A robust fuzzy approach for gene expression data clustering

Meskat Jahan · Mahmudul Hasan

Accepted: 17 August 2021 / Published online: 17 October 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract
In the big data era, clustering is one of the most popular data mining method. Most clustering algorithms have complications like automatic cluster number determination, poor clustering precision, inconsistent clustering of various datasets and parameter-dependent, etc. A new fuzzy autonomous solution for clustering named Meskat-Mahmudul (MM) clustering algorithm was proposed to overcome the complexity of parameter-free automatic cluster number determination and clustering accuracy. The Meskat-Mahmudul clustering algorithm finds out the exact number of clusters based on the average silhouette method in multivariate mixed attribute dataset, including real-time gene expression dataset and missing values, noise, and outliers. Meskat-Mahmudul Extended K-Means (MMK) clustering algorithm enhances the K-Means algorithm, which serves the purpose of automatic cluster discovery and runtime cluster placement. Several validation methods are used to evaluate clusters and certify optimum cluster partitioning and perfection. Some datasets are used to assess the performance of the proposed algorithms to other algorithms in terms of time complexity and clustering efficiency. Finally, Meskat-Mahmudul clustering and Meskat-Mahmudul Extended K-Means clustering algorithms were found superior over conventional algorithms.

Keywords FCM · K-Means · Fuzzy clustering · Clustering algorithm · Data mining

1 Introduction
Clustering is a process of recognizing similar objects and separated non-identical objects. Cluster analysis is divided into two primary groups, like hard and fuzzy clustering methods. In addition, other types of clustering methods are distribution-based clustering, connectivity-based clustering, and constraint-based clustering. Many algorithms were developed based on these categorizations. DIANA (Divisive Analysis) or AGNES (Agglomerative Nesting) methods used in hierarchical clustering. However, hierarchical clustering does not well suited for large datasets. Density-based clustering or model-based methods fall into distribution-based categories where approaches consider density instead of distances. Using density reachability and density connectivity to create a cluster by separating and differentiate regions of varying density degrees Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al. 1996) algorithm is formed. Yet, it is a slow process and required a complex algorithm to implement. Famous centroids-based hard clustering is K-Means (Han et al. 2012), K-Medoids (Gentle et al. 1991), and some extended versions of the K-means are Isodata algorithm, Forgy algorithms, etc. Fuzzy C-Means, Fuzzy K-NN, are popular soft clustering schemas. Nevertheless, we have to provide the number of cluster centroids before conducting cluster partitioning methods.

K-Means (Han et al. 2012), K-Medoids (Gentle et al. 1991), and PAM (Ng and Han 1994) are sample algorithms of partitioning clustering. The mean value of items in a cluster is used to measure similarity in the K-means algorithm. This algorithm aims to divide the dataset into K clusters, where K is a predefined number. Each cluster in each round comprises the points closest to the corresponding reference point, and the centroid of each cluster is used as the next round’s reference point. Because of these iterations, the reference point is chosen closer to the actual cluster centroid, and the clustering effect improves. Since
the PAM (Ng and Han 1994) algorithm evaluates all objects, the central point of each class is considered. For many combinations, the accuracy of the clustering results is calculated. The PAM approach works fine in a small dataset, but the result is not optimal with a giant dataset. One of the biggest weaknesses of density and demarcation approaches is the difficulty in defining the clusters prior to certain algorithms including, in particular, the form of its clusters like most algorithms.

