

# Bacterial Foraging Optimized Fuzzy C Means Clustering for efficient disease prediction

Banumathy D.<sup>1\*</sup> and Selvarajan S.<sup>2</sup>

1. Assistant Professor, Department of CSE, Paavai Engineering College, INDIA

2. Principal, Muthayammal College of Engineering, INDIA

\*banumathydhanabalanpec@paavai.edu.in

## Abstract

*An early diagnosis of disease is preferred. K-means is a famous hard clustering algorithm that splits data objects into k clusters wherein the quantity of clusters wherein the quantity of clusters, k, is determined previously as per the application purpose. Fuzzy c-means clustering is an efficient protocol, however, the arbitrary selection in center points makes the repetitive procedure forcing into local optimum solution with ease. Bacterial Foraging Optimization protocol (BFOA) on the basis of the behaviour of biologically inspired E-coli bacteria, used to find optimal solution. E-coli bacteria search for rich nutrients in the search space by using their energy per unit time. The common characteristic bacteria's are grouped together. The bacterium can communicate with each other by sending signals. In this work, Hybrid Bacterial Foraging Optimization protocol - Fuzzy C-Means Clustering (BFO-FCM) method is proposed.*

**Keywords:** K-means Clustering, Fuzzy c-means clustering (FCM), Bacterial Foraging Optimization Algorithm (BFOA) and Hybrid Bacterial Foraging Optimization Algorithm - Fuzzy C-Means Clustering (BFO-FCM).

## Introduction

In today's world, technology is a critical component of healthcare. Of course, all healthcare comprises of mortal collaboration, the solicitation of tools, or, typically, both. Technology is significant in all examinations of the organization as well as functions of healthcare service as well as system for several reasons. Knowledge is a huge module of modern healthcare cost as well as possibly the primary drive of future costs. The main regulatory framework as well as organizations are present only to cope with the overview as well as usage of safe, efficient as well as secure technology in healthcare. Improvements in healthcare knowledge are possible huge sources of economic wealth and forces for changes in the organization of healthcare.

Data Mining refers to the procedure of identifying the related information according to the input query from a large set of data sets. Clustering typically is an unsupervised procedure that groups elements together such that the clusters<sup>1</sup> designated to one cluster are more like one another than the other data points. The motivation for the choice resides in the fact that possessing a greater number of dimensions

typically results in the curse of dimensionality, wherein the performance of several standard machine-learning protocols is affected. This is almost always because of 2 pervasive impacts: empty space as well as the concentration of distance. The former is where all the higher-dimensional datasets are sparse as the quantity of points needed for denoting distributions increases exponentially with the quantity of dimensions. This results in poor density estimate for higher-dimensional data, leading to problems for density-based methods. Concentration of distances refers to a counter-intuitive characteristic of higher-dimensional data representation, wherein every distance between data points tends to turn more difficult to differentiate when dimensionality rises leading to issues with distance-based protocols.

The high-dimensional clustering can be applied in a variety of ways and can be adapted to the problem of decision making<sup>2</sup>. The decisive support system is one which provides support to the clinical process, and there are many decisive support systems which works based on symptoms provided. The symptoms based decisive support systems, generates recommendations on different diseases according to input symptoms given. Similarly, the decisive support systems can use the high dimensional data set to perform such recommendations.

The intelligent support system is one, which generates knowledge for the decision-making process. The decisive support system can use the grouped information to perform decision making. For example, by grouping the similar data points of different disease class, the similarity to the input data point can be computed. By calculating the similarity of data points, the user disease can be easily identified by the medical practitioner. Feature selection algorithm<sup>3</sup> iteratively chooses a sub-set of the original features known as a candidate sub-set and measures the optimality of the candidate sub-set utilizing evaluation functions. Thereby, features selection method decreases data dimensionality, discards non-relevant data, increases learning accuracy, as well as enhances results comprehensibility. Features selection protocols are either filter or wrapper models. The former depends on generic characteristics of the training data for selecting some attributes with no involvement of any learning protocols. Wrappers need a pre-determined learning protocol in features selection and use their performance to evaluate as well as determine the features that are chosen.

Clustering refers to the procedure of assignment of data objects to a set of disjoint groups known as clusters such that objects in all clusters are like one another more than those

from other clusters. The FCM protocol is sensitive to initializations, as well as is trapped with ease into the local optimum. Fuzzy protocols are capable of assigning data objects partially to several clusters. The degree of membership in fuzzy clusters relies on the nearness of data objects to cluster centres. The top famous clustering protocol is FCM that was proposed by Bezdek back in 1974 and is still vastly utilized.

FCM has several disadvantages because of the limitation that the sum of membership values is: First, the restriction leads to higher membership value for outlier points as well as because of this the protocol finds it hard to handle outlier point. Second, in clusters, memberships of data points depend in a direct manner on the membership value of other cluster centres that might result to non-desirable outputs. FCM also has issues when dealing with higher dimensional datasets as well as a huge quantity of prototypes. FCM is also sensitive to initializations as well as gets forced into local optimum with ease.

