distance-based clustering algorithm called the critical distance clustering algorithm	Produced reasonable clusters without specifying parameters priory
Singh (2020)	Hariss hawk optimization	Proposed chaotic sequence Hariss hawk optimization for data clustering	Helps in finding the optimized cluster centers for the considered data set
Kaur and Kumar (2022)	Water wave optimization	Proposed new metaheuristic based on water wave optimization for data clustering Improvements are incorporated to improve global best information and premature convergence. The modified search equation and decay operator have been incorporated for the aforementioned issues	Results show an improved accuracy of 4% and F-score of 7% in comparison to traditional and hybrid clustering algorithms
Lee and Perkins (2021)	Simulated annealing with Gaussian mutation and distortion equalization algorithm	Proposed clustering algorithm based on simulated annealing (SA), named SAGMDE The proposed algorithm employed dual perturbation using Gaussian Mutation (GM) along with the Distortion Equalization (DE) algorithm,	SAGMDE generates consistent and better cluster-quality results

Table 2 (continued)

Author name	Algorithm used	Objective	Results
Mansuet and Schoen (2021)	Memetic differential evolutions	Proposed memetic differential evolution (MDE) for Euclidean minimum sum of squares clustering (MSSC)	MDE gives good quality results in terms of minimizing an objective function, improving efficiency, and number of calls to the local optimization routine
Turkoglu et al. (2022)	Artificial algae algorithm	Proposed a new metaheuristic-based Artificial Algae algorithm (AAA) for data clustering	AAA efficiently explores the search space and prevents it from stuck into local optima. Results provide a high convergence rate and accurate solutions in comparison to various existing metaheuristics
Demirci et al. (2023)	Electrical search algorithm (ESA)	Proposed a new metaheuristic algorithm called the Electrical Search Algorithm (ESA). The objective is to address the initialization problem and lower and upper bounds while searching for the solution	Results are provided at a persistent rate and don't trap in local optima
Alotaibi (2022)	Meta-heuristic tabu search adaptive search memory	Proposed a new meta-heuristics algorithm called MHTSASM for data clustering. The proposed algorithm uses TS with K-means	Generates superior results in terms of f-measure, precision, and recall
Moghadam and Ahmadi (2023)	Red deer clustering algorithm (RDCA)	Proposed a new bio-inspired red deer clustering algorithm. Two-stage algorithm that adopts CLIQUE algorithm to partition the data space and then employs a search strategy for a modified version of RDA	Results show that the proposed algorithm is less sensitive to increasing the number of clusters and dimensions

Table 2 (continued)

Author name	Algorithm used	Objective	Results
Harita et al. (2024)	Montecarlo-clustering search algorithm	Proposed a heuristic search algorithm based on Montecarlo sampling and clustering strategy. The clustering strategy employs DBSCAN in the first phase and then K-means in the second phase. The proposed approach was evaluated on different test functions and tested for the Knapsack Problem	Achieved high-quality solutions surpassing 90% of the results in comparison

Table 3 Illustrates dynamic and automatic clustering

Author name	Algorithm used	Objective	Results
Yao et al. (2018)	K-prototypes algorithm	Proposed an improved algorithm to determine the initial cluster centres and update the cluster number automatically using entropy for the mixed data	Improved algorithms guarantee the unique clustering result and have good performance
Kumar et al. (2016a)	Parameter adaptive harmony search algorithm	Applied parameter adaptive harmony search to determine the number of clusters automatically from the given dataset	Results in well-separated and compact clusters
Puschmann et al. (2017)	Adaptive clustering method for dynamic IOT data streams	Presented adaptive clustering method to find the number of clusters in the data stream automatically	The proposed method provides better in terms of cluster quality
Sheng et al. (2014)	Multilocal search and adaptive niching-based genetic algorithm	Developed genetic algorithm-based clustering algorithm for automatic data clustering	Generated quality results
Sheng et al. (2016)	Adaptive multi-subpopulation competition and multiniche crowding-based memetic algorithm	Proposed a memetic algorithm based on adaptive multi-subpopulation competition and multi-niche crowding methods	Provides superior results
Xiang et al. (2015)	Dynamic shuffled differential evolution algorithm	Proposed a dynamic shuffled differential evolution algorithm for data clustering	Yields more effective results
Yang and Jiang (2018)	HMM and Bi-Weighting Scheme	Proposed hybrid meat-clustering technique to solve the initialization and model selection problems, determining the number of clusters automatically	Best performance
Zhou et al. (2019)	Symbiotic organism search	Employed symbiotic organism search (SOS). Symbiotic communication strategies have been adopted by organisms and also the initial search space has been generated randomly	Results in stable and accurate clusters
Queiroga et al. (2018)	continuous GRASP	Proposed C-GRASP-based algorithm, automatically adjusted the step size while searching	Provides good results in terms of convergence analysis and solution quality

clustering is a type of clustering approach where each data point can belong to more than one cluster with a certain degree of membership. In contrast to traditional (hard) clustering methods, such as k-means, where each data point is assigned to one and only one cluster. Fuzzy clustering is particularly useful when the boundaries between clusters are not clear-cut, or when the data itself is inherently ambiguous or overlapping. The most commonly used fuzzy clustering algorithm is Fuzzy C-Means (FCM), introduced by Jim Bezdek in 1981. FCM is an extension of the classic k-means algorithm that allows data points to have partial membership in multiple clusters. Fuzzy clustering is widely used in various applications such as pattern recognition, data analysis, image segmentation, and bioinformatics, where overlapping or ambiguous groups may exist in the data. Further, Meta-heuristic algorithms can be employed in fuzzy clustering to optimize the clustering process, particularly in terms of finding the optimal number of clusters, the best initial cluster centroids, or the optimal fuzziness parameter (m). The most common fuzzy clustering algorithm is Fuzzy C-Means (FCM), but it can suffer from limitations such as sensitivity to initial conditions and local optima. Meta-heuristic algorithms can help improve the performance of fuzzy clustering by exploring a broader search space and finding better solutions. By integrating meta-heuristic algorithms with fuzzy clustering, more robust, flexible, and efficient clustering results can be obtained in complex data environments. Table 4, highlights the recent work reported on fuzzy clustering. Fuzzy clustering is widely used in various applications such as pattern recognition, data analysis, image segmentation, and bioinformatics, where overlapping or ambiguous groups may exist in the data. Further, Meta-heuristic algorithms can be employed in fuzzy clustering to optimize the clustering process, particularly in terms of finding the optimal number of clusters, the best initial cluster centroids, or the optimal fuzziness parameter (m). The most common fuzzy clustering algorithm is Fuzzy C-Means (FCM), but it can suffer from limitations such as sensitivity to initial conditions and local optima. Meta-heuristic algorithms can help improve the performance of fuzzy clustering by exploring a broader search space and finding better solutions. By integrating meta-heuristic algorithms with fuzzy clustering, more robust, flexible, and efficient clustering results can be obtained in complex data environments. Table 4, highlights the recent work reported on fuzzy clustering.

3.1.3 Improved meta heuristic algorithm for partitional clustering

Meta-heuristic algorithms can explore the search space to determine solutions to optimization problems. But, sometimes it is not possible to explore the entire search space through a meta-heuristic algorithm. As these algorithms are not exact; so to enhance the performance of meta-heuristic algorithms, a few amendments can be made to improve the efficiency and effectiveness of meta-heuristic algorithms. These amendments can be described as using neighbourhood concepts, defining new search strategies, making the algorithmic parameters adaptive, etc. The improved meta-heuristic algorithms can be described by enhancing their efficiency, convergence speed, exploration–exploitation balance, and robustness in solving Partitional-clustering problems. It can be understood as combining different meta-heuristic algorithms according to their strengths and offset individual weaknesses. Further, integrating the local search methods with meta-heuristics can refine solutions in promising areas of the search space. Dynamically adjust the parameters of the algorithms based on feedback from the search process so that these algorithms can adapt more effectively to solve the clustering problems. Also, design the procedure for algorithms to self-adapt parameters automatically during the search. These improvements can be tailored and

Table 4 Illustrates fuzzy clustering

Author name	Algorithm used	Objective	Results
Amiri and Mahmoudi (2016)	Fuzzy cuckoo optimization algorithm	Proposed new fuzzy cuckoo optimization to determine the number of clusters	Generated optimum cluster centers
Chang et al. (2016)	FCM with sparse regularization	Employed fuzzy c-means (FCM) model with sparse regularization to determine relevant features and clustering structure in high dimensional data	FCM gives better and enhanced results
Ghorbanzadeh et al. (2016)	Adaptive neuro-fuzzy-based correlation model	Applied adaptive neuro-fuzzy-based correlation model to extract relevant features and form clusters	Efficient and effective in reducing tumour tracking errors as compared to the Cyber knife model
D'Urso and Leski (2016)	Fuzzy c-ordered-medoids clustering	Proposed fuzzy c-ordered-mediods for clustering. It also helped to detect outliers for interval-valued data	Provides state of art results than other fuzzy clustering methods
Li et al. (2016)	Density-based weighted online FCM algorithm (OWFCM) and (strwfcm) algorithm	Presented two novel data stream clustering algorithms to cluster large-scale data efficiently	Yields better quality results
Meng et al. (2016)	Vigilance parameter in fuzzy ART self-adaptable	Investigated the efficiency of fuzzy adaptive resonance theory for solving clustering problems	Yield superior results
Yang et al. (2015)	Non-dominated sorting genetic algorithm	Developed clustering algorithm based on Non-dominated sorting and fuzzy genetic algorithm for improving clustering quality of categorical data	Yields better results in terms of fuzzy compactness, separation among clusters, and computation time
Zhang et al. (2016a)	Fuzzy clustering algorithm	Presented Fuzzy clustering to cluster the data streams	The proposed algorithm is more capable of detecting concept drift using entropy theory
Xu et al. (2019)	Fuzzy rough set	Proposed fuzzy rough algorithm (FRC) for handling categorical data	Performs better on most of the datasets
Nayak et al. (2018)	Elicit TLBO and Fuzzy-C means	Integrated Elicit TLBO with Fuzzy-C means to improve the fitness value of cluster centers	Best clusters formed

Table 4 (continued)

Author name	Algorithm used	Objective	Results
Kushwaha and Pant (2018)	Fuzzy C-means	Presented a novel clustering algorithm called fuzzy magnetic optimization clustering (Fuzzy-MOC) to handle the issue of getting stuck in local optima in the case of Fuzzy-C Means	Fuzzy-MOC outperforms in terms of performance metrics like F1, purity, accuracy, and RI measure
Baykasoğlu, Gölcük and Özsoydan (2018)	Fuzzy C-means & swarm intelligence	Employed weighted superposition attraction algorithm (WSA) to enhance the performance of fuzzy-c means clustering	Indicated significant improvements over the traditional fuzzy c-means algorithm
Mikaeil et al. (2018)	Lloyd's algorithm (k-means clustering)	Employed imperialist competitive algorithm and fuzzy C-mean to classify dimensions of stones. It considers the mechanical and physical attributes of dimension stone	Dimension stones have been classified. Results have been validated through the performance of the diamond circular saw measured in terms of hourly production rates
Yan et al. (2019)	Fuzzy cluster ensemble	Presented a new fuzzy cluster ensemble method known as improved fuzzy cluster ensemble (IFCE) to improve robustness against noise	Gives stable results
Gupta and Saini (2019)	PSO, K-harmonic means	Combined PSO with KHM to overcome the problem of being stuck in the local optima of KHM, use of fuzzy logic for making PSO adaptive	Better clustering results
Su and Denoeux (2018)	Belief peaks on the notion of belief functions	Introduced belief-peaks evidential clustering (BPEC) to form the cluster center based on belief functions	BPEC method results in good performance
Kuo et al. (2021)	Sine–cosine algorithm	Proposed a fuzzy possibilistic c-ordered means (FPCOM) algorithm to overcome the influence of outliers on clustering results. Further, the proposed SCA-FPCOM algorithm combines the sine cosine algorithm (SCA) with FPCOM to handle parameter problems and initial centroids	Increased clustering performance and reduced cost of search for parameters

Table 4 (continued)

Author name	Algorithm used	Objective	Results
Singh and Srivastava (2022)	Teaching learning-based optimization and kernel fuzzy C-means algorithms	Proposed new approach using teaching learning-based optimization with Kernel Fuzzy C-means (TLBO-KFCM). Further, employed kernel function for improving separation and performing clustering. TLBO has helped in improving the compactness of clustering	TLBO-KFCM approach gives better performance
Hashemi et al. (2023)	Fuzzy C-means (FCM) and whale optimization algorithm (WOA)	Proposed FCM-WOA for handling large datasets. The initialization sensitivity issue and slow convergence issues of FCM are handled thru WOA	FCM-WOA effectively clusters large data in less time and gives better clustering results

combined in various ways depending on the specific problem and application. Research and innovation in meta-heuristic algorithms continue to evolve, and new approaches and enhancements are regularly being proposed in the academic and research communities. Hence, this section summarizes the improvements reported in original meta-heuristic algorithms for effectively solving clustering problems. Table 5, illustrates various improved metaheuristic algorithms in literature.

3.1.4 Hybrid metaheuristic algorithm for partitional clustering

Hybridization is a warm area of research to improve and enhance the performance of algorithms. A hybrid meta-heuristic algorithm combines different meta-heuristic approaches or integrates a meta-heuristic with other optimization techniques to take advantage of their respective strengths while mitigating weaknesses. In the context of clustering, a hybrid meta-heuristic algorithm can optimize cluster assignments and centroids while balancing exploration and exploitation in the search process. Hybrid meta-heuristic algorithms for Partitional clustering combine the strengths of different optimization techniques to achieve better clustering results. Partitional clustering involves dividing the dataset into disjoint clusters where each data point belongs to exactly one cluster. A hybrid meta-heuristic algorithm for Partitional clustering can enhance the clustering process by improving the selection of initial cluster centres, balancing exploration and exploitation during the search process, and increasing the algorithm's robustness and efficiency. Hybrid meta-heuristic algorithms can be fine-tuned and adapted based on the specific clustering problem and dataset characteristics. This approach can be particularly beneficial for complex clustering problems where traditional methods may struggle. By leveraging the strengths of multiple meta-heuristic approaches, hybrid algorithms can potentially outperform individual methods, offering more robust and effective solutions for clustering problems. Hence, this section aims to present various hybrid meta-heuristic algorithms reported for solving clustering problems. Table 6, illustrates various hybrid metaheuristic algorithms for clustering in literature.

