International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

# 2CHCOKC - A 2D Chaotic based Hybrid Coyote Optimization with K-Means for Clustering the Massive Datasets to Enhance the Datasearch Ability

Dr Pradeep Kumar Atulker
Assistant Professor, Department of Computer Science and Application, Govt. College Khimlasa, Sagar, (M.P.),

Dr Rajendra Gupta
Associate Professor, Department of CS & IT, Rabindranath Tagore University, Raisen (M.P.), India,

Dr Ankur Khare
Assistant Professor, Department of CS & IT, Rabindranath Tagore University, Raisen (M.P.), India,

## Abstract

Data clustering is broadly exploited across a broad spectrum of industries, including teaching, healthcare, and innumerable decentralized business settings. The miscellaneous datasets are effectively handled and analyzed through various classification procedures to enhance the eminence of essential data sharing in decentralized environments. K-Means represents a widely recognized and intuitive clustering method employed to segment a set of data values into a predetermined count of clusters. It has proven successful in addressing numerous practical clustering challenges. However, the critical aspects of choosing the initial cluster centers and congregating to local optima present significant limitations in the K-Means method. Various optimization methods are implemented to enhance the clustering's accuracy. Here, a 2D Chaotic based Hybrid Coyote Optimization with K-Means for Clustering (2CHCOKC) method is implemented and utilized to categorize a substantial amount of varied datasets, with the goal of enhancing the data retrieval proficiency and K-Means performance. The central aim of the 2CHCOKC approach is to accomplish resourceful and exact scattering of data values within a dataset. This is accomplished by utilizing 2D chaotic map for establishing populations. The addition of 2D chaotic map is anticipated to enrich the calculation and employment skills of search agents within the optimization framework for attaining the optimal and exact classification of centers and members within a cluster. The development of the 2CHCOKC methodology using MATLAB 2022a tool on five huge datasets highlights the exceptional effectiveness of the method. This is marked when comparing it to former approaches such as K-Means, ALO, GO, and CO. This comparison is established through various aspects containing F-Measure, Accuracy, Standard Deviation, Root Mean Square Error, Time Complexity, Purity Index, Intra-Cluster Distance, and Statistical Explication.



(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

**Keywords:** Chaotic Map, Clustering, Coyote Optimization, F-Measure, K-Means, Statistical Explication, Time Complexity.

## 1. Introduction

The researchers delve into the management and investigation of extensive datasets across various applications, harnessing factors such as space, execution, and database organization [1, 2]. Pre-processing is employed to mitigate data ruin, while also increasing data exactness and proficiency. A prevailing scattered design is put into action, facilitating swift and cost-effective data exploration for both sequential and parallel message passing [3, 4]. Multiple optimization strategies based on population dynamics [5], such as Grey-Wolf Optimization (GO) [6], Ant Lion Optimization (ALO) and Moth Flame Optimization (MO) are employed to thoroughly and exactly explore the abundant data volume in diverse scenarios [7, 8, 9]. Decorative cerebral and deep learning techniques utilizing neural networks are employed in the investigation of large-scale data. In this context, relevant insights are engaged for query management, targeting to sensible query recovery rates and enrich recovery intervals. Health-related data is scrutinized to identify ailments with qualitative and quantitative excuses, certifying the supreme medical interventions [10, 11].

Convex optimization is highlighted as a prominent method for extensive data analysis in distributed settings, mostly across various origins. Convex signal dispensation methods are harnessed alongside the least squares approach for the sampling and valuation of vast datasets. The adjoining operator, uncertainty management, and scattered valuation play pivotal roles in optimizing data exploration [12, 13]. Deep learning is instrumental in implementing distributed clustering on miscellaneous data through the utilization of an auto-encoder. The auto-encoder freely decides cluster totals and distributes cluster members with the aid of a deep learning framework. The results are evaluated alongside the widely recognized clustering technique, K-Means [14, 15].

Furthermore, a fusion of machine learning techniques with Convolutional Neural Network (CNN) is arranged to enrich clarity and meticulousness in the selection and management of disease categories. Parallel calculation, containing the application of auto-encoders, is utilized to determine optimal grouping tactics for healthcare supervision and informed understanding of illnesses [16]. Decision supervisory concerning illnesses is conducted through the aid of artificial intelligence, thoroughly inspected and confirmed using medical archives. A range of cancer types are analyzed employing support vector machines to certify accurate and applicable disease supervision [17]. Machine learning is further engaged alongside tensors and core operators for actual supervision and endorsement of datasets. Data endorsement is conceded out over deep learning methods, relevant across miscellaneous platforms such as Bing and self-motivated applications [18, 19].

Categorization, similar to clustering, entails organizing a multitude of datasets into different groups for various purposes. Within the medical domain, a decision support structure is engaged for the purpose of combination medical information and envisaging illnesses [20, 21]. In the realm of medical data classification, a CNN is applied in conjunction with a novel


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

loss function. The employment of a multi-pooling structure is engaged to seizure the spatial details of key characteristics within medical data, aiding in the valuation of disease brutality and determination of applicable treatments [22].

The preceding literature demonstrates the application of diverse clustering methodologies and optimization strategies such as GO [6], ALO, MO and K-Means [14, 15] across numerous datasets, with the aim of enhancing data retrieval and search proficiencies. While certain clustering methods exhibit favorable outcomes with explicit datasets, their effectiveness might differ with other datasets, as suggested by the no-free-lunch theorem. This theorem underscores that no singular clustering approach comprehensively outperforms all problems. Additionally, the convergence rate stands as a key factor in evaluating optimization techniques.

To tackle these obstacles, a fresh methodology referred to as a 2D Chaotic based Hybrid Coyote Optimization with K-Means for Clustering (2CHCOKC) method is presented. This approach is employed to group widespread datasets and represents an enhanced version of the Coyote Optimization (CO) method. In contrast to the conventional qualitative evaluations seen in earlier studies, the innovative 2CHCOKC method places emphasis on quantitatively assessing clustering methodologies designed for extensive datasets.

The key contributions of the proposed 2CHCOKC are outlined below:

1. The primary objective of 2CHCOKC is to achieve an actual scattering of data element values through miscellaneous outlets in a distributed context.

2. Data scattering is achieved through harnessing 2D chaotic map [23, 24] for population production.

3. The overarching aim is to optimize the scattering of data elements, certifying the well-organized deployment of heterogeneous assets across distributed environment.

4. The execution of the 2CHCOKC method using MATLAB 2022a tool on five widespread datasets is demonstrated through various metrics containing F-Measure, Accuracy, Standard Deviation, Root Mean Square Error, Time Complexity, Purity Index, Intra-Cluster Distance, and Statistical Explication.This comparison highlights the superiority of the 2CHCOKC method over former approaches such as K-Means, ALO, GO, and CO.

Section 1 delivers an overview, whereas Section 2 offers a comprehensive review of data clustering techniques, as well as extensive data analysis. This segment investigates plentiful parameters within these domains. In Section 3, we delve into the Coyote Optimization (CO) and K-Means method, serving as the groundwork for the proposed 2CHCOKC method elucidated in Section 4. Section 4 encompasses exhaustive details, including flowcharts and algorithms pertinent to the 2CHCOKC method. Section 5 is dedicated to the datasets utilized in the implementation, performance metrics, and the outcomes derived from implementing the 2CHCOKC method. It illuminates the practical effectiveness of the suggested method. Finally, Section 6 encapsulates the paper's conclusions, summarizing the principal discoveries and consequences of the research.


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

## 2. Literature Review

The article on Decision Support Systems (DSS) [25] delves into the augmentation of device to handle the intricate process of extracting diverse and complex data from various scattered origins. The paper presents the idea of Big Data [26, 27]—sophisticated multidimensional structures for supervising information—which supports stakeholders in employing contemporary data-driven strategies to solve problems. The predominance of Big Data in ensuring nutrition safety is highlighted, as information within the nutrition distribution sequence is detached and exhibits miscellany in terms of arrangement, quantity, and topographical origin. Overall, the study and dialogue presented aim to foster the adoption of Big Data and DSS in nutrition safety, eventually increasing risk debt and certifying nutritional well-being within the framework of climate erraticism

The process of combining data into related clusters, known as clustering [28], is extensively employed. Nevertheless, latest clustering schemes define clusters using occurrence values and deliberation, requiring precise semantic meaning. To tackle this problem, a fresh Evolutionary Clustering Algorithm (ECA) has been established, incorporating shared group placing, population centred strategies and K-Means method. The calculation encompasses the utilize of 32 miscellaneous datasets, calculating their usefulness through both intra-cluster and inter-cluster measures. In summary, the ECA scheme offers a noteworthy advancement in clustering mixed and multi-dimensional datasets, exhibiting improved cluster validation and enriched strength through a range of dataset features.

