IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 1, March 2023, pp. 180~188
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i1.pp180-188

Fuzzy C-means clustering on rainfall flow optimization technique for medical data

Antony Jaya Mabel Rani[1], C. Srivenkateswaran[1], M. Rajasekar[2], M. Arun[3]
[1]Department of Computer Science and Engineering, Kings Engineering College, Chennai, India
[2]Institute of Computer Science and Engineering, Saveetha School of Engineering, Chennai, India
[3]Department of Electronics and Communication Engineering, Panimalar Institute of Technology, Chennai, India

Article Info

Article history:

Received Jul 28, 2021
Revised Jul 27, 2022
Accepted Aug 25, 2022

Keywords:

Accuracy
Clustering
Flood optimization
Fuzzy C-means
Rainfall flow

ABSTRACT

Due to various killing diseases in the world, medical data clustering is a very challenging and critical task to handle and to take the proper decision from multidimensional complex data in an effective manner. The most familiar and suitable speedy clustering algorithm is K-means than other traditional clustering approaches. But K-means is extra sensitive for initialization of clustering centroid and it can easily surround. Thus, there is a necessity for faster clustering with an effective optimum clustering centroid. Based on that, this research paper projected an optimization-based clustering by hybrid fuzzy C-means (FCM) clustering on rainfall flow optimization technique (RFFO), which is the normal flow and behavior of rainfall flow from one position to another position. FCM clustering algorithm is used to cluster the given medical data and RFFO is used to produce optimum clustering centroid. Finally, the clustering performance is also measured for the proposed FCM clustering on RFFO technique with the help of accuracy, random coefficient, and Jaccard coefficient for medical data set and find the risk factor of a heart attack.

Corresponding Author:

Antony Jaya Mabel Rani
Department of Computer Science and Engineering, Kings Engineering College
Chennai, Tamil Nadu, India
Email: jayamabelrani@kingsedu.ac.in

1. INTRODUCTION

In a recent trend, an optimization-based clustering algorithm is used to cluster complex problems in various environments in different situations. This clustering algorithm is mainly projected on clustering the data by nearest neighboring of flow depth by the natural behavior and characteristic flow of rainwater. This hybrid optimization-based clustering algorithm also used some mathematical calculations for updating the next nearest neighbor location and velocity from one position to another position. Here land space location is used to locate all the drops of rainfall, which is considered as land space data (LSD). All drops are considered together as a flood, that is total data in the located land space (dataset). It can move from its current position to another position based on the condition of the location such as a river, drains, pond, lake, and some other storage locations.

Optimization based algorithms such as sun flower optimization (SFO) [1], rider optimization algorithm (ROA) [2], gray wolf optimization (GWO) [3], particle swarm optimization (PSO) [4] and genetic algorithm [5], are very powerful algorithms in machine learning under data mining, that is the sub branch of artificial intelligence [6], [7]. There are three different models of machine learning such as unsupervised learning, supervised learning and reinforcement learning. Depends on the input and some other features it is defined as supervisor or without supervisor or situation like failure or success [8]. Then another form of machine learning is called semi-supervised, that is the combination of with supervisor (classification) and without supervisor (clustering) [9]–[11].

The main inspiration of designing this novel proposed data clustering algorithm is to produce better optimum clustering solutions with faster convergence. To produce an optimum clustering solution this paper proposed fuzzy-C-means (FCM) based on rainfall flow optimization (RFFO) clustering algorithm for medical data. In this problem, there are three steps; i) data preprocessing, ii) feature selection and iii) data clustering by a proposed data clustering algorithm.

The proposed solution is designed based on two important algorithmic concepts. First to produce an optimum clustering solution here used FCM clustering. Then RFFO is used to find optimum clustering centroid. FCM clustering algorithm [12], [13] is the most popular clustering algorithm based on a mathematical logical model. RFFO [14] approach works based on the natural rainfall flow behavior. The open land source (OLS) is suitable for locating raindrops, which is referred to as OLSD (open land source data) [15], [16]. The raindrops are poised together and well-thought-out as a torrent, that is total fed data [17]. In some situations, and places, the water dew will not move to any other places, which is called data stagnation. When the raining, the collected raindrops (water) flow from one position to another position and get together as flood, which is called as clustering of data. The stowage space jerks raindrops by its least distance and slant by its acceleration. The total fed data (collected raindrops) is called the torrent. The water stowage is based on the minimum distance of stowage location, max depth, maximum storage size, condition of stowage location such as soil, climate, nature of wetlands, and artificial lake. Today real-world applications have various types with heterogeneous data sets with dissimilar features [18]. For solving all these complex problems this paper presented a novel optimization-based clustering algorithm, which is the FCM based RFFO algorithm. The main challenges of medical data clustering are to handle data preprocessing to find missing data, noise data, data inconsistency, and redundant data in data mining [19]. The visual representation of the projected FCM clustering algorithm based on RFFO technique for medical data is shown in Figure 1. Today medical data clustering is very vast and intricate, due to the large size of receiving data, hidden information, massive volume, and its most frequency.

