Research Article

CBO-IE: A Data Mining Approach for Healthcare IoT Dataset Using Chaotic Biogeography-Based Optimization and Information Entropy

Manish Kumar Ahirwar ,[1] **Piyush Kumar Shukla** ,[1] **and Rakesh Singhai** [2]

[1]*Department of Computer Science and Engineering, UIT-RGPV, Bhopal 462033, India*
[2]*University Institute of Technology RGPV, Shivpuri, M.P. 473551, India*

Correspondence should be addressed to Piyush Kumar Shukla; pphdwss@gmail.com

Data mining is mostly utilized for a huge variety of applications in several fields like education, medical, surveillance, and industries. The clustering is an important method of data mining, in which data elements are divided into groups (clusters) to provide better quality data analysis. The Biogeography-Based Optimization (BO) is the latest metaheuristic approach, which is applied to resolve several complex optimization problems. Here, a Chaotic Biogeography-Based Optimization approach using Information Entropy (CBO-IE) is implemented to perform clustering over healthcare IoT datasets. The main objective of CBO-IE is to provide proficient and precise data point distribution in datasets by using Information Entropy concepts and to initialize the population by using chaos theory. Both Information Entropy and chaos theory are facilitated to improve the convergence speed of BO in global search area for selecting the cluster heads and cluster members more accurately. The CBO-IE is implemented to a MATLAB 2021a tool over eight healthcare IoT datasets, and the results illustrate the superior performance of CBO-IE based on *F*-Measure, intracluster distance, running time complexity, purity index, statistical analysis, root mean square error, accuracy, and standard deviation as compared to previous techniques of clustering like K-Means, GA, PSO, ALO, and BO approaches.

1. Introduction

The big data [1, 2] is used and analyzed in several wireless applications by utilizing various characteristics like storage, processing, and maintenance of data. The preprocessing of a huge amount of data is performed before analysis to reduce the data redundancy with enhancing the data accuracy and efficiency [3–5]. The big data is processed by using various nature inspired optimization approaches like Genetic Algorithm (GA), Ant Colony Optimization (ACO), Ant Lion Optimization (ALO), and Particle Swarm Optimization (PSO) to perform optimal analysis [6].

The data mining [7–9] is a systematic procedure utilized for extracting the secret knowledge and model from an immense, multifaceted, and multidimensional dataset [10]. The association rule mining is one of the famous data mining

approaches introduced with MapReduce concept to evaluate the relationship among data elements in a huge dataset [11]. These data relationships are recognized to generate maximum profit from marketing through IoT devices in industries. The time and security are crucial issues in business, which are frequently resolved by using data mining [12–14].

The data clustering [15] is a type of data mining, in which data components are divided into various sets or groups. The data are collected from various heterogeneous resources; after that time series clustering is applied on this huge amount of data to improve the data accessibility. The industry data are distributed more precisely and accurately by clustering to predict the future aspects of market [16]. The cybercrime data are analyzed after preprocessing to perform training and testing of clustering techniques. The nature of crime is easily understood and detected by clustering the

similar type of crime patterns to reduce the crime ratio [17]. The healthcare data are also received by IoT devices in the form of data stream and divided into cluster by using stream clustering combining the cluster building and merging steps [18].

In above analysis, the clustering of data is an extremely challenging issue, while it creates baffling to select the optimal clustering strategy in nature. Additionally, it is to be an incredibly exigent attempt as every dataset is not expected to be consistent, allowing for the fact that infection form and previous medical circumstance area might differ immensely in training. Previously, having a meticulous prophecy scheme does not constantly offer a valuable height of exactness beneath entire healthcare application areas due to mostly depending on the situation utilized.

A lot of clustering algorithms are introduced on various datasets to enhance the data extraction. Several optimization approaches like GA, ALO, and PSO are also described for clustering. Although few clustering techniques realize an advanced output by means of a specified dataset, the efficiency of such clustering techniques might be complementary on other datasets. The nature of the clustering methods is steady by means of no-free-lunch theorem, but at hand no solitary clustering method survives which is able to be a cure for entire problems. This means all the problems are not resolved by any algorithm and convergence speed is also a major issue in optimization techniques, which are used for data clustering.

Here, a Chaotic Biogeography-Based Optimization approach using Information Entropy (CBO-IE) is implemented for clustering over healthcare IoT datasets, which is improved form of Biogeography-Based Optimization (BO) approach. In its place of basically performing a qualitative examination by utilizing a methodical mapping analysis regarding prior implemented works, this proposed work contributes on a quantitative examination of clustering strategy for healthcare dataset.

The contribution of implementing CBO-IE is as follows:

(1) The major intent of CBO-IE is to obtain accurate data point distribution in dataset by utilizing Information Entropy strategy and to give initial values of the population by utilizing chaos theory.

(2) Both Information Entropy and chaos theory are introduced to enhance the convergence speed of BO in exploration field globally for generating the cluster heads and cluster members more precisely.

(3) The CBO-IE is applied over eight healthcare IoT datasets and the outputs are evaluated on the basis of *F*-Measure, intracluster distance, running time complexity, purity index, statistical analysis, standard deviation, root mean square error, and accuracy as compared to previous techniques of clustering like K-Means, GA, PSO, ALO, and BO approaches.

The remainder of the paper is organized as follows: Section 2 explains the literature survey of data mining techniques, big data processing, and IoT dataset processing by utilizing numerous parameters. Section 3 describes the

BO approach, and the proposed CBO-IE is explained in Section 4 briefly according to flowchart, algorithms, and preliminaries. The datasets, performance factors, and experimental results are illustrated in Section 5, and at last conclusions are given in Section 6.

2. Literature Survey

The IoT [19] is situated to transfigure entire future consequently for lives. The information extracted from the IoT systems will be explored to recognize and organize multifaceted surroundings about users, fitting superior decision creation, larger computerization, advanced effectiveness, efficiency, exactness, and prosperity production. This work explores the accessibility of multiple popular data mining approaches for IoT information. The results are evaluated over artificial intelligence and neural network with superior accuracy, efficiency, and velocity in comparison to prior techniques [20].

The K-Means approach is applied for processing the big data with the help of Hadoop platform over Internet of Things (IoT) aspects. The data from IoT devices are combined or distributed in several groups, known as clusters, which are used in various smart applications such as medical, disaster, and industries applications. The power expenses and traffic over communication network was reduced to use only necessary information, not whole raw data. In actuality, it supplies enormous flexibility, permitting the formation of clusters having millions of Hadoop illustrations [21].

This big data has raised the concept of space and time complexity due to storing and processing huge amount of data over mobile and dynamic network. The Extensible Mark-up Language (XML) and machine learning methods apply to big data processing like classification, clustering, and preprocessing of IoT information [22, 23]. Latest IoT applications and performances remain further on an intellectual indulgence of the surroundings from information extracted by means of assorted sensors and small machines. Here a framework combines an ontology dependent description of information dispensations by means of unusual logic for pretty beneficial incident recognition by seasoning the specific classification strategy of machine learning. A road and traffic examination is performed to verify the results of framework [24].

The personal data of users are digitally distributed over network through smart IoT devices and multiple artificial techniques are introduced to control the digitally transferred information following few rules and policies. An intellectual police examination strategy is introduced, discovering the contradictory policy or agreements of users over advocate platform. An intellectual decision-making method is applied with the help of fuzzy cognitive maps [25].