The article [29] centers around strategies for mining association procedures and their utilization in assessing the excellence of institution students. These techniques are employed to unearth potential connections among various aptitudes of university students through their occurrence and awareness data. This aids educators in recognizing distinct students' assets and flaws, facilitating customized instruction. The aim is to tackle concerns linked to the elimination of associative rules and investigate the impact of incorporating such procedures in evaluating university student performance. The experiment verifies the effectiveness and practicability of the procedure for evaluating student performance, endorsing its pragmatic implementation in real-world circumstances. On the whole, the research underscores the advantages of employing association procedure mining policies in systems for evaluating university student superiority, laying the groundwork for more efficient and tailored teaching approaches.

The article presents an innovative heuristic method for data clustering utilizing the MO scheme [30], a fresh metaheuristic stimulated through intellect of moths in the ordinary world. Despite the extensive acknowledgment of the k-means approach for clustering, its progression heavily depends on initial cluster mid points and can become ensnared in local optima due to limited search skill. The MO is suggested as a promising remedy for tackling intricate problems, and its performance has been deemed satisfactory in various studies. To assess the usefulness of the MO, trials are conducted engaging UCI datasets.


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

The article [31] explores the aggregate capability to amass an extensive variety of miscellaneous data, with a specific emphasis on time-series information, which harbors a substantial reservoir of untapped insights. Former data mining tactics frequently encounter constraints when considering time-series information, especially when handling with multi-dimensional time-series that necessitate joint inspection to extract significant insights. In answer, the paper employs an alternative clustering method termed K-MDTSC (K-Multi-Dimensional Time-Series Clustering), specifically designed to tackle the grouping of multi-dimensional time-series. The results demonstrate that K-MDTSC yields satisfactory grouping results, mainly in the context of intricate artificial datasets. In a broader context, the article underscores the importance of grouping multi-dimensional time-series and showcases the usefulness of the K-MDTSC system in both simulated and real-world settings. The discoveries indicate that K-MDTSC has the potential to be a appreciated instrument for mining information from intricate multi-dimensional time-series information, especially in manufacturing uses like extrapolative preservation.

The Particle Swarm Optimization (PSO) method, as outlined in reference [32], is a extensively employed technique for computing the optimal solution across several variables in objective functions. In this investigation, a novel PSO approach is recognized with the precise aim of computing the optimal value for multidimensional objective functions. The PSO iterative process encompasses several extensive iteration steps, each comprised of two stages. During the initial phase, a developmental procedure is utilized to effectively explore the multi-dimensional parameter space. The subsequent phase employs the conventional PSO methodology to ascertain the global optimum value of the function. In sum, the article introduces a fresh PSO scheme tailored for optimizing functions, showcasing its efficacy and heightened precision through comprehensive examination and exploration. This scheme has the prospective to prove valuable across a spectrum of systematic and engineering solicitations.

The article [33] delves into the active illustration of clusters to comprehend their characteristics pertaining to size and arrangement. Since uncertain approaches alone might not provide a holistic view of cluster structures, independent theoretical investigations are necessary to achieve a comprehensive understanding of the proportioned processes and attributes of the clusters. To identify multiple minima and ultimately ascertain the global minima for the clusters, potential energy surfaces (PES) are explored. Numerous optimization methods are devised and applied to resolve regular isomers within distinct functional sets, such as the utilization of the Genetic Algorithm (GA). This paper aims to scrutinize the optimization techniques, assessment methodologies, and the accuracy of results attained through the implementation of these schemes across diverse cluster categories.

The article [34] tackles the significant issue of the Oath of Dimensionality, a challenge that hinders development across numerous domains, containing bioinformatics. In response to this obstacle, the authors present a collaborative explanation that merges a clambering based approach for reducing dimensionality with estimation scheme based on PSO scheme. The paper concentrates on attribute assortment for cancer prophecy, employing a sophisticated Spark Distributed PSO (SDPSO). The offered methodology merges Binary PSO (BPSO) with


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

the standard PSO approach to attain competent and accurate attribute choice. In conclusion, the introduced SDPSO technique demonstrates remarkable productivity and exactitude in the context of attribute assortment for cancer prophecy, underscoring its prospective to combat the Oath of Dimensionality and enhance efficiency in solicitations within the realm of bioinformatics.

## 3. Hybrid K-Means - Coyote Optimization Method

### 3.1. K-Means Method

K-Means is an extensively used clustering method grounded on a specific exploration of variances methodology. It segregates the entire dataset into T clusters through eq. (1).

$$\vartheta = \sum_{g=1}^{T} \sum_{h=1}^{|D|} d^2\left(N_g - E_h\right) \qquad (1)$$

Here,

$d^2\left(N_g - E_h\right)$ = Euclidean distance square from g$^{th}$ cluster centre to h$^{th}$ data element value

$|D|$ = Dataset Size

The K-Means method assigns all data element values to the cluster that is in proximity to the centre (whole data element values mean inside a cluster) of that cluster. Subsequently, it computes the resemblance between data element values and the centre, subsequently reassigning the data element values in the dataset to the centre of the cluster. This process of data element values reallocation to cluster centres persists until a convergence criterion is met.

### 3.2. Coyote Optimization (CO)

A CO technique represents the most recently formulated strategy relying on the coyote's adaptive behaviors within the environment and its adaptive learning. The CO ensures a balanced harmony between exploration and implementation.

The CO method commences with the utilization of the overall population (O) and the complete count of coyotes (Y) as the original solution. The social behavior inherent in coyotes is employed as the objective function, accounting for the model's cost (M_C) (as defined in eq. (2)).

$$M_C_y^{o,t} = \varphi = [\varphi_1, \varphi_2, \varphi_3, \ldots . \varphi_n] \qquad (2)$$

Here,

y = Coyote value

o = Coyotes set


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

t = Coefficients formulating Time

$\varphi_n$ = $n^{th}$ coyote cost

At first, a couple of casual coyotes is acquired as preliminary solutions in the assessment region, which are assessed using eq. (3).

$$M_C_{y,j}^{o,t} = BT_j + \tau \times \left(TT_j - BT_j\right) \quad (3)$$

Where,

$TT_j$ & $BT_j$ = Top Threshold and Bottom Threshold value of $j^{th}$ variable in assessment zone.

$\tau$ = Indiscriminate numeral ($\tau \in [0,1]$)

Each coyote's performance is measured by the target function (TAR); defined in eq. (4).

$$TAR_y^{o,t} = Fun\left(M_C_{y,j}^{o,t}\right) \quad (4)$$

This method alters the group`s location in a random manner. Additionally, agents adjust their positions by transitioning from single group to another. The process of transitioning is determined using a probabilistic design (as defined by eq. (5)).

$$O = 0.005 \times (Y)^2 \quad (5)$$

At this time, deliberate the values of $Y \leq \sqrt{200}$ & $O > 1$. The maximum number of coyotes is allowed in each group is limited to 14, aiming to diversify the methodology, i.e., promoting traditional shifts among coyotes. The best solution from each iteration is referred to as the Chi ($\chi$) coyote and is computed through eq. (6).

$$\chi^{o,t} = M_C_y^{o,t} \quad \forall \quad minimum\left(TAR_y^{o,t}\right) \quad (6)$$

The common features of coyotes in terms of Cultural Conversion (CC) are appraised via eq. (7).

$$CC_j^{o,t} = \begin{cases} Q_{\frac{Y+1}{2},j}^{o,t} & ,Y = \ odd\ numeral \\ \frac{1}{2}\left(Q_{\frac{Y}{2},j}^{o,t} + Q_{\frac{Y}{2}+1,j}^{o,t}\right) & Otherwise \end{cases} \quad (7)$$

Where,

$Q^{o,t}$ = Coyotes social state of $j^{th}$ variable on behalf of group o at time t.

The CO also takes into account the Coyotes' Life Cycle (CLC), integrating a fusion of ecological elements and the communal conduct of the parent coyotes (expressed in eq. (8)).


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

$$CLC_j^{o,t} = \begin{cases} M_C_{a_1,j}^{o,t} & , a_j < B_d \ \ or \ \ j = j_1 \\ M_C_{a_2,j}^{o,t} & , a_j \geq B_d + B_r \ \ or \ \ j = j_2 \\ \omega_j & , Otherwise \end{cases} \quad (8)$$

Here,

$a_j$ = Capricious number ($a_j \in [0,1]$)

$a_1$ & $a_2$ = Capricious coyotes in group o.

