# Clustering by Hybrid K-Harmonic Means and Biogeography Based Optimization Algorithm for Medical Data

A. Jaya Mabel Rani, A. Pravin

***Abstract***: ***In today trendy world hybrid based optimized data clustering is unique and imperative clustering tool in the area of data mining, which is dynamic research of actual creation problems. The oldest and furthermost commonly used popular clustering technique is the K-means(KM) algorithm, which is very complex and for the initialization of the cluster centroid and it will easily go for premature converge. This initialization problem of K-means can be evaded by built in boost function of K-Harmonic Means, which is centroid based clustering algorithm and also unresponsive for collection of initial partition clustering , but it can easily go for pre-matured conjunction in local optima. To avoid this convergence problem, this proposed algorithm uses Boosting K-harmonic means(KHM) algorithm with BBO to produce more precise, robust, better clustering solution in few number of iterations, evade conning in local optima and simply convergence to relate with Harmonic Means, BBO algorithms. Biogeography based algorithm works with the concept of emigration and immigration of inhabitants from one location to another location, Which has high computation cost. For avoiding this high computation cost in this hybrid optimization technique Biogeography-Based Optimization (BBO) is integrated with K-Harmonic means algorithm to produce optimum and effective clustering solution with faster convergence. BBO is universal optimization methods to solve utmost of the optimization problem, which is an production based generation of evolutionary algorithm (EA)that augments a function by stochastically and re peatedly improving the clustering solution of quality, or fitness function. The experimental results of this paper shown as the projected method is very resourceful and faster to afford better clustering solution in less number of repetitions for medical data.***

***Keywords : BBO, Clustering, Data mining, K-means, Optimized intelligent tool.***

## I. INTRODUCTION

In current movement there is definite and increasing attention towards medical health data clustering due to more different types of disease in this world. In future all over world people will be affected by more killing disease and skin disease due to high radiation of towers. So we are need of proper data clustering algorithm to predict the proper clustering solution.

**Revised Manuscript Received on April 30, 2020.**
***** Correspondence Author
**A. Jaya Mabel Rani***, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India. ajayamabelrani@gmail.com
**Dr. A. Pravin**, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India. Email: xyz2@blueeyesintlligence.org



For resolving this type of real or actual time problems or huge set of data the prevailing algorithms (K_means and its variance algorithms) do not support.

This may be gone for early convergence or stagnation or inactivity of data. K-means [1], [2] algorithm is recorded prevalent and

straightforward clustering algorithm. This can be easily implemented and applied in most of the environment. But it is very complex for starting of initialization of initial seed data, which is not suitable for enormous data, its complication cultivates in linear manner. In order to avoid initialization problem here used K-harmonic means and to avoid premature convergence here used BBO algorithm with K-Harmonic means algorithms. Anyway K-Harmonic means is sooner than BBO algorithm [3].

The full detailed description of this paper designed in the succeeding manner. The related work about this paper is analyzed in the section of 2. In the part of 3 and 4 explicated the K-means and boosting K-Harmonic Means algorithm, in section 5 offerings BBO Algorithm, in section 6 suggests KHM+BBO algorithm, and in section 7 described tentative medical data sets and clustered results. In the last and final segment 8 completes with the work of this paper.

### A. Clustering Concept

Clustering is the main area of unsupervised learning of data mining[4]. Organization of documents, topic extraction in automatic manner [5], retrieval of formation from large storage of data [6], pattern recognition are also main area of data mining, which is main area of machine learning. Clustering is the main area of unsupervised based machine learning, which is the process of grouping or subdividing the data objects with from top to bottom resemblance, but are same distinct to further clusters. Commonly the clustering algorithms divided into different types based on the application, types of data, size of data, etc, which are earliest partitioning clustering methods, grid based clustering methods, density based clustering methods, hierarchical clustering methods and model based clustering methods, fuzzy based classification[17]-[20] or hybrid clustering methods[13]-[16]. Here found partitioned clustering for huge dataset due to their less cost supplies. K-Means and its variants clustering algorithms are the best-known and popular partitioning clustering algorithm [9], due to its easily implementation and efficiency in terms of low computation and execution cost. The main shortcoming of this local optimization of K-Means algorithm is premature convergence in less number of iterations. In order to solve critical problem, here used BBO algorithm[10] to produce optimum solution with faster convergence to attain the nearby total optimum.