4 Objective function, performance metric and dataset

This section describes various objective functions, performance metrics, and datasets used to solve clustering problems.

4.1 Objective function

Clustering is an unsupervised technique that can be applied for data exploration. Clustering aims to find a group of data, known as clusters. An objective function is required to find these groups of data. The objective function is a distance-based function that can measure the distance between data and clusters. Hence, the objective function in clustering aims to determine the quality of clusters. This can be described in terms of cluster compactness. The cluster compactness can be defined as the total distance of each cluster data to the cluster centroid. There are a lot of objective functions presented in the literature for effective clustering. Without these, the clustering cannot be performed. For effective clustering, it is necessary to pick the appropriate clustering objective. Table 7 depicts the well-known clustering objective reported for the clustering task. It is seen that Euclidean distance is a

Table 5 Improved metaheuristic

Author name	Algorithm used	Objective	Results
Alam et al. (2015)	Evolutionary PSO, hierarchical PSO	Proposed a clustering algorithm based on evolutionary PSO and hierarchical PSO for effective clustering	Gives more accurate and efficient results than other techniques
Allab et al. (2017)	Semi-NMF-PCA algorithm	Proposed a new algorithm based on the Semi-NMF-PCA algorithm for data clustering and dimension reduction	Provides more accurate results
Bahrololoum et al. (2015)	Gravity based algorithm	Proposed a gravity-based algorithm for clustering problems to reduce noise effect and enhance clustering quality	A competitive and effective algorithm for solving clustering problems
Banharnsakun (2017)	MapReduce-based artificial bee colony algorithm	Presented MapReduce-based artificial bee colony algorithm to optimize large-scale data	Capable of clustering large-scale data and also maintaining the quality of clustering
Cruz et al. (2016)	Bee algorithm	Proposed a capture algorithm for data clustering problems	More competitive and effective than other compared clustering algorithms
Han et al. (2017)	Bird flock gravitational search algorithm	Proposed Bird Flock Gravitational Search Algorithm to address premature convergence GSA algorithm	Gives more precise results
Kumar and Singh (2018)	Improved cat swarm optimization algorithm	Introduced an improved cat swarm optimization algorithm to improve the convergence for solving data clustering problems	ICSO algorithm gives better results in contrast to compared algorithms
Noorbehbahani et al. (2015)	Mixed self-organizing incremental neural network algorithm	Proposed self-organizing incremental neural network algorithm for mixed data clustering	A competitive and efficient algorithm for mixed data clustering
Ozturk et al. (2015)	Improved discrete binary artificial bee colony algorithm	Proposed an improved discrete binary artificial bee colony algorithm for dynamic clustering	Provides superior results
Siddiqi & Sait (2017)	Greedy and evolutionary algorithm	A proposed new algorithm based on a greedy approach and a few heuristics to determine the optimized centroid of clusters with a fixed number of clusters	Gives better clustering results and requires a lesser number of evaluations

Table 5 (continued)

Author name	Algorithm used	Objective	Results
Yu et al. (2016)	The new tree-based incremental overlapping clustering algorithm	Proposed an algorithm to tackle overlapping and incremental clustering	The performance of the proposed algorithm is far better than other compared algorithms using F-measure and NMI indices
Yuwono et al. (2014)	Rapid centroid estimation algorithm	Developed a rapid centroid estimation algorithm to reduce the computational complexity and simplify update rules	Gives more accurate and improved results
Tsai et al. (2019)	Coral reef optimization	Proposed coral reef optimization with substrate layers (CRO-SL to find the optimized clusters for big data	Gives better clustering results than compared clustering algorithms
Boushaki et al. (2018)	Cuckoo search algorithm	Proposed a new quantum chaotic cuckoo search algorithm (QCCS) to improve the search exploitation and diversification	Enhanced searching strategy for handling the boundary values
Abualigah et al. (2018a)	Krill herd algorithm	Proposed an improved Krill Herd algorithm to balance between exploration and exploitation	Provides better results
Das et al. (2018b)	Bee colony optimization	Proposed modified Bee Colony Optimization (MBCO) approach for data Clustering. Also, proposed MKCLUST and KMCLUST for obtaining the global optima and different solutions	Proposed algorithms are competent for data clustering
Liu et al. (2019)	Path-based clustering algorithm	Optimized a criterion function for Path-based clustering algorithms to improve the clusters	IPC outperforms when compared with clustering algorithms
Cho and Nyunt (2020)	Differential evolution and quasi-opposition based learning (QBL) strategy	Proposed enhanced DE for faster convergence and to generate quality cluster results. Mutation strategy adopted with QBL strategy for initial solution selection	Increased convergence rate and quality cluster results

Table 5 (continued)

Author name	Algorithm used	Objective	Results
Ahmadi et al. (2021)	Grey wolf optimizer	Proposed modified grey wolf optimizer to improve exploration and exploitation for producing the best solution with improved local search ability	Lower intra-cluster distance and lowest average error rate of 11.2% as compared to other clustering techniques
Rahnema and Gharehchopogh (2020)	Artificial bee colony algorithm, whale optimization algorithm	Proposed Artificial Bee colony (ABC) based on whale optimization algorithm (ABCWOA). Employed elite memory and random memory concepts	ABCWOA has a positive effect on the population and contains less noise
Kumar and Kaur (2022)	Bat Algorithm	Proposed enhanced cooperative co-evolution method for addressing issues in Bat algorithm (BA). Proposed three variants BA-C, BA-CN, and BA-CNE	Effective clustering results are given
Asadi-Zonouz et al. (2022)	Unconscious search algorithm, K-means	Proposed modified unconscious search (US) by hybridizing with K-means. Firstly, modification in initial population generation is considered. Secondly, the local search of the US is replaced through a heuristic search method	Generated 0.176% better quality solutions
Abualigah et al. (2023)	Augmented arithmetic optimization algorithm	Proposed Augmented arithmetic Optimization algorithm (AAOA) by integrating opposition-based learning and Levy flight in arithmetic optimization algorithm (AOA). The objective is to improve exploration and exploitation trends	Boosts the search mechanism and provides improved results
Shoal et al. (2023)	Enhanced gray wolf optimization (GWO) algorithm	Proposed enhanced Gray wolf Optimization by introducing a differential perturbation operator. Select three random omega wolves that assist the three leader wolves of the original GWO algorithm to provide diverse and quality solutions	The effectiveness of the proposed algorithm has been evaluated using a pairwise Wilcoxon signed-rank test and the Friedman and Nemenyi hypothesis test

Table 5 (continued)

Author name	Algorithm used	Objective	Results
Singh et al. (2023)	Enhanced whale optimization algorithm	Developed an enhanced whale optimization algorithm (EWOA) for addressing the clustering issues. The enhanced version utilizes the advantages of Water Wave optimization, tabu search, and neighbourhood search mechanism for handling slower convergence rates and problems of local optima	Superior results are provided in terms of average intra-cluster distance and f-measure
Premkumar et al. (2024)	Augmented weighted K-means Grey wolf optimizer	Presented Augmented weighted K-means Grey wolf optimizer for handling data clustering issues. The proposed algorithms enhanced the grey wolf optimizer using a new weight factor and the K-means algorithm to increase diversity and avoid premature convergence	The proposed algorithm provides 35% better results on various test functions and clustering datasets as compared to the original grey wolf optimizer
Patel et al. (2023)	Local neighbor spider monkey optimization	To handle the problems of premature convergence, stuck in local optima, a local neighbour spider monkey optimization algorithm is proposed. The local Leader Phase of the spider monkey optimization algorithm has been improved thru its neighbour solution and a chaotic operator has been added to search the global leader phase in the enhanced algorithm	Provides improved results in terms of accuracy and f-measure
Salih et al. (2023)	Multi-population black hole algorithm (MBHA)	Proposed and enhanced version of Black Hole algorithm (BHA) as the multi-population algorithm to overcome the issues of BHA	The proposed MBHA gives a high convergence rate and overcomes the problems of BHA

Table 5 (continued)

Author name	Algorithm used	Objective	Results
Tekieh et al. (2024)	MapReduce clustering model	Proposed an improved artificial bee colony algorithm based on a MapReduce clustering model (MR-CWABC). The algorithm utilizes the Weighted-average artificial bee colony (WABC) algorithm Implemented for the Hadoop framework to find optimal samples to preserve the compactness & separation of clusters	Implemented for clustering Big data. Results show improvements of 7.13, 7.71, and 6.77 percentile of average f-measure

Table 6 Hybrid approaches

Author name	Algorithm used	Objective	Results
Elyasigomari et al. (2015)	Cuckoo optimization algorithm, genetic algorithm	Proposed an innovative gene selection approach using the shuffle method before cancer classification using COA-GA	Optimized clustering enhanced the accuracy of gene selection and classification
Kumar et al. (2015)	CulsiVAT algorithm	Proposed culsiVAT algorithm for large datasets. No requirement for initialization and data sampling	Faster and gives accurate results in contrast to algorithms in comparison
Kumar and Sahoo (2015a)	Improved cat swarm optimization, K-Harmonic means algorithm	Developed hybrid algorithm using improved CSO and K-Harmonic means to handle local optima and convergence issues	Improvement in convergence speed of CSO and escape local optima in the KHM method
Kumar and Sahoo (2015b)	Magnetic charge system search, particle swarm optimization	Proposed hybridization of magnetic charge system search and particle swarm optimization (MCSS–PSO) for efficient data clustering	Enables search to escape from local optima and explores more promising solution direction
Kumar and Sahoo (2016)	K-harmonic means, cat swarm optimization	Integrated CSO and KHM techniques for efficient data clustering	Attains advantages of both the algorithms and performs better
Nguyen et al. (2015)	Kernel interval-valued fuzzy clustering, multiple kernel interval-valued fuzzy c-means clustering	Developed hybrid clustering approach using single and multiple kernel interval-valued and fuzzy c-means. to enhance clustering results	Optimized results
Ozbakõr and Turna (2017)	Ions motion optimization, weighted superposition attraction algorithm	Adopted Ions motion optimization and Weighted superposition attraction algorithm for solving clustering problems	Competitive solution approach for clustering problems
Pakrashi and Chaudhuri (2016)	Heuristic Kalman algorithm, K-means algorithms	Developed an improved clustering algorithm based on heuristic Kalman algorithm and k-means algorithms to handle partitional clustering problems effectively	The hybrid algorithm developed performs better than its components

Table 6 (continued)

Author name	Algorithm used	Objective	Results
Serapião et al. (2016)	Swarm clustering algorithm	Propose Swarm clustering algorithm combination of K-means, K-harmonic means and fish school search, called FSS-SCA. Helps to exploit the search capability and prevent from getting stuck in local optima	FSS-SCA gives promising results for clustering
Abualigah, Khader and Hanandeh (2018b)	Krill Herd algorithm	Presented an enhanced hybrid krill herd algorithm with a harmony search algorithm for solving data clustering problems	The proposed hybridization (Harmony-KHA) is fast, efficient, and suitable for solving data clustering problems
Kuo et al. (2018a)	Kernel intuitionistic fuzzy c-means (KIFCM) algorithm, PSO, ABC, GA	Proposed three hybrid algorithms, PSO-KIFCM, ABC-KIFCM, and GA-KIFCM algorithms to overcome the problem of KIFCM	Proposed algorithms accomplished better accuracy
Gupta et al. (2018)	Automatic clustering, fuzzy relationships, and differential evolution	Proposed hybrid model for forecasting low dimensional numerical data	The proposed method is accurate and provides the lowest MAPE and MSE in comparison to the existing methods
Aljarah et al. (2020)	Grey wolf optimizer, Tabu Search	Proposed hybrid algorithm GWOTS using GWO and TS	Results showed improved performance in terms of SSE, entropy, and purity
Kuo et al. (2018b)	Growing self-organizing map, bee colony optimization algorithm	Proposed hybrid algorithm using growing self-organizing map (GSOM) algorithm and bee colony optimization based on self-organizing map (BCOSOM)	Finds better solution in comparison
Tinós et al. (2018)	GA and NK	Presented NK hybrid genetic algorithm (GA) for clustering. Proposed new mutation operator (PX) to help explore information	Results validation with the Wilcoxon Signed Rank Test gives a remarkable performance in comparison

Table 6 (continued)

Author name	Algorithm used	Objective	Results
Jadhav and Gomathi (2018)	Grey wolf optimizer and whale optimization	Proposed hybridized WGC using Wolf Optimization Algorithm (WOA) and Exponential Grey Wolf Optimization (EGWO) for data clustering	Performs better in the context of Jaccard coefficient, rand coefficient, minimum value of MSE, and maximum value for fitness
Sharma and Chhabra (2019)	PSO	Proposed encoded hybrid algorithm (PSOPC) using PSO for global search and polygamous approach for crossover	PSOPC outperforms other approaches in the context of within-cluster distance; cluster quality measures and convergence speed to find near-optimal solutions and can generate compact clusters
Lakshmi et al. (2018)	K-means, crow search method	Proposed hybridized crow search and K-means to find the global optimum solution	ANOVA and statistical tests show good performance
Bouyer and Hatamlou (2018)	An improved Cuckoo search algorithm, modified PSO, K-harmonic means	Integrated K-Harmonic means with improved Cuckoo search and modified PSO for fast convergence leading to global optimum, not falling into local optima	Produced stable clusters with high-quality
Mageshkumar et al. (2019)	ACO & ALO algorithm	Proposed a hybrid ACO-ALO Algorithm by combining Ant Lion Optimization and Ant Colony Optimization to increase the performance for solving data clustering problems. Cauchy's mutation operator has been adopted	The proposed ACO–ALO algorithm outperforms concerning traditional clustering algorithms for intra-cluster distance measure
Rathore et al. (2018)	Ensemble-based clustering algorithm	Proposed a hybrid approach new fast clustering algorithm named FensiVAT to handle the high dimensionality and huge volumes of data	Gives accurate and fastest results, and clusters large volumes of high dimensional data in less time