BFOA is an evolutionary optimization method that owes its inspiration to the foraging activity of the E.Coli bacteria. The biologic features of the BFOA as well as its motile activities / decision-making methods. As a heuristic technique, BFOA is formulated for tackling non-gradient optimizing issues as well as for handling complicated as well as nondifferentiable objective function. Search of hyper space is carried out via 3 major operations, which are chemotaxis, reproduction as well as elimination dispersal activities. Bacterium searches for nutritional food in a fashion for maximizing energy got every unit time. Bacteria also interconnect with other through signals. Bacteria take foraging decision after consideration of 2 preceding factors. The procedure, through which looking for nutrient through bacteria occurs via small steps known as chemotaxis. The primary notion of BFO is the simulation of chemotaxis measure of virtual bacterium in the complicated exploratory space<sup>5</sup>.

In this work, a hybrid BFOA-FCM is proposed for efficient disease prediction. Section 2 stated the literatures used associated with the work. Section 3 describes the detailed of techniques used. Section 4 explains the results obtained and discussed the obtained results. In the end, section 5 presents the conclusion.

### Related Works

Lee and Lin<sup>6</sup> proposed a Strategy-Adaptation-Based Bacterial Foraging Optimization (SABFO) protocol for solving the optimization of complicated issues. The suggested SABFAO protocol takes up the strategic method into chemotaxis step of the conventional BFOA. The suggested technique ensures every bacteria swims on varying run length as well as improves bacteria diversity also. 5 optimization issues of non-linear benchmarks were utilized for verifying the performance of SABFO. Simulations reveal that SABFO gets improved global optimum solutions than other techniques.

FCM is a widely-utilized clustering method as well as an effective one, in comparison to clustering protocols. Grover<sup>7</sup> presented a research on several fuzzy clustering protocols like FCM PCM, FPCM, as well as PFCM, with their corresponding benefits as well as disadvantages.

Izakian et al<sup>8</sup> proposed a hybridized fuzzy clustering technique on the basis of FCM as well as fuzzy PSOA (FPSOA) that makes usage of the benefits of both protocols. Outputs of experiments reveal that the suggested technique is effective as well as shows promising outputs.

Sharma & SSCET<sup>9</sup> investigated a programmable technique for analysing diabetes illnesses on the basis of BFOA as well as ANN. ANN is a strong tool for solving the issues like classifications as well as predictions. The performance is extremely reliant on architecture as well as weight. For gaining higher efficacy in ANN, adequate architecture as well as learning protocol is preferred. The simulations on the data set for verifying the promise of BFOA show that BFOANN is more precise than conventional ANN with regard to classifications.

Yahya et al<sup>3</sup> suggested an adaptive variant of GA which may be employed for features selection in higher dimensional data. The suggested method has its basis in varying length representation strategy as well as a set of altered as well as suggested genetic operator. For assessing the efficacy of the suggested method, cue phrase selection is employed it for comparing the performance with several ranking methods that were employed for the job. The outputs reveal experiment evidence for the efficacy of the suggested method in features selection in higher dimensional data.

Wan et al<sup>10</sup> presented a novel clustering method that is grounded in the analysis of BF. It is an optimizing method for the clustering problems wherein a set of bacteria forage for converging to particular positions as final cluster centres through minimization of fitness functions. Quality of the method was assessed on various popular benchmarks. In comparison to famous clustering technique like K-means, ACO, PSO etc. experiments proved that the suggested technique was an efficient clustering method which handled datasets with several cluster sizes, densities as well as several dimensions.

Chen et al<sup>11</sup> presented a novel evolutionary learning protocol for optimizing the variables of neural fuzzy classifiers. This novel evolutionary learning protocol has its basis in a hybridized BFOA as well as PSO. It is hence known as BFPSOA. The suggested BFPSO technique carries out local searches via chemotaxis movement operation of bacteria foraging while the global searches over the whole search space is accomplished by particles swarm operator. NFC model utilizes functional link neural network as the subsequent component of fuzzy rule. The research utilizes orthogonal polynomial as well as linear independent function in functional expansions of functional link NNs. In

the end, the suggested NN fuzzy classifier with BFPSOA is taken up in various classification applications. Experiments and their results prove that the suggested NFC-BFPSO technique outperforms other techniques.

**Methodology**

**Wisconsin Dataset:** Table 1 shows the attributes of Wisconsin Dataset<sup>12</sup>.

**Table 1**  
**Data feature information**

# Feature	Domain
1. Sample code number	id number
2. Clump Thickness	1 - 10
3. Uniformity of Cell Size	1 - 10
4. Uniformity of Cell Shape	1 - 10
5. Marginal Adhesion	1 – 10
6. Single Epithelial Cell Size	1 – 10
7. Bare Nuclei	1 – 10
8. Bland Chromatin	1 – 10
9. Normal Nucleoli	1 – 10
10. Mitoses	1 – 10
11. Class 2 for benign	2 for benign 4 for malignant

**Bacterial Foraging Optimization Algorithm (BFOA):**

BFOA is a part of a domain of bacterial optimization protocols as well as swarm optimization protocols, as well as more broadly in the domains of computational intelligence as well as meta-heuristics. It is associated with other bacterial optimization protocols like bacteria chemotaxis protocol, as well as other SI protocols like ACO as well as PSOA. The BFOA owes its inspiration to the swarm foraging activity of the bacteria like E.Coli as well as M. Xanthus. Particularly speaking, BFO owes inspiration to chemotactic behaviour of bacteria which perceives chemical gradient in the environment (like nutrient) as well as move away from particular signals.