Thus, an adjustment in classical BBO algorithm is combined with K-Harmonic Means [11], [12] strategy to produce faster convergence and little more better solution. BBO is naturally used to optimize multidimensional actual valued functions, also can be used on intermittent functions.

## II. RELATED WORK

The main aim of the clustering algorithm is determining of similarity to find how closed two data patterns to one another. There are thousands and thousands clustering algorithms are framed in various research papers including traditional or old clustering methods to today's different optimization techniques. KM is the traditional and greatest simple clustering technique. Because of its sensitivity and prior knowledge about number of clusters and premature convergence of local optimization, here in need of global optimization technique. For that here taken Biogeography Based Optimization techniques are used for global optimization. Anirban Mukhopadhyay, Lopamudra Dey Microarray [13] focused the hybrid PSO +KM algorithm for solving of real time problems or huge set of data. This algorithmic procedure can obtain the rewards of above two (BBO and K-Means) algorithms of fast convergence and to avoid premature convergence. To solve complex and discontinuous types of problems such as clustering of medical data, the individual algorithms will not produce optimum clustering solution directly. In order to solve this problem here used Biogeography-Based Optimization technique for initial clustering of data, then this algorithm is combined with K-Harmonic means algorithm for better solution.

## III. K- MEANS CLUSTERING ALGORITHM

The K-Means clustering algorithm is centroid based partitioning clustering procedure [16], which uses the mean value $C_i$ to the cluster. The mean or average value of the cluster is referred as centroid or middling of the data facts. The distance between an object P ϵ $C_i$ and $C_i$ is measured by dist(P , $C_i$), where dist(X ,Y) is the Euclidean distance of two data points X , Y. The cluster center distance is considered by the equation (1).

$$Dis(X_i, Y_{ij}) = \sqrt{\sum_{i=1}^{D}(X_i - Y_{ji})^2} \qquad (1)$$

The algorithm procedural steps of K-Means is given below algorithm1.

Initially select the random values -*K* objects from the *n* number of data set *D*, which are simply called as cluster mean or center. Then using Euclidean distance measure find distance between data objects and cluster mean, to assign each of an item to the group, which is greatest simile to the objects. This process will be repeated till there is no alteration in cluster or number of iteration within cluster variation. Then each cluster find new mean or average from the previous iteration, and reassign the new cluster center from all objects using the newly updated. This will be continued until the stable assignment of iteration. In this K-Means partition clustering method, number of clusters represented by the mutable of K , the data set D, which covers the input of n objects and a set of output as K clusters.

Step 1: Erratically elect initial cluster center roughly of the data objects based on number of clusters or partition K from the data set D.

Step 2: Reallocate each item to the cluster to which the data item is most similar, by mean value using the equation (2).

$$y_j = \frac{1}{n_j}\left[\sum_{\forall X_{p\epsilon c_j}} X_p\right] \qquad (2)$$

Step 3: Continue step 2 until stopping condition is not fulfilled or until no change of data points.

Alg. 1. The K-means partitioning algorithm.

The major hindrance of the K-means algorithm is more subtle of the initialization to cluster centroid, which may go for pre-matured or earlier convergence before go to the local optima. So to overcome all these problems we are in need of built in boost function, which is given in the algorithm of K-Harmonic means.