Therefore, we can summarize the drawbacks of partitioning algorithms as follows: (1) Number of clusters K determination; (2) The PAM approach works well for a small dataset, but it is not suitable for a large dataset; (3) If the data contain noise and outliers, it is prone to errors. To overcome these, we come up with the idea of automatic clustering. Hence, we proposed two approaches named Meskat-Mahmudul clustering (MM clustering) and Meskat-Mahmudul Extended K-Means (MMK clustering) clustering algorithm for non-parametric automatic cluster number determination and clustering to solve the above-stated problems accordingly. In MM clustering, we will use average silhouette-based cluster number determination techniques, and for clustering, we adopt the FCM algorithm. In addition, in MMK clustering, we used the K-Means clustering schema for cluster partitioning and the average silhouette method for cluster number determination. The dataset used in this paper includes Wisconsin Breast Cancer, Iris, Leukemia, Motorcycle, etc. These datasets all come from the UCI database (Dua and Graff 2017). Seven validation criteria were used for evaluating outcomes. Findings show that the recommended methods are more enhanced compared with the other state-of-the-art algorithms. Finally, the proposed methods outperform previous research. Figure 1 illustrates the overview of the working procedure.

Enhancing existing algorithms, hybridizing diverse algorithms, and presenting unique algorithms are the three categories of conceptual studies reported in the literature. These three fields are all quite active, with a significant number of algorithms and solutions. As per the No Free Lunch principle (Wolpert and Macready 1997), there exists hardly any optimization algorithm that can handle all optimization issues. That’s why experts don’t employ any single algorithm. As a result, existing algorithms must be improved or new algorithms must be proposed in order to effectively tackle major issues or provide responses to new challenges. This is what prompted us to develop new approaches.

In brief, the contribution of this work is as follows:

1. We present an autonomous solution for clustering, which includes automated cluster number determination using the average silhouette method and clustering accordingly;
2. We incorporate a preprocessing mechanism into the proposed algorithms to manage unnecessary noise and outliers in the dataset;
3. Proposed methods are parameter-free and do not require any manual intervention in the process;

The remaining of this research includes: The backgrounds of the automated clustering techniques are described in Sect. 2. The suggested algorithms structure is depicted in Sect. 3. In Sect. 4, algorithms performance is analyzed using standard evaluation processes. Section 5 presents the outcomes of the proposed algorithms. There are also comparisons to state-of-the-art procedures. Lastly, Sect. 6 brings the research to a close.

2 Literature review

Tao Lei et al. (2018) proposed a significantly fast and robust Fuzzy C-Means clustering algorithm. They introduced local spatial information to an objective function of the FCM algorithm for image segmentation based on morphological reconstruction for image detail-preservation and noise-immunity and membership filtering (FRFCM). Singa and Yang (2020) highlight a novel U K-means clustering algorithm that automatically finds the optimal number of clusters without providing any parameter selection and initialization. The classic computational methods used to assess numbers using algebraic operators (division, multiplication, subtraction, and addition). Abualigah et al. (2021a) employ those basic operators as optimization techniques to select the optimum element out of a pool of potential options (solutions) based on particular requirements. Dhanachandra et al. (2015) provide a methodology for image segmentation using K-means and subtractive clustering algorithms. The initial centers produced using a subtractive cluster, and these centers were used in the k-means algorithm to segment the image. Finally, a medial filter was added to the segmented image to eliminate any unwanted regions. Cai et al. (2020) discovered that the cluster centers cannot be calculated automatically and that the selected cluster centers which fall into a local optimum and the random selection of the parameter cut-off distance value when they examined the quick search and find of density peaks clustering (DPC) algorithm (Rodriguez and Laio 2014). PDPC (Cai et al. 2020) clustering algorithm is proposed to address these issues. Particle swarm optimization (PSO) (Eberhart and Kennedy 1995) is implemented because of its simple idea and strong global searchability, which allows it to find the best solution in a short amount of time. The Aquila
Optimizer (AO) is an innovative population-based optimization method proposed in this study (Abualigah et al. 2021b), which is motivated first by Aquila’s instincts while grabbing prey. Jinyin et al. (2017) proposed a cluster center fast determination clustering algorithm that can adapt the dc parameter and automatically get the optimal density radius and select the cluster center in the clustering process. In cloud computing, a novel adaptive task allocation strategy (Abualigah et al. 2021) to achieve the shortest duration of job transfer on existing resources which extend the discoverability of the multi-verse scheduler by combining it with evolutionary algorithms. Nayak et al. (2015) gather some drawbacks of the FCM algorithm, including (1) point-based membership (Chao et al. 2013), (2) information loss when calculating similarity degree, (3) slow convergence rate, (4) sensitivity to initialization, etc. Hou et al. (2016) propose clustering as image segmentation techniques using a new parameter-free algorithm based on DSets-DBSCAN. Reddy et al. (2017a) using a hybrid Fuzzy ANN-based classifier for the diagnosis of diabetes. Also, Reddy et al. (2017b) forecasting heart disease by applying the Hybrid OFBAT combined with a Rule-Based Fuzzy Logic Model. Another work on heart disease classification was presented by Reddy et al. (2020). Table 1 provides a rundown of the preceding discussion. Chen (2015) developed the DC-MDACC algorithm which divides the dataset into graded dominant, numerical dominant and mixed attribute types. MH algorithm (Jahan and Hasan 2020) also provides the solution for automatic cluster detection and identification. Jahan et al. (2019) show a comparative analysis of some traditional algorithms. Abualigah et al. (2021) cover a variety of SCA versions and deployments which have surfaced throughout the research after presenting an outline of the SCA algorithm. To reduce the makespan value and improve resource consumption, Abualigah et al. (2021) developed a multi optimization built on the suggested MHDA.