BFO owes its inspiration to the foraging activity of E.coli bacteria discovered in the intestine. The benefit of the protocol is the parallel search capability of SI that results in escape from local minima. In complicated optimization issues, BFO is difficult to converge to the optimum solution as well as the performance is extremely dependent on the chemotactic step lengths<sup>6</sup>.

The E. coli bacteria is the typical bacterium present in the intestines of both humans as well as animals. They have two patterns of movement: swim as well as tumble, which is performed through rotating a set of tensile flagella. The

movement of the bacteria assist in avoiding toxic substances and push the body towards regions with more nutrients when foraging. As per modern biological studies, the E. coli bacteria grows rapidly in favourable environments and is capable of carrying out reproduction every 20 minutes. BFO utilizes 3 dominant methods: chemotaxis, reproduction, as well as elimination-dispersal, for solving the optimization problem.

Assume a chemotaxis step is a tumble after a tumble, or a run after a tumble. Assume j is the index for the chemotaxis step. Assume k is the index for the reproduction step. Assume l is the index of elimination-dispersal event. Furthermore, assume:

- p: Dimension of search space,
- S: Overall quantity of bacteria in the population,
- N<sub>c</sub>: The quantity of chemotactic steps,
- N<sub>s</sub>: The swimming length.
- N<sub>re</sub>: The quantity of reproduction steps,
- N<sub>ed</sub>: The quantity of elimination-dispersal events,
- P<sub>ed</sub>: Elimination-dispersal probability,
- C (i): The size of the step taken in an arbitrary direction as mentioned by the tumble.

**Chemotaxis:** This technique mimics the motion of E. coli cells by way of swimming as well as tumbling with flagella. E. coli bacteria is capable of transferring in 2 different ways<sup>9</sup>. It is capable of swimming for a duration of time in a single direction or it could tumble & switch between the 2 modes of operation for its whole lifespan. Let θ<sup>i</sup> (j, k, l) represent the ith bacteria at jth chemotactic, kth reproductive as well as lth elimination-dispersal step. C (i) specify the extent of the steps taken in an arbitrary direction mentioned by the tumble (run length units). Then in computational chemotactic movement of the bacteria can be given by equation (1):

$$\theta^i(j+1,k,l) = \theta^i(j,k,l) + \frac{C(i)\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}} \tag{1}$$

Wherein Δ denotes a vector in the arbitrary direction wherein the elements are between [-1, 1].

**Swarming:** It is an group behaviour that is noted in various motile species of bacterium that includes E. coli as well as S. typhimurium, wherein intricate as well as stable spatio-temporal pattern (swarm) is created in semi-solid nutrient media. A set of E. coli cells structure themselves in a travelling ring through moving up the nutrient gradient when situated amongst a semi-solid matrix with one nutrient chemo effector. The cells after stimulation by a higher level of succinate, release an attractant aspartate that assists in aggregation into groups as well as thereby fly as concentric pattern of swarm with extreme bacterial density. The cell-to-cell signalling in E. coli swarms could be denoted by (2).

$$\begin{aligned}
 J_{cc}(\theta, P(j, k, l)) &= \sum_{i=1}^S J_{cc}(\theta, \theta^i(j, k, l)) \\
 &= \sum_{i=1}^S [-d_{attractant} \exp(-w_{attractant} \sum_{m=1}^p (\theta_m - \theta_m^i)^2)] + \\
 &\sum_{i=1}^S [-h_{repellant} \exp(-w_{repellant} \sum_{m=1}^p (\theta_m - \theta_m^i)^2)]
 \end{aligned} \tag{2}$$

Wherein  $J_{cc}(\theta, P(j, k, l))$  is the objective function value for addition to the actual objective function value (to be decreased to a minimum) to give a time-variant objective function, S represents the overall quantity of bacterium, p refers to the quantity of parameters to be optimized, that are in every bacteria, while  $\theta = [\theta_1, \theta_2, \dots, \theta_p]^T$  represents a point in the p-dimensional search domain.  $d_{attractant}, w_{attractant}, h_{repellant}, w_{repellant}$  represent various coefficients which ought to be selected with care.

**Reproduction stage:** After evolution through several phases of chemotaxis bacteria arrived at this stage, wherein it is split into 2 forms. The healthy half substitutes the other one that gets discarded because of poorer foraging abilities. This substitution leads to a constant bacteria population.

**Elimination and dispersal stage:** Unpredictable events occur like bacteria in the area being killed or groups of bacteria dispersed to novel environments that might considerably modify the smooth procedure of evolution. From a broader view, elimination as well as dispersal are a component of the population-level long-distance motile behavior.

Bacterial Foraging (BF) has its basis in a computational intelligence method which is not highly impacted by size as well as non-linearity of the issue and converges to the optimum solution for several issues, wherein the analytical techniques have failed to converge. Furthermore, it has benefits like lesser operational burden, global convergence, lesser computation time requirements, as well as capable of handling more objective functions.

**Fuzzy C-Means Clustering Algorithm (FCM):** FCM clustering protocol is vulnerable to few disadvantages because of the restrictions that the sum of the membership values of data points xi in every cluster should be the same as the one in (^). First, the restriction leads to higher membership value for outlier point and because of this, the protocol finds it hard to handle outlier point. Second, in a cluster, the membership of data point depends directly on the membership value of other cluster centers that might result in non-favorable results. FCM also has issues in dealing with higher dimensional datasets, as well as a huge quantity of prototypes. FCM is also sensitive to initialization, as well as is forced into local optimal value with ease<sup>7</sup>.