The business is directly or indirectly related to customer behaviour, evaluated by trust mining. The trust of sellers is evaluated by dynamic clustering, where data is modified day by day over huge network area. The quick development of online shopping details illustrates a growing issue for customers who are comprised to select faithful sellers and

successful seller selections from numerous existing record beginning e-commerce usage areas to enhance the business. A dynamic clustering method is adopted to calculate the trust of customers and differentiate the buying equivalence of customers for predicting the customer behaviour. The real-time and artificial datasets are utilized to perform clustering and evaluate the results based on accuracy [26].

The medical field [27] also well utilizes the mining techniques like classification to identify the mental health and evaluate the risk to manage the system. This classification is done by decision trees to discover and classify the patients according to their mental health. Here, a relative calculation of a huge array classifier is associated with several areas like tree, group, neural, possibility, categorization, and policy dependent classifiers. The linear and random classifiers are introduced to perform disease classification [28].

The cancer disease in unstructured format is well classified to discover several types of cancer disease. It generates valuable results in terms of risk aspects, treatment, and management. The text mining is well exploited in cancer disease prediction like lung, breast, and ovarian cancer. The data about cancer patients is collected from several heterogeneous resources and text mining is performed to provide various cancer related information like survival, treatment, and risk of disease [29].

The geographical data is collected from several heterogeneous resources and data mining is initialized over the spatial data to manage the disaster. The natural hazards are previously identified before happening; this is achieved by spatial data mining over huge geographic information. The several geographical information such as soil, ocean, earth's crust, and air quality is used to employ the mining process with the help of PSO and fuzzy logics. The clusters are decided on the basis of spatial information and natural behaviours in environments [30].

The data mining based educational information extraction is a new era of research, in which distributed educational data is compiled and used for several purposes. The students, teachers, and other staffs have played an important role in educational data mining. The educational data is collected from several resources and saved in tabular format. These student data can be further analyzed by university and examined by data mining model to evaluate the results of students and maintain the records in appropriate manner [31].

The classification is also performed over data streams, combining the text, music, and video information with the help of decision trees and naive structure. These data streams also combined the human activities like movement of hands and legs [32, 33]. There are various key issues of classification like infinite size, perception development, and perception flow and characteristic assessment. Various researchers mainly solved the issues of infinite size and perception development, but the perception flow and characteristic assessment are not taken into consideration by researchers. Here, these two issues are removed in implementing a classification method over big data and parameters like error rate and running time are obtained to provide superior efficiency of the method [34].

3. The Biogeography-Based Optimization (BO) Approach

A Biogeography-Based Optimization (BO) is a metaheuristic approach based on island biogeography premise, which relates with the migration, evolution, and destruction of the castes in a habitat. A Habitat Suitability Index (HSI) is utilized to generate the best solution globally and share the characteristics with week habitat. The population is enhanced by taking the optimal solutions from prior iteration from emigrating [35] to immigrating habitats by using the migrating Suitability Index Variables (SIVs) in BO migrating process. A fresh characteristic is obtained in complete solution area using fitness function [36] and replaced every habitat's SIVs arbitrarily and stochastically to enhance the population [37] miscellany and searching strength in BO mutation process.

3.1. Migration Process. In migration stochastic process, several parents can be utilized for a particular offspring and every habitat (H_a) is modified by receiving the SIVs from a superior HSI habitat in population (P^S). The migration rates are straightly dependent on the caste number in a habitat for enhancing [38] the habitat miscellany and linear migration is evaluated by using the following equation:

$$H_a(\text{SIV}) = H_b(\text{SIV}), \tag{1}$$

where $H_b(\text{SIV})$ is a superior HSI b^{th} habitat (H_b) to transfer the SIV value to the ath habitat (H_a).

The emigration rate (α) and immigration rate (β) are evaluated for c castes in habitat by using the two following equations:

$$\alpha_c = \frac{G_c}{N_{\text{maximum}}}, \tag{2}$$

$$\beta_c = M \times (1 - \alpha_c). \tag{3}$$

Here, G and M are highest emigration and immigration rates, respectively, and N_{maximum} is maximum realizable number of castes utilized by habitat.

There are six models developed for migration phase, out of which sinusoidal model performed superior migration as compared to the other five models. This model is evaluated for a^{th} habitat by using the two following equations:

$$\beta_a = \frac{M}{2}\left(1 + \cos\left(\frac{a\pi}{N_{\text{maximum}}}\right)\right), \tag{4}$$

$$\alpha_a = \frac{G}{2}\left(1 - \cos\left(\frac{a\pi}{N_{\text{maximum}}}\right)\right). \tag{5}$$

3.2. Mutation Process. A mutation is a stochastic operator enhancing the population miscellany to obtain optimal solution [39]. The mutation process is utilized for updating at least one arbitrary chosen SIV of a solution with the help

of mutation rate (T_a) and the previous possibility of subsistence (S_a) by using the following equation:

$$T_a = T_{\text{maximum}}\left(1 - \frac{S_a}{S_{\text{maximum}}}\right), \qquad (6)$$

where T_{maximum} is the highest mutation rate term illustrated by user and S_{maximum} is the highest possibility of castes count.

After that the mutation operator is improved by utilizing Cauchy model to provide optimal exploration strength of BO in huge searching area by reducing potential limitation of mutation strategy. The Cauchy distribution is evaluated with the help of probability density function, which is formulized by the following equation:

$$F(X; 0, 1) = \frac{1}{\pi(1 + X^2)}. \qquad (7)$$

Hence, the Cauchy mutation equation is described by using the following equation:

$$H_a(\text{SIV}) = \text{minimum}(H_a(\text{SIV})) + (\text{maximum}(H_a(\text{SIV})) - \text{minimum}(H_a(\text{SIV}))) \times \pi \times F(H_a(\text{SIV}); 0, 1), \qquad (8)$$

where $H_a(\text{SIV})$ is a^{th} habitat. $F(H_a(\text{SIV}); 0, 1)$ is Cauchy distribution (Algorithm 1).

4. The Chaotic BO Approach Using Information Entropy (CBO-IE)

An updated BO approach is proposed using the chaotic behaviour and Information Entropy. The BO approach is well suitable for exploration and exploitation in huge searching space and generates the efficient results for numeric optimization [40]. Still, the numeric optimization is reasonably different from data clustering mechanism. Here, some intrinsic characteristics are investigated and a chaotic BO is implemented using Information Entropy to apply the BO for data clustering.

4.1. Element Information Entropy. The data points are distributed in multidimensional area in data clustering strategy. The confusion degree of a probabilistic variable is calculated by utilizing Information Entropy; after that, this is utilized for elements measurements to calculate the distribution. The entropy (E_a) is evaluated to isolate the element values with estimation of every value to its closer integer by using the following equation:

$$E_a = - \sum_{b=\text{min}_a}^{\text{max}_a} R_b \log(R_b), \qquad (9)$$

where E_a is the entropy of ath element ($a = 1, 2, 3, ..., L$) and L is number of elements in dataset. min_a and max_a are lowest and highest integers after discretization of element values. R_b is b^{th} element integer percentage value.

The migration procedure of BO population is guarded by using element entropy; hence, maximum entropy is evaluated for every element in dataset by the following equation:

$$\text{maximum}(E_a) = -\log\left(\frac{1}{\text{max}_a - \text{min}_a + 1}\right), \qquad (10)$$

where maximum (E_a) is maximum entropy of ath element in dataset.

At last, the normalized entropy for every element is calculated by the following equation:

$$\text{nomalized}(E_a) = \frac{E_a}{\text{maximum}(E_a)}, \qquad (11)$$

where normalized (E_a) is normalized entropy of ath element in dataset.