$j_1$ & $j_2$ = Capricious developmental variables.

$\omega_j$ = Capricious value within the scope of developmental variable.

$B_d$ & $B_r$ = Distribute and Relationship prospects illustrating the cultural conversion of coyotes in group (eq. (9) & eq. (10)).

$$B_d = \frac{1}{m} \quad (9)$$

$$B_r = \frac{1}{2}(1 - B_d) \quad (10)$$

Where,

m = Dimensions of Variable.

The process of equating CLC is outlined in algorithm 1.

**Algorithm 1: Coyotes` Life Cycle (CLC)**

Choose u and v

If (u = 1)

CLC survives and coyote in v end

Else if (u > 1)

CLC survives and ancient coyote in v end

Else

CLC end

END if

Where,

u = Overall group coyotes

v = Coyotes worst outputs.



(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

The likelihood of coyote elimination is set at 10%. The cultural conversion between groups is elucidated using two expressions ($\lambda_1$&$\lambda_2$) (eq. (11) & eq. (12)).

$$\lambda_1 = \chi^{o,t} - M_C_{e_1}^{o,t} \quad (11)$$

$$\lambda_2 = CC^{o,t} - M_C_{e_2}^{o,t} \quad (12)$$

Here,

$\lambda_1$ = Culture separation among frontrunner (Chi) and acquired coyote ($e_1$).

$\lambda_2$ = Culture separation amoung group culture designing and acquired coyote ($e_2$).

Eq. (13) is employed to alter the social behavior based on the interaction between the frontrunner and the group.

$$LM_C_y^{o,t} = M_C_y^{o,t} + a_1 \times \lambda_1 + a_2 \times \lambda_2 \quad (13)$$

Where,

$a_1$ and $a_2$ = Capricious values in the range of [0, 1].

The modified cost is computed using eq. (14) and eq. (15) ultimately with the aid of adjustment formulas.

$$LTAR_y^{o,t} = Fun\left(LM_C_y^{o,t}\right) \quad (14)$$

$$M_C_y^{o,t+1} = \begin{cases} LM_C_y^{o,t} & , LTAR_y^{o,t} < TAR_y^{o,t} \\ M_C_y^{o,t} & , Otherwise \end{cases} \quad (15)$$

A notable aspect of these methodologies is their capacity to escape from local optima.

### 3.3. Hybrid Coyote Optimization with K-Means (HCOK)

The usefulness of the K-Means clustering technique is elevated through the utilization of the Hybrid Coyote Optimization with KMeans (HCOK). At the start, the K-Means method is employed to conduct clustering on the provided datasets, partitioning them into designated clusters based on the smallest Euclidean distances. Subsequently, the optimized centres for whole clusters are computed utilizing the CO approach. In this process, all clusters are adjusted as a coyote, and all data element values are prepared as a search agent. Consequently, all coyote modifies its position based on its fitness value, aiming to diminish intracluster distances mean total. For all cluster, the coyote assesses its optimal position in order to determine the optimal centres.


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

## 4. 2CHCOKC - A 2D Chaotic based Hybrid Coyote Optimization with K-Means for Clustering

The CO method is altered utilizing 2D chaotic dynamics. The practical outcomes for numerical optimization are achieved through the CO method, which is well-suited for assessment and execution in an extensive search space. Nevertheless, the data clustering approach significantly differs from numerical optimization. Currently, multiple crucial characteristics are accustomed, and the employment of 2D chaotic dynamics for data clustering is executed via a 2CHCOKC method.

### 4.1. 2D Chaotic based Population Production of CO Method

An modified CO is formulated employing 2D chaotic dynamics to mitigate unwise convergence catastrophe. The incorporation of 2D chaotic in the CO method is designed to address the concern of regional optima by employing arbitrary numbers.

This map exhibits significantly more intricate chaotic dynamics and a faster rate of iteration compared to its 1D counterpart, a characteristic lacking in high dimensional chaotic systems. The mathematical representation of the 2D chaotic dynamic is provided below (eq. (16) & eq. (17)):

$$J_{\theta+1} = tan^{-1}\left[\frac{r}{10 * q * K_\theta} + \tan\left(q * \pi * J_\theta\right)\right] \qquad (16)$$

$$K_{\theta+1} = tan^{-1}\left[\frac{r}{q * J_\theta} + \tan\left(10 * q * \pi * K_\theta\right)\right] \qquad (17)$$

Here,

q & r = Coefficients $q, r \in (-\infty, +\infty)$.

Within the framework of 2D chaotic, the $\tan^{-1}$ function is employed for map compression. While the use of the $\tan^{-1}$ function as a monotonic operation ensures an additional identical scattering of the chaotic arrangement. Furthermore, the chaotic system encompasses a broader chaotic system, rendering it better fit for data clustering..

The preliminary population (O) of the CO are established by evaluating the 2D chaotic dynamics (eq. (16) to eq. (17)), with the goal of amplifying the efficiency of the CO method over the appropriate exploitation of an extensive solution zone.

### 4.2. The Overall 2CHCOKC Method for Data Clustering

The integration of 2D chaotic focussed assortment of the population is engaged to attain an optimal outcome for data clustering within the 2CHCOKC framework (as depicted in Fig. 1 and outlined in Algorithm 2). To initiate the process, the 2D chaotic dynamics (eq. (16) and eq. (17)) are engaged to construct the population (O) within the CO method. All individuals within this population are signified as a matrix indicated by $W_x = W \times T$, where ($W_x$ =


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

Separate attributes, W = Whole elements in the dataset, T = Whole cluster centers). The locations of the T cluster centers are organized in a matrix, through all centers connected to its corresponding collection of W attributes. The preliminary values of these definite matrices are casually and consistently generated inside the range defined by the minor and superior element values extant in the dataset, over an recapitulation threshold (G). Subsequently, the 2CHCOKC method is realized to estimate the entire individuals fitness values (eq. (2) to eq. (15)).

**Algorithm 2: The Overall 2CHCOKC Method**

| **Algorithm** | **Number of Operations** |
|---|---|
| **START** | |
| Select T capricious values as cluster centres | 1 |
| **WHILE** end condition is not attained DO | G+1 |
| **FOR** each data element value | G*(W+1) |
| Compute Euclidean distance from cluster centres to the data element values (eq. (1)) | G*W*T |
| Assign the data element values to least distance cluster | G*W |
| **END FOR** | |
| Measure the data element values average in all cluster | G*T |
| Assign the average values as latest cluster centres | G*T |
| **END WHILE** | |
| Return T Clusters | 1 |
| **FOR** all clusters | T+1 |
| Initialize population (O) and Coyotes (Y) (eq. (3)) | T |
| Measure the coyote's performance (Target function) (eq. (4)) | T*O*Y |
| **WHILE** end condition is not attained **DO** | T*(G+1) |
| **FOR** every coyote set o **DO** | T*G*Y |
| Describe the χ coyote of o coyote set (eq. (6)) | T*G*O |
| Evaluate the Cultural Conversion (CC) of coyote set (eq. (7)) | T*G*O |


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

| | |
|---|---|
| **FOR** every y coyote of coyote set o **DO** | T*G*(Y+1) |
| Modify the social behaviour (eq. (13)) | T*G*Y |
| Calculate new coyote performance (target function) (eq. (14)) | T*G*Y |
| Modify Model Cost (eq. (15)) | T*G*Y |
| **END FOR** | |
| Develop Coyote Life Cycle (CLC) (eq. (8) & Algorithm 1) | T*G*O |
| **END FOR** | |
| Transition between coyote set (eq. (5)) | T*O |
| **END WHILE** | |
| Choose the optimal coyote | T |
| **END FOR** | |
| **STOP** | |