3 Proposed algorithms

Detaching heterogeneous data and assemble them into a homogeneous set based on similarity is the goal of clustering. Earlier, partition-based clustering depends on pre-specified cluster numbers, which is one of the main drawbacks. Here, we propose a hybrid method for the automatic clustering of data. Optimum cluster number selection based on the average silhouette method is integrated into the proposed algorithms. Following are the proposed approaches:

**Approach 1:** Meskat-Mahmudul (MM) Clustering Algorithm.

**Approach 2:** Meskat-Mahmudul Extended K-Means (MMK) Clustering Algorithm.

In the following sections, we described approaches.
Meskat-Mahmudul (MM) clustering algorithm

Meskat-Mahmudul clustering is a soft clustering algorithm. This is a platform that includes essential clustering operations, such as deciding the number of subgroups and grouping. Several popular existing approaches use some sort of dependencies and require specific information assumptions to understand the dataset’s structure or attributes. To address this, Meskat-Mahmudul clustering offers a robust algorithm to calculate the number of clusters. Algorithms steps are shown below:

1. The first step is to normalize the dataset for removing unwanted noise and outliers.

   Phase 1: Determine the structure beneath the surface.

2. Phase: Determine the number of clusters at algorithm execution time.

3. Phase: By using Fuzzy C-Means algorithm to classify data.

Following is the MM algorithm. Here, steps 2 to step 5 are used to compute the cluster number using average silhouette methods and step 6 to step 8 perform data element clustering using FCM. The pictorial representation of the MM clustering algorithm is provided in Fig. 2.

\[
U_i = \{1 \ldots n\} \text{ be the dataset, } \bar{a_i} \text{ is average distance, } \bar{b_i} \text{ is minimum average distance, } s_i \text{ is silhouette value, } U' \text{ is the normalized dataset using z-score method} \]

\[
\text{avg}
\]

\[
\text{silhouette} = \text{mean} \left( \bar{s}_i \right)
\]

\[
x_{\max} = \text{max} \left( \bar{s}_i \right)
\]

\[
\text{assign cluster number, } k = x_{\max}
\]

\[
\text{for } j = 1 \text{ to } k, \text{ compute fuzzy membership } [29],
\]

\[
\mu_{ij}^{(1)} = \frac{1}{\sum_{p=1}^{k} (D_{ij} - D_{ip})^{2m-1}}
\]

\[
\text{for } j = 1 \text{ to } k, \text{ calculate the fuzzy centers } [30],
\]

\[
V_j = \frac{\sum_{i=1}^{n} (\mu_{ij}^m \bar{x}_i)}{\sum_{i=1}^{n} (\mu_{ij}^m)}
\]

\[
\text{repeat step 6 to 7, until } \|U^{(0)} - U^{(1)}\| < \varepsilon
\]