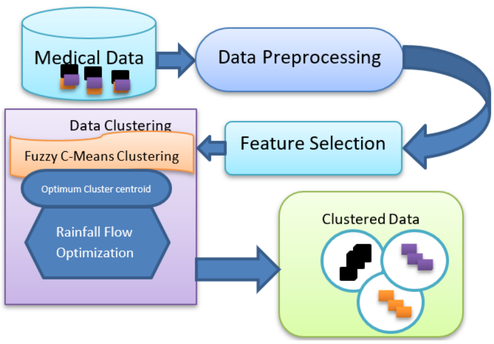


Figure 1. Visual representation for FCM on RFFO algorithm

2. LITERATURE REVIEW

In this literature review given five different latest existing medical data-based clustering methods with its drawback. In 2018 Yelipe *et. al* designed an imputation based on class-based clustering, which is simply called IM-CBC to identify and evaluate the similarity between the two medical records. This paper used Euclidean distance to find the similarity between the clusters with fuzzy similarity functions. Then, classification is also done with the help of classification methods, such as SVM, C4.5, or k-nearest neighbors (KNN). Here the performance is given as high accuracy. At the same time, this method is not examined fuzzy measures for data classification and predicting the results based on given medical data [20]. Then in 2018 Das *et al* proposed a modified bee colony optimization (MBCO) technique for clustering the medical data with the combination of K-means clustering algorithm with chaotic theory for faster convergence. This method compared with other clustering methods and shown MBCO produced faster convergence. But this hybrid method does not adapt for multi-objective functions and is not processed for high-frequency data streams [21]. Then in 2019 Chauhan *et al.* given a two-step clustering technique to analyze the patient's disorders by using different variables to and determine the earlier stage of the liver disease from the hidden knowledge [22]. In 2020 Yu *et.al.* [23] designed medical data clustering and feature extraction by using immune evolutionary algorithm under cloud computing for big data. Here the final results produced the better accuracy of data classification, improve the performance for medical data. Here the final results produced the better clustering solution by optimum clustering centroid. Then this algorithm is also compared with the existing algorithm [14].

3. THE PROPOSED CLUSTERING METHOD

By getting advantages of both traditional clustering and global optimization-based technique for optimum centroid here, proposed hybrid FCM based on RFFO techniques for medical data. RFFO can produce optimum clustering cenroid. Based on the optimized clustering centroid FCM clustering algorithm can produce better clustering solution.

3.1. Introduction about clustering

In clustering, the collection of data items is grouped into a set of disjoint classes, which is the sub branch of unsupervised learning in machine learning [24], [25]. Here are different forms of clustering algorithms from traditional clustering algorithms to global optimization-based clustering techniques. This paper used optimization-based clustering by using hybrid FCM based RFFO clustering algorithm.

3.2. Fuzzy C-means clustering

FCM clustering algorithm is also simple clustering under fuzzy logic. (It can have the value 0 and 1), that is mathematical logical model-based partitioning clustering [26]. The core objective of the FCM algorithm is the minimum cost function *OFCM* using Euclidean distance by (1).

$$O_{FCM} = O(W,V) = \sum_{i=1}^{X}\sum_{j=1}^{C}(W_{ij})^d \left\|B_i - C_j\right\|^2 \quad (1)$$

Where fuzzification degree, here $i=1, \ldots, X$, and $j=1, \ldots, C$ that is membership matrix. Then B_i is the *ith* dimension of the given data, and *the jth* dimension of the cluster center is C_j. Then the cluster center will be updated by (2);

$$C_j = \frac{\sum_{i=1}^{X} W_{ij}^d \cdot B_i}{\sum_{i=1}^{X} W_{ij}^d} \quad (2)$$

Here the updated fuzzy membership matrix is calculated by (3);

$$W_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\|B_i - C_j\|}{\|B_i - C_k\|}\right)^{2/d-1}} \quad (3)$$

3.3. Pseudocode for FCM clustering

Consider $B = \{B_1, B_2, B_3, \ldots B_n\}$ for the data point sets and the cluster centers $C=\{C_{e1}, C_{e2}, C_{ej}\}$. Initially the centers of each cluster are selected randomly, then the fuzzy membership value will be the computer, W_{ij} by (3), calculate center Cj for the fuzzy cluster center by (2), reiterate steps 2 and 3 till the defined number of iterations or if it is less than the given threshold value or there is no improvement.

$$\left\|(W_{ij})^{k+1} - (W_{ij})^k\right\| < \mathcal{E} \quad (4)$$

Here, the iteration step k, then the expiry condition. This FCM iteration stops when the value of the partition matrix is less than, which is definite as 0.0001. FCM is a little more beneficial than K-means. But it has also some shortcomings than global optimization-based clustering. The drawback of the FCM is a sensibility for initialization of cluster centroid and premature convergence.