(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

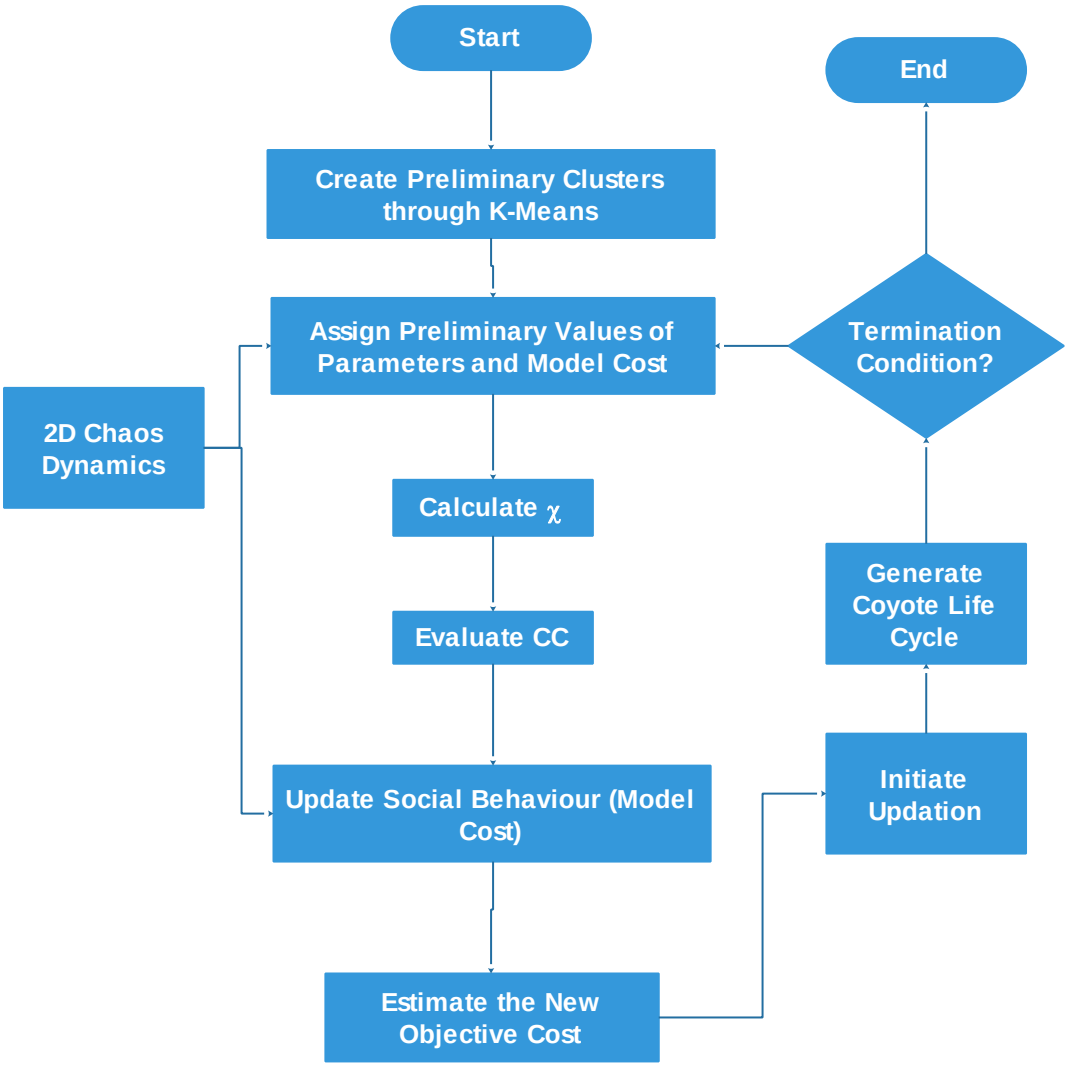



Fig. 1: The flowchart of 2CHCOKC Method

## 5. Results and Analysis

This section provides a brief description of the five massive datasets (as shown in Table 1) and the assessments metrics for 2CHCOKC method. All of these methodologies are employed in the MATLAB 2022a software, working on a Windows 10 environment. Validation was performed using five complete datasets. The outcomes for the 2CHCOKC method are achieved over 400 iterations, encompassing various aspects containing F-Measure, Accuracy, Standard Deviation, Root Mean Square Error, Time Complexity, Purity Index, Intra-Cluster Distance, and Statistical Explication. A comprehensive evaluation is executed against former methodologies such as K-Means, ALO, GO and CO conducted over 35 separate runs.

### 5.1. Datasets

The 2CHCOKC method is applied to five noteworthy, massive datasets (HTRU, Dry Bean, Codon_Usage, Online_Shoppers_Intention, Credit Card Clients) (Table 1) sourced from the UCI repository.



(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

**Table 1: Datasets**

| Sr. No. | Datasets | Overall Elements | Overall Attributes | Overall Clusters |
|---|---|---|---|---|
| 1 | HTRU | 12330 | 9 | 2 |
| 2 | Dry Bean | 13611 | 17 | 7 |
| 3 | Codon_Usage | 13028 | 69 | 9 |
| 4 | Online_Shoppers_Intention | 12330 | 18 | 2 |
| 5 | Credit Card Clients | 30000 | 25 | 2 |

### 5.2. Assessment Metrics

The 2CHCOKC method is assessed in various aspects containing F-Measure, Accuracy, Standard Deviation, Root Mean Square Error, Time Complexity, Purity Index, Intra-Cluster Distance, and Statistical Explication.

#### 5.2.1. Standard Deviation

The fixed data clustering inside the average value range is reached utilizing a geometrical attribute identified as Standard Deviation (η). The most precise categorization is accomplished by exploiting the tiniest value of η (eq. (18)).

$$\eta = \sqrt{\frac{\sum(S - \bar{S})}{|D|}} \qquad (18)$$

Here,

S = Dataset Elements Points

$\bar{S}$ = Average Value of Dataset Elements

**Table 2: Standard Deviation**

| Datasets | K-Means | ALO | GO | CO | 2CHCOKC |
|---|---|---|---|---|---|
| HTRU | 0.3245 | 0.2174 | 0.2019 | 0.1627 | 0.0892 |
| Dry Bean | 0.4623 | 0.3182 | 0.2435 | 0.1278 | 0.1028 |
| Codon_Usage | 0.1526 | 0.2637 | 0.3425 | 0.1426 | 0.0927 |
| Online_Shoppers_Intention | 0.3452 | 0.2172 | 0.2736 | 0.1728 | 0.1126 |
| Credit Card Clients | 0.4567 | 0.4253 | 0.3254 | 0.2232 | 0.1627 |

As per the data presented in Table 2, the 2CHCOKC method identified the minimum η value among the five massive datasets. In relation to Standard Deviation (η), the 2CHCOKC outperforms CO, GO, ALO and K-Means by 16%, 32%, 34%, and 61% correspondingly through whole five datasets (Fig. (2)).


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

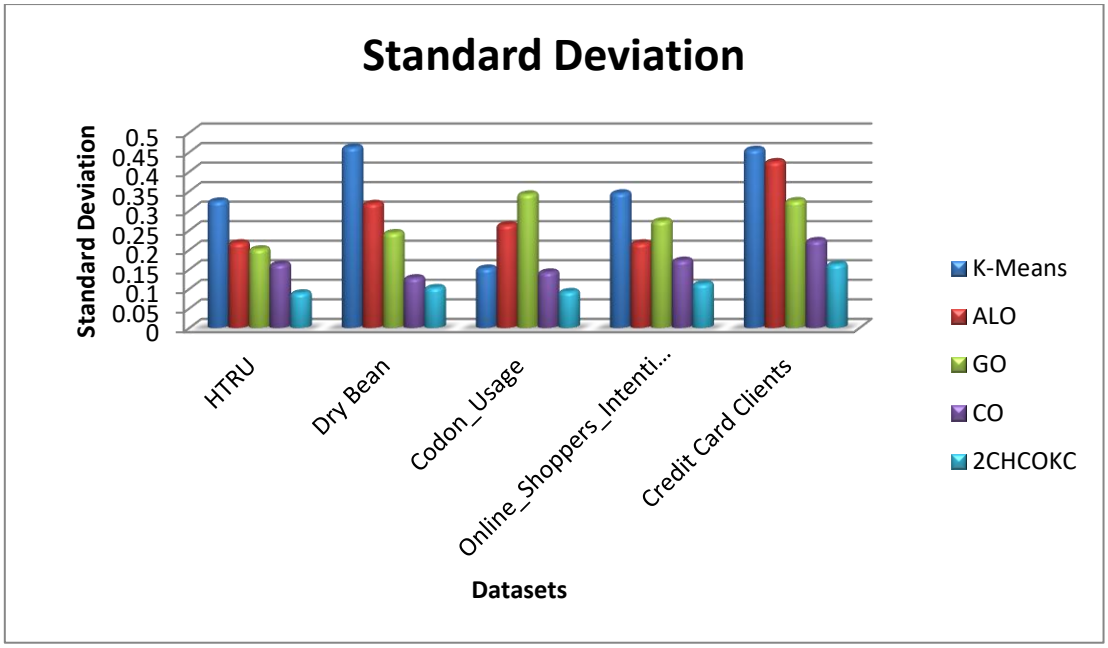



**Fig. 2: Standard Deviation**

### 5.2.2. Intra-cluster Distance

To initiate the process, the distances among data elements inside a cluster are calculated. Subsequently, the intra-cluster distance is identified as the average of these computed distances. The most optimal clustering is attained by picking the tiniest value of intra-cluster distance. By calculating the average of distances among cluster center and overall data elements inside the cluster, the distance mean is achieved. This practice is continued for every cluster. In the end, the average intra-cluster distance is deduced by amalgamating the average distances from whole clusters.