## IV. K-HARMONIC MEANS ALGORITHM

K-Harmonic Means is the next improvement of K-Means (KM), which customs the Harmonic means (HM) values as a detachments from every data point to the cluster centres. K-Harmonic means uses the summation of all over data points of the vocal function. Let *X= {x1,x2,x3, . . . }* be the data to be clustered. *N= {Na/a=1, 2, . . .K}* be K centres. Then the association function mem(mi/xi) is demarcated by the intention of data point *xi* that fits to centre mi. The output measure of K-harmonic Means is calculated by the summation function *d(x, M)* for all over *x*, by equation (3).

$$f(X,N)=\sum_{x \epsilon X} d(x,N) \qquad (3)$$

The *d(x,N)* is designed using equation (4).

$$d(x,N) = HA\{||x-N||^2\}/ n€N = |N| \Big/ \sum_{n€N} \frac{1}{||x-n||^2} \qquad (4)$$

The Harmonic average is calculated for 1 to K numbers as

$$\{a1\,, a2, a3,\ .\ \ .\ \ .\ \ ak\} \qquad (5)$$

This can be defined as the following equation (6).

$$HA(\{a1, a2, a3,\ .\ \ .\ \ .\ \ ak\}) = K \Big/ \sum_{k=1}^{k} \frac{1}{ak} \qquad (6)$$

When the value of *ak* is minor , the sung middli-,.7ng may also have small. So here used *MIN()* function for harmonic average function *HA()*. This goods is used as a recital fitness function for calculating the quality of the clustering. The fitness function of K-Harmonic Means is restrained by the given equation (1). Here *d(x,N)* $= \|x - n\|^2$, has no favourite increment function of its repeated manner optimization. In this K-Harmonic means algorithm uses the weighting function *w(xi)*, which defines how much inspiration of data point *xi* has in re-calculating the centre restrictions in the following repetition. The weighted function with K-Harmonic means clustering algorithm 2 is given below.

**Weighted function with K-Harmonic means clustering algorithm:**

1. Erratically choose the preliminary cluster centres.
2. Estimate recital fitness function using equation (3) with biased function value $w_p$, where $w_p>2$.
3. Compute membership function $mem(n_l/x_i)$ for each data point $x_i$ according each center $n_i$ by the equation (7).

$$mem(n_l/x_i) = \frac{\|x_{i-}\ \ n_l\|^{-w_p-2}}{\sum_{l=1}^{k}\|x_{i-}\ \ n_l\|^{-w_p-2}} \qquad (7)$$

4. Then compute its weight $w(x_i)$, for each and every data point $x_i$ by the equation (8).

$$w(x_i) = \frac{\|x_i - n_l\|^{-w_p-2}}{\left(\sum_{l=1}^{k}\|x_i - n_l\|^{-w_p}\right)^2} \quad (8)$$

5. Recalculate its position from every facts points *xi* each centre $n_l$, affording by its membership function and weights by the equation (9)

$$n_l = \frac{\sum_{i=1}^{n} mem(n_l/x_i)w(x_i)x_i}{\sum_{i=1}^{n} mem(n_l/x_i)w(x_i)} \quad (9)$$

6. The above steps 2 to 5 will be repeatedly until predefined number of iterations or there is no change in *d(x,N).*

7. Finally allot data point $x_i$ to the cluster *l* with the greatest of $mem(n_l/x_i)$.

Alg. 2. K-Harmonic Means Algorithm

## V. BIOGEOGRAPHY OPTIMIZATION ALGORITHM

Biogeography based optimization(BBO) is designed by Dan simon in the year of 2008, which is constructed by the natural performance of animals, birds human's behavior of immigration and emigration and other living things. BBO is the simple species migration process of the of biological species of the space based on various time sequence. The basic concept of Biogeography is species migrate between "islands" via wreckage, wind, flying, swimming, or any other transport etc. In BBO, each and every individual species are taken as an individual "locale'' or environment with HSI, which are simple called it as dependent variables and some islands are more suitable for habitation than others are Suitability Index Variables(SIVs), which are independent variables. Habitability of each class type based on the features such as rainfall, changing of climate, topography, diversity of vegetation, temperature, etc.