Meskat-Mahmudul extended K-means (MMK) clustering algorithm

Meskat-Mahmudul Extended K-Means (MMK) is an enhancement of the K-Means clustering. In the MMK algorithm, we propose to combine the strategies for autonomously selecting the number of clusters with of standard K-Means algorithm. MMK is a hard clustering procedure. The below are phases of this automated system:

Phase 1: Scale the data.
Phase 2: Investigate the internal structure.
Phase 3: Methods for automatically estimating the cluster number.
Phase 4: Conduct clustering using the K-Means clustering technique.
The MMK algorithm is given below. Here, from steps 2 to 5, it computes the number of clusters, and from steps 6 to 10, it performs clustering. Figure 3 depicts the algorithm flowchart.

### MM Extended K-Means Clustering Algorithm

1. for \( i=1 \) to \( n \),
   
   \[
   U' = \frac{U - \mu}{\sigma} \quad / \text{ normalize dataset using z-score method} / \]

2. for \( i=1 \) to \( N \),
   
   \[
   s_i = \frac{(b_i - a_i)}{\text{max}(a_i, b_i)} \quad (28) \quad \text{and determine, } \quad \overline{s}_i = \text{mean} (s_i)
   \]

3. obtain average silhouette value,
   
   \[
   \text{avg}_{\text{silhouette}} = \text{mean} (\overline{s}_i)
   \]

4. find the maximum average silhouette values,
   
   \[
   x_{\text{max}} = \text{max} (\overline{s}_i)
   \]

5. assign cluster number, \( k = x_{\text{max}} \)

6. random selection of \(^k\) cluster centers

7. for \( i=1 \) to \( n \),
   
   Calculate and search the minimal distance between a data element and cluster centers and assign the data point to the nearest cluster center

8. for \( i=1 \) to \( n \),
   
   reappraise the cluster centers using
   
   \[
   v_i = \frac{1}{k_i} \sum_{j=1}^{k_i} U_i
   \]

9. for \( i=1 \) to \( n \),
   
   reappraise the data elements distance and update the cluster centers.

10. Repeat steps 7 to 9, until no data elements will not be reassigned.

### 3.1 Determination of number of clusters

We used optimization techniques for determining the number of clusters. Average silhouette values are used for cluster number selection. First, we observe the silhouette values for over a range of cluster numbers. Then, we obtain the average silhouette values from them. From the literature, we know that ranges of silhouette value vary from 1 to \(-1\), and a high silhouette value is used to measure well-matched data element to its self-cluster and imperfectly matched to other clusters. The precise clustering depends on high silhouette values for most of the data elements. The standard of clustering can be measured using average silhouette values of various cluster numbers. From the average silhouette values, find the maximum average silhouette value and its corresponding cluster number. This cluster number is the desired optimized number of clusters. Below is the outcome of optimizing techniques for cluster number determination for all datasets. These are shown in Figs. 4, 5, 6 and 7.

### 3.2 MM and MM extended K-means clustering algorithm

MM and MM Extended K-Means clustering algorithms share a common approach for determining the number of...
clusters where the MM algorithm uses a fuzzy approach, and the MMK algorithm is a hard algorithm.

### 4 Performance analysis among clustering algorithms

#### 4.1 Algorithms performance evaluation based on execution time

A comparative analysis of algorithms performance based on execution time provided below. Algorithms are ordered from fuzzy algorithm to crips algorithm. Here, algorithms are compared for every selected dataset altogether. Results are summarized in Fig. 8. In Fig. 8, we consider four datasets, and seven algorithms' performance is tested on them. In the WBC dataset, MM algorithms take acceptable execution time, and the MMK algorithm also takes less time than the rest of the hardcore algorithms. In the Motorcycle dataset, all fuzzy algorithms take almost the same time to execute. MMK algorithms need more time to execute because the MMK algorithm used an optimization mechanism for cluster number selection. In the leukemia dataset, both the MM algorithm and MMK algorithm run in the shortest possible time than others. In the Iris dataset, the MM algorithm takes less time and the MMK algorithms take adequate time.