Equation (11) is repeated for all data elements in dataset and normalized entropy set (normalized (E)) is generated as follows:

Normalized (E) = (normalized (E_1), normalized (E_2),..........., normalized (E_L)).

4.2. Information Entropy-Based Migration Process in the BO Approach. Here, two mechanisms are introduced in migration to improve the performance of BO approach. The first one is original migration process of BO explained in the previous section. The second one is updated mechanism of original migration in which P particulars are elected arbitrarily from present population ($1 \leq P \leq P^S$) and the one optimal particular is assigned as reference particular (H_{ref}). The direction of migration is guided by H_{ref} and Information Entropy is utilized to provide equivalency between population miscellany and convergence speed. The higher Information Entropy of element indicates the maximum uncertainty, which slows down the convergence speed. Hence, the speed is enhanced for element by transferring it to the location based on reference particular, globally with maximum possibility, as compared to the least Information Entropy element.

4.3. Chaos-Based Population Selection of the BO Approach. The chaos method is extremely related to initial circumstances and successfully utilized for arbitrary number generation using logistic function. The chaotic system is used in the following equation:

$$\gamma_{f+1} = \mu \times \gamma_f \times (1 - \gamma_f), \qquad (12)$$

where μ represents constants in the range of [1, 4]. γ represents variables ($\gamma \in [0, 1]$, $f = 0, 1, \ldots$).

The population of BO is initialized by utilizing chaos function (equation (12)), which has improved the efficiency of BO with proper use of huge solution area.

4.4. The Complete CBO-IE Approach for Data Mining. The Information Entropy-based migration and chaos-based population selection are combined in Chaotic Biogeography-Based Optimization using Information Entropy (CBO-IE) to solve the data mining, that is, data clustering in optimal way (Figure 1). Firstly, the population (P^S) of BO approach is initialized with the help of chaos function

```
 START
Initialize the parameters G = M = 1, T_maximum = 1, P^S and Max_iteration
Initialize the populations (arbitrary group of habitats) H_1, H_2,........, H_P^s
Evaluate fitness value (HSI) for every habitat
WHILE (ending condition is not found)
    Evaluate α_a, β_a and T_a for every habitat
    Obtain a random ∈ (0, 1)//Migration
    FOR every habitat (from minimum to maximum HSI values)
        Choose habitat H_a (SIV) stochastically proportional to β_a
        IF random < β_a and H_a (SIV) choose, then
            Choose habitat H_b (SIV) stochastically proportional to α_a
            IF random < α_a and H_b (SIV) choose, then
                H_a (SIV) = H_b (SIV)
            END IF
        END IF
    END FOR//MUTATION
    Choose H_a (SIV) with the help of mutation possibility proportional S_a
    IF random < T_a then
        Arbitrary change the SIVs in H_a (SIV)
    END IF
    Evaluate HSI value
END WHILE
STOP
```

ALGORITHM 1: Biogeography-Based Optimization.

(equation (12)) and every particular is denoted as a vector with size $L_x = L \times K$ (L is number of elements in dataset, K is number of cluster centroids, and L_x is dimension of particulars). The K centroids locations are fixed into vector, in which 1st centroid relates with 1st L attributes, 2nd centroid relates to 2nd L attributes, and so on. The primary particular vector values are generated arbitrarily and consistently between lower and higher elements values in existing dataset within maximum number of iterations (Mi). Then, CBO-IE is applied to calculate the fitness values of entire particulars by using equations (1) to (8) (Algorithm 2).

5. Results and Analysis

In this section, the healthcare IoT datasets and performance factors for proposed CBE-IE approach are explained briefly. The entire approaches are implemented with the help of MATLAB 2021a tool using Core i3-3110M processor having Windows 8 operating system with 2 GB RAM and analyzed over 8 healthcare IoT datasets (Table 1). The proposed CBO-IE approach has evaluated the results for 500 iterations in terms of intracluster distance, purity index, standard deviation, root mean square error, accuracy, and F-Measure as compared to prior algorithms like K-Means, ACO, PSO, ALO, and BO over 30 independent runs.

5.1. Healthcare IoT Datasets.
The proposed CBO-IE approach is applied on 8 dissimilar healthcare IoT datasets taken from UCI repository (Table 1). The datasets are thoracic surgery, breast cancer, cryotherapy, liver patient, heart patient, chronic kidney disease, diabetic retinopathy, and blood transfusion. The 470 instances of thoracic surgery dataset with

17 attributes are divided into 2 classes (survive and not survive), the 699 instances of breast cancer dataset with 10 attributes are distributed into 2 classes (benign and malignant), the 90 instances of cryotherapy dataset with 7 attributes are divided into 2 classes (success and failure), the 583 instances of liver patient dataset with 10 attributes are distributed into 2 classes (liver patient and not), the 299 instances of heart patient dataset with 12 attributes are divided into 2 classes (heart failure and not), the 400 instances of chronic kidney disease with 24 attributes are divided into 2 classes (disease found and not), the 1151 instances of diabetic retinopathy dataset with 18 attributes are distributed into 2 classes (diabetic retinopathy and not), and the 748 instances of blood transfusion dataset with 4 attributes are divided into 2 classes (donating blood and not).

5.2. Performance Factors.
The performance of proposed CBO-IE approach is evaluated in terms of intracluster distance, purity index, standard deviation, root mean square error, accuracy, and F-Measure.

5.2.1. Intracluster Distance.
Firstly, the distances between data points are evaluated within a cluster. After that, the average of these distances is generated representing an intracluster distance. The optimal clustering is achieved with least intracluster distance. For a cluster, the mean distance is evaluated between a centroid and total data points. This step is continuously performed for every cluster and at last mean value of entire clusters' intracluster distances is obtained.

Table 2 illustrates that the CBO-IE generates least intracluster distance value for all IoT datasets. The CBO-IE obtains 90% superior results than BO, 92% superior results
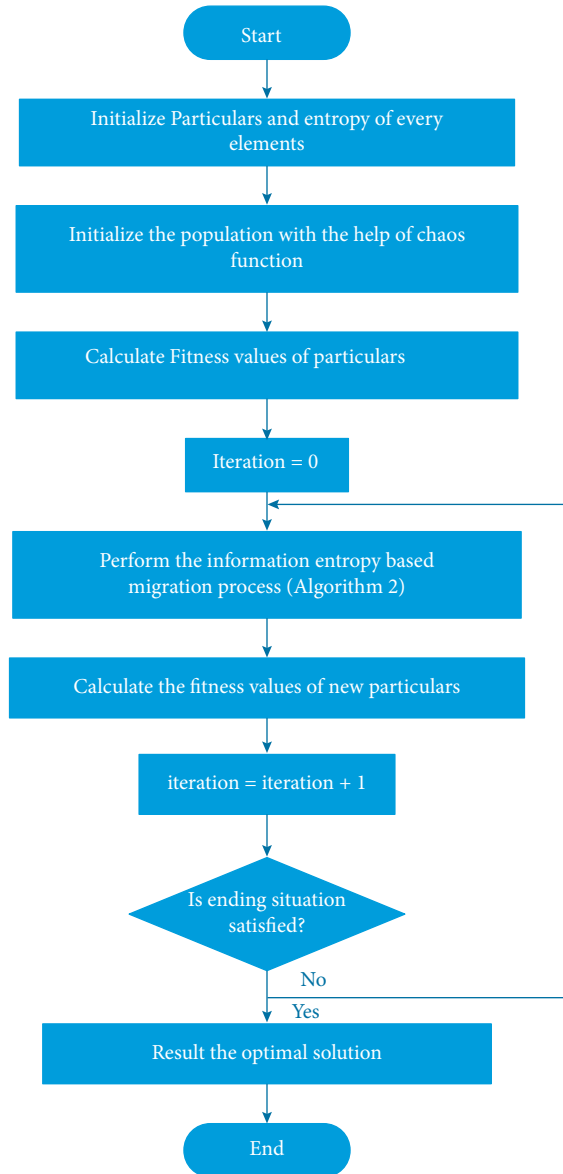
FIGURE 1: Flowchart of the CBO-IE approach.

than ALO, 94% superior results than PSO, 96% superior results than GA, and 98% superior results than K-Means in terms of intracluster distance for entire healthcare IoT datasets. The average ranking of all approaches is evaluated on the basis of minimum to maximum intracluster distance (from 1 to 6).