**Table 3: Average Rank (Intra-Cluster Distance)**

| Dataset | K-Means | ALO | GO | CO | 2CHCOKC |
|---|---|---|---|---|---|
| HTRU | 97.37 (5) | 71.36 (3) | 79.36 (4) | 57.38 (2) | 43.28 (1) |
| Dry Bean | 846.37 (4) | 935.37 (5) | 743.48 (3) | 648.37 (2) | 523.46 (1) |
| Codon_Usage | 7.836 (3) | 8.374 (4) | 9.372 (5) | 7.483 (2) | 4.372 (1) |
| Online_Shoppers _Intention | 10.368 (4) | 12.364 (5) | 9.327 (3) | 8.382 (2) | 6.372 (1) |
| Credit Card Clients | 65.387 (5) | 45.373 (3) | 56.497 (4) | 36.682 (2) | 23.282 (1) |
| **Average Rank ($V_\emptyset$)** | 4.2 | 4 | 3.8 | 2 | 1 |

Based on the data provided in Table 3, the 2CHCOKC method is identified the tiniest intra-cluster distance among the five massive datasets. In terms of intra-cluster distance, the 2CHCOKC demonstrates a superior efficiency compared to CO, GO, ALO and K-Means by 17%, 36%, 39%, and 49% respectively through all five datasets (Fig. (3)). The strategies are



(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

arranged in a hierarchy on the basis of the mean of their intra-cluster distances, spanning from the lowest to the highest (ranging from 1 to 4.2).

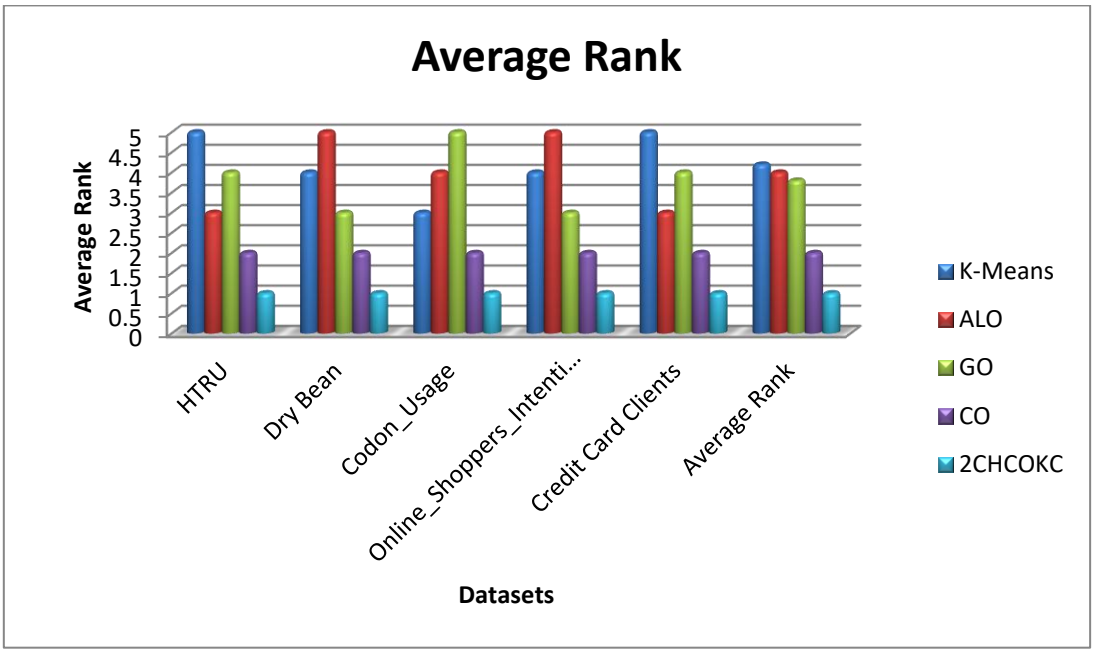



**Fig. 3: Average Rank (Intra-Cluster Distance)**

### 5.2.3. Root Mean Square Error (RS)

The RS value measures the dissimilarity between expected and calculated values. The optimal clustering is managed by exploiting the tiniest RS value (eq. (19)).

$$RS = \sqrt{\frac{1}{|D|}\sum_{h=1}^{|D|}\left(E_h - \hat{E}_h\right)^2} \qquad (19)$$

Where,

$E_h$ & $\hat{E}_h$ = h[th] element calculated and expected values in dataset.

**Table 4: Root Mean Square Error (RS)**

| Datasets | K-Means | ALO | GO | CO | 2CHCOKC |
|---|---|---|---|---|---|
| HTRU | 0.304 | 0.235 | 0.215 | 0.198 | 0.132 |
| Dry Bean | 0.327 | 0.274 | 0.263 | 0.187 | 0.103 |
| Codon_Usage | 0.374 | 0.232 | 0.264 | 0.164 | 0.123 |
| Online_Shoppers_Intention | 0.312 | 0.204 | 0.236 | 0.183 | 0.164 |
| Credit Card Clients | 0.332 | 0.251 | 0.217 | 0.194 | 0.173 |

The information presented in Table 4 demonstrates that the 2CHCOKC is produced the lowest RS value among the five massive datasets. In terms of Root Mean Square Error (RS), the 2CHCOKC outperforms CO, GO, ALO, and K-Means by margins of 18%, 61%, 65%, and 78%, respectively, through whole five datasets (Fig. (4)).


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

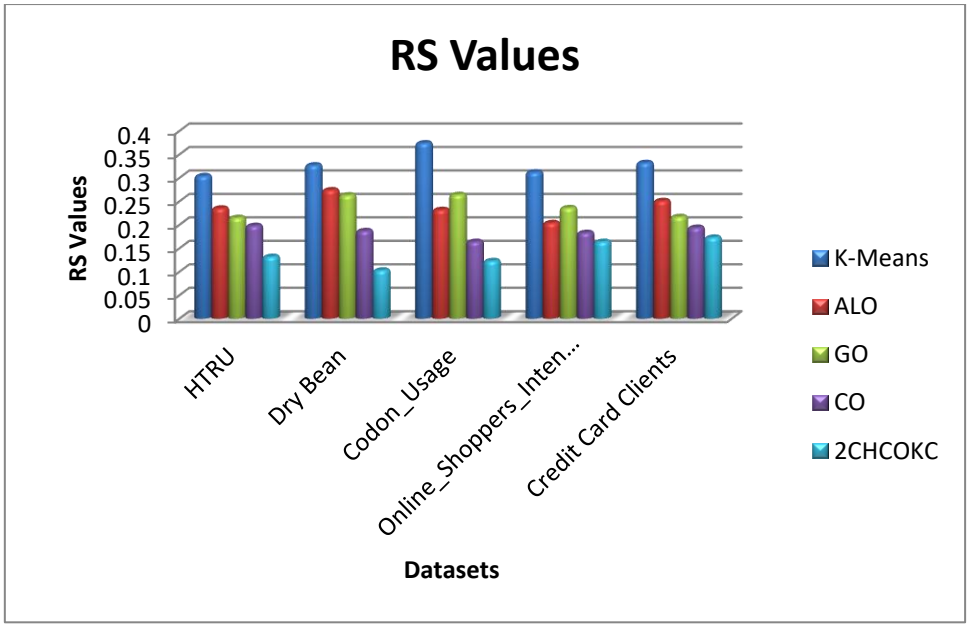



**Fig. 4: Root Mean Square Error**

### 5.2.4. F-Measure

Initially, Precision (C) and Recall (L) are measured to regain the information (eq. (20) & eq. (21)). Following this, C and L are joined to assess the F-Measure (M) (eq. (22) & eq. (23)).

$$C(\gamma,\delta) = \frac{|A_{\delta\gamma}|}{|A_{\delta}|} \qquad (20)$$

$$L(\gamma,\delta) = \frac{|A_{\delta\gamma}|}{|A_{\gamma}|} \qquad (21)$$

$$M(\gamma,\delta) = \frac{2 \times C(\gamma,\delta) \times L(\gamma,\delta)}{C(\gamma,\delta) + L(\gamma,\delta)} \qquad (22)$$

$$M = \sum_{\gamma=1}^{T} \frac{|A_{\gamma}|}{|D|} max\{M(\gamma,\delta)\} \qquad (23)$$

Where,

$|A_{\gamma}|$= $\gamma^{th}$ class size.