A good optimistic solution of the habitat is similar to high HSI , and a poor or worst solution specifies less HSI of an island. High HSI solutions and low HIS solution will be shared their features between them. So High HSI can provide the information to low HSI and low HSI can get the information from high HSI. Each and every individual species of BBO has own immigration or arrival *(k)* and expatriation or relocation rate (*l*). The best output solution has highest *l* and lowest k and so on , which are number of classes in the environment's functions. This situation is given in the following figure 1.

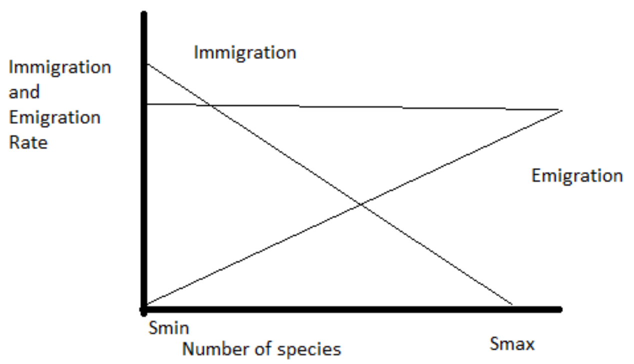


**Fig. 1.Entry "I" and Exit "E" rate based on amount of species**

This Entry and exit rate can be calculated by following equations 3 and 4.

$$\lambda_A = I\left(1 - \frac{a}{n}\right) \quad (10)$$

$$\mu_A = E\left(\frac{a}{n}\right) \quad (11)$$

Here I is the extreme imaginable arrival rate; E is the extreme imaginable exit rate; *A* specifies the number of classes of the $a^{th}$ different and *n* is the number of classes. Now, ponder the superior case *E=I* in figure 1. In this situation;

$$\lambda_A + \mu_A = E . \quad (12)$$

λ is the probability of individual's feature solution is replaced in terms of immigration.

μ is the probability of individual's feature solution is replaced in terms of emigration. In Mathematically entry and exit concept may be characterized as a probabilistic method. The actual species *Sp* can deviations from time *t* to time *t+Δt*, which are consider the probability *Ps*, given in the following equation (13).

$$P_s(t+\Delta t) = P_s(t)(1-\lambda_s\Delta t - \mu_s\Delta t) + P_{s-1}\lambda_{s-1}\Delta t + P_{s+1}\mu_{s+1}\Delta t \quad (13)$$

At time *t* *Δt* is small, then the value of *Δt* will become as 0, ie., Δt→ 0. (14)

$$\text{Prob}(x_j) \text{ is selected for emigration} = \frac{\mu_j}{\sum_{i=1}^{N}\mu_i}$$

For $j = 1, \ldots, N$ , Here N is the number of applicants from the total population,

The steps for Biogeography-Based Optimization is given below:

Step 1: First Initialize the important restrictions of BBO such as number of SIVs, population size, habitat probability modification(P*mod*)consider as 1, mutation probability (MP) consider the value as 0.01, then the maximum arrival rate and exit rates are rumored as 1, finally the elitism parameter value consider as 2.

Step 2*:* Calculate fitness function by setting each SIV habitat as randomly and assume as initial result to the given problem.

Step 3: From the given entry and exit rate find the HSI value for each habitat, which is consider as performance fitness function.

Step 4: Then recognize the elite locales based on the HSI values, which is the least performance fitness value.

Step 5: Compute *S*, $\lambda$, and $\mu$ for each data set using equations (3) and (4).

Step 6: Perform migration operation using entry and exit rate for each non-elite habitats, and recompute HSI for each habitat.

Step 7: Now make mutation operation for each non elite locales by updating probability of each locale using (5) and re-calculate HSI value of each new habitat set.

Step 8: Finally dismiss the progression, until there is no modification or the determined number of iteration steps are reached. Otherwise go to step 2 for further proceeding process.