Figure 2 represents the better rank of proposed CBO-IE with least intracluster distance values against other approaches, K-Means, GA, PSO, ALO, and BO, for all IoT datasets.

5.2.2. Standard Deviation. The stiff data clustering in the region of average value is described by a numerical characteristic called standard deviation (D^S). The best clustering is achieved with less standard deviation, which is obtained by the following equation:

$$D^S = \sqrt{\frac{\sum (V - \overline{V})}{|H|}}, \tag{13}$$

where $|H|$ is dataset length, V is dataset values (points), and $\overline{V}$ is dataset average value.

Table 3 describes that the CBO-IE obtains less standard deviation value for all IoT datasets. The CBO-IE generates 93% better quality outcomes than BO, 95% better quality outcomes than PSO and ALO, 97% better quality outcomes than GA, and 98% better quality outcomes than K-Means in terms of standard deviation for entire healthcare IoT datasets.

Figure 3 shows the better standard deviation of proposed CBO-IE against other approaches, K-Means, GA, PSO, ALO, and BO, for all IoT datasets.

```
Algorithm                                                          Number of Executions

START

FOR a = 1 to P^S DO                                                (P^S+1)

        Choose the optimal one from P arbitrary particulars as H_ref    P^S

        FOR b = 1 to L_x DO                                        P^S*(L_x+1)

                IF random < normalized (E_a) then                 P^S*L_x

                        H_a(SIV) = H_ref(SIV)

                ELSE

                        H_a(SIV) = H_b(SIV)

                END IF

        END FOR

END FOR

STOP
```

ALGORITHM 2: Information Entropy-based migration process in the BO approach.

TABLE 1: Healthcare IoT datasets.

Sr. no.	Healthcare IoT dataset	No. of instances	No. of attributes	No. of clusters
1	Thoracic surgery	470	17	2
2	Breast cancer	699	10	2
3	Cryotherapy	90	7	2
4	Liver patient	583	10	2
5	Heart patient	299	12	2
6	Chronic kidney disease	400	24	2
7	Diabetic retinopathy	1151	18	2
8	Blood transfusion	748	4	2

TABLE 2: Average ranking of all algorithms based on intracluster distance.

Dataset	K-Means	GA	PSO	ALO	BO	CBO-IE
Thoracic surgery	3.364 (5)	0.7569 (4)	0.4256 (4)	0.2547 (3)	0.0657 (2)	0.00652 (1)
Breast cancer	26.265 (5)	0.8634 (4)	0.3126 (4)	0.2614 (3)	0.0874 (2)	0.00728 (1)
Cryotherapy	24.3640 (5)	0.4562 (4)	0.2864 (4)	0.2167 (3)	0.0725 (2)	0.00548 (1)
Liver patient	35.3245 (5)	0.5863 (4)	0.1567 (4)	0.1321 (3)	0.0635 (2)	0.00457 (1)
Heart patient	14.2366 (5)	0.9362 (4)	0.7265 (4)	0.6492 (3)	0.0623 (2)	0.00832 (1)
Chronic kidney disease	46.3625 (5)	0.5682 (4)	0.3568 (4)	0.2864 (3)	0.0742 (2)	0.00786 (1)
Diabetic retinopathy	0.3658 (5)	0.2526 (4)	0.1737 (4)	0.1421 (3)	0.0967 (2)	0.00758 (1)
Blood transfusion	0.6745 (5)	0.5684 (4)	0.2375 (4)	0.2068 (3)	0.0854 (2)	0.00852 (1)
Average ranking (AK$_\theta$)	5	4	4	3	2	1

5.2.3. Purity Index. The correctness of clustering strategy is known as purity, in which accurate classification is performed over data elements. Hence, entire points of an isolated class can be accurately allocated to an isolated cluster. Purity index (I^P) is generated with the help of purity by equations (14) and (15). Maximum purity is achieved with maximum I^P value nearer to 1.

$$\text{purity}(R_z) = \frac{\text{maximum}\left(|R_{wz}|\right)}{|R_z|}, \tag{14}$$

$$I^P = \sum_{z=1}^{K} \frac{\left(|R_z|\text{Purity}(R_z)\right)}{|H|}, \tag{15}$$
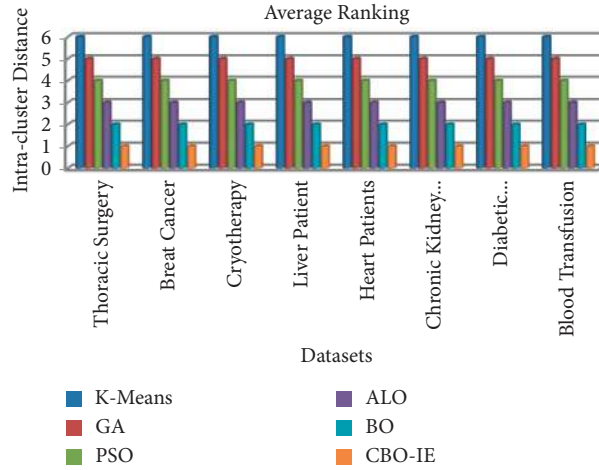
FIGURE 2: Average ranking based on intracluster distance.

TABLE 3: Standard deviation for entire IoT datasets.

Dataset	K-Means	GA	PSO	ALO	BO	CBO-IE
Thoracic surgery	0.116857	0.03984	0.012036	0.01452	0.007856	0.000524
Breast cancer	0.002578	0.168742	0.12542	0.15713	0.003657	0.000746
Cryotherapy	0.0025143	0.165875	0.001254	0.001674	0.001036	0.000845
Liver patient	0.256323	0.17589	0.147856	0.12312	0.096875	0.000986
Heart patient	0.212678	0.096852	0.042574	0.021563	0.003687	0.000251
Chronic kidney disease	0.020233	0.103625	0.09854	0.065214	0.006574	0.000865
Diabetic retinopathy	0.086574	0.186523	0.142657	0.102647	0.086478	0.0008657
Blood transfusion	0.213658	0.196875	0.185632	0.152461	0.010365	0.000236
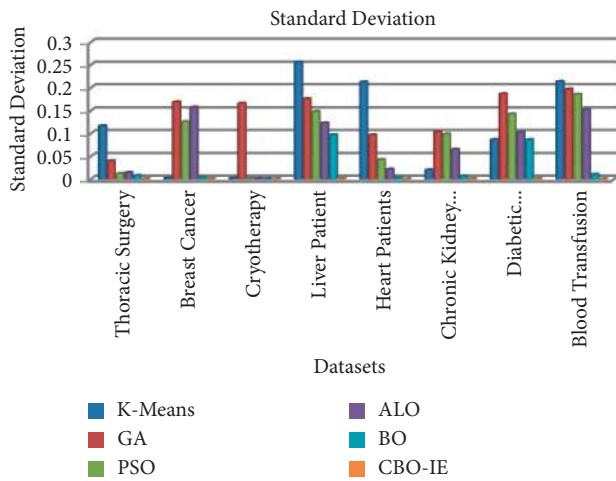
FIGURE 3: Standard deviation for all IoT datasets.