3.4. Procedure for proposed hybrid FCM based RFFO

FCM depend on the primary membership matrix values. Based on probability distribution, the candidate data is selected, which is performed by random initialization. The algorithmic steps for hybrid FCM based RFFO is shown as;

- Step 1: Initialization. Initialize Fmax, Maximum iteration numbers (Itmax) acceleration coefficient *AC1* and, *AC2*, Flood best ($F_b = \propto$), Depth best ($D_b = \propto$). Then initial cluster centroids will be selected randomly,

$$C = \{C_{e1}, C_{e2}, \dots C_{ej}\} \quad (5)$$

– Step 2: Evaluate fitness function. Here each and every iteration does, calculate the fitness function Ffit by using (6).

$$F_{fit} = O_{FCM}/K \quad (6)$$

O_{FCM} is calculated by (1), that is the objective function, then total number of clusters are represented by "K" for FCM algorithm, with minimum cost **n** with the help of distance measure as Euclidean distance formula. Again, centroid will be calculated for cluster by using (2); For finding optimum centroid of the cluster use RFFO using (6).

– Step 3: Velocity and position updation. By (7) and (8) RFFO's position and velocity are updated. The updated position is calculated by,

$$P_i(t+1) = P_i(t) + Y_i(t+1) \quad (7)$$

Then the updated velocity is calculated by;

$$Y_i(t+1) = Y_i(t) + C1 * \lambda 1(Db - Pi(t)) + C2 * \lambda 2(Fb - Pi(t)) \quad (8)$$

Here $Y_i(t)$ is computed by using the (9);

$$Y_i(t) = (\text{Kci} * \Delta\text{x}) / \text{WP} \quad (9)$$

The gradient Δx**,** that is M, computed by the (10);

$$M = (a2 - a1)/(b2 - b1) \quad (10)$$

Here there are two slope coordinates such as *(a1, b1)* and *(a2, b2)*, then $P_i(t)$ is the present particle's position at *t.* Particle's next position will be updating $P_i(t+1)$ at *(t+1*). W_P is water absorbency, that is overall suckled data. D_b denotes personal best location. Then F_b is the globally best solution. The hydraulic conductivity (K_{ci}) value from 0.8 to 0.95, and the capillary constant *AC1*, *AC2* coefficient, which is *2.0* and the values of random variables ranging from X_1 *to* X_n considered from 0 to 1.

– Step 4: Defining the optimum centroid by RFFO. RFFO method is used to define optimum clustering centroid to produce better clustering solution.
– Step 5: Termination Condition. Iterate the steps 2 to 3 until the extreme or determined number of iteration count reaches.

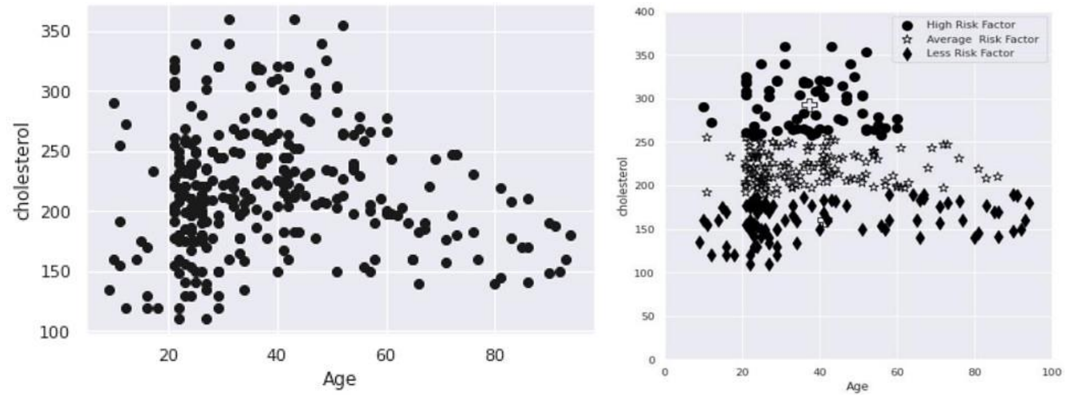
4. RESEARCH METHOD

The research method is designed based on two important algorithmic concepts. First to produce an optimum clustering solution here used FCM clustering. Then RFFO is used to find optimum clustering centroid. Today real-world applications have various types with heterogeneous data sets. For solving all these complex problems this paper presented a novel optimization-based clustering algorithm, which is the FCM based RFFO algorithm. The main problem of medical data clustering are to handle data preprocessing to find missing data, noise data, data inconsistency, and redundant data in data mining. This fuzzy clustering is implemented using the Python 3.8.6 in Windows 10 operating system, intel i5 core processor. For this experimentation taken 300 persons real medical checkup data to predict the symptoms of heart disease. The experimental results also compared with existing methods and shown the performance measure based on accuracy, Jaccard coefficient and random coefficient.