## VI. PROPOSED HYBRID K-HM+ BBO ALGORITHM.

In this proposed paper used the optimized medical data clustering system by Hybrid K-HM+ BBO algorithm. In this hybridization techniques model used two important computation steps. In the first step, the whole inhabitants is separated as dual portions and it is grown with methods of migration and mutation.
Then executes two various techniques in simultaneously and selects the best entities from each scheme replacing after a user demarcated the best amount of repetitions. Here the largest best fitness value is selected for the individuals.

In the second computation step is just like a parallel technique. In this parallel processing technique combine the advantages of above two K-HM algorithm and BBO algorithms and avoid the disadvantages of these two algorithms. The major concept of this second step is the combination of K-Harmonic means and BBO algorithms, which integrated algorithm can runs K-Harmonic means for some number of iterations then make another some number of iterations for BBO algorithm. So we can get benefits of both K-harmonic means and BBO algorithms like optimizes solution and faster convergence. The simple architecture for proposed K-Harmonic means and BBO algorithm is given in the following figure 2. In this figure first the medical data set consider as a input data, which is preprocessed and in the second step of feature selection is done. After done the feature selection the proposed hybrid algorithm is applied for proper clustering solution. Then finally the clustered solution are considered as a final clusters. The flow diagram for this proposed system is given in the following figure 3. In THIS figure first generate the initial population and evaluate fitness function. Then check whether the iterations are reached or not. If it is over come to end state else apply hybrid algorithm. This hybrid algorithm steps are as given as follows:

Step1: Select a population of size n randomly and initialize the population.

Step 3: Perform either K-Harmonic means and BBO algorithms.

Step 4: Generate output of the algorithms.

Step 5: Evaluate fitness for each individual.

Step 6: Perform these transitions until termination condition is reached.