Table 6 (continued)

Author name	Algorithm used	Objective	Results
Kaur et al. (2020)	Chaotic optimization and flower pollination algorithm	Proposed hybrid Chaotic FPA to improve the efficiency of minimizing cluster Integrity	CFPA and BHA perform better than other algorithms based on cluster integrity. CFPA and CSA were superior concerning the execution time. CFPA and FPA converge earlier than the other algorithms. More stable results were produced by CFPA and BHA
Abasi et al. (2020)	Multiverse optimizer and K-means	Proposed hybrid algorithm H-MVO using multi-verse optimizer (MVO) and k-means	Improved global search ability, optimized cluster partitions
Pacifico and Ludermir (2021)	K-means, group search optimization	Employed K-means as a local search strategy for Group Search Optimization. Proposed FMKGSO, MKGSO, and TMKGSO hybrid algorithms	Limited steps are required for execution by combining GSO with K-Means. Robust approach for exploration and exploitation. Yields good computational cost performance
Hu et al. (2023)	*k*-Means clustering algorithm based on Lévy flight trajectory (Lk-means)	Proposed Levy flight trajectory K-means algorithm to avoid premature convergence and avoid the falling of k-means into local optima. The diversity of the clusters has also been increased with the proposed approach	The ability to handle large data, results in uniformly distributed cluster centroids and, also provides enriched search ability
Qtaish et al. (2024)	Hybrid capuchin search algorithm (HCSA)	To handle local optima issues and initialization sensitivity issues of K-means, a hybrid Capuchin Search algorithm by integrating the Chameleon Swarm algorithm with CSA	Results show the increased capacity for exploration and exploitation and give better clustering results in comparison
Haeri Boroujeni and Pashaei (2023)	Chimp optimization algorithm (ChOA), generalized normal distribution algorithm (GNDA), and opposition-based learning (OBL)	Proposed hybrid Chimp Optimization Algorithm (ChOA), Generalized Normal Distribution Algorithm (GNDA), and Opposition-Based Learning (OBL) to solve clustering problems	Achieves optimized clustering results and can handle large & complex datasets

Table 6 (continued)

Author name	Algorithm used	Objective	Results
Barshandeh et al. (2022)	Artificial jellyfish search algorithm (JS) and marine predator algorithm (MPA)	Presented a neoteric LA-based hybrid optimization algorithm for global optimization problems. The proposed algorithm utilized an artificial Jellyfish search algorithm (JS) and Marine Predator Algorithm (MPA) and has been implemented for solving clustering problems	The proposed algorithms have achieved 86.85% results on test functions and provide better clustering results in terms of convergence rate, dispersion, and SSE metric

Table 7 List of objective functions

Objective function	References
Euclidean distance	Alam et al. (2015), Amiri and Mahmoudi (2016), Bahrololoum et al. (2015), Banharnsakun (2017), Cruz et al. (2016), Santos and Zárate (2015), Gutierrez-Rodríguez et al. (2015), Han et al. (2017), Kumar et al. (2016a), Kumar et al. (2016b), Kumar and Sahoo (2014), Kumar and Sahoo (2015a), Kumar and Sahoo (2015b), Kumar and Sahoo (2016), Kumar and Singh (2018), Kushwaha et al. (2018), Nguyen et al. (2015), Özbakır and Turna (2017), Ozturk et al. (2015), Pakrashi and Chaudhuri (2016), Serapião et al. (2016), Sheng et al. (2014), Siddiqi and Sait (2017), Tang et al. (2016), Xiang et al. (2015), Yu et al. (2016), Yuwono et al. (2014), Zhang et al. (2016c), Xu et al. (2019), Alswaitti et al. (2018), Zhou et al. (2019), Boushaki et al. (2018), Abualigah et al. (2018b), Abualigah et al. (2018b), Das et al. (2018b), Kuo et al. (2018b), Mikaeil et al. (2018), Pimentel and Carvalho (2019), Narayana and Vasumathi (2018), Su and Denoeux (2018), Rathore et al. (2018), Kuwil et al. (2019), Abasi et al. (2020), Cho and Nyunt (2020), Ahmadi et al. (2021), Kaur and Kumar (2022), Lee and Perkins (2021), Pacifico and Ludermir (2021), Kumar and Kaur (2022), Hu et al. (2023), Abualigah et al. (2023), Qtaish et al. (2024), Demirci et al. (2023), Singh et al. (2023), Haeri Boroujeni and Pashaei (2023), Barshandeh et al. (2022), Premkumar et al. (2024), Patel et al. (2023), Alotaibi (2022), Salih et al. (2023), Moghadam and Ahmadi (2023)
Indexing	Pohl et al. (2016), Zhu and Ma (2018)
Gaussian distance	Ghorbanzadeh et al. (2016), Zhang et al. (2016a)
Gene distance	Elyasigomari et al. (2015)
Hamming distance	Ghorbanzadeh et al. (2016)
FCM	Nayak et al. (2018), Kushwaha and Pant (2018), Baykasoğlu et al. (2018), Kuo et al. (2018a), Singh and Srivastava (2022), Hashemi et al. (2023)
Mean square error	Gupta et al. (2018), Bouyer and Hatamlou (2018), Kuo et al. (2021), Asadi-Zonouz et al. (2022), Hu et al. (2023)
NKCV2 function	Tinós et al. (2018)
Manhattan distance	Salem et al. (2018)
Cluster integrity	Kaur et al. (2020)
DB index	Tekieh and Beheshti (2024)

widely adopted and popular objective function for clustering problems. Table 7, illustrates the objective functions studied during this survey.

4.2 Performance metrics

The performance metrics are used to evaluate the performance of the clustering algorithm. The performance metrics should be independent and reliable measures that can assess and compare the experimental results of the clustering algorithm. Based on comparison, the validity of a clustering algorithm is described. In general, to evaluate the performance of the clustering, two evaluations are used i.e. external evaluation and internal evaluation. The external evaluation contains the information of the dataset. The internal evaluation can be described as the evaluation of the dataset itself. Performance metrics like accuracy, f-measure, normalized mutual information, and rand index are commonly used in external evaluation. Performance metrics like the Davies-Bouldin index, Silhouette index, Dunn index, and Entropy are used for internal evaluation. This paper also focuses on different performance metrics reported for clustering algorithms to assess the performance. It is seen that 42 performance metrics are reported in the literature. Table 8 illustrates the

Table 8 List of different performance measures reported in the literature

References	Performance measure	Formula representation
Kumar et al. (2016b), Das et al. (2018a), Abualigah et al. (2018a), Mikaeil et al. (2018), Sharma and Chhabra (2019), Abasi et al. (2020), Shial et al. (2023), Alotaibi (2022), Tekieh and Beheshti (2024)	Precision	$\text{Precision(i,j)} = \frac{N_{ij}}{N_j}$
Kumar et al. (2016b), Das et al. (2018a), Abualigah et al. (2018a), Abasi et al. (2020), Alotaibi (2022), Tekieh and Beheshti (2024)	Recall	$\text{Recall(i,j)} = \frac{N_{ij}}{N_i}$
Kumar et al. (2016b), Sharma and Chhabra (2019)	G-measure	$\text{G(i,j)} = \sqrt{Precision(i,j) * Recall(i,j)}$
Allab et al. (2017), Jing et al. (2015), Kushwaha et al. (2018), Yu et al. (2016), Zhang et al. (2016b), Nazari et al. (2019), Lakshmi et al. (2018), Liu et al. (2019), Salem et al. (2018), Lee and Perkins (2021), Moghadam and Ahmadi (2023)	Normalized mutual information (NMI)	$\text{NMI} = \frac{MI(Truelabel,C)}{((T(Truelabel)+T(C))/2}$
Kushwaha et al. (2018), Leski (2016), Meng et al. (2016), Sheng et al. (2014), Zhang et al. (2016b), Nayak et al. (2018), Kushwaha and Pant (2018), Jadhav and Gomathi (2018), Lakshmi et al. (2018), Liu et al. (2019), Kumar and Kaur (2022)	Rand index	$\text{Rand Index} = (\text{TP} + \text{TN})/\text{N}$
Kushwaha et al. (2018), Meng et al. (2016), Yuwono et al. (2014), Alswaitti et al. (2018), Kushwaha and Pant (2018), Abualigah et al. (2018a), Aljarah et al. (2020), Lakshmi et al. (2018), Abasi et al. (2020), Turkoglu et al. (2022)	Purity	$\frac{1}{ni} * \max_j * n_{ij}$
Allab et al. (2017), Yao et al. (2018), Cruz et al. (2016), Kushwaha et al. (2018), Özbakır and Turna (2017), Yu et al. (2016), Zhang et al. (2016c), Das et al. (2018a), Deb et al. (2018), Alswaitti et al. (2018), Kushwaha and Pant (2018), Yang and Jiang (2018), Abualigah et al. (2018a), Sharma and Chhabra (2019), Rathore et al. (2018), Salem et al. (2018), Xie et al. (2019), Kaur and Kumar (2022), Kumar and Kaur (2022), Qtaish et al. (2024), Shial et al. (2023), Patel et al. (2023)	Accuracy	$\text{Acc} = \sum_{i=1}^{n} \delta(Truelabel, map(c))/n$
Santos and Zárate (2015),	NCC	$\text{NCC}(R) = \sum_{t=1}^{T} (S_{intra}(R_t) + D_{inter}(R_t))$

Table 8 (continued)

References	Performance measure	Formula representation
Santos and Zárate (2015), Meng et al. (2016), Yuwono et al. (2014), Zhang et al. (2016a), Abualigah et al. (2018a), Aljarah et al. (2020), Salem et al. (2018), Abasi et al. (2020), Hashemi et al. (2023)	Entropy	$H_{lt} = -\sum_{n}^{V_l} p(u)_{lt} ln(p(u)_{lt})$
Santos and Zárate (2015), Jing et al. (2015), Moghadam and Ahmadi (2023)	Compactness	$CpS = \sum_{i=1}^{c} dm(i)(\frac{m_i}{m})$
Santos and Zárate (2015), Leski (2016), Pohl et al. (2016), Puschmann et al. (2017), Lakshmi et al. (2018), Bouyer and Hatamlou (2018), Kuo et al. (2021)	Silhouette index	$SHIt(t) = \frac{\sum_{j=1}^{m_t} SHI(j)}{m_t}$
Chang et al. (2016), Han et al. (2017), Kumar and Sahoo (2015b), Özbakır and Turna (2017), Das et al. (2018a), Boushaki et al. (2018), Abualigah et al. (2018a), Abualigah et al. (2018b), Narayana and Vasumathi (2018), Abasi et al. (2020), Ahmadi et al. (2021), Haeri Boroujeni and Pashaei (2023), Alotaibi (2022), Salih et al. (2023)	Error rate	$ER = \left[\left(\sum_{i=1}^{n-1} \sum_{l=i+1}^{n} \lvert A_{il} - B_{il} \rvert \right) / \left(\frac{n(n-1)}{2} \right) \right] * 100$
Bijari et al. (2018), Qiao et al. (2019), Deb et al. (2018), Gupta and Saini (2019), Xie et al. (2019)	Fitness	$Fitness = \sum_{n=1}^{N} \sum_{k=1}^{K} I_{nk}(d_n) * (\lvert\lvert d_n - BC_k \rvert\rvert^2)$
Gutierrez-Rodríguez et al. (2015), Kumar and Sahoo (2014), Kumar and Sahoo (2016), Kumar and Singh (2018), Noorbehbahani et al. (2015), Tan et al. (2016), Yu et al. (2016), Das et al. (2018a), Alswaitti et al. (2018), Kushwaha and Pant (2018), Abualigah et al. (2018a), Jadhav and Gomathi (2018), Lakshmi et al. (2018), Bouyer and Hatamlou (2018), Gupta and Saini (2019), Salem et al. (2018), Xie et al. (2019), Abasi et al. (2020), Kaur and Kumar (2022), Qtaish et al. (2024), Shial et al. (2023), Singh et al. (2023), Patel et al. (2023), Alotaibi (2022)	F-measure	$F = 2 \frac{precision * recall}{precision + recall}$
Chang et al. (2016), Gebru et al. (2016), Pakrashi and Chaudhuri (2016), Pohl et al. (2016), Bouyer and Hatamlou (2018), Hu et al. (2023), Moghadam and Ahmadi (2023)	Davies-Bouldin index	$DB = \frac{1}{K} \sum_{k=1}^{k} R_k$

Table 8 (continued)