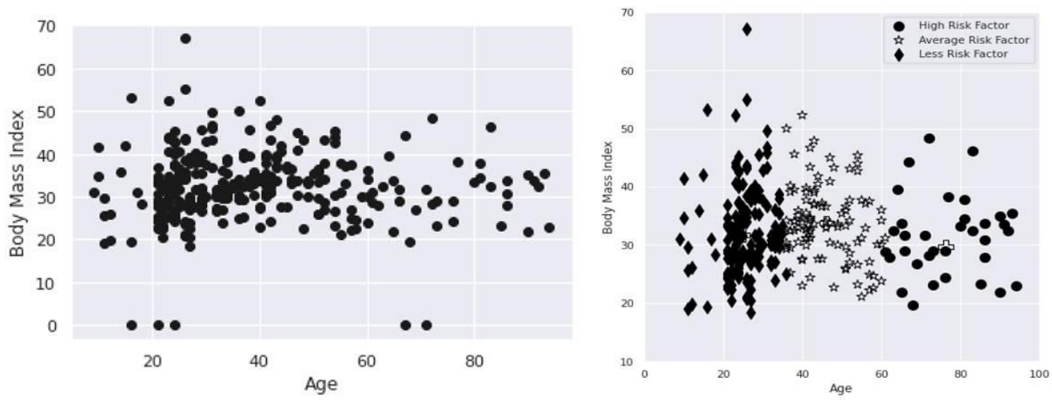
5. EXPERIMENTAL RESULTS AND DISCUSSION

For the medical data clustering, heart disease data has been taken for experimentation to analyze and forecast the risk factors of heart disease. Heart disease data were collected from Johnson Jims, a staff nurse from Kuwait based on the reference. From this, we can provide various suggestions for each type of clustering. For low symptoms of heart disease, we can provide suggestions to take healthy food, and doing exercise then average risk factors, can provide the suggestions such as food diet, walking distance, exercise to do, and any medicine to take. Then for high risk of heart disease can suggest more concentrate on a food diet, walking, regular medical checkup, and exercise to do. Figure 2 detailed about 2-dimensional (2D) view of clustering results for cholesterol vs age, body mass index vs age and glucose level vs age. Then in Figures 2(a) to 2(c) are shown the simulation results in the 2D model using FCM based RFFO with different features of medical data in the sense early phase and ending phase of the clusters.

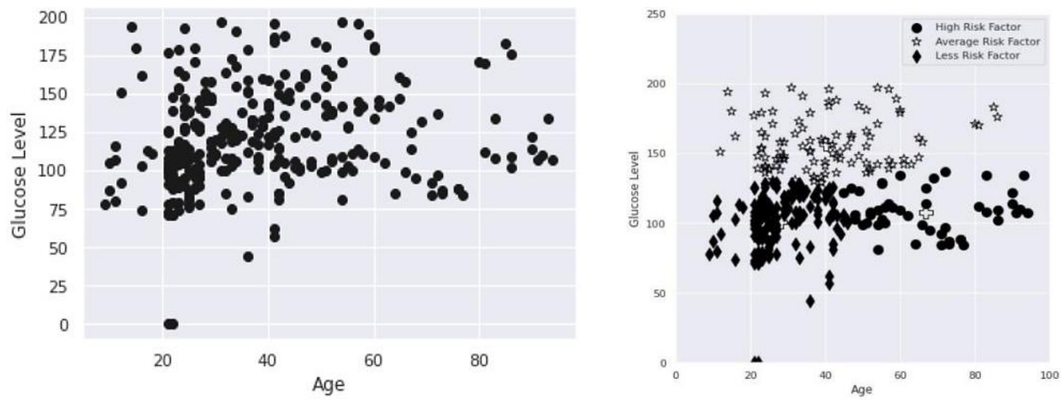
The Figure 2(a) shows the 2D simulation result for age vs cholesterol. Here green color shows less symptoms of heart disease risk factor, then the blue color denotes the average risk factor of heart disease and finally the red color shows the high-risk factor of heart disease. Similar that the Figure 2(b) shows the 2D simulation result for age vs body mass index, and the Figure 2(c) figured for age vs glucose level in the form of 2D simulation.