TABLE 4: Purity index for entire IoT datasets.

Dataset	K-Means	GA	PSO	ALO	BO	CBO-IE
Thoracic surgery	0.83	0.86	0.89	0.89	0.91	0.93
Breast cancer	0.86	0.88	0.90	0.89	0.93	0.95
Cryotherapy	0.84	0.87	0.90	0.91	0.93	0.96
Liver patient	0.86	0.87	0.88	0.86	0.90	0.92
Heart patient	0.80	0.83	0.86	0.86	0.88	0.92
Chronic kidney disease	0.86	0.87	0.90	0.89	0.91	0.93
Diabetic retinopathy	0.88	0.90	0.91	0.91	0.93	0.95
Blood transfusion	0.76	0.78	0.81	0.83	0.86	0.88

where Purity (R_z) is purity of zth cluster. $|R_z|$ is z^{th} cluster length. $|R_{wz}|$ is the number of data elements of wth class allocated to z^{th} cluster.

Table 4 explains that the CBO-IE generates maximum purity index value for all IoT datasets. The CBO-IE obtains 5% superior outputs than BO, 7% superior outputs than ALO, 8% superior outputs than PSO, 12% superior outputs than GA, and 15% superior outputs than K-Means in terms of purity index for entire healthcare IoT datasets.

Figure 4 represents the better quality purity index of proposed CBO-IE against other approaches, K-Means, GA, PSO, ALO, and BO, for all IoT datasets.

5.2.4. F-Measure. Firstly Precision (prsn) and Recall (rel) are evaluated to repossess the information by equations (16) and (17). After that, both are combined to formalize the *F*-Measure (FM) by utilizing equations (18) and (19).
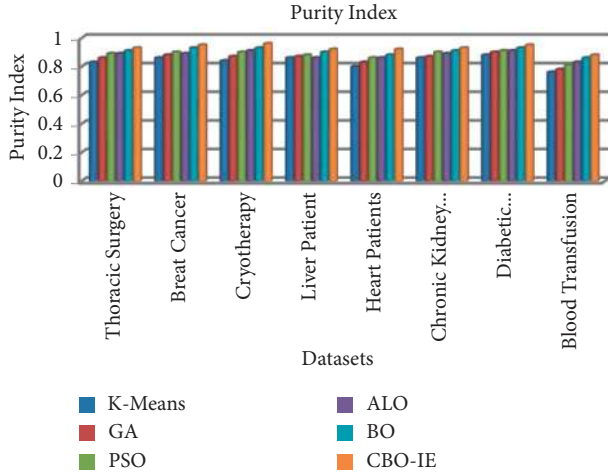
FIGURE 4: Purity index for all IoT datasets.

$$\text{prsn}\,(w, z) = \frac{|R_{wz}|}{|R_z|}, \qquad (16)$$

$$\text{rel}\,(w, z) = \frac{|R_{wz}|}{|R_w|}, \qquad (17)$$

$$\text{FM}\,(w, z) = \frac{2 \times \text{prsn}\,(w, z) \times \text{rel}\,(w, z)}{\text{prsn}\,(w, z) + \text{rel}\,(w, z)}, \qquad (18)$$

$$FM = \sum_{w=1}^{K} \frac{|R_w|}{|H|} \text{maximum}\{FM\,(w, z)\}, \qquad (19)$$

where $|R_w|$ is wth class length.

Table 5 describes that the CBO-IE obtains higher F-Measure value for all IoT datasets. The CBO-IE generates 4% better outputs than BO, 7% better outputs than ALO, 8% better outputs than PSO, 13% better outputs than GA, and 16% better outputs than K-Means in terms of F-Measure for entire healthcare IoT datasets.

Figure 5 shows the better F-Measure of proposed CBO-IE against other approaches, K-Means, GA, PSO, ALO, and BO, for all IoT datasets.

5.2.5. Root Mean Square Error (RMSE). The RMSE is defined as a divergence between predicted values and experimental (calculated) values. The best clustering is achieved with minimum RMSE values of datasets. It is evaluated by utilizing the following equation:

$$\text{RMSE} = \sqrt{\frac{1}{|H|} \sum_{j=1}^{|H|} \left(V_j - \widehat{V}_j \right)^2}, \qquad (20)$$

where V_j is j^{th} element value in dataset and $\widehat{V}_j$ is j^{th} element predicted value in dataset.

Table 6 explains that the CBO-IE generates least RMSE value for all IoT datasets. The CBO-IE generates 8% better results than BO, 78% better results than ALO, 81% better

TABLE 5: *F*-Measure for entire IoT datasets.

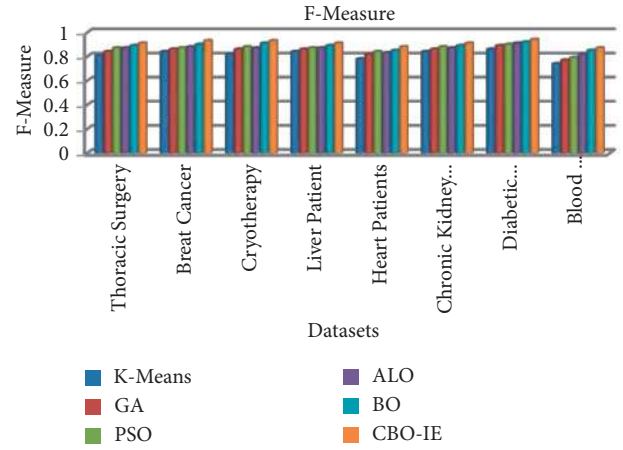| Dataset	K-Means	GA	PSO	ALO	BO	CBO-IE
Thoracic surgery	0.81	0.84	0.87	0.87	0.89	0.91
Breast cancer	0.84	0.86	0.87	0.88	0.90	0.93
Cryotherapy	0.82	0.86	0.88	0.87	0.91	0.93
Liver patient	0.84	0.86	0.87	0.87	0.89	0.91
Heart patient	0.78	0.82	0.84	0.83	0.85	0.88
Chronic kidney disease	0.84	0.86	0.88	0.87	0.89	0.91
Diabetic retinopathy	0.86	0.89	0.90	0.91	0.92	0.94
Blood transfusion	0.74	0.77	0.79	0.82	0.85	0.87

FIGURE 5: *F*-Measure for all IoT datasets.

TABLE 6: RMSE values for entire IoT datasets.

Dataset	K-Means	GA	PSO	ALO	BO	CBO-IE
Thoracic surgery	0.537	0.327	0.157	0.124	0.0124	0.0034
Breast cancer	0.567	0.386	0.168	0.264	0.0317	0.0075
Cryotherapy	0.621	0.374	0.234	0.364	0.0421	0.0066
Liver patient	0.437	0.361	0.261	0.213	0.0652	0.0054
Heart patient	0.537	0.412	0.201	0.186	0.0145	0.0063
Chronic kidney disease	0.610	0.422	0.341	0.267	0.0254	0.0047
Diabetic retinopathy	0.517	0.414	0.367	0.257	0.0374	0.0087
Blood transfusion	0.438	0.314	0.120	0.265	0.0451	0.0071

results than PSO, 87% better results than GA, and 95% better results than K-Means in terms of RMSE for entire healthcare IoT datasets.