The protocol designated a membership value to data items for clusters ranging between 0 and 1, as well as fuzzification

variable between [1, n] that defines the degree of fuzziness in clusters. FCM offers a technique of clustering which permits a data item to be a part of 2 or more clusters, and this strategy is often utilized in patterns recognition applications. It has its basis in the decreasing of objective functions to a minimum as in equation (3):

$$J_{mf} = \sum_{j=1}^N \sum_{k=1}^c \mu_{jk}^{mf} \|x_j - c_k\|^2 \tag{3}$$

Wherein:  $m_f$  denotes real numbers bigger than one so that  $1 \leq m_f < \infty$ ,  $\mu_{jk}$  represents the degree of membership of  $x_i$  in the cluster j while  $c_k$  represents the centre of the cluster. In FCM, the membership matrix U is permitted to not merely zero as well as one but also the components with values ranging from zero to one, this matrix fulfilling the restriction below as in equation (4):

$$\sum_{j=1}^c \mu_{jk} = 1, \forall k = 1, 2, \dots, n \tag{4}$$

Fuzzy partitioning is performed via iterative optimizing of objective function given previously with updating of membership  $\mu_{jk}$  as well as cluster centres  $c_k$  are as in equation (5):

$$\begin{aligned}
 \mu_{jk} &= \frac{1}{\sum_{p=1}^c \left[ \frac{\|x_j - c_k\|^2}{\|x_j - c_p\|^2} \right]^{\frac{2}{m-1}}} \\
 c_k &= \frac{\sum_{j=1}^N \mu_{jk}^{mf} \cdot x_j}{\sum_{j=1}^N \mu_{jk}^m}
 \end{aligned} \tag{5}$$

The FCM comprises the following steps:

1. The membership matrix U is set with random values ranging from zero to one so that restrictions are satisfied.
2. Calculate fuzzy cluster centers  $c_k$ , where  $k=1, \dots, C$ .
3. Calculate objective function and terminate if it is lesser than a particular threshold or enhancement for a few cycles is lesser than a particular level.
4. Compute a new membership matrix U.
5. Repeat back to step 2.
6. This cycle will stop if  $\|U^{(k+1)} - U^k\| < \xi$  wherein  $\xi$  signifies a stopping condition ranging from zero to one wherein k signifies the iteration step.

**Hybrid BFOA-FCM:** The social foraging activity of E. coli bacteria has been utilized for solving optimizing issues in medical data set. This work suggests a hybridized method which involves Bacterial Foraging (BF) algorithms as well as FCM protocols for Wisconsin and broad institute dataset. First illustrate the suggested technique utilizing the performance of the protocol is examined with focus on chemotactic steps, lifetime of the bacteria and the clusters. Outcomes of the outputs illustrated the suggested method was extremely effective as well as can be expanded with ease for other global optimization issues<sup>13</sup>.

After completing all the processes, the created result is provided to FCM as input. The optimum value of BFOA via various medical data set is provided as an input to FCM. The

goal of FCM is the finding of cluster centres which decrease dissimilitude functions to a minimum. Membership matrix (U) is arbitrarily set as (6):

$$\sum U_{ij} = 1 \tag{6}$$

Wherein i signifies the quantity of clusters; j signifies the image data points. Dissimilitude functions are computed with (7):

$$C_i = \sum_{j=1}^n U_{ij} = \sum_{j=1}^n \sum_{i=1}^c U_{ij} J^n \tag{7}$$

Wherein  $U_{ij}$  are between zero as well as one,  $C_i$  signifies the centroid of cluster I,  $D_{ij}$  signifies the Euclidean distance between ith as well as centroid ( $C_i$ ) and jth data point. To calculate Euclidean distance ( $d_{ij}$ ) in (8):

$$D_{ij} = CCp - Cn \tag{8}$$

Wherein CCp signifies the Cluster centre value, Cn signifies the Current neuron, that is, quantity of clusters is calculated by (9):

$$C = (N / 2) \tag{9}$$

N= quantity of pixels in image.

### Results and Discussion

The parameters used are: The size of the population was 100 individuals in the three case of dimension problem, where in case 1: n=10 dimensions, case 2: n=25 dimensions and case 3: n=50 dimensions. Table 2 and 3 shows the Results for optimized HBFO-FCM Clustering (Wisconsin Dataset) and optimized HBFO-FCM Clustering (Broad Dataset) respectively. Figure 1 to 6 shows the Classification Accuracy, Specificity, Sensitivity, F Measure, Fowles Mallow Index and Dice Coefficient for optimized HBFO-FCM in Wisconsin Dataset respectively. Figure 7 to 12 shows the Classification Accuracy, Specificity, Sensitivity, F Measure, Fowles Mallow Index and Dice Coefficient for optimized HBFO-FCM in Broad Institute Dataset respectively.