(a)



(b)

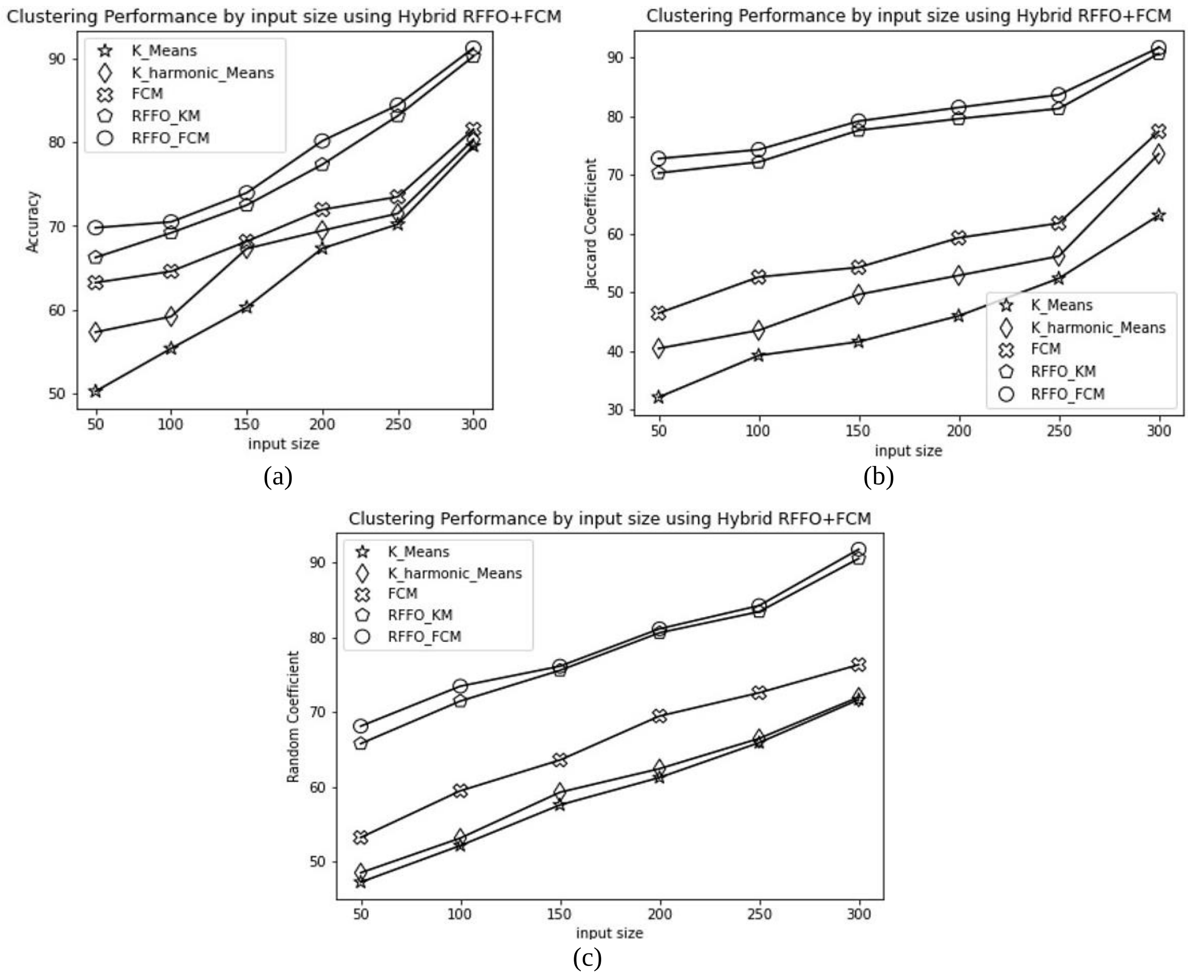


(c)

Figure 2. 2D clustering result for (a) age vs cholesterol, (b) age vs body mass index, and (c) age vs glucose level using FCM based RFFO

5.1. Comparative study analysis

Figures 3(a) to 3(c), are shown the qualified comparative study analysis by using input size for the above performance measures. The input size of the comparative study analysis is varying from 50 to 300. The study analysis is shown based on accuracy measure Jaccard coefficient and random coefficient. When input size is 50, the corresponding accuracy values are computed by existing K-means, K-harmonic means (KHM), FCM, K-means+RFFO and proposed RFFO-based FCM. Likewise, accuracy, Jaccard coefficient and random coefficients are also calculated for the input size 300.