Figure 6 shows the better RMSE of proposed CBO-IE against other approaches, K-Means, GA, PSO, ALO, and BO, for all IoT datasets.

5.2.6. Accuracy. It evaluates the division of clusters that are accurate (i.e., it evaluates proportion of verdicts that are exact) and describes the portion of clusters in the prevailing group:
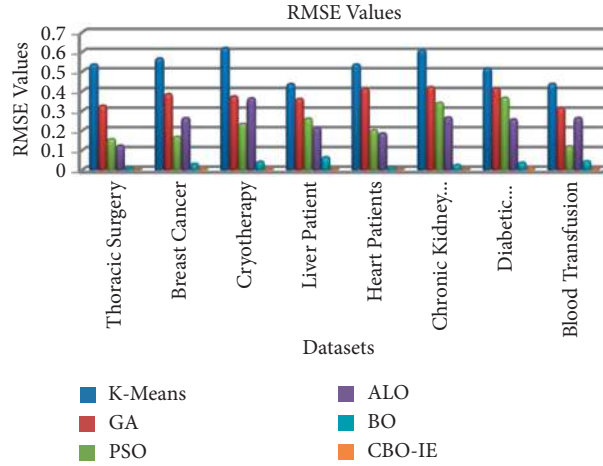
FIGURE 6: RMSE for all IoT datasets.

TABLE 7: Accuracy values for entire IoT datasets.

Dataset	K-Means	GA	PSO	ALO	BO	CBO-IE
Thoracic surgery	0.80	0.83	0.86	0.86	0.88	0.90
Breast cancer	0.83	0.85	0.86	0.87	0.89	0.92
Cryotherapy	0.81	0.85	0.87	0.86	0.90	0.92
Liver patient	0.83	0.85	0.86	0.86	0.88	0.90
Heart patient	0.77	0.81	0.83	0.82	0.84	0.87
Chronic kidney disease	0.83	0.85	0.87	0.86	0.88	0.90
Diabetic retinopathy	0.85	0.88	0.89	0.90	0.91	0.93
Blood transfusion	0.73	0.76	0.78	0.81	0.84	0.86

$$\text{accuracy} = \frac{\sum_{z=1}^{K} R_z}{|H|} * 100\%. \qquad (21)$$

Table 7 represents that the CBO-IE obtains higher accuracy value for all IoT datasets. The CBO-IE has 3% superior outputs than BO, 5% superior outputs than ALO, 6% superior outputs than PSO, 9% superior outputs than GA, and 13% superior outputs than K-Means in terms of accuracy for entire healthcare IoT datasets.

Figure 7 represents the better accuracy of proposed CBO-IE against other approaches, K-Means, GA, PSO, ALO, and BO, for all IoT datasets.

The Information Entropy is used with BO approach for clustering to provide accurate and efficient distribution of data points in dataset. The chaos theory is utilized to initialize the population of BO to improve the searching capability of habitat, which helps in cluster member selection process more precisely. These two strategies, Information Entropy and chaos theory, have improved the performance of BO to provide the best selection of cluster heads and their members optimally. So, CBO-IE generates superior results than BO, ALO, PSO, GA, and K-Means clustering approaches.

5.3. Running Time Complexity of the CBO-IE Approach.
The number of executions is directly related to time complexity of clustering approaches to run. The
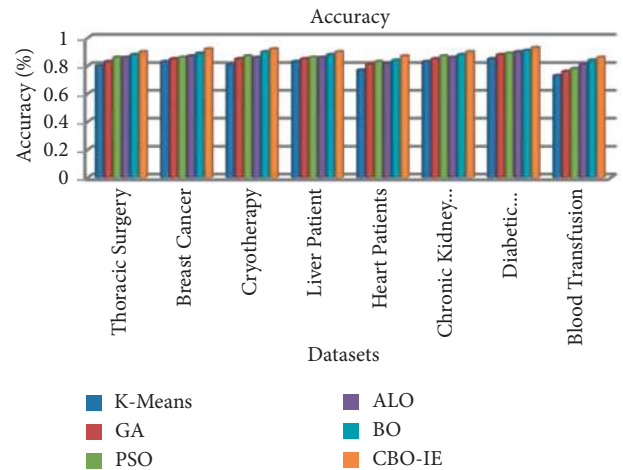

FIGURE 7: Accuracy for all IoT datasets.

complexity is calculated by utilizing few circumstances as follows: P^S is population size, L is number of elements in dataset, L_x represents dimensions of particulars (elements), K is number of cluster centroids, and Mi represents maximum iterations. The cost of every execution is assumed as 1 unit. The Cost of all Executions (CE) is obtained to utilize Algorithm 3 by the following equations:

Algorithm	Number of Executions
START	
Initialize the parameters G=M=1, $T_{maximum} = 1$ and Mi	1
Calculate the information entropy for all elements and	
obtain the entropy vector normalized(E)	L
Initialize the population P^S of BO by using chaos function (eq. 8) and	
habitats symbolized as H_a $(1 \leq a \leq P^S)$	1
Evaluate fitness value (HSI) for every habitat	P^S
WHILE (ending condition is not found)	Mi+1
Evaluate α_a, β_a and T_a for every habitat	Mi*P^S
Obtain a $random\ \epsilon\ (0,1)$ //**Migration**	Mi*P^S
Perform the Information Entropy based Migration Process (Algorithm 2)	Mi*$(2*P^S*L_x+3*P^S+1)$
FOR every Habitat (from minimum to maximum HSI values)	Mi*K*(P^S+1)
Choose habitat H_a(SIV) stochastically proportional to β_a	Mi*K*P^S
IF random < β_a and H_a(SIV) choose, then	Mi*K*P^S
Choose habitat H_b(SIV) stochastically proportional to α_a	
IF random < α_a and H_b(SIV) choose, then	Mi*K*P^S
H_a(SIV) = H_b(SIV)	
END IF	
END IF	
END FOR // **MUTATION**	
Choose H_a(SIV) with the help of mutation possibility proportional S_a	Mi*K*L
IF random < T_a then	Mi*K*L
Arbitrary change the SIVs in H_a(SIV)	
END IF	
Evaluate HSI value	Mi*K*L
END WHILE	
STOP	

ALGORITHM 3: The complete CBO-IE approach.

TABLE 8: Holm strategy outputs.

J	Approaches	z value	p value	$(\lambda/(A^C - j))$	Hypothesis
1	K-Means	-6.3245	0.00001	0.020	Rejected
2	GA	-5.0596	0.00001	0.025	Rejected
3	PSO	-3.7947	0.0001	0.033	Rejected
4	ALO	-2.5298	0.0057	0.05	Rejected
5	BO	-1.3649	0.0861	0.10	Rejected

$$CE = 1 + L + 1 + P^S + Mi + 1 + Mi \times P^S + Mi \times P^S + Mi \times \left(2 \times P^S \times L_x + 3 \times P^S + 1\right)$$
$$+ Mi \times K \times \left(P^S + 1\right) + Mi \times K \times P^S + Mi \times K \times P^S + Mi \times K \times P^S \qquad (22)$$
$$+ Mi \times K \times L + Mi \times K \times L + Mi \times K \times L,$$

$$CE = 2 \times Mi \times P^S \times L_x + 4 \times Mi \times K \times P^S + 3 \times Mi \times K \times L + 5 \times Mi \times P^S + Mi \times K + L + P^S + 2 \times Mi + 3. \qquad (23)$$

Supposing that all circumstances are equal in equation (23) in worst case, equation (24) is generated as follows:

$$CE = 9n^3 + 6n^2 + 4n + 3. \qquad (24)$$

The running time complexities of clustering approaches are O(n^3) for CBO-IE, O(n^3) for BO, O(n^3) for ACO, O(n^3) for PSO, O(n^3) for ALO, and O(n^2) for K-Menas in worst case. Hence, all are executable in polynomial time.