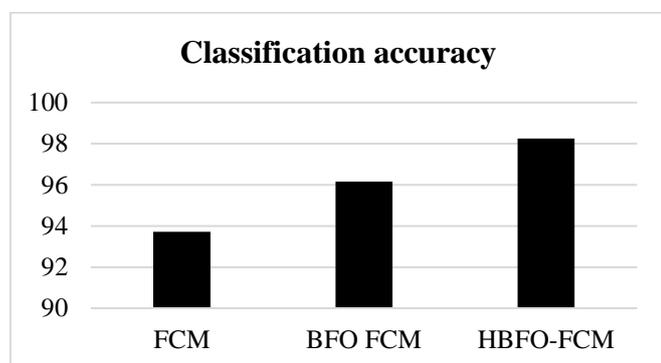


Figure 1: Classification Accuracy for optimized HBFO-FCM in Wisconsin Dataset

From the figure 1, it can be observed that the HBFO-FCM has higher classification accuracy in Wisconsin dataset by 4.73% for FCM and by 2.16% for BFO-FCM method.

From the figure 2, it can be observed that the HBFO-FCM has higher Specificity in Wisconsin dataset by 5.15% for FCM and by 2.54% for BFO-FCM method for abnormal.

And also HBFO-FCM has higher Specificity in Wisconsin dataset by 3.73% for FCM and by 1.23% for BFO-FCM method for normal.

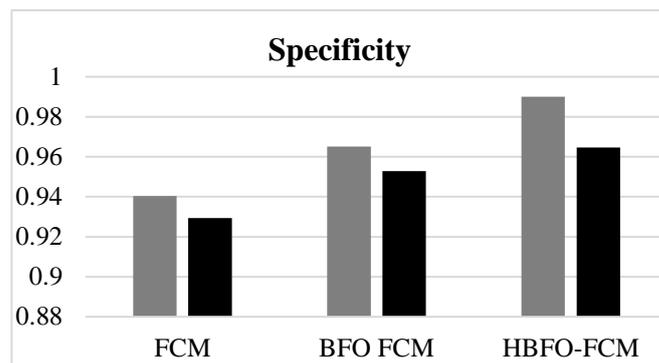


Figure 2: Specificity for optimized HBFO-FCM in Wisconsin Dataset

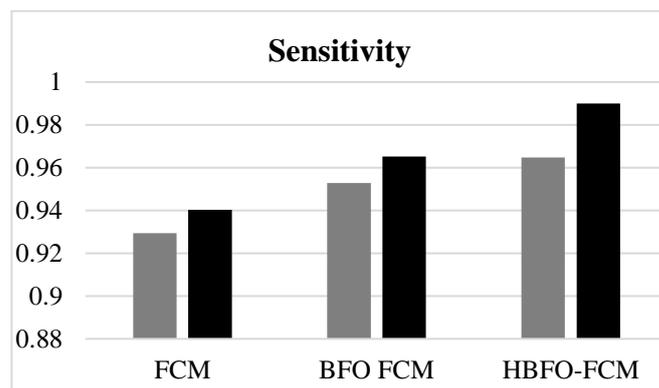


Figure 3: Sensitivity for optimized HBFO-FCM Clustering in Wisconsin Dataset

From the figure 3, it can be observed that the HBFO-FCM has higher Sensitivity in Wisconsin dataset by 3.72% for FCM and by 1.23% for BFO-FCM method for normal. And also HBFO-FCM has higher Sensitivity in Wisconsin dataset by 5.15% for FCM and by 2.54% for BFO-FCM method for abnormal.

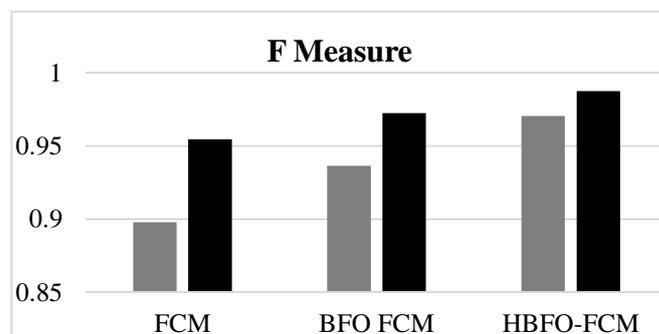
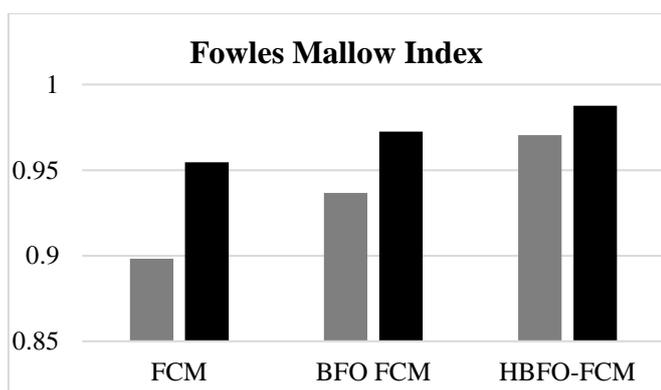


Figure 4: F Measure for optimized HBFO-FCM Clustering in Wisconsin Dataset

From the figure 4, it can be observed that the HBFO-FCM has higher F Measure in Wisconsin dataset by 7.78% for FCM and by 3.57% for BFO-FCM method for normal. And also HBFO-FCM has higher F Measure in Wisconsin dataset by 3.40% for FCM and by 1.54% for BFO-FCM method for abnormal.