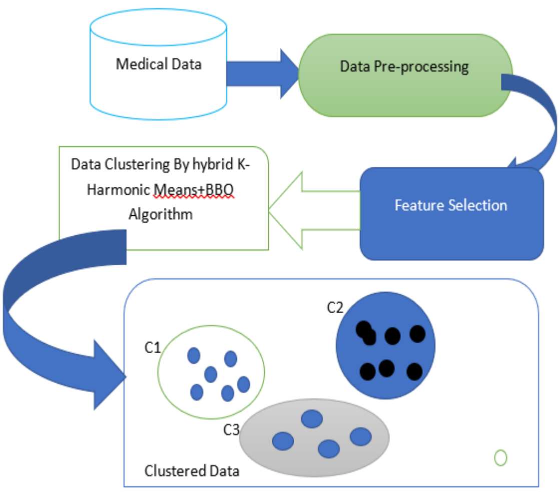


**Fig. 2. Simple Architecture of proposed System**

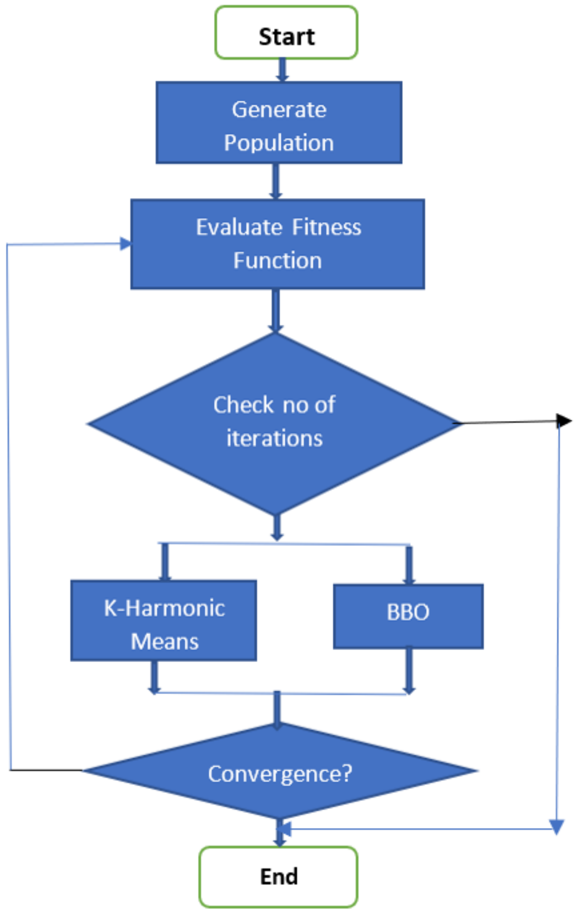


**Fig. 3. Flow Diagram for proposed K-Harmonic Means and BBO Algorithm.**

## VII. RESULTS BASED ON TENTATIVE DATA SETS

For the result of the experiment, here used four various data set to relate the performance of the simple K-Means algorithm, simple K-harmonic means algorithms, Individual BBO algorithm and hybrid K-Harmonic means and BBO algorithms with different set of data.

### A. Proposed Results

For the proposed results here taken 50 iterations. In this 50 iterations K-HM algorithms converged within 20 iterations with most document datasets. For an easy experimental 50 iteration taken for KHM and BBO approaches, which is first executed for 25 iterations, then used this as the early cluster centroid for BBO clustering algorithm, then Biogeography Based (BBO) algorithm performs for next 25 iterations to produce the closing clustering result.

The quality of the cluster is evaluated by fitness equation, which is based on the value of the mean distance of documents and the cluster centroid. The smaller the distance value, the more compact the clustering solution. The Table 1 shows the experimental results of various existing and proposed algorithms such as the K-means, K-Harmonic means, BBO and Hybrid K-Harmonic means and BBO algorithms. The following Table I shows the recorded average value of ADVDC and standard division.

The characteristics features and convergence behavior clustering algorithms and average ADVDC values of each and every iteration are noted separately for KHM algorithm BBO algorithm and KHM+BBO algorithms with datasets.

**Table- I: Performance judgement of KHM, BBO and proposed K-Harmonic means+ BBO Algorithms.**

| Data Set | | ADVDC value | | |
|---|---|---|---|---|
| | | *KHM Algorithm* | *BBO Algorithm* | *K-HM +BBO* |
| Data set1 | Euclidian Distance | 7.9898±0.091 | 5.998±0.9456 | 4.1956±1.0405 |
| | Cosine Similarity | 7.999±0.0950 | 0.6344±0.4161 | 7.587±0.5014 |
| Data set2 | Euclidian Distance | 7.125±0.0966 | 6.29562±0.912 | 4.764±2.0034 |
| | Cosine Similarity | 8.001±0.211 | 9.231±0.398 | 7.234±0.021 |
| Data set3 | Euclidian Distance | 4.666±0.102 | 4.004±0.198 | 2.456±0.676 |
| | Cosine Similarity | 4.998±0.112 | 5.650±0.3965 | 4.0215±0.2431 |
| Data set4 | Euclidian Distance | 8.98±0.1097 | 9.221±1.102 | 5.994±2.543 |
| | Cosine Similarity | 9.22±0.4602 | 11.874±0.565 | 8.567±0.2237 |

Figure 4 shows the convergence performances of the above given algorithms on the four different dataset for similarity measure using the Euclidian distance. In this Figure the K-means algorithm converges quickly but prematurely with high quantization error. Then K-harmonic means algorithm and finally convergence hybrid K_harmonic means BBO algorithm with less quantization error. Then K-harmonic means algorithm and finally convergence hybrid K_harmonic means BBO algorithm with less quantification fault as given in table1, the ADVDC value of the K-means algorithm is suddenly condensed from 10.9 to 7.82 within 10 iterations and fixed at 7.82.