#### 4.2 Algorithms performance evaluation based on validation indexes

Validation played an essential role in assessing cluster efficiency. The efficiency of clusters has been tested by using screening indicators: Classification Entropy (CE), Partition Coefficient (PC), Xie and Beni’s Index (XB), Partition Index (SC), and Separation Index (S), Dunn’s Index (DI), Alternative Dunn Index (ADI). For evaluation of cluster partitions, the minimum separation measurement done by the separation index (S), with the smallest values indicating a true optimal partition. Cluster partition coefficient (PC) indexing is used to assess the quantity of overlapping between clusters, and its high value ensures cluster accuracy. The Dunn Index (DI) is also an internal assessment parameter for evaluating which cluster is well separated. Better clustering results demonstrated by higher DI and ADI values. The fuzziness of cluster partitions measured by CE. The strong output indicated by low CE and SC values. The XB index acknowledges the compactness of the whole cluster, with the smallest value suggesting the optimal number of clusters. Figure 9 demonstrates the algorithm’s performance evaluation based on the above-selected approaches.
5 Result & discussion

Validating the clustering algorithm is crucial for the gene expression dataset. Cluster validation is an essential criterion for assessing a cluster’s quality. Here, we analyzed the total results of the algorithms for the selected four datasets. Figure 10 illustrates the validation results for all algorithms chosen for Wisconsin Breast Cancer (WBC) dataset. The upper limit of PC (Bezdek et al. 1984) indexing is one, and the lower limit is 1/k, where k is the cluster number. Therefore, the high value of PC is always desirable, and for the WBC dataset, K-Means, K-Mediods, and MM Extended K-Means algorithms show the highest value, and the rest gives high value. The small value of CE in MM and FCM algorithms indicates fuzzy partitions, and others algorithm gives zero value, which recommended hard partitions in clustering algorithms. Low SC value of MM, FCM, GG, and GK point out good clustering algorithms.

Similarly, the low S index of FCM, MM, GG, GK algorithms stipulates satisfactory cluster partitioning. However, DI and ADI value remains same for all soft and hard algorithms. Upon analyzing the above-stated performance, we concluded that MM and MMK algorithms performance is satisfactory.

Figure 11 shows the validation results at a glance of the algorithms based on validation indexes for the Irises dataset. The PC indexing values of K-Means, K-Mediods, and MM Extended K-Means algorithms are maximum which points to the rigid cluster partitioning, and the average value of PC in FCM, GG, and MM clustering
algorithms represents fuzzy partitioning. As MM is a fuzzy clustering algorithm, it secures its fuzzy partitioning according to CE indexing along with FCM, GG, and GK algorithms. Low values of SC, S, and XB values provided by FCM, MM, and GG algorithms. The highest value of ADI and DI is provided by MMK, GG, and MM clustering algorithms. We summarized that two developed algorithms (MM and MMK) performance is satisfactory in comparing other literature for the Irises dataset.

Figure 12 summarizes the algorithms performance analysis on the leukemia dataset. It shows that the performance of MM and MMK algorithms is meet the requirement of standard literature. MM algorithm secures its position by satisfying the low value of CE, SC, and S indexing. On the other hand, the MMK algorithm provides the maximum value in PC indexing and the lowest value of S indexing. GK algorithm gives the upper limit of PC, DI, and ADI values. K-Means algorithm shows the highest PC value and lowest S values.

In Fig. 13, MM algorithms give low value in SC and S indexing and top value in ADI indexing. GK algorithms show the lowest value of XB and CE indexing. K-Means,
K-Mediods, and MMK algorithms provide unity values in PC indexing. The MM clustering algorithm solves the problems that include automatic cluster number identification, cluster placement at runtime, and fuzzy partitioning. It specifies the number of clusters and separates them appropriately. For nonlinear samples, the MM clustering methodology operates well. The MM Extended K-Means approach employs a dynamic cluster number decision to perform hard clustering. It fits nicely on a dataset that is distinguishable. The MM and MM Extended K-Means clustering algorithms choose the precise cluster number and deliver optimal partitioning and clustering based on all of the data. The MM clustering algorithm takes a lot less time to run. The MM clustering algorithm achieves automated cluster numbering prediction without post-cluster analysis and efficient clustering and higher accuracy. While it takes longer than the MM clustering algorithm, the MM Extended K-Means clustering algorithm outperforms other hard clustering algorithms. MM and MM Extended K-Means clustering algorithms are acceptable by considering validation performance.