5.4. Statistical Examination Measurement. A statistical examination is functioned to obtain the extension of significance dissimilarities in the effectiveness of clustering techniques. Here, a nonparametric Friedman Examination (FE) is utilized to discover dissimilarities amid the group of serial appropriate variables. The entire clustering techniques provide equally effective explaining in the null hypothesis (Y_0).

The Friedman Examination (FE) is formulized by

$$FE = \frac{12 \times N^{DS}}{A^C\left(A^C + 1\right)}\left[\sum_{\theta=1}^{5}\left(AK_\theta\right)^2 - \frac{A^C \times \left(A^C + 1\right)^2}{4}\right], \qquad (25)$$

where N^{DS} is number of datasets, AK_θ is average rank of θth approach, and A^C is number of clustering approaches.

The FE critical value is 2.04925 taken from F-distribution table [41] with (A^C-1) and (A^C-1)*(N^{DS}-1) freedom degree, which is obtained between (6–1) = 5 and (6-1)*(8-1) = 35 for applying 6 clustering approaches (A^C = 6) over 8 datasets (N^{DS} = 8) with $\lambda = 0.10$ (confidence stage). The calculated FE value is higher as compared to the critical value for null hypothesis (Y_0) rejection, if Y_0 is not accepted. The evaluated FE value is 17.14 for initiating 6 clustering approaches (A^C = 6) over 8 datasets (N^{DS} = 8) with $\lambda = 0.10$. Hence, the calculated FE is higher as compared to the critical FE, and then Y_0 is rejected. So, it is to be summarized that entire clustering approaches are not equally effective.

Therefore, a post hoc examination is performed using Holm strategy. The proposed CBE-IE is analyzed statistically against other clustering approaches in this examination. Firstly the z value is obtained from equation (24) and after

that probability (p) is generated with the help of z value and normal distribution table [42]. At last, p_j value is analyzed with $(\lambda/(A^C - j))$ (Table 8):

$$Z = \frac{AK_j - AK_\theta}{\delta}, \qquad (26)$$

where

$$\delta = \sqrt{\frac{A^C\left(A^C + 1\right)}{6 \times N^{DS}}}. \qquad (27)$$

Table 8 illustrates that the $(\lambda/(A^C - j))$ value is higher as compared to p_j value; this indicates the hypothesis rejection for entire cases. Hence, the proposed CBO-IE is superior in clustering as compared to the K-Means, GA, PSO, and BO approaches according to the above analysis.

6. Conclusion

Various fields like medicine, education, and industries utilize data mining for their useful applications. The grouping of data is performed on data clustering, which is a specific task in data mining to examine the database efficiently. In this work, a Chaotic Biogeography-Based Optimization approach using Information Entropy (CBO-IE) is proposed to obtain data clusters for healthcare IoT datasets. The Information Entropy is introduced with a BO approach to generate a better distribution of data points in the dataset and chaos theory is utilized to initialize the population of BO approach. The Information Entropy and chaos theory are combined with BO approach to generate optimal cluster heads and cluster members with enhanced convergence speed of BO in huge search area. The MATLAB 2021a tool is used to implement the CBO-IE for eight healthcare IoT datasets and the outcomes describe the better quality efficiency of CBO-IE on the basis of F-Measure, intracluster distance, running time complexity, purity index, statistical analysis, standard deviation, root mean square error, and accuracy as compared to previous techniques of clustering like K-Means, GA, PSO, ALO, and BO approaches. The Friedman Examination and Holm strategy are introduced to

perform statistical analysis of proposed CBO-IE against previous techniques of clustering like K-Means, GA, PSO, ALO, and BO approaches, which represent that the CBO-IE generates 90% accurate results. In future, the proposed technique will be anticipated to estimate and be authorized with huge databases under big data. Additionally, a cross-layer communication will be anticipated to be offered and legalized in IoT structural design in the future.

Data Availability

The data that support the findings of this study are available upon request from the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

[1] A. Haque, L. Khan, and M. Baron, "SAND: semi-supervised adaptive novel class detection and classification over data stream," in *Proceedings of the 13th AAAI Conference on Artificial Intelligence (AAAI-16)*, pp. 1652–1658, Phoenix, AZ, USA, February 2016.

[2] N. Mishra, H. K. Soni, S. Sharma, and A. K. Upadhyay, "Development and analysis of artificial neural network models for rainfall prediction by using time-series data," *I. J. Intelligent Systems and Applications*, vol. 1, pp. 16–23, 2018.

[3] H. N. Dai, R. C. W. Wong, H. Wang, Z. Zheng, and A. V. Vasilakos, "Big data analytics for large scale wireless networks: challenges and opportunities," *ACM*, vol. 52, pp. 1–46, 2019.

[4] A. H. Sabry, W. Z. W. Hasan, M. N. Mohtar, R. M. K. R. Ahmad, and H. R. Harun, "Planter pressure repetability data analysis for health adult based on EMED system," *Malaysian Journal of Fundamental and Applied Sciences*, vol. 14, no. 1, pp. 96–101, 2018.

[5] A. Mensikova and C. A. Mattmann, "Ensemble sentiment analysis to identify human trafficking in web data," in *Proceedings of the ACM Workshop on Graph Techniques for Adversarial Activity Analytics (GTA)*, pp. 1–6, ACM, New York, Ny, USA, 2018.

[6] S. S. Gill and R. Buyya, "Bio-inspired algorithms for big data analytics: a survey, taxonomy and open challenges," *Big Data Analytics for Intelligent Healthcare Management*, Elsevier, Amsterdam, Netherlands, pp. 1–17, 2019.

[7] A. Verma, I. Kaur, and N. Arora, "Comparative analysis of informatio extraction techniques for data mining," *Indian Journal of Science and Technology*, vol. 9, no. 11, pp. 1–18, 2016.

[8] B. R. Manju, A. Joshuva, and V. Sugumaran, "A data mining study for condition monitoring on wind turbine blades using hoeffding tree algorithm through statistical and histogram features," *International Journal of Mechanical Engineering & Technology*, vol. 9, no. 1, pp. 1061–1079, 2018.

[9] I. Aytekin, E. Eyduran, K. Karadas, R. Akshan, and I. Keskin, "Prediction of fattening final live weight from some body measurements and fattening period in young bulls of crossbred and exotic breeds using MARS data mining algorithm," *Journal of Zoology*, vol. 50, no. 1, pp. 189–195, 2018.

[10] S. Ahmad and R. Varma, "Information extraction from text messages using data mining techniques," *Malaya Journal of Matematik*, vol. 5, no. 1, pp. 26–29, 2018.

[11] N. E. Oweis, M. M. Fouad, S. R. Oweis, S. S. Owais, and V. Snasel, "A novel mapreduce lift assiciation riule mining algorithm (MRLAR) for big data," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 8, pp. 1–7, 2016.

[12] C. W. Tsai, C. F. Lai, M. C. Chiang, and L. T. Yang, "Data mining for internet of things: a survey," *Communications Surveys and Tutorials, IEEE*, vol. 16, no. 1, pp. 77–97, 2014.