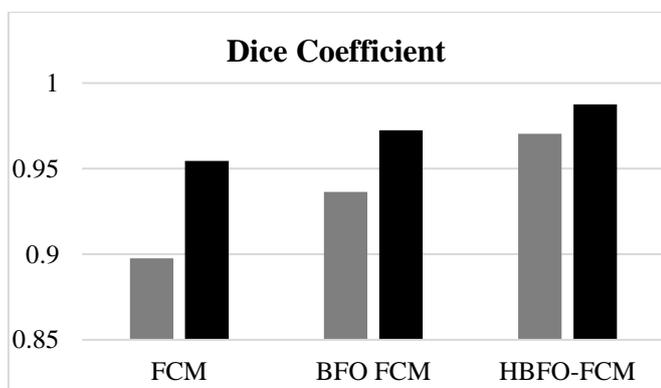
**Table 2**  
**Summary of Results for optimized HBFO-FCM Clustering (Wisconsin Dataset)**

	FCM	BFO FCM	HBFO-FCM
Classification accuracy	93.71	96.15	98.25
Specificity for Abnormal	0.9403	0.9652	0.99
Specificity for Normal	0.9294	0.9529	0.9647
Sensitivity for Normal	0.9294	0.9529	0.9647
Sensitivity for Abnormal	0.9403	0.9652	0.99
F measure for normal	0.8977	0.9364	0.9704
F measure for abnormal	0.9545	0.9724	0.9875
Fowles Mallow Index for Normal	0.8982	0.9366	0.9704
Fowles Mallow Index for Abnormal	0.9546	0.9725	0.9875
Dice Coefficient for Normal	0.8977	0.9364	0.9704
Dice Coefficient for Abnormal	0.9545	0.9724	0.9876



**Figure 5: Fowles Mallow Index for optimized HBFO-FCM Clustering in Wisconsin Dataset**

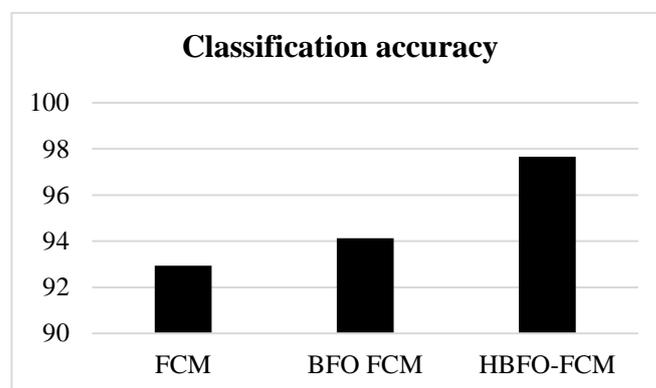
From the figure 5, it can be observed that the HBFO-FCM has higher Fowles Mallow Index in Wisconsin dataset by 7.73% for FCM and by 3.54% for BFO-FCM method for normal. And also HBFO-FCM has higher Fowles Mallow Index in Wisconsin dataset by 3.39% for FCM and by 1.53% for BFO-FCM method for abnormal.



**Figure 6: Dice Coefficient for optimized HBFO-FCM Clustering in Wisconsin Dataset**

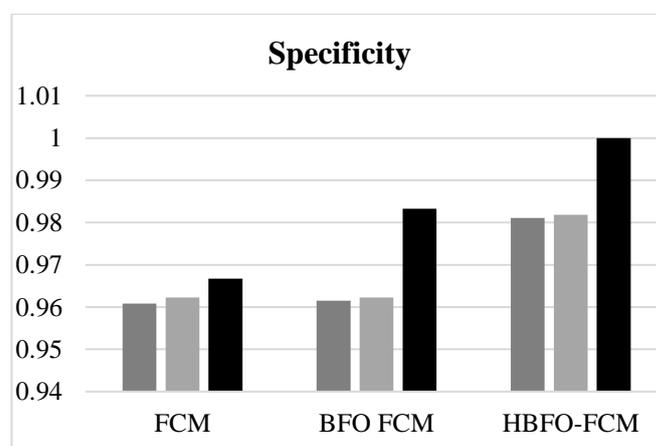
From the figure 6, it can be observed that the HBFO-FCM has higher Dice Coefficient in Wisconsin dataset by 7.78% for FCM and by 3.57% for BFO-FCM method for normal.

And also HBFO-FCM has higher Dice Coefficient in Wisconsin dataset by 3.41% for FCM and by 1.55% for BFO-FCM method for abnormal.



**Figure 7: Classification Accuracy for optimized HBFO-FCM in Broad Institute Dataset**

From the figure 7, it can be observed that the HBFO-FCM has higher classification accuracy in broad institute dataset by 4.94% for FCM and by 3.68% for BFO-FCM method.