References	Performance measure	Formula representation
Alam et al. (2015), Ferrari and Castro (2015), Kumar and Sahoo (2014), Xiang et al. (2015), Zhang et al. (2016c), Gupta and Saini (2019), Mageshkumar et al. (2019), Singh (2020), Ahmadi et al. (2021), Pacifico and Ludermir (2021), Kumar and Kaur (2022), Qtaish et al. (2024), Singh et al. (2023), Haeri Boroujeni and Pashaei (2023), Salih et al. (2023)	Intra and inter cluster distance	$J1 = \sum_{j=1}^{k} \sum_{\forall xi \in Zj} \|xi - zj\|^2$ $J2 = \sum_{i=1}^{k} \sum_{j=i+1}^{k} \|zi - zj\|^2$
Chang et al. (2016), Yuwono et al. (2014), Rathore et al. (2018)	Dunn index	$DI = (\min 1 \le k \le l \le K\ \delta(Gk, Gl) / \max 1 \le m \le K\ m)$
Yuwono et al. (2014)	CH index	$CH = \frac{T_r(S_B)/(n_c - 1)}{T_r(S_w)/(n_j - n_c)}$
Noorbehbahani et al. (2015), Xu et al. (2019), Nayak et al. (2018), Jadhav and Gomathi (2018)	Jaccard coefficient	$J = \frac{\|SS\|}{\|SS\| + \|SD\| + \|DS\|}$
Noorbehbahani et al. (2015), Xu et al. (2019), Nayak et al. (2018)	Fowlkes and Mallows measure	$FM = \sqrt{\frac{\|SS\|}{\|SS\| + \|SD\|} * \frac{\|SS\|}{\|SS\| + \|DS\|}}$
Nguyen et al. (2015)	True positive rate	$TPR = \frac{TP}{TP + FN}$
Nguyen et al. (2015)	False positive rate	$FPR = \frac{FP}{TN + FP}$
Yao et al. (2018), Nazari et al. (2019), Qiao et al. (2019), Deb et al. (2018), Kuo et al. (2018a), Kuo et al. (2018b), Narayana and Vasumathi (2018), Narayana and Vasumathi (2018), Rathore et al. (2018), Kaur et al. (2020), Abasi et al. (2020), Rahnema and Gharehchopogh (2020), Lee and Perkins (2021), Pacifico and Ludermir (2021), Asadi-Zonouz et al. (2022)	Computation time/convergence time	Minimum number of iterations required for fast convergence
Bijari et al. (2018), Nayak et al. (2018), Bouyer and Hatamlou (2018), Kaur et al. (2020), Singh (2020), Kaur and Kumar (2022), Asadi-Zonouz et al. (2022), Demirci et al. (2023)	Friedman's statistical test	$R_j = 1/(N) \sum_i y_i^j$
Nayak et al. (2018)	Hubert's statistics (C)	$T = (\frac{1}{M}) \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} X(i,j) * Y(i,j)$
Baykasoğlu et al. (2018)	Non-parametric sign test	

Table 8 (continued)

References	Performance measure	Formula representation
Zhou et al. (2019), Lakshmi et al. (2018), Kuwil et al. (2019), Rahnema and Gharehchopogh (2020), Barshandeh et al. (2022)	ANOVA test	A one-way ANOVA is used to find out whether the means of groups are significantly different from one another or whether each group is relatively the same
Tsai et al. (2019), Boushaki et al. (2018), Aljarah et al. (2020), Kaur et al. (2020), Cho and Nyunt (2020), Lee and Perkins (2021), Pacifico and Ludermir (2021), Turkoglu et al. (2022), Barshandeh et al. (2022)	Sum of squared errors (SSE)	$\mathrm{SSE} = \sum_{j=1}^{k} \sum_{i-1}^{\lvert c_j \rvert} \sigma(C_j, r_i)2$
Boushaki et al. (2018), Pimentel and Carvalho (2019)	Student's *T*-test	$t = \frac{mean-\mu}{s/\sqrt{n}}$
Senthilnath et al. (2019), Alswaitti et al. (2018), Das et al. (2018b), Shial et al. (2023)	Percentage of error/misclassification rate percentage (MCR)	PE or MCR = s/n*100 where s is the number of misclassified objects and n is the size of the test data Set
Gupta et al. (2018), Jadhav and Gomathi (2018), Hu et al. (2023), Premkumar et al. (2024)	Mean square error (MSE)	$\mathrm{MSE} = \sum_{i=1}^{n} (Oi - Ai)2$
Gupta et al. (2018), Premkumar et al. (2024)	Mean absolute percentage error (MAPE)	$\mathrm{MAPE} = 1/N \sum \lvert\lvert (Oi - Ai)/\rvert\rvert Ai * 100\%$
Tinós et al. (2018), Abualigah et al. (2023), Demirci et al. (2023), Barshandeh et al. (2022)	Wilcoxon signed rank test	$\mathrm{W} = \sum_{1}^{N} [sign(x2, i - x1, i).Ri]$
Su and Denoeux (2018)	Creedal rand index	$\rho_s(M, M\prime) = 1 - \sum_{i<j} \frac{\delta(m_{ij}, m\prime_{ij})}{n(n-1)/2}$
Tinós et al. (2018), Narayana and Vasumathi (2018), Liu et al. (2019), Lee and Perkins (2021), Kuo et al. (2021)	Adjusted rand index (ARI)	$ARI = \frac{Index - ExpectedIndex}{MaxIndex - ExpectedIndex}$
Liu et al. (2019)	Adjusted mutual information (AMI)	$AMI(U, V) = \frac{MI(U,V) - E\{MI(U,V)\}}{max\{H(U), H(V)\} - E\{MI(U,V)\}}$ where H(U) and H(V) is entropy associated with partition U and V respectively, MI(U,V) is mutual information between two partitions e{MI(U,V) is the expected mutual Information
Gupta and Saini (2019)	Paired *T*-test	$T = \frac{mean}{Stdev}$
Rathore et al. (2018)	Chi-square distance	$\chi^2(A, B) = 1/(2) \sum_{i=1}^{f} \frac{(A_i - B_i)2}{A_i + B_i}$

Table 8 (continued)

References	Performance measure	Formula representation
Xie et al. (2019), Asadi-Zonouz et al. (2022), Qtaish et al. (2024)	Average sensitivity, average specificity	
Xu et al. (2019)	Czekanowski–Dice index	$CD = \frac{2.SS}{2.SS+DS+Sd}$
Xu et al. (2019)	Kulczynski index	$K = \frac{1}{2}.(\frac{SS}{SS+SD} + \frac{SS}{SS+DS})$
Hu et al. (2023)	Xie Beni index (XB)	$XB = \frac{MSE}{d_{min}}$
Hu et al. (2023), Moghadam and Ahmadi (2023)	Separation index (S)	$S = \frac{1}{\sum_{j=1;i\neq j}^{k} \lvert D_i\rvert\lvert D_j\rvert} \sum_{j=1;i\neq j}^{k} \lvert D_i\rvert\left\lvert D_j\right\rvert \,\lvert\lvert D_i\rvert - \left\lvert D_j\right\rvert\rvert$
Turkoglu et al. (2022)	V-measure	$V - measure = 2.\frac{HS.CS}{HS+CS}$ where HS isHomogeneityy Score and CSisthee completeness score

performance metrics reported in the literature. It is observed that widely adopted performance metrics are NMI, rand index, accuracy, entropy, f-measure, and error rate. Figure 5 presents a dynamic 3D pie chart, offering a visual representation of key aspects related to clustering algorithm performance assessment. The chart portrays an intricate interplay of various metrics, each contributing to the evaluation of clustering algorithms. As the pie chart rotates, viewers can observe the distribution and significance of different performance metrics within the clustering domain. Additionally, the performance metrics prevalent in the literature, shed light on the diversity and breadth of assessment criteria utilized by researchers. Among these metrics, certain indicators emerge as particularly prominent and widely embraced within the research community. Noteworthy examples include Normalized Mutual Information (NMI), Rand Index, Accuracy, Entropy, F-measure, and Error Rate. Their prevalence underscores their significance in gauging the effectiveness and efficiency of clustering algorithms across various applications and scenarios.

4.3 Dataset

The dataset also plays an important role in validating the performance of clustering algorithms. Clustering is an unsupervised method. Therefore, when a clustering algorithm is implemented no class information is given. The objects are assigned to different clusters based on the objective function. Some external evaluations are used to assess the performance of the clustering algorithm. These evaluations require the class information (cluster information). Moreover, some datasets are linearly separable, whereas some others are non-linearly separable. The performance of the clustering algorithm may be affected due to the above-mentioned properties of data. Another point, the simulation results of the clustering algorithm also depend on attribute types, dimensions of the dataset, size of data, etc. This

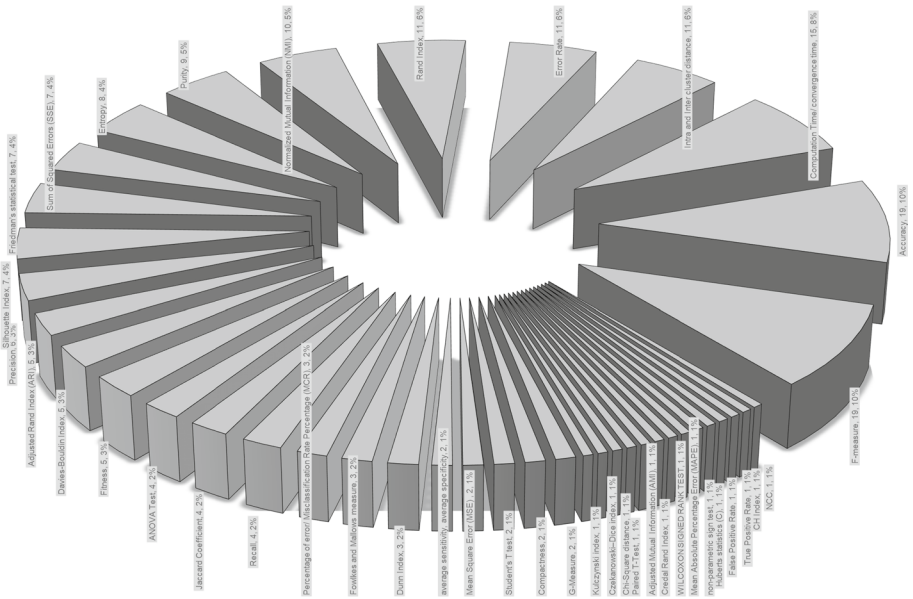

Fig. 5 3-D pie chart for performance measures

study also highlights the various datasets that are used to evaluate the performance of clustering algorithms. It is seen that forty datasets are reported in the literature to evaluate the performance of the clustering algorithms. Table 9 demonstrates the list of these datasets. It is also revealed that iris, wine, glass, CMC, vowel, cancer, breast cancer, and thyroid datasets are widely used datasets to evaluate the performance of clustering algorithms.

Figure 6 showcases a dynamic 3D pie chart, providing a comprehensive overview of the datasets commonly utilized in assessing clustering algorithm performance. The chart captures the diversity and breadth of datasets employed in clustering research, offering insights into the range of scenarios and applications where these algorithms are applied. Each segment of the pie chart represents a specific dataset, with the size of the segment corresponding to the relative frequency or significance of its usage in clustering algorithm evaluation. Notably, the chart underscores the prevalence of certain datasets such as iris, wine, glass, CMC, vowel, cancer, breast cancer, and thyroid, which emerge as widely adopted benchmarks for assessing clustering algorithms. This visualization serves as a valuable reference for researchers and practitioners, providing a visual depiction of the dataset landscape and highlighting key datasets that have become standard benchmarks within the clustering community. By presenting this information in a visually accessible format, Fig. 5 facilitates a deeper understanding of the datasets employed in clustering research and their role in algorithm evaluation.

5 Issues and challenges

This section summarizes the various issues that can be addressed through meta-heuristic algorithms. It is observed that large numbers of meta-heuristic algorithms are taken into consideration to solve the clustering problems effectively.

5.1 Issues in partitional clustering

In Partitional clustering, various meta-heuristic algorithms are applied to solve clustering problems effectively. The main reasons for adopting the meta-heuristic algorithm for Partitional clustering are listed.

(i) To determine near-optimal solutions for Partitional clustering problems.
(ii) To evaluate optimal centroid for effective clustering.
(iii) To determine similar patterns in categorical data.
(iv) To handle heterogeneous data.
(v) To determine subspace clusters in the dataset.
(vi) To handle multimodal and heterogeneous data for effective clustering.
(vii) To perform clustering of high dimensional data.
(viii) To handle the educational data mining.

5.2 Issues in dynamic and automatic clustering

From the extensive literature survey, it is inferred that some meta-heuristic algorithms are also adopted in the field of dynamic and automatic clustering. The main reasons for applying the meta-heuristic algorithm are listed.

Table 9 List of datasets adopted to evaluate simulation results

Data sets	Referred papers
Iris	Alam et al. (2015), Amiri and Mahmoudi (2016), Bahrololoum et al. (2015), Banharnsakun (2017), Bijari et al. (2018), Chang et al. (2016), Cruz et al. (2016), Gutierrez-Rodríguez et al. (2015), Han et al. (2017), Kumar et al. (2015), Kumar et al. (2016a), Kumar et al. (2016b), Kumar and Sahoo (2014), Kumar and Sahoo (2015a), Kumar and Sahoo (2015b), Kumar and Sahoo (2016), Kumar and Singh (2018), Kushwaha et al. (2018), Leski (2016), Li et al. (2016), Montgomery et al. (2016), Nguyen et al. (2015), Özbakır and Turna (2017), Ozturk et al. (2015), Pakrashi and Chaudhuri (2016), Santi et al. (2016), Serapião et al. (2016), Sheng et al. (2016), Siddiqi and Sait (2017), Tang et al. (2016), Vo et al. (2016), Xiang et al. (2015), Yang et al. (2015), Yuwono et al. (2014), Zhang et al. (2016c), Das et al. (2018a), Deb et al. (2018), Alswaitti et al. (2018), Nayak et al. (2018), Kushwaha and Pant (2018), Baykasoğlu et al. (2018), Zhou et al. (2019), Queiroga et al. (2018), Tsai et al. (2019), Boushaki et al. (2018), Abualigah et al. (2018b), Das et al. (2018b), Kuo et al. (2018a), Aljarah et al. (2020), Kuo et al. (2018b), Tinós et al. (2018), Jadhav and Gomathi (2018), Sharma and Chhabra (2019), Lakshmi et al. (2018), Bouyer and Hatamlou (2018), Gupta and Saini (2019), Mageshkumar et al. (2019), Su and Denoeux (2018), Zhu and Ma (2018), Xie et al. (2019), Singh (2020), Abasi et al. (2020), Cho and Nyunt (2020), Ahmadi et al. (2021), Kaur and Kumar (2022), Lee and Perkins (2021), Pacifico and Ludermir (2021), Kumar and Kaur (2022), Kuo et al. (2021), Asadi-Zonouz et al. (2022), Hu et al. (2023), Turkoglu et al. (2022), Abualigah et al. (2023), Qtaish et al. (2024), Demirci et al. (2023), Shial et al. (2023), Singh et al. (2023), Hashemi et al. (2023), Haeri Boroujeni and Pashaei (2023), Barshandeh et al. (2022), Premkumar et al. (2024), Patel et al. (2023), Alotaibi (2022), Salih et al. (2023), Moghadam and Ahmadi (2023)