(c)

Figure 3. The comparative study analysis for the performance measures (a) accuracy b) Jaccard coefficient, and (c) Rand coefficient based on input size

Here the Figure 3(a) shows the comparative study analysis by using accuracy measure, which is computed for existing models such as K-means, KHM, FCM, RFFO+K-means and proposed FCM based RFFO at the input size 50 are 50.263%, 57.327%, 63.23%, 66.2387% and 69.748 respectively. Similar that, accuracy is calculated by using existing K-means, KHM, FCM, RFFO+K-means and proposed FCM based RFFO at the input size 300 are 79.545%, 80.321%, 81.532%, 90.234% and 91.234% respectively. The Figure 3(b) shows the comparative study analysis by using Jaccard coefficient, which is computed for existing models such as K-means, KHM, FCM, RFFO+K-means and proposed FCM based RFFO at the input size 50 are 32.123%, 40.437%, 46.438%, 70.297% and 72.748% respectively. Similar that, Jaccard coefficient is calculated by using existing K-means, KHM, FCM, RFFO+K-means and proposed FCM based RFFO at the input size 300 are 63.09%, 73.555%, 77.3825%, 90.626% and 91.614% respectively. The Figure 3(c) shows the comparative study analysis by using random coefficient, which is computed for existing models such as K-means, KHM, FCM, RFFO+K-means and proposed FCM based RFFO at the input size 50 are 47.231%, 48.487%, 53.208%, 65.767% and 68.101% respectively. Similar that, random coefficient is calculated by using existing K-means, KHM, FCM, RFFO+K-means and proposed FCM based RFFO at the input size 300 are 71.653%, 71.985%, 76.326%, 90.534% and 91.767% respectively.

5.2. Performance metric

The performance measure for the hybrid FCM based RFFO clustering algorithm is employed by accuracy, random coefficient, and Jaccard coefficient which are given in the following section, by using accuracy data quality can be calculated from true positives ACT^p, true negatives ACT^n, false positives ACF^p and false negatives ACF^n;

$$\text{Accuracy} = \frac{ACT^p + ACT^n}{ACT^p + ACT^n + ACF^p + ACF^n} \tag{11}$$

Here ACT^p , ACT^n , ACF^p, ACF^n are the parameters.

Jaccard coefficient measure is used to find similarities by comparing two data clusters;

$$\text{Jack}(U, V) = \frac{|U \cap V|}{|U \cup V|} \tag{12}$$

Here U and V are two different clusters.
Random coefficient is the third performance measure, which is used to find the ratio of correct decision. Rand coefficient is calculated to estimate the right clustered pairs and the equation to compute and coefficient is as,

$$\text{Random Coefficient} = \frac{Correct\ similar\ pairs + Correct\ dissimilar\ pairs}{Total\ possible\ pairs} \tag{13}$$

5.3. Comparative analysis table based on performance measure

The given Table 1 analyze the above three performance measures. The maximal accuracy Jaccard coefficient and random coefficient for the proposed FCM based RFFO are 91.234%, 89.614%, and 92.767%. Here the maximal accuracy is acquired by proposed RFFO-based FCM with accuracy of 91.234%, whereas the accuracy of existing K-means, KHM, FCM and K-means based RFFO are 79.545%, 80.231%, 81.534% and 90.166% respectively. Likewise, the input size for the maximal Jaccard coefficient and random coefficient also given in that Table 1.

Table 1. Comparative analysis

Input	Comparative metrics	K-means	KHM	FCM	KM+RFFO	FCM+RFFO
Input size	Accuracy (%)	79.545	80.321	81.534	90.166	91.234
	Jaccard coefficient (%)	63.09	73.555	77.382	87.626	89.614
	Random coefficient (%)	71.653	71.985	76.326	90.534	92.767

6. CONCLUSION

Thus, the paper proposed an optimization-based clustering algorithm with the name of a hybrid RFFO based FCM clustering algorithm for medical data. Here heart disease-based medical data has been taken and projected the model with optimal data clustering. The final data clustering was done by FCM based RFFO algorithm for medical data. The proposed success is achieved for FCM based RFFO algorithm with maximal accuracy 91.234%, Jaccard coefficient 89.614% and Rand coefficient 92.767%. The main advantages of this hybrid optimization-based clustering algorithm combine the advantages of both algorithms like fast convergence of traditional clustering FCM algorithm and to produce better centroid by using optimization-based method RFFO. So, this hybrid algorithm can avoid premature convergence and it can also produce optimum centroid. In the future, this model can be extended by multi-objective functions for a more effective and better clustering centroid. This will help the doctors to take proper decisions from the immense needs with huge data size.