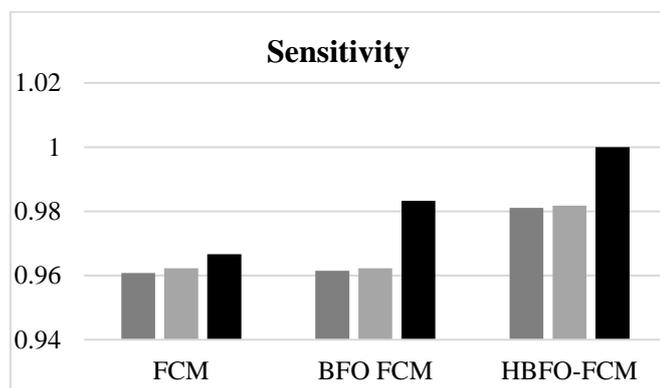


**Figure 8: Specificity for optimized HBFO-FCM Clustering in Broad Institute Dataset**

**Table 3**  
**Summary of Results optimized HBFO-FCM Clustering (Broad Dataset)**

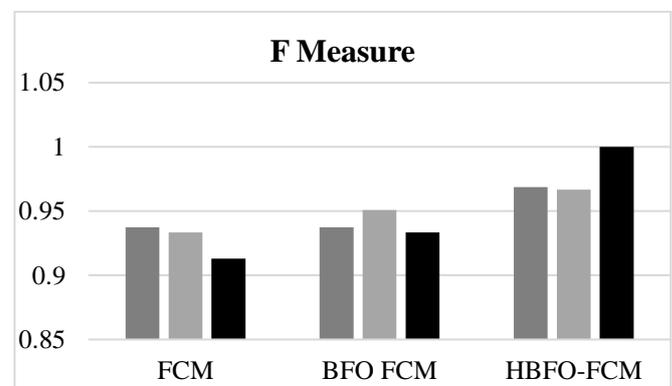
	FCM	BFO FCM	HBFO-FCM
Classification accuracy	92.94	94.12	97.65
Specificity for Normal	0.9608	0.9615	0.9811
Specificity for DCIS	0.9623	0.9623	0.9818
Specificity for IDC	0.9667	0.9833	1
Sensitivity for Normal	0.9375	0.9375	0.9688
Sensitivity for DCIS	0.9333	0.9667	0.9667
Sensitivity for IDC	0.913	0.913	1
F measure for normal	0.9375	0.9375	0.9688
F measure for DCIS	0.9333	0.9508	0.9667
F measure for IDC	0.913	0.9333	1
Fowles Mallow Index for Normal	0.9375	0.9375	0.9688
Fowles Mallow Index for DCIS	0.9333	0.951	0.9667
Fowles Mallow Index for IDC	0.913	0.9335	1
Dice Coefficient for Normal	0.9375	0.9375	0.9688
Dice Coefficient for DCIS	0.9333	0.9508	0.9667
Dice Coefficient for IDC	0.913	0.9333	1

From the figure 8, it can be observed that the HBFO-FCM has higher Specificity in broad institute dataset by 2.09% for FCM and by 2.02% for BFO-FCM method for normal. HBFO-FCM has higher Specificity in broad institute dataset by 2% for FCM and by 2% for BFO-FCM method for DCIS. HBFO-FCM has higher Specificity in broad institute dataset by 3.39% for FCM and by 1.68% for BFO-FCM method for IDC.



**Figure 9: Sensitivity for optimized HBFO-FCM Clustering in Broad Institute Dataset**

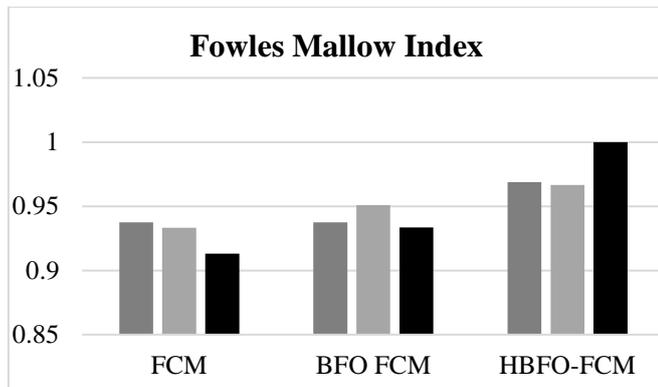
From the figure 9, it can be observed that the HBFO-FCM has higher Sensitivity in broad institute dataset by 3.28% for FCM and by 3.28% for BFO-FCM method for normal. HBFO-FCM has higher Sensitivity in broad institute dataset by 3.52% for FCM and by 0% for BFO-FCM method for DCIS. HBFO-FCM has higher Sensitivity in broad institute dataset by 9.10% for FCM and by 9.10% for BFO-FCM method for IDC.



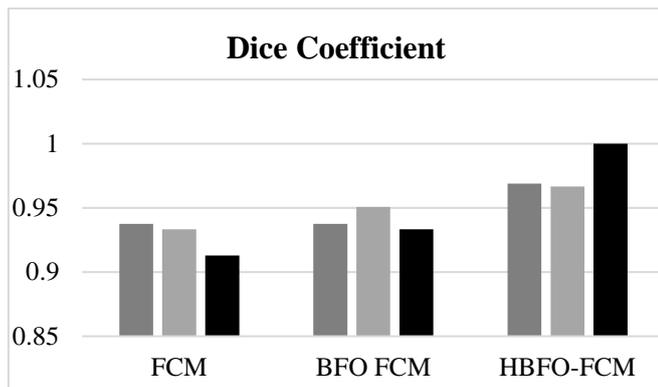
**Figure 10: F Measure for optimized HBFO-FCM Clustering in Broad Institute Dataset**

From the figure 10, it can be observed that the HBFO-FCM has higher F Measure in broad institute dataset by 3.28% for FCM and by 3.28% for BFO-FCM method for normal. HBFO-FCM has higher F Measure in broad institute dataset by 3.52% for FCM and by 1.66% for BFO-FCM method for DCIS. HBFO-FCM has higher F Measure in broad institute dataset by 9.10% for FCM and by 6.90% for BFO-FCM method for IDC.