In the end, the MM clustering algorithm takes the least amount of time compared to other standardized algorithms, and the MMK algorithm finishes the job faster than different hard clustering algorithms. In all four datasets, MM clustering and MMK clustering approaches work well.
Fig. 6 Determination of the number of clusters for Irises dataset

Fig. 7 Determination of the number of clusters for Motorcycle dataset

Fig. 8 Comparative study of algorithms based on execution time
Algorithms Performance Evaluation based on Partition Coefficient (PC)

Algorithms Performance Evaluation based on Dunn's Index (DI)

Algorithms Performance Evaluation based on Partition Index (SC)

Algorithms Performance Evaluation based on Separation Index (S)

Evaluation of Algorithms based on Alternative Dunn Index (ADI)

Fig. 9 Evaluation of algorithms based on PC, CE, SC, S, XB, DI, ADI for the selected dataset
6 Conclusion

The philosophy behind the MM clustering algorithm and the MM Extended K-Means clustering algorithm is to detect the exact clusters and classify them correspondingly naturally. On both linear and nonlinearly separated datasets, MM and MM Extended K-Means algorithms were used. By validating the cluster and conducting performance analysis, the developed algorithms results compared to other avant-garde algorithms. The distinction focused on execution time and validation indexes, and this is an excellent way to measure which clustering algorithm is ideal for this type of dataset. The outcome of the MM Extended K-Means clustering algorithm on discriminant datasets is stronger than the MM clustering algorithm, and on nonlinearly separable datasets, MM algorithms triumph over others. Both the MM and MM Extended K-Means clustering algorithms achieve the requirements of correctly selecting the number of clusters efficiently. They provide faster and more reliable results than most of the other clustering algorithms.

We concentrate above on gene expression datasets and the methodologies put to the test in a monolithic architecture. We expect to work with large and real-time micro-
array gene expression datasets in future and incorporate them in distributed, and parallel systems and perhaps even upgrade the algorithms. Consequently, innovation will be used to aid in the classification of micro-array gene expression information.