REFERENCES

[1] G. F. Gomes, S. S. da Cunha, and A. C. Ancelotti, “A sunflower optimization (SFO) algorithm applied to damage identification on laminated composite plates,” *Engineering with Computers*, vol. 35, no. 2, pp. 619–626, May 2019, doi: 10.1007/s00366-018-0620-8.
[2] D. Binu and B. S. Kariyappa, “RideNN: a new rider optimization algorithm-based neural network for fault diagnosis in analog circuits,” *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 1, pp. 2–26, Jan. 2019, doi: 10.1109/TIM.2018.2836058.
[3] A. N. Jadhav and N. Gomathi, “DIGWO: hybridization of dragonfly algorithm with improved grey wolf optimization algorithm for data clustering,” *Multimedia Research*, vol. 2, no. 3, pp. 1–11, Jul. 2019, doi: 10.46253/j.mr.v2i3.a1.

[4] N. Zemmal *et al.*, "Particle swarm optimization-based swarm intelligence for active learning improvement: application on medical data classification," *Cognitive Computation*, vol. 12, no. 5, pp. 991–1010, Aug. 2020, doi: 10.1007/s12559-020-09739-z.

[5] B. Gao, X. Li, W. L. Woo, and G. Y. Tian, "Physics-based image segmentation using first order statistical properties and genetic algorithm for inductive thermography imaging," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2160–2175, May 2018, doi: 10.1109/TIP.2017.2783627.

[6] S. N. Ghazavi and T. W. Liao, "Medical data mining by fuzzy modeling with selected features," *Artificial Intelligence in Medicine*, vol. 43, no. 3, pp. 195–206, Jul. 2008, doi: 10.1016/j.artmed.2008.04.004.

[7] A. Jaya Mabel Rani and A. Pravin, "Multi-objective hybrid fuzzified PSO and fuzzy C-means algorithm for clustering CDR data," in *Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing, ICCSP 2019*, Apr. 2019, pp. 94–98, doi: 10.1109/ICCSP.2019.8698080.

[8] J. Han, M. Kambar, and J. Pei, *Data mining: concepts and techniques*, 3rd Ed. Elsevier Inc., 2012.

[9] X. Li and N. Ye, "A supervised clustering and classification algorithm for mining data with mixed variables," *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, vol. 36, no. 2, pp. 396–406, Mar. 2006, doi: 10.1109/TSMCA.2005.853501.

[10] B. S. Chandana, K. Srinivas, and R. K. Kumar, "Clustering algorithm combined with hill climbing for classification of remote sensing image," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 4, no. 6, pp. 923–930, Dec. 2014, doi: 10.11591/ijece.v4i6.6608.

[11] Y. T. Quek, W. L. Woo, and L. Thillainathan, "IoT load classification and anomaly Wwarning in ELV DC picogrids using hierarchical extended k-nearest neighbors," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 863–873, Feb. 2020, doi: 10.1109/JIOT.2019.2945425.

[12] G. Parthasarathy and D. C. Tomar, "A novel approach for mining frequent itemsets in medical image databases," *International Journal of Pharmacy and Technology*, vol. 8, no. 3, pp. 18126–18135, Sep. 2016.

[13] Y. Liu, K. Xiao, A. Liang, and H. Guan, "Fuzzy C-means clustering with bilateral filtering for medical image segmentation," in *Proceeding of International Conference on Hybrid Artificial Intelligence Systems (HAIS 2012)*, vol. 7208 LNAI, no. PART 1, Springer Berlin Heidelberg, 2012, pp. 221–230.

[14] A. Jaya Mabel Rani and A. Pravin, "Rainfall flow optimization based K-means clustering for medical data," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 17, Sep. 2021, doi: 10.1002/cpe.6308.

[15] Y. Qin, "Urban flooding mitigation techniques: a systematic review and future studies," *Water (Switzerland)*, vol. 12, no. 12, p. 3579, Dec. 2020, doi: 10.3390/w12123579.

[16] A. K. Saini, S. S. Chauhan, and A. Tiwari, "Creeping flow of Jeffrey fluid through a swarm of porous cylindrical particles: Brinkman–Forchheimer model," *International Journal of Multiphase Flow*, vol. 145, 2021, doi: 10.1016/j.ijmultiphaseflow.2021.103803.

[17] H. Ma and D. Simon, "Analysis of migration models of biogeography-based optimization using Markov theory," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 6, pp. 1052–1060, Sep. 2011, doi: 10.1016/j.engappai.2011.04.012.

[18] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, 1995, pp. 1942–1948, doi: 10.1109/icnn.1995.488968.

[19] Y. Zhong, S. Zhang, and L. Zhang, "Automatic fuzzy clustering based on adaptive multi-objective differential evolution for remote sensing imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 5, pp. 2290–2301, Oct. 2013, doi: 10.1109/JSTARS.2013.2240655.