From the figure 11, it can be observed that the HBFO-FCM has higher Fowles Mallow Index in broad institute dataset by 3.28% for FCM and by 3.28% for BFO-FCM method for normal. HBFO-FCM has higher Fowles Mallow Index in broad institute dataset by 3.52% for FCM and by 1.64% for BFO-FCM method for DCIS. HBFO-FCM has higher Fowles Mallow Index in broad institute dataset by 9.10% for FCM and by 6.88% for BFO-FCM method for IDC.



**Figure 11: Fowles Mallow Index for optimized HBFO-FCM Clustering in Broad Institute Dataset**



**Figure 12: Dice Coefficient for optimized HBFO-FCM Clustering in Broad Institute Dataset**

From the figure 12, it can be observed that the HBFO-FCM has higher Dice Coefficient in broad institute dataset by 3.28% for FCM and by 3.28% for BFO-FCM method for normal. HBFO-FCM has higher Dice Coefficient in broad institute dataset by 3.52% for FCM and by 1.66% for BFO-FCM method for DCIS. HBFO-FCM has higher Dice Coefficient in broad institute dataset by 9.10% for FCM and by 6.90% for BFO-FCM method for IDC.

## Conclusion

Recently, bacterial foraging behaviour, bacterial chemotaxis as a strong resource of possible engineering applications as well as computation models have obtained greater interest. In this work, the implementation of a hybrid intelligent system data mining technique such as BFO algorithm to solve new practical issues, the detailed description of the method, as well as the illustration of various applications solved by the new method. This work also focuses on exploring the possibilities of employing this hybrid intelligent system data mining method to environmental as well as biological applications. The two domains attract a lot of attention in recent times, not only due to the complexity of the issue, but also due to the large amounts of data which is available and constantly on the rise. Result show that the HBFO-FCM has higher classification accuracy in Wisconsin dataset by 4.73% for FCM and by 2.16% for BFO-FCM method. The HBFO-FCM has higher classification accuracy

in broad institute dataset by 4.94% for FCM and by 3.68% for BFO-FCM method.

## References

1. Tomašev N., Radovanović M., Mladenčić D. and Ivanović M., Hubness-based fuzzy measures for high-dimensional k-nearest neighbor classification, In International Workshop on Machine Learning and Data Mining in Pattern Recognition, Springer Berlin Heidelberg, 16-30 (2011)
2. Chitra R. and Seenivasagam V., Heart Attack Prediction System Using Fuzzy C Means Classifier, *IOSR Journal of Computer Engineering*, **14**, 23-31 (2013)
3. Yahya A.A., Osman A., Ramli A.R. and Balola A., Feature selection for high dimensional data: an evolutionary filter approach, *Journal of Computer Science*, **7**(5), 800-820 (2011)
4. Binsy Thomas and Madhu Nashipudimath, Comparative Analysis of Fuzzy Clustering Algorithms In Data Mining, *International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE)*, **1**(7), 221-225 (2012)
5. Beenu S.K., Image segmentation using improved bacterial foraging algorithm, *Int. J. Sci. Res.*, **2**(1), 63-69 (2013)
6. Lee C.L. and Lin C.J., A Novel Strategy Adaptation Based Bacterial Foraging Algorithm for Numerical Optimization, *International Journal of Swarm Intelligence and Evolutionary Computation*, **5**(1), 1-6 (2016)
7. Grover N., A study of various Fuzzy Clustering Algorithms, *International Journal of Engineering Research (IJER)*, **3**(3), 177-181 (2014)
8. Izakian H., Abraham A. and Sná V., Fuzzy clustering using hybrid fuzzy c-means and fuzzy particle swarm optimization. In Nature & Biologically Inspired Computing, NaBIC 2009, World Congress, IEEE, 1690-1694 (2009)
9. Sharma R.P.M. and Sscet B., Diabetes Prediction by using Bacterial Foraging Optimization Algorithm and Artificial Neural Network, *International Journal of Computer Science and Information Technology & Security*, **6**(1), 557-564 (2016)
10. Wan M., Li L., Xiao J. Wang C. and Yang Y., Data clustering using bacterial foraging optimization, *Journal of Intelligent Information Systems*, **38**(2), 321-341 (2012)
11. Chen C.H., Su M.T., Lin C.J. and Lin C.T., Hybrid of bacterial foraging optimization and particle swarm optimization for evolutionary neural fuzzy classifier, *Int. J. Fuzzy Syst.*, **16**(3), 422-433 (2014)
12. Salama G.I., Abdelhalim M. and Zeid M.A.E., Breast cancer diagnosis on three different datasets using multi-classifiers, *Breast Cancer (WDBC)*, **32**(569), 2 (2012)
13. Kim D.H., Abraham A. and Cho J.H., A hybrid genetic algorithm and bacterial foraging approach for global optimization, *Information Sciences*, **177**(18), 3918-3937 (2007)

14. Abraham A., Bacterial Foraging Optimization Algorithm: Theoretical Foundations, Analysis and Applications, *Foundations of Computational Intelligence*, **3**, 23-55 (2009)

15. Mo H. and Yin Y., Image segmentation based on bacterial foraging and fcm algorithm, *International Journal of Swarm Intelligence Research*, **2(3)**, 16-28 (2012).