$|A_{\delta}|$= $\delta^{th}$ cluster size.

$|A_{\delta\gamma}|$= Whole data elements of $\gamma^{th}$ class entrusted to $\delta^{th}$ cluster.

**Table 5: F-Measure**

| Dataset | K-Means | ALO | GO | CO | 2CHCOKC |
|---|---|---|---|---|---|
| HTRU | 0.72 | 0.78 | 0.76 | 0.85 | 0.91 |



(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

| | | | | | |
|---|---|---|---|---|---|
| Dry Bean | 0.73 | 0.8 | 0.78 | 0.86 | 0.9 |
| Codon_Usage | 0.75 | 0.81 | 0.8 | 0.86 | 0.91 |
| Online_Shoppers_Intention | 0.71 | 0.76 | 0.79 | 0.82 | 0.9 |
| Credit Card Clients | 0.73 | 0.78 | 0.81 | 0.85 | 0.91 |

As per Table 5, the 2CHCOKC is identified the utmost M value for the five massive datasets. In terms of F-Measure (M), the 2CHCOKC achieves a performance surpassing CO, GO, ALO, and K-Means by 8%, 15%, 19%, and 23% correspondingly through whole five datasets (Fig. 6).

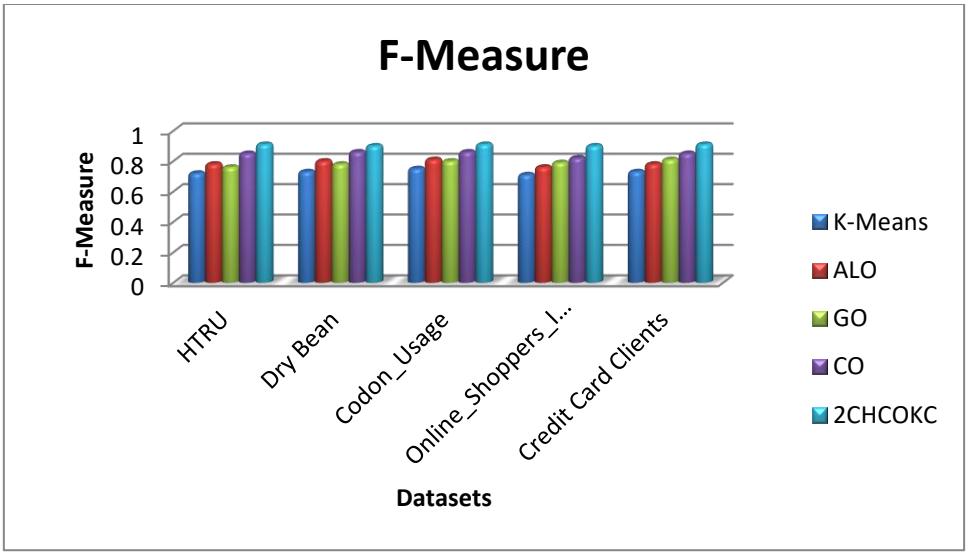



**Fig. 5: F-Measure**

### 5.2.5. Purity Index

Purity (U) functions as an indicator of the proficiency of clustering methods, showcasing the accuracy of formed clusters for data elements. As a result, whole elements attributed to a particular cluster can be exclusively assigned to that cluster. The Purity Index (I) is measured employing eq. (24) and eq. (25) derived from purity (U). Supreme purity is reached when the I value approaches 1.

$$U(A_\delta) = \frac{max(|A_{\delta\gamma}|)}{|A_\delta|} \qquad (24)$$

$$I = \sum_{\delta=1}^{T} \frac{(|A_\delta| U(A_\delta))}{|D|} \qquad (25)$$

Where,

$U(A_\delta)$ = $\delta^{th}$ cluster purity.



(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

**Table 6: Purity Index**

| Dataset | K-Means | ALO | GO | CO | 2CHCOKC |
|---|---|---|---|---|---|
| HTRU | 0.74 | 0.8 | 0.78 | 0.87 | 0.93 |
| Dry Bean | 0.75 | 0.82 | 0.8 | 0.88 | 0.91 |
| Codon_Usage | 0.76 | 0.84 | 0.82 | 0.89 | 0.92 |
| Online_Shoppers_Intention | 0.73 | 0.78 | 0.81 | 0.85 | 0.92 |
| Credit Card Clients | 0.75 | 0.79 | 0.83 | 0.87 | 0.93 |

The information presented in Table 6 indicates that the 2CHCOKC is produced the highest I value among the five massive datasets. In accordance with the Purity Index (I), the 2CHCOKC outperforms CO, GO, ALO and K-Means by margins of 7%, 14%, 17%, and 25%, correspondingly, through whole five datasets (Fig. (6)).

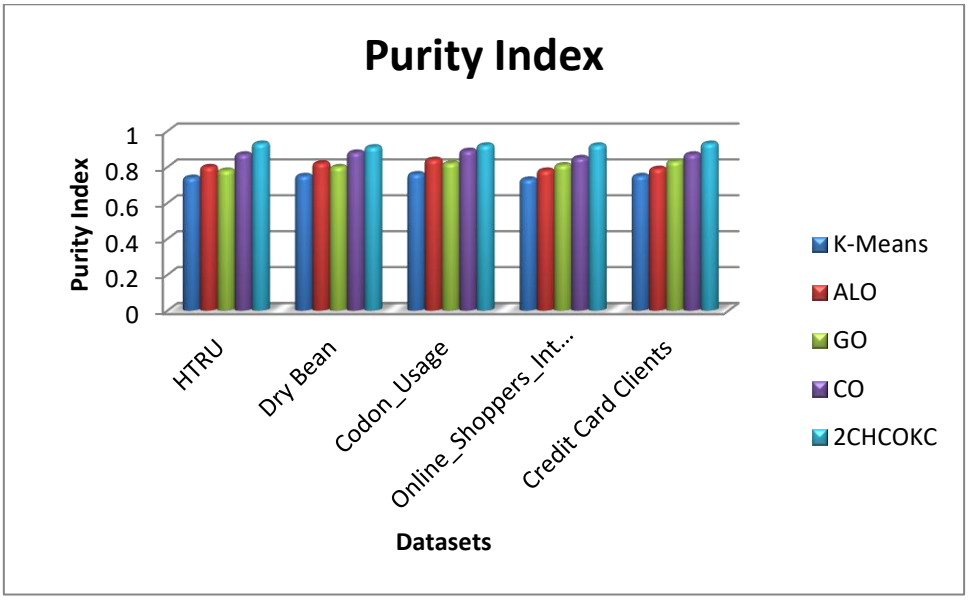



**Fig. 6: Purity Index**

### 5.2.6. Accuracy

The Accuracy (R) is measured by considering the arrangement of accurate clusters (i.e., estimating the proportion of correct determinations) and specifying the section of clusters within the predominant cluster (eq. (26)).

$$R = \frac{\sum_{\delta=1}^{T} A_{\delta}}{|D|} * 100\% \qquad (26)$$

**Table 7: Accuracy (%)**

| Dataset | K-Means | ALO | GO | CO | 2CHCOKC |
|---|---|---|---|---|---|
| HTRU | 75 | 85 | 81 | 91 | 94 |
| Dry Bean | 74 | 81 | 78 | 86 | 92 |
| Codon_Usage | 76 | 80 | 78 | 87 | 92 |
| Online_Shoppers_Intention | 73 | 79 | 80 | 86 | 91 |



(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

| | | | | | |
|---|---|---|---|---|---|
| Credit Card Clients | 77 | 82 | 83 | 88 | 93 |

The information in Table 7 indicates that the 2CHCOKC is achieved the highest R (%) value among the five massive datasets. In terms of Accuracy (R), the 2CHCOKC surpasses CO, GO, ALO and K-Means by margins of 9%, 15%, 19%, and 26%, respectively, through whole five datasets (Fig. (7)).