Table 9 (continued)

Data sets	Referred papers
Wine	Alam et al. (2015), Amiri and Mahmoudi (2016), Bahrololoum et al. (2015), Banharnsakun (2017), Bijari et al. (2018), Cruz et al. (2016), Ferrari and Castro (2015), Gebru et al. (2016), Gutierrez-Rodríguez et al. (2015), Han et al. (2017), Kumar et al. (2016a), Kumar et al. (2016b), Kumar and Sahoo (2015b), Kumar and Sahoo (2016), Kumar and Singh (2018), Nguyen et al. (2015), Pakrashi and Chaudhuri (2016), Santi et al. (2016), Serapião et al. (2016), Sheng et al. (2014), Xiang et al. (2015), Yuwono et al. (2014), Zhang et al. (2016c), Das et al. (2018a), Alswaitti et al. (2018), Nayak et al. (2018), Kushwaha and Pant (2018), Baykasoğlu et al. (2018), Zhou et al. (2019), Queiroga et al. (2018), Tsai et al. (2019), Boushaki et al. (2018), Abualigah et al. (2018b), Das et al. (2018b), Kuo et al. (2018a), Aljarah et al. (2020), Kuo et al. (2018b), Jadhav and Gomathi (2018), Sharma and Chhabra (2019), Lakshmi et al. (2018), Bouyer and Hatamlou (2018), Gupta and Saini (2019), Mageshkumar et al. (2019), Su and Denoeux (2018), Zhu and Ma (2018), Xie et al. (2019), Singh (2020), Abasi et al. (2020), Cho and Nyunt (2020), Kaur and Kumar (2022), Lee and Perkins (2021), Pacifico and Ludermir (2021), Kumar and Kaur (2022), Kuo et al. (2021), Asadi-Zonouz et al. (2022), Hu et al. (2023), Turkoglu et al. (2022), Qtaish et al. (2024), Demirci et al. (2023), Singh et al. (2023), Hashemi et al. (2023), Haeri Boroujeni and Pashaei (2023), Barshandeh et al. (2022), Premkumar et al. (2024), Patel et al. (2023), Alotaibi (2022), Salih et al. (2023), Moghadam and Ahmadi (2023)

Table 9 (continued)

Data sets	Referred papers
Glass	Alam et al. (2015), Amiri and Mahmoudi (2016), Bahrololoum et al. (2015), Bijari et al. (2018), Cruz et al. (2016), Han et al. (2017), Senthilnath et al. (2019), Senthilnath et al. (2019), Kumar et al. (2016a), Kumar et al. (2016b), Kumar and Sahoo (2014), Kumar and Sahoo (2015a), Kumar and Sahoo (2015b), Kumar and Sahoo (2016), Kumar and Singh (2018), Kushwaha et al. (2018), Özbakır and Turna (2017), Pakrashi and Chaudhuri (2016), Serapião et al. (2016), Sheng et al. (2014), Siddiqi and Sait (2017), Tang et al. (2016), Xiang et al. (2015), Yuwono et al. (2014), Alswaitti et al. (2018), Nayak et al. (2018), Baykasoğlu et al. (2018), Queiroga et al. (2018), Abualigah et al. (2018b), Das et al. (2018b), Kuo et al. (2018a), Aljarah et al. (2020), Kuo et al. (2018b), Tinós et al. (2018), Sharma and Chhabra (2019), Lakshmi et al. (2018), Gupta and Saini (2019), Mageshkumar et al. (2019), Singh (2020), Abasi et al. (2020), Cho and Nyunt (2020), Ahmadi et al. (2021), Kaur and Kumar (2022), Rahnema and Gharehchopogh (2020), Lee and Perkins (2021), Pacifico and Ludermir (2021), Kumar and Kaur (2022), Kuo et al. (2021), Asadi-Zonouz et al. (2022), Abualigah et al. (2023), Singh et al. (2023), Hashemi et al. (2023), Barshandeh et al. (2022), Patel et al. (2023), Patel et al. (2023), Alotaibi (2022), Salih et al. (2023), Moghadam and Ahmadi (2023)
Haberman	Kumar et al. (2016b), Zhang et al. (2016c), Deb et al. (2018), Alswaitti et al. (2018), Baykasoğlu et al. (2018), Zhou et al. (2019), Abualigah et al. (2018b), Aljarah et al. (2020), Sharma and Chhabra (2019), Lakshmi et al. (2018), Shial et al. (2023), Patel et al. (2023)
CMC	Alam et al. (2015), Amiri and Mahmoudi (2016), Yao et al. (2018), Yao et al. (2018), Banharnsakun (2017), Bijari et al. (2018), Kumar et al. (2016b), Kumar and Sahoo (2014), Kumar and Sahoo (2015a), Kumar and Sahoo (2015b), Kumar and Sahoo (2016), Kumar and Singh (2018), Kushwaha et al. (2018), Noorbehbahani et al. (2015), Pakrashi and Chaudhuri (2016), Tang et al. (2016), Xiang et al. (2015), Das et al. (2018a), Nayak et al. (2018), Kushwaha and Pant (2018), Zhou et al. (2019), Boushaki et al. (2018), Das et al. (2018b), Sharma and Chhabra (2019), Lakshmi et al. (2018), Bouyer and Hatamlou (2018), Gupta and Saini (2019), Abasi et al. (2020), Kaur and Kumar (2022), Kumar and Kaur (2022), Asadi-Zonouz et al. (2022), Abualigah et al. (2023), Qtaish et al. (2024), Singh et al. (2023), Haeri Boroujeni and Pashaei (2023), Barshandeh et al. (2022), Alotaibi (2022), Salih et al. (2023)

Table 9 (continued)

Data sets	Referred papers
Vowel	Alam et al. (2015), Amiri and Mahmoudi (2016), Banharnsakun (2017), Bijari et al. (2018), Kumar et al. (2016a), Kumar and Sahoo (2014), Kumar and Sahoo (2015b), Kushwaha et al. (2018), Tang et al. (2016), Xiang et al. (2015), Kushwaha and Pant (2018), Queiroga et al. (2018), Das et al. (2018b), Kuo et al. (2018b), Bouyer and Hatamlou (2018), Kaur and Kumar (2022), Kumar and Kaur (2022), Asadi-Zonouz et al. (2022), Abualigah et al. (2023), Salih et al. (2023)
Crude oil	Kushwaha et al. (2018), Serapião et al. (2016), Xiang et al. (2015), Kushwaha and Pant (2018), Kumar and Kaur (2022)
Balance	Bahrololoum et al. (2015), Han et al. (2017), Kushwaha et al. (2018), Serapião et al. (2016), Siddiqi and Sait (2017), Alswaitti et al. (2018), Nayak et al. (2018), Kushwaha and Pant (2018), Baykasoğlu et al. (2018), Zhou et al. (2019), Queiroga et al. (2018), Xie et al. (2019), Ahmadi et al. (2021), Kaur and Kumar (2022), Kumar and Kaur (2022), Turkoglu et al. (2022), Qtaish et al. (2024)
Thyroid	Bahrololoum et al. (2015), Han et al. (2017), Kumar and Sahoo (2014), Kumar and Sahoo (2015b), Kushwaha et al. (2018), Özbakır and Turna (2017), Xiang et al. (2015), Alswaitti et al. (2018), Kushwaha and Pant (2018), Sharma and Chhabra (2019), Bouyer and Hatamlou (2018), Xie et al. (2019), Cho and Nyunt (2020), Kaur and Kumar (2022), Kumar and Kaur (2022), Singh et al. (2023)
Cancer	Bahrololoum et al. (2015), Bijari et al. (2018), Han et al. (2017), Kumar and Sahoo (2015a), Kumar and Sahoo (2016), Kumar and Singh (2018), Özbakır and Turna (2017), Pakrashi and Chaudhuri (2016), Tang et al. (2016), Xiang et al. (2015), Das et al. (2018a), Alswaitti et al. (2018), Nayak et al. (2018), Kushwaha and Pant (2018), Zhou et al. (2019), Queiroga et al. (2018), Tsai et al. (2019), Boushaki et al. (2018), Das et al. (2018b), Kuo et al. (2018a), Aljarah et al. (2020), Sharma and Chhabra (2019), Lakshmi et al. (2018), Bouyer and Hatamlou (2018), Gupta and Saini (2019), Ahmadi et al. (2021), Asadi-Zonouz et al. (2022), Abualigah et al. (2023), Singh et al. (2023), Haeri Boroujeni and Pashaei (2023), Alotaibi (2022), Salih et al. (2023)
Zoo	Noorbehbahani et al. (2015), Özbakır and Turna (2017), Yang et al. (2015), Mageshkumar et al. (2019), Shial et al. (2023)
Votes	Özbakır and Turna (2017), Salem et al. (2018)
Credit	Bahrololoum et al. (2015), Han et al. (2017), Noorbehbahani et al. (2015), Özbakır and Turna (2017), Ahmadi et al. (2021)

Table 9 (continued)

Data sets	Referred papers
Heart	Cruz et al. (2016), Han et al. (2017), Özbakır and Turna (2017), Serapião et al. (2016), Siddiqi and Sait (2017), Nayak et al. (2018), Kushwaha and Pant (2018), Baykasoğlu et al. (2018), Zhou et al. (2019), Aljarah et al. (2020), Ahmadi et al. (2021), Kaur and Kumar (2022), Turkoglu et al. (2022), Patel et al. (2023)
Madelon	Kaur and Datta (2015)
Banknote	Siddiqi and Sait (2017), Yu et al. (2016), Jadhav and Gomathi (2018), Turkoglu et al. (2022), Barshandeh et al. (2022)
Page blocks	Yu et al. (2016), Pacifico and Ludermir (2021), Hu et al. (2023)
Waveform	Gutierrez-Rodríguez et al. (2015), Siddiqi and Sait (2017), Yu et al. (2016), Su and Denoeux (2018), Pacifico and Ludermir (2021)
Breast cancer	Alam et al. (2015), Cruz et al. (2016), Jing et al. (2015), Kumar et al. (2016a), Kumar and Sahoo (2014), Nguyen et al. (2015), Serapião et al. (2016), Sheng et al. (2014), Yuwono et al. (2014), Zhang et al. (2016c), Aljarah et al. (2020), Lakshmi et al. (2018), Salem et al. (2018), Xie et al. (2019), Cho and Nyunt (2020), Kaur and Kumar (2022), Rahnema and Gharehchopogh (2020), Barshandeh et al. (2022), Premkumar et al. (2024), Moghadam and Ahmadi (2023)
Wisconsin	Kumar and Sahoo (2014), Nguyen et al. (2015), Ozturk et al. (2015), Zhang et al. (2016c), Qiao et al. (2019), Nayak et al. (2018), Kushwaha and Pant (2018), Zhou et al. (2019), Queiroga et al. (2018), Tsai et al. (2019), Boushaki et al. (2018), Kuo et al. (2018a), Su and Denoeux (2018), Xie et al. (2019), Kaur and Kumar (2022), Rahnema and Gharehchopogh (2020), Lee and Perkins (2021), Kumar and Kaur (2022), Turkoglu et al. (2022), Shial et al. (2023)
Liver disease	Kumar and Sahoo (2014), Kushwaha and Pant (2018), Aljarah et al. (2020), Kaur and Kumar (2022), Rahnema and Gharehchopogh (2020), Kumar and Kaur (2022), Turkoglu et al. (2022), Shial et al. (2023), Barshandeh et al. (2022), Patel et al. (2023)
Letter recognition	Gebru et al. (2016), Pacifico and Ludermir (2021)
Ionosphere	Gutierrez-Rodríguez et al. (2015), Serapião et al. (2016), Siddiqi and Sait (2017), Qiao et al. (2019), Kushwaha and Pant (2018), Queiroga et al. (2018), Tinós et al. (2018), Rahnema and Gharehchopogh (2020), Pacifico and Ludermir (2021), Kumar and Kaur (2022), Turkoglu et al. (2022), Shial et al. (2023)
Mammographic	Gutierrez-Rodríguez et al. (2015), Shial et al. (2023)

Table 9 (continued)

Data sets	Referred papers
Seeds	Amiri and Mahmoudi (2016), Gutierrez-Rodríguez et al. (2015), Alswaitti et al. (2018), Zhou et al. (2019), Boushaki et al. (2018), Abualigah et al. (2018b), Aljarah et al. (2020), Su and Denoeux (2018), Abasi et al. (2020), Pacifico and Ludermir (2021), Abualigah et al. (2023), Qtaish et al. (2024), Demirci et al. (2023), Shial et al. (2023), Patel et al. (2023)
Segment	Chang et al. (2016), Gutierrez-Rodríguez et al. (2015), Sheng et al. (2016), Siddiqi and Sait (2017)
Sonar	Cruz et al. (2016), Gutierrez-Rodríguez et al. (2015), Serapião et al. (2016), Siddiqi and Sait (2017), Qiao et al. (2019), Salem et al. (2018), Xie et al. (2019), Turkoglu et al. (2022), Premkumar et al. (2024)
Diabetes	Bahrololoum et al. (2015), Cruz et al. (2016), Han et al. (2017), Serapião et al. (2016), Siddiqi and Sait (2017), Yuwono et al. (2014), Aljarah et al. (2020), Ahmadi et al. (2021), Kaur and Kumar (2022), Pacifico and Ludermir (2021)
Vertebral column	Zhang et al. (2016c), Kuo et al. (2021)
Leukaemia	Allab et al. (2017), Elyasigomari et al. (2015), Jing et al. (2015)
Yeast	Allab et al. (2017), Chang et al. (2016), Senthilnath et al. (2019), Queiroga et al. (2018), Kuo et al. (2018b), Singh (2020), Lee and Perkins (2021), Pacifico and Ludermir (2021), Hu et al. (2023)
E. coli	Bahrololoum et al. (2015), Han et al. (2017), Siddiqi and Sait (2017), Tinós et al. (2018), Bouyer and Hatamlou (2018), Xie et al. (2019), Cho and Nyunt (2020), Ahmadi et al. (2021), Lee and Perkins (2021), Pacifico and Ludermir (2021), Kuo et al. (2021), Asadi-Zonouz et al. (2022), Hu et al. (2023), Qtaish et al. (2024), Shial et al. (2023)
Image segment	Senthilnath et al. (2019), Sheng et al. (2016), Siddiqi and Sait (2017), Pacifico and Ludermir (2021), Moghadam and Ahmadi (2023)
Subcellcycle	Sheng et al. (2014)
Libra movement	Chang et al. (2016)
Gesture segmentation	Chang et al. (2016), Deb et al. (2018)
Dermatology	Bahrololoum et al. (2015), Han et al. (2017), Ozturk et al. (2015), Qiao et al. (2019), Alswaitti et al. (2018), Kushwaha and Pant (2018), Ahmadi et al. (2021), Kaur and Kumar (2022), Rahnema and Gharehchopogh (2020), Hu et al. (2023), Qtaish et al. (2024)
Morro Bay	Ozturk et al. (2015)
Lena	Ozturk et al. (2015)
Mandril	Ozturk et al. (2015)
Semeion, MF, IS, FCT, MNIST, ODR,LS, ISOLE, USPS,	Nazari et al. (2019)
HV	Das et al. (2018a), Das et al. (2018b)