[20] U. R. Yelipe, S. Porika, and M. Golla, "An efficient approach for imputation and classification of medical data values using class-based clustering of medical records," *Computers and Electrical Engineering*, vol. 66, pp. 487–504, Feb. 2018, doi: 10.1016/j.compeleceng.2017.11.030.

[21] P. Das, D. K. Das, and S. Dey, "A modified bee colony optimization (MBCO) and its hybridization with K-means for an application to data clustering," *Applied Soft Computing Journal*, vol. 70, pp. 590–603, Sep. 2018, doi: 10.1016/j.asoc.2018.05.045.

[22] R. Chauhan, N. Kumar, and R. Rekapally, "Predictive data analytics technique for optimization of medical databases," in *Advances in Intelligent Systems and Computing*, vol. 742, Springer Singapore, 2019, pp. 433–441.

[23] J. Yu, H. Li, and D. Liu, "Modified immune evolutionary algorithm for medical data clustering and feature extraction under cloud computing environmentjournal of healthcare engineering," *Journal of Healthcare Engineering*, vol. 2020, pp. 1–11, Jan. 2020, doi: 10.1155/2020/1051394.

[24] A. Al-Shammari, R. Zhou, M. Naseriparsaa, and C. Liu, "An effective density-based clustering and dynamic maintenance framework for evolving medical data streams," *International Journal of Medical Informatics*, vol. 126, pp. 176–186, Jun. 2019, doi: 10.1016/j.ijmedinf.2019.03.016.

[25] V. Kumar, "Implementation of data mining techniques for information retrieval," National University of Science and Technology "NSIT-MISIS," Jaipur, india, 2018.

[26] Y. Zhang, Y. Du, X. Li, S. Fang, and F. Ling, "Unsupervised subpixel mapping of remotely sensed imagery based on fuzzy C-means clustering approach," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 5, pp. 1024–1028, May 2014, doi: 10.1109/LGRS.2013.2285404.

BIOGRAPHIES OF AUTHORS

Antony Jaya Mabel Rani received the M.E degree in Computer Science and Engineering from, Rajarajeshwari Engineering college, Anna University, Chennai, India in 2007. She is pursuing her Ph.D in Computer Science and Engineering at Sathyabama University Chennai, India. She has 18 Years of teaching experience. She has participated and presented many Research Papers in International and National Conferences and also published papers in International and National Journals. Her area of interests includes artificial intelligence, data mining, machine learning, deep learning and big data. She can be contacted at email: ajayamabelrani@gmail.com.

Dr. C. Srivenkateswaran received the B.E degree in Electronics & Communication Engineering from University of Madras, Chennai, India in 1996, M.E degree in Computer and Communication Engineering from Anna University, Chennai, India in 2009 and Ph.D degree in Information and Communication Engineering at Anna University, Chennai, India in 2019. He works currently as an Associate Professor for the Department of Computer Science and Engineering at Kings Engineering College; Chennai and he has 23 Years of teaching & Industries experience. He has participated and presented many Research Papers in International and National Conferences and also published papers in International and National Journals. His area of interests includes cyber security, ethical hacking and big data. He can be contacted at email: srivenkateswaran@kingsedu.ac.in.

Dr. M. Rajasekar received the B.E degree in Computer Science and Engineering from Sardhar Raja College Engineering, Anna University, Chennai, India in 2007, M.E degree in Computer Science and Engineering from Veltech engineering college, Anna University, Chennai, India in 2010 and Ph.D degree in Information and Communication Engineering at Anna University, Chennai, India in 2021. He works currently as an Associate Professor for the Department of Computer Science and Engineering at Institute of Computer Science and Engineering, Saveetha School Of Engineering, SIMATS, Chennai and he has 13 Years of teaching experience. He has participated and presented many Research Papers in International and National Conferences and also published many papers in International and National Journals. His area of interests includes image processing, artificial intelligence, machine learning and network. He can be contacted at email: mrajasekarcse@gmail.com.

Mr. M. Arun received his M.E., Degree in Applied Electronics from College of Engineering Guindy, Anna University Chennai. He is pursuing his Ph.D in Antenna Domain at sathyabama Institute of Science and Technology, India. He is working as an Assistant Professor ECE, Panimalar Institute of Technology. He is also serving as Executive Committee member of IETE Chennai Centre. He is serving as Vice Chairman of IEEE Madras YP & IEEE TEMS society, Secretary in IEEE EMC society, Treasurer, IEEE COMSOC Madras Chapter & Ex-com Member of IETE Chennai Center. He has got 13 Years of experience in Teaching Field. He received IEEE Outstanding Student Branch Counselor and Branch Chapter Advisor Award from IEEE MGA for the year 2019. He also received best student branch counselor award for the year 2015-20 from IEEE Madras Section. He can be contacted at email: arunmemba@ieee.org.