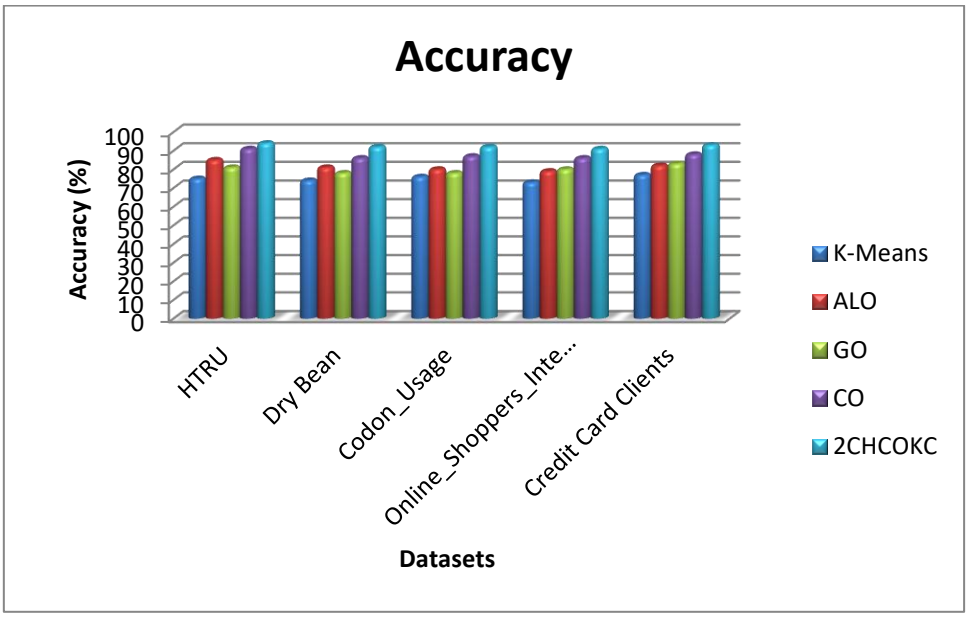



**Fig. 7: Accuracy**

The application of K-Means in combination with the CO approach for clustering guarantees exact and proficient distribution of data points inside the dataset. The inclusion of 2D chaotic behavior is integrated into the first population creation of the CO approach to enrich the search and employment skills of coyotes inside the group, a technique substantially engaged for the scattering of cluster members. As a result, the 2CHCOKC method creates cluster centers and members of superiority. Therefore, 2CHCOKC produces superior results compared to the CO, GO, ALO and K-Means methodologies.

### 5.3. 2CHCOKC method Time Complexity

The time complexity is measured through executing the clustering and counting the overall operations. The cost per step is agreed at 1 unit for every operation. The entire Execution Cost (C_X) is computed through algorithm 2 in conjunction with eq. (27) and eq. (28).

$$C_X = 1 + G + 1 + G * (W + 1) + G * W * T + G * W + G * T + G * T + 1 + T + 1 + T + T * O * Y + T * (G + 1) + T * G * Y + T * G * O + T * G * O + T * G * (Y + 1) + T * G * Y + T * G * Y + T * G * Y + T * G * O + T * O + T \quad (27)$$

$$C_X = G * W * T + T * O * Y + 5 * T * G * Y + 3 * T * G * O + 2 * G * W + 4 * G * T + T * O + 2 * G + 4 * T + 4 \quad (28)$$

Recognize that whole factors are indistinguishable in eq. (28) under worst-case conditions; thus, eq. (29) is formulated in the following manner:



(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

$$C_X = 10N^3 + 7N^2 + 6N + 4 \quad (29)$$

In the most unfavorable situations, for the 2CHCOKC, CO, GO, and ALO methods, the time complexities amount to $O(N^3)$, whereas for K-Means, it stands at $O(N^2)$. Consequently, all methods demonstrate a polynomial time complexity.

### 5.4. Statistical Explication

The statistical explication is conducted to determine the extent of significant differences among the effects of clustering approaches. In this situation, a nonparametric Friedman Valuation (FV) is performed to examine variations among the array of consecutive significant variables. Each clustering method is evenly effective, corroborating the null hypothesis ($H_0$).

The Friedman Valuation (FV) is measured utilizing eq. (30).

$$FV = \frac{12 * \alpha}{\beta * (\beta + 1)} \left[ \sum_{\emptyset=1}^{5} (V_\emptyset)^2 - \frac{\beta * (\beta + 1)^2}{4} \right] \quad (30)$$

Where,

α = Whole Datasets

β = Whole Clustering Methods

$V_\emptyset$= $\emptyset^{th}$ Method Average Rank

The FV critical value of 2.33274 is measured using F-distribution reference [35] utilizing degrees of freedom (α - 1) and (α – 1) * (β – 1), within the interval of 4 (for (5 - 1)) and 16 (for (5 – 1) * (5 – 1)) to accommodate the application of 5 clustering methods (α = 5) on 5 datasets (β = 5) with a significance level of μ = 0.10. The computed FV score surpasses the critical value for null hypothesis ($H_0$) rejection; otherwise, $H_0$ is embraced. The calculated FV value (equation (27)) is 36.16 for the utilization of 5 clustering methods (α = 5) across 5 datasets (β = 5) with a significance level of μ = 0.10. Thus, the determined FV value surpasses the critical FV threshold, leading to the refusal of $H_0$.

Therefore, a subsequent examination is conducted using the Holm methodology, wherein the suggested 2CHCOKC is subjected to statistical assessment with other clustering methods within this analysis. Initially, the Z-score is calculated through eq. (31) and eq. (32), followed by the computation of Probability (P) using the Z-score and a standard normal distribution reference [36]. Finally, the $H_\rho$ value (as provided in Table 8) is determined using the formula $\mu/(\alpha - \rho)$.

$$Z = \frac{V_\rho - V_\emptyset}{\varepsilon} \quad (31)$$

Where,


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

$$\varepsilon = \sqrt{\frac{\alpha * (\alpha + 1)}{6 \times \beta}} \qquad (32)$$

**Table 8: Holm Methodology Analysis**

| $\rho$ | Methods | Z-value | H-value | $\mu/(\alpha - \rho)$ | Hypothesis |
|---|---|---|---|---|---|
| 1 | K-Means | -3.2 | 0.00069 | 0.025 | Rejected |
| 2 | ALO | -3 | 0.00135 | 0.033 | Rejected |
| 3 | GO | -2.8 | 0.00256 | 0.05 | Rejected |
| 4 | CO | -1 | 0.15866 | 0.10 | Accepted |

Based on Table 8, the $\mu/(\alpha - \rho)$ ratio surpasses the $H_\rho$ value, implying the null hypothesis rejection across all scenarios. As a result of the aforementioned analysis, the suggested 2CHCOKC and CO methods demonstrate superior clustering proficiency when compared to the GO, ALO and K-Means methods.

## 6. Conclusions and Future Work

data clustering plays a crucial role across diverse industries, including education, healthcare, and decentralized business environments. Utilizing various classification techniques, different datasets are effectively managed and analyzed to enhance the quality of data sharing in decentralized setups. While K-Means stands as a well-established and intuitive clustering method, it faces challenges related to selecting initial cluster centers and converging to local optima, which can hinder its effectiveness. To address these limitations and enhance clustering accuracy, optimization methods come into play. In this context, the innovative 2D Chaotic based Hybrid Coyote Optimization with K-Means for Clustering (2CHCOKC) method is introduced. Its primary objective is to achieve precise and efficient distribution of data values within datasets. This is accomplished by leveraging a 2D chaotic map for population initialization. The incorporation of this chaotic map aims to enhance the capabilities of search agents within the optimization framework, leading to more accurate and optimal classification of both cluster centers and members.

The implementation of the 2CHCOKC method using MATLAB 2022a on massive datasets showcases its exceptional efficacy when compared to existing approaches like K-Means, ALO, GO, and CO. This comparison spans various critical metrics, including F-Measure, Accuracy, Standard Deviation, Root Mean Square Error, Time Complexity, Purity Index, Intra-Cluster Distance, and Statistical Explication. The success of the 2CHCOKC method underscores its potential to significantly improve clustering outcomes, making it a promising avenue for advancing clustering techniques in various real-world applications. Moving ahead, the envisaged approach is anticipated to project and authenticate using extensive databases within the realm of big data. Furthermore, consistent communication is anticipated to furnish and corroborate within the framework of the Internet of Things (IoT) in the times ahead.



(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

## Declaration

**Funding:** Not Applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Availability of data:** Not Applicable.

**Code Availability:** The Coding is developed with MATLAB 2022a environment.

**Author Contributions:** Conceptualization was originated by Pradeep Kumar Atulker and Ankur Khare; methodology and implementation of proposed work is done by Pradeep Kumar Atulker; validation is done by Rajendra Gupta and Ankur Khare; data curation is done by Rajendra Gupta and Pradeep Kumar Atulker; writing—original draft preparation, and writing—review and editing are performed by Pradeep Kumar Atulker, Rajendra Gupta and Ankur Khare.