Table 9 (continued)

Data sets	Referred papers
Forest type mapping, firm teacher Clave direction classification data set	Qiao et al. (2019)
Parkinsons data set,	Qiao et al. (2019), Su and Denoeux (2018)
Turkiye student	Qiao et al. (2019)
Evaluation general (TSEG)	Qiao et al. (2019)
Mice protein	Deb et al. (2018), Lee and Perkins (2021)
Transfusion (BTSCD)	Alswaitti et al. (2018), Pacifico and Ludermir (2021)
Landsat	Alswaitti et al. (2018), Pacifico and Ludermir (2021)
Lenses	Nayak et al. (2018), Baykasoğlu et al. (2018)
Hayesroth	Nayak et al. (2018), Baykasoğlu et al. (2018), Zhu and Ma (2018)
Robot Navigation	Nayak et al. (2018), Baykasoğlu et al. (2018)
Spect heart	Nayak et al. (2018), Baykasoğlu et al. (2018)
Artificial dataset Two circle, Two moon	Nayak et al. (2018)
Jain	Kushwaha and Pant (2018), Singh (2020), Turkoglu et al. (2022)
Aggregation	Kushwaha and Pant (2018), Tinós et al. (2018), Liu et al. (2019), Zhu and Ma (2018), Singh (2020), Kuo et al. (2021), Turkoglu et al. (2022)
High-dimensional gene expression data set	Kushwaha and Pant (2018)
Magic	Baykasoğlu et al. (2018), Patel et al. (2023)
Syn control	Yang and Jiang (2018), Lee and Perkins (2021), Kumar and Kaur (2022)
HMM-generated dataset, Gun-Point, CBF, Face, OSU Leaf, Swedish leaf, 50Words, trace, Two pattern, wafe, face(four), Lightning-2, Lightning-7, ECG, Adiac, Yoga, CAVIAR database	Yang and Jiang (2018)
unbalance	Queiroga et al. (2018), Liu et al. (2019), Su and Denoeux (2018)
Abalone	Queiroga et al. (2018), Tsai et al. (2019), Pacifico and Ludermir (2021)
Artset1, artset2,artset3, artset4, a1	Queiroga et al. (2018)
D31	Queiroga et al. (2018), Kuo et al. (2018b), Liu et al. (2019), Singh (2020)
s1	Queiroga et al. (2018), Zhu and Ma (2018)
HTRU2, Spambase, User locations Finland	Tsai et al. (2019)
Blood	Boushaki et al. (2018), Turkoglu et al. (2022), Haeri Boroujeni and Pashaei (2023), Barshandeh et al. (2022)
Technical reports, web pages, TREC, MEDLINE, 20newsgroup	Abualigah et al. (2018a)
Flame	Kuo et al. (2018a), Liu et al. (2019), Singh (2020), Turkoglu et al. (2022)
TAE	Allab et al. (2017), Yao et al. (2018), Kuo et al. (2018a), Qtaish et al. (2024)

Table 9 (continued)

Data sets	Referred papers
Forecasting, the year-wise enrollments of the University of Alabama, Lahi (crop) production, the monthly amount of outpatients visiting a hospital(inventory demand), the population of India from years 1930–2000	Gupta et al. (2018)
Australian, Planning to relax, Tic-tac-toe	Aljarah et al. (2020)
R15	Kuo et al. (2018b), Tinós et al. (2018), Liu et al. (2019), Su and Denoeux (2018), Singh (2020), Kuo et al. (2021)
Compound	Tinós et al. (2018), Singh (2020), Kuo et al. (2021)
Path-based	Tinós et al. (2018), Singh (2020), Turkoglu et al. (2022)
US Census 1990	Rathore et al. (2018)
12 Quarries with famous dimension stones	Mikaeil et al. (2018)
Adult	Xu et al. (2019), Narayana and Vasumathi (2018)
Chess	Narayana and Vasumathi (2018)
Mushroom	Narayana and Vasumathi (2018), Salem et al. (2018)
Connect-4	Narayana and Vasumathi (2018)
Statlog	Yao et al. (2018), Hu et al. (2023), Abualigah et al. (2023)
Golub1999v1, Golub1999v2, Amstrong2002, Chowdary2006, Nutt2003, pomeroy2002, Chen2002, Khan2001	Yan et al. (2019)
t4	Liu et al. (2019)
S2, S4, A3,	Su and Denoeux (2018)
Pima	Su and Denoeux (2018), Cho and Nyunt (2020)
MiniBooNE, MNIST, ACT	Rathore et al. (2018)
4k2	Zhu and Ma (2018)
ALL_IDB2, Ozone	Xie et al. (2019)
Faculty data, patient data	Kuwil et al. (2019)
Image segmentation, vehicle, crop type	Senthilnath et al. (2019), Hu et al. (2023)
Wechat sport user	Yao et al. (2018)
Student, Germany, Thoracic, Nursery, Car	Xu et al. (2019)
Spiral	Singh (2020)
Horse	Ahmadi et al. (2021)
Hepatitis	Rahnema and Gharehchopogh (2020), Demirci et al. (2023), Shial et al. (2023), Barshandeh et al. (2022)
LSVT voice rehabilitation	Lee and Perkins (2021)
Banknote authentication, Optical recognition, Pen-based recognition of handwritten digits, t), Ten Synthetic datasets: Disp01, Disp02, Disp03, Disp04, Disp05, Prox01, Prox02, Prox03, Prox04, Prox05	Pacifico and Ludermir (2021)
Stamps, Breast Tissue	Kuo et al. (2021)
Wireless indoor localization	Hu et al. (2023)
East cell cycle	Hu et al. (2023)
Aniso	Turkoglu et al. (2022)
Appendicitis	Turkoglu et al. (2022)

Table 9 (continued)

Data sets	Referred papers
Diagnosis II	Turkoglu et al. (2022)
Iris2D	Turkoglu et al. (2022)
Moons	Turkoglu et al. (2022)
Mouse	Aljarah et al. (2020), Turkoglu et al. (2022)
Smiley	Turkoglu et al. (2022)
Var density	Turkoglu et al. (2022)
Vertebral 2, Vertebral 3	Turkoglu et al. (2022)
Hill-Valley, Hungarian	Qtaish et al. (2024)
WDBC	Shial et al. (2023), Premkumar et al. (2024)
Indian liver patient	Shial et al. (2023)
LR	Das et al. (2018a), Das et al. (2018b), Singh et al. (2023)
ISOLET	Singh et al. (2023)
KDD CUP'99	Rathore et al. (2018), Xie et al. (2019), Hashemi et al. (2023)
2D15	Hashemi et al. (2023)
Divorce predictors	Barshandeh et al. (2022)
WBDC	Shial et al. (2023)
HTRU2	Premkumar et al. (2024), Patel et al. (2023)
Emission	Premkumar et al. (2024)
Ukraine-Russia war	Premkumar et al. (2024)
Alcohol QCM sensor	Moghadam and Ahmadi (2023)

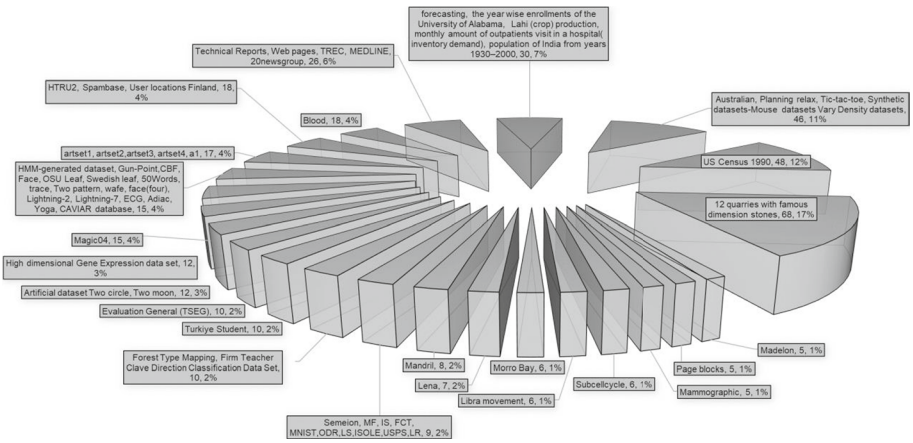

Fig. 6 3-D pie chart for datasets adopted to evaluate simulation results

(i) To enhance the convergence rate of algorithms.
(ii) To avoid stagnation and premature convergence.
(iii) To develop an optimization strategy for dynamic clustering.
(iv) To handle dynamic streams automatically.

5.3 Issues in fuzzy clustering

In the field of fuzzy clustering, some meta-heuristic algorithms are also reported. These algorithms aim to improve the quality of solutions, especially for fuzzy clustering problems. The issues handled by these algorithms are listed.

(i) To generate optimum cluster centres using the fuzzy membership function.
(ii) To handle high-dimensional dataset.
(iii) To determine relevant features in case of high dimensional data.
(iv) To develop accurate prediction models.
(v) To improve the quality of solutions.
(vi) To handle data streams in an effective manner.

5.4 Issues in improved meta heuristic algorithm for clustering

This subsection demonstrates various issues related to the performance of the meta-heuristic algorithm and the need to improve these algorithms for efficiently solving clustering problems. The various shortcomings associated with meta-heuristic algorithms and successfully addressed through improved versions of meta-heuristic algorithms. The main reasons to improve the meta-heuristic algorithms are listed.

(i) To overcome the slow convergence rate of meta-heuristic algorithms.
(ii) To avoid premature convergence problem.
(iii) To reduce noise effect and improve quality of solutions.
(iv) To handle clustering in a hierarchical manner.
(v) To reduce computational cost.
(vi) To effective trade-off between local search and global search.
(vii) To tackle overlapping and incremental clustering.
(viii) To handle constraints in an effective manner.

5.5 Issues in hybrid meta heuristic algorithm for clustering

The issues that can be addressed through hybrid meta-heuristic algorithms are listed.

(i) To overcome the shortcomings of traditional clustering algorithms like local optima and improve the quality of results.
(ii) To remove infeasible solutions generated during execution.
(iii) To handle local optima and convergence issues of meta-heuristic algorithm.
(iv) To improve search mechanisms of algorithms.
(v) To effectively handle exploration and exploitation processes.
(vi) To address the initialization issues of clustering algorithms.
(vii) To explore more promising solutions for clustering problems.
(viii) To explore solution search space in an effective and efficient manner.
(ix) To generate a neighbourhood solution.

6 Conclusion

In this survey, a large number of meta-heuristic algorithms are analysed concerning clustering applications. It is inferred that clustering problems can be classified in terms of Partitional, dynamic, and fuzzy clustering. A diversity of algorithms are reported in the literature to solve clustering problems effectively and efficiently. Some algorithms address issues related to performance, population diversity, local optima, search strategies, neighbourhood solutions, number of clusters, optimized cluster centres, etc. This paper presents a survey of high-repute publications in a particular period (2015–2021). These articles are categorized into Partitional, dynamic & automatic, and fuzzy clustering. Moreover, they are further classified into meta-heuristic, improved meta-heuristic, and hybrid meta-heuristic algorithms. Before the literature survey, several research questions are designed for an effective and efficient survey. The major contributions of this literature survey to the scientific community are.

RQ 1 What are the various meta-heuristic techniques available for clustering problems?

Answer: Large numbers of meta-heuristic algorithms employed to solve clustering problems are analysed. Several new algorithms are developed to solve these problems (CSS, MCSS, Bird flock algorithm, Electromagnetic force based algorithm, Magnetic optimization algorithm, Gravity algorithm, Big Bang Big Crunch algorithm). It is observed that these algorithms provide significant results in contrast to PSO, SA, TS, ACO, GA, and K-means etc. It is also observed that a smaller number of algorithms are based on traditional mathematical models. All recently developed algorithms are inspired by some natural phenomenon like the Big Bang Big Crunch, well-established laws like gravity law, and swarm behaviour (cuckoo optimization inspired through cuckoo's behaviour). Tables 2, 3, 4, 5, 6 summarizes various algorithms.

RQ 2 How to handle automatic data clustering?

Answer: Dynamic & Automatic clustering problems are an active area of research due to online, web, and social mining. In these problems, the number of clusters is undefined, and clusters are designed according to the nature of the data. It is observed that several single-objective clustering algorithms are proposed to address the dynamic clustering problem. Again, these algorithms are based on natural phenomena (swarm behaviour). A few multi-objective algorithms are developed to handle dynamic clustering problems. Hence, it can be concluded that a lot of attention soon will be formed in this direction.