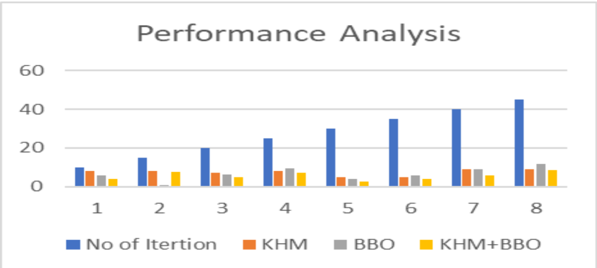


**Fig. 4. Performance analysis of KHM, BBO and KHM+BBO**

The ADVDC value of BBO method's vastly converged from 10.9 to 6.07 within 30 iterations. This ADVDC reduction of BBO is not as sharp as in K-means and K-harmonic means algorithms and becomes smoothly after 30 iterations. The Bar chart indicates more number of iteration is executed, but its execution speed will be lesser than KHM algorithm due to is computation complexity. The average performance of KHM+BBO approach significantly improves. Here among the 50 iterations, the first 25 iterations, the KHM and the remaining 25 number of iterations will be done for BBO algorithms After KHM's 25 iterations, the ADVDC value has a sharp reduction with the value reduced BBO algorithm reduced from 6.257 to 4.657.

**B.** *Algorithm performance evaluation:-*

The performance analysis of the above all algorithms are measured based on two measuring methods such as Entropy based measure and Correlation based measure.

**C.** *Discussion*

The given hybrid clustering algorithm is not a novel clustering technique, but we can improve the efficiency and performance by this hybrid method. However, K-Means clustering method is used for generating initial seed of the clustering in most of the hybrid clustering method and any other optimal clustering algorithms used for generating final optimum clustering solution. The above table I shows, high compact clustering result for the experiment of hybrid KHM+BBO algorithm. Here lowest average distance value has been considered among all the distance values. In the experimental result 25 number of iterations taken for KHM algorithm, but which is not enough for the BBO clustering method for finding optimal clustering solution. The output result of KHM used to get initial seed of the BBO algorithm to locate the optima clustering with a less average distance.

Thus ensuring convergence reliability of proposed K-Harmonic means +BBO algorithm and further it is observed, that the K-Harmonic means+ BBO has much better convergence than individual K-means, K-Harmonic means and BBO based algorithms.

## VIII. CONCLUSION

In this paper, offered a medical data clustering by hybrid K-Harmonic Means+BBO algorithm, that which can be achieved the rewards of both clustering algorithms. In the general BBO algorithm, can get a globalized penetrating for the optimal clustering, with more number of iteration and computation complexity than the traditional methods such as K-means and K-harmonic means, etc. Like K-Means algorithm, K-Harmonic means also meet or convergence faster than the BBO optimization technique, but can be jammed in a limited best area. The hybrid KHM+BBO algorithm syndicates the skill of the globalized searching of BBO algorithm and the fast convergence meet of the and K-Harmonic means algorithm, which can be avoided the weakness of both KHM and BBO algorithms.

## REFERENCES

1. A.D. Thakare1,Nikita Muthiyan2, Deepali Nangde3, Dipti Patil4, Minal Patil5"Clustering Of News Articles to Extract Hidden Knowledge" International Journal of Emerging Technology and Advanced Engineering ISSN 2250-2459, Volume 2, Issue 11, November 2012.
2. Selim, S. Z. And Ismail, M. A, "K-means type algorithms: A generalized convergence theorem and characterization of local optimality" IEEE Transaction. Pattern Anal. Mach. Intell, pp. 81–87, 1984.
3. Haping, M.A.; Dan Simon.: Blended biogeography –based optimization for constrained optimization. Engineering applications of artificial intelligence, Elsevier, Vol.24, (2011)517-625
4. Osama Abu Abbas, "Comparison between Data Clustering Algorithms". International Arab journal of information Technology, vol 5, no 3, July 2007.
5. Berkhin, P. "Survey of clustering data mining techniques" Accrue Software Research Paper. 2002.