## References

1. R. J and K. K., "Diabetes data classification using whale optimization algorithm and back propagation neural network", Int. Res. J. Pharm., Vol. 8(11), pp. 219-222, 2017. http://dx.doi.org/10.7897/2230-8407.0811242
2. 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 (107), pp. 1-21, 2021. (https://doi.org/10.3390/math9020197)
3. S. Kiruthika, EVRM. Kalaimani, R. Sudha and K. Suganthi, "ACO Feature selection and Novel Black Widow meta-heuristic Learning rate optimized CNN for Early diagnosis of Parkinson`s disease", Turkish Journal of Computer and Mathematics Education, Vol. 12 (7), pp. 809-817, 2021.
4. D. Giordano, M. Mellia and T. Cerquitelli, "K-MDTSC: K-Multi-Dimensional Time-Series Clustering Algorithm", Electronics, MDPI, Vol. 10 (1166), pp. 1-21, 2021. (https://doi.org/10.3390/electronics10101166)
5. A. N. Habowski, T. J. Habowski and M. L. Waterman, "GECO: gene expression clustering optimization app for non-linear data visualization of patterns", BMC Bioinformatics, Vol. 22 (29), pp. 1-13, 2021. 5
6. A. Khare, R. Gupta and P. K. Shukla, "A Grey Wolf Optimization Algorithm (GWOA) for Node Capture Attack to Enhance the Security of Wireless Sensor Network", International Journal of Scientific & Technology Research, Vol. 9 (3), pp. 206-209, 2020. 6
7. D. Sonntag, and H. J. Profitlich, "An architecture of open-source tools to combine textual information extraction, faceted search and information visualisation", Artificial Intelligence In Medicine, Elsevier, Vol. 93, pp-13-28, 2019. (https://doi.org/10.1016/j.artmed.2018.08.003) 7
8. X. Cui, Y. Li, J. Fan, T. Wang and Y. Zheng, "A Hybrid Improved Dragonfly Algorithm for Feature Selection", IEEE Access, Vol. 8, pp. 155619-155629, 2020. 8


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

9. F. L. Vinmalar and A. K. Kombaiya, “An Improved Dragonfly Optimization Algorithm based Feature Selection in High Dimensional Gene Expression Analysis for Lung Cancer Recognition”, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol. 9 (8), pp. 896-908, 2020. 9
10. M. Daouda, and M. Mayo, “A survey of neural network-based cancer prediction models from microarray data”, Artificial Intelligence in Medicine, Elsevier, Vol. 97, pp-204-214, 2019. (https://doi.org/10.1016/j.artmed.2019.01.006) 10
11. X. Cui, Y. Li, J. Fan, T. Wang and Y. Zheng, “A Hybrid Improved Dragonfly Algorithm for Feature Selection”, IEEE Access, Vol. 8, pp. 155619-155629, 2020. 11
12. H. Chantar, M. Tubishat, M. Essgaer and S. Mirjalili, “Hybrid Binary Dragonfly Algorithm with Simulated Annealing for Feature Selection”, SN Computer Science, Vol. 2(295), pp. 1-11, 2021. 12
13. 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 (6076), pp. 1-23, 2020. (doi:10.3390/s20216076) 13
14. V. M. Herrera, T. M. Khoshgoftaar, F. Villanustre and B. Furht, “Random forest implementation and optimization for Big Data analytics on LexisNexis’s high performance computing cluster platform”, Journal of Big Data, pp-1-36, 2019. 14
15. 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, MDPI, Vol. 4 (9), pp. 1-16, 2020. (doi:10.3390/bdcc4020009) 15
16. Y. Denga, A. Sanderb, L. Faulstichb, and K. Deneckea, “Towards automatic encoding of medical procedures using convolutional neural networks and autoencoders”, Artificial Intelligence in Medicine, Vol. 93, pp-29-42, 2019. (https://doi.org/10.1016/j.artmed.2018.10.001) 16
17. N. Shahid, T. Rappon, and W. Berta, “Applications of artificial neural networks in health care organizational decision-making: A scoping review”, PLOS ONE, pp-1-22, 2019. (https://doi.org/10.1371/journal.pone.0212356) 17
18. S. Mitta, “Cognitive Computing Architectures for Machine (Deep) Learning at Scale”, Proceedings, MDPI, Vol. 1 (186), 2019. 18
19. T. B. Nun and T. Hoefler, “Demystifying Parallel and Distributed Deep Learning: An In-Depth Concurrency Analysis”, cs.LG, pp-1-47, 2019. 19
20. B. A. Tama and S. Lim, “A Comparative Performance Evaluation of Classification Algorithms for Clinical Decision Support Systems”, Mathematics, MDPI, Vol. 8 (1814), pp. 1-25, 2020. (doi:10.3390/math8101814) 20
21. Y. Yao, H. Lei, T. Gedeon and L. Zheng, “Large-scale Training Data Search for Object Re-identification”, arXiv, cs.CS, pp. 1-13, 2023. 21
22. B. Hea, Y. Guana, and R. Dai, “Classifying medical relations in clinical text via convolutional neural networks”, Artificial Intelligence In Medicine, Elsevier Vol. 93, pp-43-49, 2019. (https://doi.org/10.1016/j.artmed.2018.05.001) 22
23. A. Khare, P. Shukla and S. Silakari, “Secure and Fast Chaos based Encryption System using Digital Logic Circuit”, I. J. Computer Network and Information Security, MECS, Vol. 6, pp. 25-33, 2014. 23


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

24. A. Khare, P. K. Shukla, M. A. Rizvi and S. Stalin, “An Intelligent and Fast Chaotic Encryption using Digital Logic Circuits for Ad-Hoc and Ubiquitous Computing”, Entropy, MDPI, Vol. 18 (201), pp. 1-27, 2016. 24
25. G. Talari, E. Cummins, C. McNamara and J. O`Brien, “State of the art review of Big Data and web-based Decision Support Systems (DSS) for food safety risk assessment with respect to climate change”, Trends in Food Science & Technology, Elsevier, Vol. 126, pp. 192-204, 2022. 25
26. L. Abualigah, A. H. Gandomi, M. A. Elaziz, H. A. Hamad, M. Omari, M. Alshinwan and A. M. Khasawneh, “Advances in Meta-Heuristic Optimization Algorithms in Big Data Text Clustering”, Electronics, MDPI, Vol. 10 (101), pp. 1-30, 2021. 26
27. S. L. M. Mosharraf and M. A. Adnan, “Improving lookup and query execution performance in distributed Big Data systems using Cuckoo Filter”, Journal of Big Data, Vol. 9 (12), pp. 1-30, 2022. 27
28. B. A. Hassan and T. A. Rashid, “Multi-disciplinary Ensemble Algorithm for Clustering Heterogeneous Datasets”, Neural Computing and Applications, pp. 1-30, 2021. 28
29. J. Lei, “Association Rule Mining Algorithm in College Students` Quality Evaluation System”, Journal of Electrical and Computer Engineering, Hindawi, Vol. 2022, pp. 1-9, 2022. 29
30. T. Singh, N. Saxena, M. Khurana, D. Singh, M. Abdalia and H. Alshazly, “Data Clustering Using Moth-Flame Optimization Algorithm”, Sensors, MDPI, Vol. 21 (4086), pp. 1-19, 2021. 30
31. D. Giordano, M. Mellia and T. Cerquitelli, “K-MDTSC: K-Multi-Dimensional Time-Series Clustering Algorithm”, Electronics, MDPI, Vol. 10 (1166), pp. 1-21, 2021. 31
32. G. Li, J. Sun, M. N. A. Rana, Y. Song, C. Liu and Z. Y. Zhu, “Optimizing High-Dimensional Functions with an Efficient Particle Swarm Optimization Algorithm”, Mathematical Problems in Engineering, Hindawi, Vol. 2020, pp. 1-10, 2020. 32
33. R. Srivastava, “Application of Optimization Algorithms in Clusters”, Frontiers in Chemistry, pp. 1-17, 2021. 33
34. K. Tadist, F. Mrabti, N. S. Nikolov, A. Zahi and S. Najah, “SDPSO: Spark Distributed PSO-based approach for feature selection and cancer disease prognosis”, Journal of Big Data, Vol. 8 (19), pp. 1-22, 2021. 34
35. F Distribution Table, 2018 Mar 18. Retrieved from http://www.socr.ucla.edu/applets.dir/f_table.html.
36. Normal Distribution Table. Retrieved from http://math.arizona.edu/~rsims/ma464/standardnormaltable.pdf.


(This work is licensed under a Creative Commons Attribution 4.0 International License.)