RQ 3 How to handle high dimensional data (problems) with clustering?

Answer: At present, a large number of data is generated, and this volume is increasing exponentially. This data contains meaningful patterns, but it is not an easy task to explore and analyze these patterns. So, to handle large data problems and extract meaning, several meta-heuristic clustering algorithms are proposed. A few are integrated with Hadoop (a parallel architecture) to retrieve and process data much faster than traditional approaches. Some ensemble clustering methods can handle high-dimensional data. It is seen that lack of multi-objective clustering methods to handle the aforementioned issues.

RQ 4 What are the main reasons for hybridizing the clustering algorithms?

Answer: Many improved and hybridized versions of algorithms are proposed. An algorithm is either improved/hybridized due to shortcomings associated with it or to avoid shortcomings related to problems being solved. Through the literature survey, it is observed that several shortcomings are associated with algorithm and clustering problems. These are local optima, convergence rate, population diversity, boundary constraints, neighbourhood solution structure, the effective trade-off between local and global searches of the algorithm, solution search mechanism, solution search equations, and dependence on random functions. It is also observed that hybridization is an active area of research and hybridization of an algorithm can improve its performance. Hence, to overcome the aforementioned problems, an algorithm can either be improved or hybridized to obtain significant and optimized results. Till date, there is no generic algorithm for solving all types of clustering problems and data (categorical, nominal, numeric, text, and binary).

RQ 5 What objective functions, performance measures, and datasets are adopted to evaluate the performance of clustering algorithms?

Answer: Large numbers of performance measures are employed to evaluate the performance of clustering algorithms. Table 8 contains performance measures, which are reported in the literature. It is observed that NMI, rand index, accuracy, inner and inter-cluster distance, and F-measure are widely adopted performance measures. Table 7 summarizes objective functions to find closeness between data objects. Ten objective functions are reported in the literature, Euclidean Distance is a widely adopted objective function. To evaluate performance various datasets reported in the literature are summarized in Table 9. It is analysed that Iris, Wine, Glass, Haberman, CMC, Vowel, and Breast cancer are the most significant (benchmark) datasets for evaluation. Highlights of the survey are listed.

- 130 SCI and/or Scopus (Free) articles are included from 70 journals that are published (2015-2024).
- Euclidean distance is adopted as a significant distance to determine closeness between data objects.
- It is analysed that partitional clustering is a widely adopted problem.
- Improved and enhanced meta-heuristic algorithms are hybrid algorithms for effective and efficient clustering of data.
- It is analysed that hybrid meta-heuristic algorithms are the more significant approach to handling various clustering problems.
- Fuzzy and Automatic data clustering is a new and active area of research.
- Lack of work reported on multi-objective data clustering, which leads to a scope in this direction.

In this survey, we have undertaken a comprehensive analysis of various meta-heuristic algorithms in the context of clustering applications. Our investigation has shed light on the diverse landscape of clustering problems, which can be classified into Partitional, dynamic, and fuzzy clustering categories. Through an extensive review of the literature published between 2015 and 2024, we have identified a multitude of algorithms that address key challenges associated with clustering, including performance, population diversity, local optima, and search strategies. Our survey has revealed the emergence of

several novel meta-heuristic techniques for solving clustering problems, such as CSS, MCSS, Bird flock algorithm, Electromagnetic force-based algorithm, Magnetic optimization algorithm, Gravity algorithm, and Big Bang big crunch algorithm. These algorithms have demonstrated promising results compared to traditional methods like PSO, SA, TS, ACO, GA, and K-means, showcasing the effectiveness of leveraging natural phenomena and established laws as inspiration for algorithm design.

Additionally, we have explored the ongoing research efforts in dynamic and automatic clustering, which are driven by the growing demand for real-time data analysis in domains like online, web, and social mining. While single-objective clustering algorithms have made significant strides in addressing dynamic clustering challenges, there remains a need for the development of multi-objective algorithms to handle the complexity of evolving datasets more effectively. Furthermore, our survey has highlighted the importance of addressing the challenges posed by high-dimensional data in clustering. With the exponential growth of data volumes, there is a pressing need for meta-heuristic clustering algorithms capable of handling large-scale datasets efficiently. Integration with parallel architectures like Hadoop and the exploration of ensemble clustering methods represent promising avenues for addressing these challenges in the future. While our survey has provided valuable insights into the state-of-the-art in clustering, it is essential to acknowledge certain limitations inherent in our study. From a theoretical standpoint, the complexity of clustering problems and the diversity of datasets make it challenging to devise a one-size-fits-all solution. Moreover, practical limitations, such as computational resources and algorithm scalability, may impact the applicability of certain clustering techniques in real-world scenarios.

Moving forward, future research in clustering should focus on addressing these limitations and exploring new avenues for improvement. One promising direction is the development of hybrid meta-heuristic algorithms that combine the strengths of different optimization techniques to overcome the shortcomings of individual approaches. Additionally, there is a need for more extensive benchmarking of clustering algorithms using diverse datasets and performance metrics to ensure robustness and generalizability of results. In conclusion, our survey has provided valuable insights into the state-of-the-art meta-heuristic clustering algorithms and identified key areas for future research. By addressing the challenges posed by clustering in the era of big data, we can unlock new opportunities for knowledge discovery and decision-making in various domains.

Author contributions A.B wrote the main manuscript text and C. prepared all figures & tables. All authors reviewed the manuscript.

Funding This study was not funded by any organization.

Data availability This is a survey paper and data is not associated with the manuscript.

Declarations

Conflict of interest There is no conflict of interest.

Informed consent Informed consent was obtained from all individual participants included in the study.

Research involving human participants and/or animals This article does not contain any studies with human participants and animals performed by any of the authors.

References

Abasi AK, Khader AT, Al-Betar MA, Naim S, Alyasseri ZAA, Makhadmeh SN (2020) A novel hybrid multi-verse optimizer with K-means for text documents clustering. Neural Comput Appl 32:17703–17729

Abbasi S, Choukolaei HA (2023) A systematic review of green supply chain network design literature focusing on carbon policy. Decis Anal J 6:100189

Abualigah LM, Khader AT, Hanandeh ES (2018a) Hybrid clustering analysis using improved krill herd algorithm. Appl Intell 48(11):4047–4071

Abualigah LM, Khader AT, Hanandeh ES (2018b) A hybrid strategy for krill herd algorithm with harmony search algorithm to improve the data clustering1. Intell Decis Technol 12(1):3–14

Abualigah L, Elaziz MA, Yousri D, Al-qaness MA, Ewees AA, Zitar RA (2023) Augmented arithmetic optimization algorithm using opposite-based learning and lévy flight distribution for global optimization and data clustering. J Intell Manuf 34(8):3523–3561

Ahmadi R, Ekbatanifard G, Bayat P (2021) A modified grey wolf optimizer based data clustering algorithm. Appl Artif Intell 35(1):63–79

Alam S, Dobbie G, Rehman SU (2015) Analysis of particle swarm optimization-based hierarchical data clustering approaches. Swarm Evol Comput 25:36–51

Aljarah I, Mafarja M, Heidari AA, Faris H, Mirjalili S (2020) Clustering analysis using a novel locality-informed grey wolf-inspired clustering approach. Knowl Inform Syst 62:507–539

Allab K, Labiod L, Nadif M (2017) A semi-NMF-PCA unified framework for data clustering. IEEE Trans Knowl Data Eng 29(1):2–16

Alotaibi Y (2022) A new meta-heuristics data clustering algorithm based on tabu search and adaptive search memory. Symmetry 14(3):623

Alswaitti M, Ishak MK, Isa NAM (2018) Optimized gravitational-based data clustering algorithm. Eng Appl Artif Intell 73:126–148

Amiri E, Mahmoudi S (2016) Efficient protocol for data clustering by fuzzy cuckoo optimization algorithm. Appl Soft Comput 41:15–21

Asadi-Zonouz M, Amin-Naseri MR, Ardjmand E (2022) A modified unconscious search algorithm for data clustering. Evol Intel 15(3):1667–1693

Bahrololoum A, Nezamabadi-pour H, Saryazdi S (2015) A data clustering approach based on the universal gravity rule. Eng Appl Artif Intell 45:415–428

Banharnsakun A (2017) A MapReduce-based artificial bee colony for large-scale data clustering. Pattern Recogn Lett 93:78–84

Barshandeh S, Dana R, Eskandarian P (2022) A learning automata-based hybrid MPA and JS algorithm for numerical optimization problems and its application on data clustering. Knowl-Based Syst 236:107682

Baykasoğlu A, Gölcük İ, Özsoydan FB (2018) Improving fuzzy c-means clustering via quantum-enhanced weighted superposition attraction algorithm. Hacettepe J Math Stat 48(3):859–882

Bijari K, Zare H, Veisi H, Bobarshad H (2018) Memory-enriched big bang–big crunch optimization algorithm for data clustering. Neural Comput Appl 29:111–121

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

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

Chang X, Wang Q, Liu Y, Wang Y (2016) Sparse regularization in fuzzy $ c $-means for high-dimensional data clustering. IEEE Trans Cybern 47(9):2616–2627

Cho PPW, Nyunt TTS (2020) Data clustering based on modified differential evolution and quasi-opposition-based learning. Intell Eng Syst 13(6):168–178
Cruz DPF, Maia RD, da Silva LA, de Castro LN (2016) BeeRBF: a bee-inspired data clustering approach to design RBF neural network classifiers. Neurocomputing 172:427–437
Das P, Das DK, Dey S (2018a) A new class topper optimization algorithm with an application to data clustering. IEEE Trans Emerg Top Comput 8(4):948–959
Das P, Das DK, Dey S (2018b) A modified bee colony optimization (MBCO) and its hybridization with k-means for an application to data clustering. Appl Soft Comput 70:590–603
Deb S, Tian Z, Fong S, Wong R, Millham R, Wong KK (2018) Elephant search algorithm applied to data clustering. Soft Comput 22(18):6035–6046
Demirci H, Yurtay N, Yurtay Y, Zaimoğlu EA (2023) Electrical search algorithm: a new metaheuristic algorithm for clustering problem. Arab J Sci Eng 48(8):10153–10172
dos Santos TR, Zárate LE (2015) Categorical data clustering: What similarity measure to recommend? Expert Syst Appl 42(3):1247–1260
Elyasigomari V, Mirjafari MS, Screen HR, Shaheed MH (2015) Cancer classification using a novel gene selection approach by means of shuffling based on data clustering with optimization. Appl Soft Comput 35:43–51
Emrouznejad A, Abbasi S, Sıcakyüz Ç (2023) Supply chain risk management: a content analysisbased review of existing and emerging topics. Supply Chain Anal 3:100031
Ferrari DG, De Castro LN (2015) Clustering algorithm selection by meta-learning systems: a new distance-based problem characterization and ranking combination methods. Inform Sci 301:181–194
Gebru ID, Alameda-Pineda X, Forbes F, Horaud R (2016) EM algorithms for weighted-data clustering with application to audio-visual scene analysis. IEEE Trans Pattern Anal Mach Intell 38(12):2402–2415
Ghorbanzadeh L, Torshabi AE, Nabipour JS, Arbatan MA (2016) Development of a synthetic adaptive neuro-fuzzy prediction model for tumor motion tracking in external radiotherapy by evaluating various data clustering algorithms. Technol Cancer Res Treat 15(2):334–347
Gupta Y, Saini A (2019) A new swarm-based efficient data clustering approach using KHM and fuzzy logic. Soft Comput 23(1):145–162
Gupta C, Jain A, Tayal DK, Castillo O (2018) ClusFuDE: forecasting low dimensional numerical data using an improved method based on automatic clustering, fuzzy relationships and differential evolution. Eng Appl Artif Intell 71:175–189
Gutierrez-Rodríguez AE, Martínez-Trinidad JF, García-Borroto M, Carrasco-Ochoa JA (2015) Mining patterns for clustering on numerical datasets using unsupervised decision trees. Knowl-Based Syst 82:70–79
Haeri Boroujeni SP, Pashaei E (2023) A hybrid chimp optimization algorithm and generalized normal distribution algorithm with opposition-based learning strategy for solving data clustering problems. Iran J Comput Sci 65:1–37
Hahsler M, Bolaños M (2016) Clustering data streams based on shared density between micro-clusters. IEEE Trans Knowl Data Eng 28(6):1449–1461
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
Harita M, Wong A, Suppi R, Rexachs D, Luque E (2024) A metaheuristic search algorithm based on sampling and clustering. IEEE Access 12:15493
Hashemi SE, Gholian-Jouybari F, Hajiaghaei-Keshteli M (2023) A fuzzy C-means algorithm for optimizing data clustering. Expert Syst Appl 227:120377
Hu H, Liu J, Zhang X, Fang M (2023) An effective and adaptable K-means algorithm for big data cluster analysis. Pattern Recogn 139:109404
Jadhav AN, Gomathi N (2018) WGC: hybridization of exponential grey wolf optimizer with whale optimization for data clustering. Alex Eng J 57(3):1569–1584
Jing L, Tian K, Huang JZ (2015) Stratified feature sampling method for ensemble clustering of high dimensional data. Pattern Recogn 48(11):3688–3702
Kannan R, Vempala S, Vetta A (2004) On clusterings: good, bad and spectral. J ACM (JACM) 51(3):497–515
Kaur A, Datta A (2015) A novel algorithm for fast and scalable subspace clustering of high-dimensional data. J Big Data 2(1):17
Kaur A, Kumar Y (2022) A new metaheuristic algorithm based on water wave optimization for data clustering. Evol Intel 15(1):759–783
Kaur A, Pal SK, Singh AP (2020) Hybridization of chaos and flower pollination algorithm over k-means for data clustering. Appl Soft Comput 97:105523