6. Anderberg, M. R., “Cluster Analysis for Applications” Academic Press, Inc., New York,1973.
7. Jain A. K., Murty M. N., and Flynn P, “Data Clustering: A Review” ACM Computing Survey, Vol. 31, No. 3, pp. 264-323.,1999.
8. Jiawei Han Michaline Kamber, Jian Pei ,”Data Mining Concepts and techniques”,3/e, Elsevier Inc,2012.
9. Sandeep Rana1*, Sanjay Jasola1, Rajesh Kumar2,” A hybrid sequential approach for data clustering using K-Means and particle swarm optimization algorithm”,International Journal of Engineering, Science and Technology, Vol. 2, No. 6, 2010, pp. 167-176, 2010.
10. D.Simon, “Biogeography-Based Optimization,” IEEE Transactions on Evolutionary Computation, Vol. 12, No. 6, (2008),pp. 702-713.
11. A. Jaya Mabel Rani, Dr. Latha Parthipan, “FuzzifiedPSO and K-Harmonic Means Algorithmfor Electrical Data Clustering”, 2013 International Conference on Recent Trends in Information Technology (ICRTIT), pp. 546-550, 2013.
12. Bin Zhang, “Generalized K-Harmonic Means - Boosting in unsupervised Learning” Hewlett-Packard Laboratories, 2000.
13. Anirban Mukhopadhyay, Lopamudra DeyMicroarray “Gene Expression Data Clustering using PSO based K-means Algorithm”, International Conference on Advanced Computing, Communication and Networks,2011.
14. Mehdi Neshat, Shima , Farshchian Yazdi,Daneyal Yazdani and Mehdi Sargolzaei, “A New Cooperative Algorithm Based on PSO and Means for Data Clustering”, Journal of Computer Science, ISSN 1549 ,3636 ,8 (2): 188-194, 2012.
15. A. Pakrashi, B. B. Chaudhuri, "A kalman filtering induced heuristic optimization based partitional data clustering," Information Sciences 2016.
16. Das, P., Das, D.K. and Dey, S., "A modified Bee Colony Optimization (MBCO) and its hybridization with k-means for an application to data clustering," Applied Soft Computing, vol.70, pp.590-603, 2018.
17. Bhubaneswar, Mishra.D, Satapathy.S.K, “Particle Swarm Optimization Based Fuzzy frequent Pattern Mining from Gene Expression Data” IEEE Explore Digital Library , 2011.
18. Donghui Chen, Zhijing Liu, Zonghu Wang, “A novel fuzzy clustering algorithm based on Kernel method and ParticlenSwarm Optimization”, Journal of Convergence Information Technology(JCIT), Volume7, Number3, February 2012.
19. Ghazavi, S.N. and Liao, T.W., "Medical data mining by fuzzy modeling with selected features," Artificial Intelligence in Medicine, vol.43, no.3, pp.195-206, 2008.
20. N.M Jothi Swaroopan., P. Somasundaram , “Fuzzified PSO Algorithmfor dc-opf of Interconnected Power System” ,Journal of Theoretical and Applied Information Technology, 2010.
21. Sami Ayram, Tommi Kainen “Introduction to partitioning-based clustering methods with a robust example” Software and Computational Engineering, 2006.


## AUTHORS PROFILE

**Jaya Mabel Rani.A** is an Assistant Professor of jeppiaar maamallan engineering college at sriperumpudur, Tamilnadu, India. She has more than fourteen years of teaching experience and completed her M.E in computer science and engineering at Anna university, Chennai. She is pursuing her Ph.D in the department of Computer science and Engineering at Sathyabama Institute of Science and technology, Chennai,Tamilnadu,India under the guidance of Dr. A.Pravin. She has published several papers at national and international conferences and refereed journals. Her areas of research interest are data mining, machine learning, and Artificial Intelligence.

**Dr.A.Pravin** received the B.E degree in Computer Science & Engineering from Bharath Niketan Engineering College, Madurai Kamaraj University, Madurai, India in 2003 , M.E degree in Computer Science & Engineering from Sathyabama University, Chennai, India in 2005 and Ph.D degree in Computer Science & Engineering at Sathyabama University, Chennai, India in 2014. He works currently as an Associate Professor for the Department of Computer Science and Engineering at Sathyabama Institute of Science and Technology, Chennai and he has 14 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 Software Engineering, Data mining , Internet of Things and Big data.