[13] F. Chen, P. Deng, J. Wan, D. Zhang, A. V. Vasilakos, and X. Rong, "Data mining for Internet of things: literature review and challenges," *International Journal of Distributed Sensor Networks*, vol. 2015, Article ID 365372, 14 pages, 2015.

[14] S. B. Patel, M. N. Patel, and S. M. Shah, "FEA-HUIM: fast and efficient algorithm for high utility item-set mining using novel data structure and pruning strategy," *National Conference on Advanced Research Trends in Information and Technologies (NCARTICT), IJSSET*, vol. 4, no. 2, pp. 1–7, 2018.

[15] S. Bhaskaran, R. Marappan, and B. Santhi, "Design and analysis of a cluster-based intelligent hybrid recommendation system for E-learning applications," *Mathematics, MDPI*, vol. 9, no. 107, pp. 1–21, 2021.

[16] D. Giordano, M. Mellia, and T. Cerquitelli, "K-MDTSC: K-Multi-Dimensional time-series clustering algorithm," *Electronics, MDPI*, vol. 10, no. 1166, pp. 1–21, 2021.

[17] R. Mothukuri and B. B. Rao, "Cluster Analysis of cyber crime data using R," *International Journal of Computer Science and Mobile Applications*, vol. 6, no. 2, pp. 62–70, 2018.

[18] A. N. Onaizah, "A novel data stream clustering algorithm in healthcare IoT," *International Journal of Research in Applied Natural and Social Sciences (IJRANSS)*, vol. 5, no. 6, pp. 51–60, 2017.

[19] R. Krishnamurthi, A. Kumar, D. Gopinathan, A. Nayyar, and B. Qureshi, "An overview of IoT sensor data processing, fusion, and analysis techniques," *Sensors, MDPI*, vol. 20, pp. 1–23, 6076.

[20] F. Alam, R. Mehmood, I. Katib, and A. Albeshri, "Analysis of eight data mining algorithms for smarter Internet of things (IoT)," in *Proceedings of the International Workshop on Data Mining in IoT Systems (DaMIS)*, pp. 437–442, Elsevier, London, UK, September 2016.

[21] A. M. C. Souza and J. R. A. Amazonas, "An outlier detect algorithm using big data processing and Internet of things architecture," in *Proceedings of the International Workshop on Big Data and Data Mining Challenges on IoT and Pervasive Systems (BigD2M)*, pp. 1010–1015, Elsevier, London, UK, June 2015.

[22] D. Gil, M. Johnsson, H. Mora, and J. Szymanski, "Review of the complexity of managing big data of the Internet of things," *Complexity*, vol. 2019, Article ID 4592902, 12 pages, 2019.

[23] M. S. Mahdavinejad, M. Rezvan, M. Barekatain, P. Adibi, P. Barnaghi, and A. P. Sheth, "Machine Learning for Internet of things data analysis: a survey," *Digital Communications and Networks*, vol. 4, pp. 161–175, 2018.

[24] M. Ruta, F. Scioscia, G. Loseto, A. Pinto, and E. D. Sciascio, "Machine learning in the Internet of things: a semantic-enhanced approach," *Semantic Web*, vol. 10, pp. 1–21, 2017.

[25] K. Demertzis, K. Rantos, and G. Drosatos, "A dynamic intelligent policies analysis mechanism for personal data processing in the IoT ecosystem," *Big Data And Cognitive Computing*, vol. 4, no. 9, pp. 1–16, 2020.

[26] N. Joseph and B. S. Kumar, "Top-K competitor trust mining and customer behavior investigation using data mining technique," *Journal of Network Communications and Emerging Technologies (JNCET)*, vol. 8, no. 2, pp. 26–30, 2018.

[27] M. P. Dooshima, E. N. Chidozie, B. J. Ademola, O. O. Sekoni, and I. P. Adebayo, "A predictive model for the risk of mental illness in Nigeria using data mining," *International Journal of Immonology*, vol. 6, no. 1, pp. 5–16, 2018.

[28] B. A. Tama and S. Lim, "A comparative performance evaluation of classification algorithms for clinical decision support systems," *Mathematics, MDPI*, vol. 8, no. 1814, pp. 1–25, 2020.

[29] C. C. N. Wang, I. S. Chang, P. C. Y. Sheu, and J. J. P. Tsai, "Application of semantic computing in cancer on secondary data analysis," in *Proceedings of the 2nd International Conference on Robotic Computing, IEEE*, pp. 407–413, Laguna Hills, CA, USA, February 2018.

[30] K. Ravikumar and A. R. Kannan, "Spatial data mining for prediction of natural events and disaster management based on fuzzy logic using hybrid PSO," *Taga Journal*, vol. 14, pp. 858–878, 2018.

[31] B. M. M. Alom and M. Courtney, ""Educational data mining: a case study perspectives from primary to university education in Australia," *International Journal of Information Technology and Computer Science*, vol. 2, pp. 1–9, 2018.

[32] S. Celik and O. Yilmaz, "Prediction of body weight of Turkish tazi dogs using data mining techniques: classification and regression tree (CART) and multivariate adaptive regression splines (MARS)," *Journal of Zoology*, vol. 50, no. 2, pp. 575–583, 2018.

[33] W. Ugulino, D. Cardador, K. Vega, E. Velloso, R. Miliditi, and H. Fuks, "Wearable computing: accelerometers data classification of body postures and movements," in *Proceedings of the 21st Brazilian Symposium on Artificial Intelligence, Advances in Artificial Intelligence, Lecture Notes in Computer Science*, pp. 52–61, Springer, Stuttgart, Germany, September 2012.

[34] M. B. Chandak, "Role of big-data in classification and noval class detection in data streams," *Journal of Big Data*, Springer, vol. 23, no. 5, pp. 1–9, 2016.

[35] M. Gupta, K. K. Gupta, and P. K. Shukla, "Session key based fast, secure and lightweight image encryption algorithm," *Multimedia Tools and Applications*, vol. 80, pp. 10391–10416, 2021.

[36] H. S. Pannu, D. Singh, and A. K. Malhi, "Multi-objective particle swarm optimization-based adaptive neuro-fuzzy inference system for benzene monitoring," *Neural Computing & Applications*, vol. 31, pp. 2195–2205, 2019.

[37] D. Singh, J. Singh, and A. Chhabra, "High availability of clouds: failover strategies for cloud computing using integrated check pointing algorithms," in *Proceedings of the 2012 International Conference on Communication Systems and Network Technologies*, pp. 698–703, Rajkot, India, May 2012.

[38] D. Singh and V. Kumar, "A comprehensive review of computational dehazing techniques," *Archives of Computational Methods in Engineering*, vol. 26, pp. 1395–1413, 2019.

[39] R. Gupta, "Performance analysis of anti-phishing tools and study of classification data mining algorithms for a novel anti-phishing system," *International Journal of Computer Network and Information Security*, vol. 12, pp. 70–77, 2015.

[40] R. Bhatt, P. Maheshwary, P. Shukla, P. Shukla, M. Shrivastava, and S. Changlani, "Implementation of fruit fly optimization algorithm (FFOA) to escalate the attacking efficiency of node capture attack in wireless sensor networks (WSN)," *Computer Communications*, vol. 149, pp. 134–145, 2020, ISSN 0140-3664.

[41] F Distribution Table, http://www.socr.ucla.edu/applets.dir/f_table.html, 2018.

[42] Normal Distribution Table: http://math.arizona.edu/~rsims/ma464/standardnormaltable.pdf.