Kumar Y, Kaur A (2022) Variants of bat algorithm for solving partitional clustering problems. Eng Comput 38(Suppl 3):1973–1999
Kumar Y, Sahoo G (2014) A charged system search approach for data clustering. Progress Artif Intell 2(2–3):153–166
Kumar Y, Sahoo G (2015a) A hybrid data clustering approach based on improved cat swarm optimization and K-harmonic mean algorithm. AI Commun 28(4):751–764
Kumar Y, Sahoo G (2015b) Hybridization of magnetic charge system search and particle swarm optimization for efficient data clustering using neighborhood search strategy. Soft Comput 19(12):3621–3645
Kumar Y, Sahoo G (2016) A hybridise approach for data clustering based on cat swarm optimisation. Int J Inform Commun Technol 9(1):117–141
Kumar Y, Singh PK (2018) Improved cat swarm optimization algorithm for solving global optimization problems and its application to clustering. Appl Intell 48:2681–2697
Kumar D, Bezdek JC, Palaniswami M, Rajasegarar S, Leckie C, Havens TC (2015) A hybrid approach to clustering in big data. IEEE Trans Cybern 46(10):2372–2385
Kumar V, Chhabra JK, Kumar D (2016a) Automatic data clustering using parameter adaptive harmony search algorithm and its application to image segmentation. J Intell Syst 25(4):595–610
Kumar V, Chhabra JK, Kumar D (2016b) Data clustering using differential search algorithm. Pertanika J Sci Technol 24(2):295
Kuo RJ, Lin TC, Zulvia FE, Tsai CY (2018a) A hybrid metaheuristic and kernel intuitionistic fuzzy c-means algorithm for cluster analysis. Appl Soft Comput 67:299–308
Kuo RJ, Rizki M, Zulvia FE, Khasanah AU (2018b) Integration of growing self-organizing map and bee colony optimization algorithm for part clustering. Comput Ind Eng 120:251–265
Kuo RJ, Lin JY, Nguyen TPQ (2021) An application of sine cosine algorithm-based fuzzy possibilistic c-ordered means algorithm to cluster analysis. Soft Comput 25(5):3469–3484
Kushwaha N, Pant M (2018) Fuzzy magnetic optimization clustering algorithm with its application to health care. J Ambient Intell Human Comput 1:1–10
Kushwaha N, Pant M, Kant S, Jain VK (2018) Magnetic optimization algorithm for data clustering. Pattern Recogn Lett 115:59–65
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
Lakshmi K, Visalakshi NK, Shanthi S (2018) Data clustering using K-means based on crow search algorithm. Sādhanā 43(11):190
Lee J, Perkins D (2021) A simulated annealing algorithm with a dual perturbation method for clustering. Pattern Recogn 112:107713
Leski JM (2016) Fuzzy c-ordered medoids clustering for interval-valued data. Pattern Recogn 58:49–67
Li Y, Yang G, He H, Jiao L, Shang R (2016) A study of large-scale data clustering based on fuzzy clustering. Soft Comput 20(8):3231–3242
Li T, De la Prieta Pintado F, Corchado JM, Bajo J (2017) Multi-source homogeneous data clustering for multi-target detection from cluttered background with misdetection. Appl Soft Comput 60:436–446
Liu Q, Zhang R, Hu R, Wang G, Wang Z, Zhao Z (2019) An improved path-based clustering algorithm. Knowl-Based Syst 163:69–81
Mageshkumar C, Karthik S, Arunachalam VP (2019) Hybrid metaheuristic algorithm for improving the efficiency of data clustering. Clust Comput 22(1):435–442
Mansueto P, Schoen F (2021) Memetic differential evolution methods for clustering problems. Pattern Recogn 114:107849
Meng L, Tan AH, Wunsch DC (2016) Adaptive scaling of cluster boundaries for large-scale social media data clustering. IEEE Trans Neural Netw Learn Syst 27(12):2656–2669
Mikaeil R, Haghshenas SS, Haghshenas SS, Ataei M (2018) Performance prediction of circular saw machine using imperialist competitive algorithm and fuzzy clustering technique. Neural Comput Appl 29(6):283–292
Moghadam P, Ahmadi A (2023) A novel two-stage bio-inspired method using red deer algorithm for data clustering. Evol Intell 17:1–18
Montgomery D, Addison PS, Borg U (2016) Data clustering methods for the determination of cerebral auto regulation functionality. J Clin Monit Comput 30(5):661–668
Narayana GS, Vasumathi D (2018) An attributes similarity-based K-medoids clustering technique in data mining. Arab J Sci Eng 43(8):3979–3992
Nayak J, Naik B, Kanungo DP, Behera HS (2018) A hybrid elicit teaching learning based optimization with fuzzy c-means (ETLBO-FCM) algorithm for data clustering. Ain Shams Eng J 9(3):379–393
Nazari A, Dehghan A, Nejatian S, Rezaie V, Parvin H (2019) A comprehensive study of clustering ensemble weighting based on cluster quality and diversity. Pattern Anal Appl 22(1):133–145

Nguyen DD, Ngo LT, Pham LT, Pedrycz W (2015) Towards hybrid clustering approach to data classification: multiple kernels based interval-valued fuzzy C-means algorithms. Fuzzy Sets Syst 279:17–39
Noorbehbahani F, Mousavi SR, Mirzaei A (2015) An incremental mixed data clustering method using a new distance measure. Soft Comput 19(3):731–743
Özbakır L, Turna F (2017) Clustering performance comparison of new generation meta-heuristic algorithms. Knowl-Based Syst 130:1–16
Ozturk C, Hancer E, Karaboga D (2015) Dynamic clustering with improved binary artificial bee colony algorithm. Appl Soft Comput 28:69–80
Pacifico LD, Ludermir TB (2021) An evaluation of k-means as a local search operator in hybrid memetic group search optimization for data clustering. Nat Comput 20(3):611–636
Pakrashi A, Chaudhuri BB (2016) A Kalman filtering induced heuristic optimization based partitional data clustering. Inform Sci 369:704–717
Patel VP, Rawat MK, Patel AS (2023) Local neighbour spider monkey optimization algorithm for data clustering. Evol Intel 16(1):133–151
Pimentel BA, de Carvalho AC (2019) A new data characterization for selecting clustering algorithms using meta-learning. Inform Sci 477:203–219
Pohl D, Bouchachia A, Hellwagner H (2016) Online indexing and clustering of social media data for emergency management. Neurocomputing 172:168–179
Premkumar M, Sinha G, Ramasamy MD, Sahu S, Subramanyam CB, Sowmya R, Derebew B (2024) Augmented weighted K-means grey wolf optimizer: an enhanced metaheuristic algorithm for data clustering problems. Sci Rep 14(1):5434
Puschmann D, Barnaghi P, Tafazolli R (2017) Adaptive clustering for dynamic IoT data streams. IEEE Internet Things J 4(1):64–74
Qiao S, Zhou Y, Zhou Y, Wang R (2019) A simple water cycle algorithm with percolation operator for clustering analysis. Soft Comput 23(12):4081–4095
Qtaish A, Braik M, Albashish D, Alshammari MT, Alreshidi A, Alreshidi EJ (2024) Optimization of K-means clustering method using hybrid capuchin search algorithm. J Supercomput 80(2):1728–1787
Queiroga E, Subramanian A, Lucídio dos Anjos FC (2018) Continuous greedy randomized adaptive search procedure for data clustering. Appl Soft Comput 72:43–55
Rahnema N, Gharehchopogh FS (2020) An improved artificial bee colony algorithm based on whale optimization algorithm for data clustering. Multim Tools Appl 79(43):32169–32194
Rathore P, Kumar D, Bezdek JC, Rajasegarar S, Palaniswami M (2018) A rapid hybrid clustering algorithm for large volumes of high dimensional data. IEEE Trans Knowl Data Eng 31(4):641–654
Safarinejadian B, Hasanpour K (2016) Distributed data clustering using mobile agents and EM algorithm. IEEE Syst J 10(1):281–289
Salem SB, Naouali S, Chtourou Z (2018) A fast and effective partitional clustering algorithm for large categorical datasets using a k-means based approach. Comput Electr Eng 68:463–483
Salih SQ, Alsewari AA, Wahab HA, Mohammed MK, Rashid TA, Das D, Basurra SS (2023) Multipopulation black hole algorithm for the problem of data clustering. PLoS ONE 18(7):e0288044
Santi É, Aloise D, Blanchard SJ (2016) A model for clustering data from heterogeneous dissimilarities. Eur J Oper Res 253(3):659–672
Schaeffer SE (2007) Graph clustering computer. Sci Rev 1(1):27–64
Senthilnath J, Kulkarni S, Suresh S, Yang XS, Benediktsson JA (2019) FPA clust: evaluation of the flower pollination algorithm for data clustering. Evol Intell 14:1–11
Serapião AB, Corrêa GS, Gonçalves FB, Carvalho VO (2016) Combining K-means and K-harmonic with fish school search algorithm for data clustering task on graphics processing units. Appl Soft Comput 41:290–304
Sharma M, Chhabra JK (2019) An efficient hybrid PSO polygamous crossover based clustering algorithm. Evol Intell 14:1–19
Sheng W, Chen S, Fairhurst M, Xiao G, Mao J (2014) Multilocal search and adaptive niching based memetic algorithm with a consensus criterion for data clustering. IEEE Trans Evol Comput 18(5):721–741
Sheng W, Chen S, Sheng M, Xiao G, Mao J, Zheng Y (2016) Adaptive multisubpopulation competition and multiniche crowding-based memetic algorithm for automatic data clustering. IEEE Trans Evol Comput 20(6):838–858
Shial G, Sahoo S, Panigrahi S (2023) An enhanced GWO algorithm with improved explorative search capability for global optimization and data clustering. Appl Artif Intell 37(1):2166232
Siddiqi UF, Sait SM (2017) A new heuristic for the data clustering problem. IEEE Access 5:6801

Singh T (2020) A chaotic sequence-guided Harris hawks optimizer for data clustering. Neural Comput Appl 32:17789–17803
Singh S, Srivastava S (2022) Kernel fuzzy C-means clustering with teaching learning based optimization algorithm (TLBO-KFCM). J Intell Fuzzy Syst 42(2):1051–1059
Singh H, Rai V, Kumar N, Dadheech P, Kotecha K, Selvachandran G, Abraham A (2023) An enhanced whale optimization algorithm for clustering. Multim Tools Applic 82(3):4599–4618
Su ZG, Denoeux T (2018) BPEC: belief-peaks evidential clustering. IEEE Trans Fuzzy Syst 27(1):111–123
Tan PN, Steinbach M, Kumar V (2016) Introduction to data mining. Pearson Education India
Tang D, Dong S, He L, Jiang Y (2016) Intrusive tumor growth inspired optimization algorithm for data clustering. Neural Comput Appl 27(2):349–374
Tekieh R, Beheshti Z (2024) A MapReduce-based big data clustering using swarm-inspired meta-heuristic algorithms. Sci Iranica 31:737
Tinós R, Zhao L, Chicano F, Whitley D (2018) NK hybrid genetic algorithm for clustering. IEEE Trans Evol Comput 22(5):748–761
Tsai CW, Chang WY, Wang YC, Chen H (2019) A high-performance parallel coral reef optimization for data clustering. Soft Comput 23:9327–9340
Turkoglu B, Uymaz SA, Kaya E (2022) Clustering analysis through artificial algae algorithm. Int J Mach Learn Cybern 13(4):1179–1196
Vo TNC, Nguyen HP, Vo TNT (2016) Making kernel-based vector quantization robust and effective for incomplete educational data clustering. Vietnam J Comput Sci 3(2):93–102
Xiang WL, Zhu N, Ma SF, Meng XL, An MQ (2015) A dynamic shuffled differential evolution algorithm for data clustering. Neurocomputing 158:144–154
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
Xu S, Liu S, Zhou J, Feng L (2019) Fuzzy rough clustering for categorical data. Int J Mach Learn Cybern 10(11):3213–3322
Yan Y, Nguyen T, Bryant B, Harris FC Jr (2019) Robust fuzzy cluster ensemble on cancer gene expression data. Proc Int Conf 60:120–128
Yang Y, Jiang J (2018) Adaptive Bi-weighting toward automatic initialization and model selection for HMM-based hybrid meta-clustering ensembles. IEEE Trans Cybern 49(5):1657–1668
Yang CL, Kuo RJ, Chien CH, Quyen NTP (2015) Non-dominated sorting genetic algorithm using fuzzy membership chromosome for categorical data clustering. Appl Soft Comput 30:113–122
Yao X, Ge S, Kong H, Ning H (2018) An improved clustering algorithm and its application in wechat sports users analysis. Procedia Comput Sci 129:166–174
Yu H, Zhang C, Wang G (2016) A tree-based incremental overlapping clustering method using the three-way decision theory. Knowl-Based Syst 91:189–203
Yuwono M, Su SW, Moulton BD, Nguyen HT (2014) Data clustering using variants of rapid centroid estimation. IEEE Trans Evol Comput 18(3):366–377
Zhang B, Qin S, Wang W, Wang D, Xue L (2016a) Data stream clustering based on Fuzzy C-Mean algorithm and entropy theory. Signal Process 126:111–116
Zhang H, Raitoharju J, Kiranyaz S, Gabbouj M (2016b) Limited random walk algorithm for big graph data clustering. J Big Data 3(1):26
Zhang QH, Li BL, Liu YJ, Gao L, Liu LJ, Shi XL (2016c) Data clustering using multivariant optimization algorithm. Int J Mach Learn Cybern 7(5):773–782
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
Zhu E, Ma R (2018) An effective partitional clustering algorithm based on new clustering validity index. Appl Soft Comput 71:608–621

Authors and Affiliations

Arvinder Kaur[1] · Yugal Kumar[2] · Jagpreet Sidhu[2]

✉ Jagpreet Sidhu
Jagpreet.pu@gmail.com

Arvinder Kaur
er.arvinderdhillon@gmail.com

Yugal Kumar
yugalkumar.14@gmail.com

[1] Department of Information Technology, Chandigarh Engineering College-CGC, Landran, Punjab, India

[2] School of Technology Management and Engineering, NMIMS, Chandigarh, India