**Declarations**

**Conflict of interest** Author declares that they have no conflict of interest.
References

Abualigah L, Diabat A, Mirjalili S, AbdElaziz M, Gandomi AH (2021a) The arithmetic optimization algorithm. Comput Methods Appl Mech Eng. https://doi.org/10.1016/j.cma.2020.113609
Abualigah L, Youusi D, AbdElaziz M, Ewees AA, Al-eganess MAA, Gandomi AH (2021b) Aquila optimizer: a novel meta-heuristic optimization algorithm. Comput Ind Eng 157:107250. https://doi.org/10.1016/j.cie.2021.107250
Abualigah L, Ali D, AbdElaziz M (2021) Intelligent workflow scheduling for Big Data applications in IoT cloud computing environments. Cluster Comput. https://doi.org/10.1007/s10586-021-03291-7
Abualigah L, Ali D (2021) Advances in sine cosine algorithm: a comprehensive survey. Artif Intell Rev 54:2567–2608. https://doi.org/10.1007/10109-3008(84)90020-7
Abualigah L, Alkhrabsheh M, AbualigahAligah L, Abualigah L, Alkhrabsheh M (2021) Amended hybrid multi-verse optimizer with genetic algorithm for solving task scheduling problem in cloud computing task scheduling Multi-verse optimizer genetic algorithm Hybrid method. J Supercomput. https://doi.org/10.1007/s11227-021-03915-0
Bezdek JC, Ehrlich R, Full W (1984) FCM: the fuzzy c-means clustering algorithm. Comput Geosci 10:191–203. https://doi.org/10.1016/0098-3004(84)90020-7
Cai J, Wei H, Yang H, Zhao X (2020) A novel clustering algorithm based on DPC and PSO. IEEE Access 8:88200–88214. https://doi.org/10.1109/access.2020.2992903
Chao Xu, Zhang P, Li B, Dinghai Wu HF (2013) Vague C-means clustering algorithm - sciencedirect. Pattern Recognit Lett 34:505–510
Chen JY, He HH (2015) Research on density-based clustering algorithm. IEEE Trans Fuzzy Syst 23:539–555. https://doi.org/10.1109/TFUZZ.2018.2796074
Dua D, Graff C (2017) {UCI} Machine learning repository
Eberhart R, Kennedy J (1995) New optimizer using particle swarm theory. In: proceedings of the international symposium on micro machine and human science
Ester M, Kriegel H-P, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: proceedings of the 2nd international conference on knowledge discovery and data mining
Gentle JE, Kaufman L, Rousseeuw PJ (1991) Finding groups in data: an introduction to cluster analysis. Biometrics 47(2):788
Han J, Kamber M, Pei J (2012) Data mining: concepts and techniques
Hou J, Gao H, Li X (2016) DSets-DBSCAN: a parameter-free clustering algorithm. IEEE Trans Image Process. https://doi.org/10.1109/TIP.2016.2559803
Jahan M, Hasan M (2019) Performance analysis and benchmarking of clustering algorithms with gene datasets. In: 1st international conference on advances in science, engineering and robotics technology 2019, ICASERT 2019
Jahan M, Hasan M (2020) A novel fuzzy clustering approach for gene classification. Int J Adv Comput Sci Appl. https://doi.org/10.14569/IJACSA.2020.0110809
Lei T, Jia X, Zhang Y, He L, Meng H, Nandi AK (2018) Significantly fast and robust fuzzy C-means clustering algorithm based on morphological reconstruction and membership filtering. IEEE Trans Fuzzy Syst 26:3027–3041. https://doi.org/10.1109/TFUZZ.2018.2796074
Ng RT, Han J (1994) Efficient and effective clustering methods for spatial data mining. In: Proceedings 20th international conference very large data bases pp. 144–155
Reddy GT, Khare N (2017a) Hybrid firefly-bat optimized fuzzy artificial neural network based classifier for diabetes diagnosis. Int J Intell Eng Syst 10:18–27. https://doi.org/10.22266/ijies2017.0831.03
Reddy GT, Khare N (2017b) An efficient system for heart disease prediction using hybrid OFBAT with rule-based fuzzy logic model. J Circuits Syst Comput 26:1–21. https://doi.org/10.1142/S021812661750061X
Reddy GT, Reddy MPK, Lakshmana K, Rajput DS, Kaluri R, Srivastava G (2020) Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. Evol Intell 13:185–196. https://doi.org/10.1007/s10645-019-00327-1
Rodriguez A, Laio A (2014) Clustering by fast search and find of density peaks. Science. https://doi.org/10.1126/science.1242072
Rousseau J-P (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53–65. https://doi.org/10.1016/0377-0427(87)90125-7
Sinaga KP, Yang MS (2020) Unsupervised K-means clustering algorithm. In: remote sensing Proc vol J 57:539–555. https://doi.org/10.1007/978-81-322-2208-8
Srivastava G (2020) Artificial neural network based classifier for diabetes diagnosis. Int J Adv Comput Sci Appl. https://doi.org/10.1109/access.2020.2988796
Tilson L V., Excell PS, Green RJ (1988) A generalisation of the Fuzzy c-Means clustering algorithm. In: remote sensing Proc IGRASS ’88 Symposium Edinburgh, vol 3, pp. 1783–1784. doi: https://doi.org/10.1109/igars.1988.569600
Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. IEEE Trans Evol Comput. https://doi.org/10.1109/4235585893

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.