Iraqi Journal for Computer Science and Mathematics

Manuscript 1321

Multi Features Data Clustering Using Novel Statistical Method with Application to Color Images

Husham Y. A. Alameen

Ali Karah Bash

Follow this and additional works at: https://ijcsm.researchcommons.org/ijcsm

 Part of the Computer Engineering Commons

IRAQI JOURNAL FOR COMPUTER SCIENCE AND MATHEMATICS 2025;6:678–696

Scan the QR to view
the full-text article on
the journal website

ORIGINAL STUDY

Multi Features Data Clustering Using Novel Statistical Method with Application to Color Images

Husham Y. A. Alameen [a], **Ali Karah Bash** [b,c,*]

[a] College of Computer Science and Mathematics, University of Mosul, Postal code: 41002, Mosul, Iraq
[b] Department of Electrical and Electronic Engineering, Faculty of Engineering, Hasan Kalyoncu University, Şahinbey, Postal code: 27110, Gaziantep, Türkiye
[c] Due to dual citizenship, the author's name can be represented in two different forms: Ali Karabaş and Ali Karah Bash.

ABSTRACT

The efficiency and performance of the color image clustering algorithms are determined by various factors, including accuracy, data size, speed, and reliability (the absence of randomness in the results). Some applications, like microscopes analyzing images of biological objects or telescopes observing planetary motion prioritize accuracy over execution time. In contrast, surveillance cameras and moving object tracking prioritize speed and reliability over accuracy. This study introduces a novel algorithm that balances these four factors by clustering data with multiple features linked through specific relationships. The proposed algorithm has been practically applied to RGB color images. Traditional clustering methods, such as K-means, K-medoids, and Fuzzy C-means (FC-means), often prioritize arithmetic operations (e.g., means), which can deviate from true values, reducing calculation accuracy and resulting in outliers. Another limitation of traditional clustering methods is the irregularity in their outcomes, which reduces reliability due to the inherent randomness in selecting the initial point. These methods also perform poorly when handling large and multidimensional datasets. In contrast, the suggested clustering method employs a completely different approach. It identifies the Relationship Between Features Clustering (RBFC) of a specific object, represented across several matrices. It reduces their dimensions into a one-dimensional matrix representing the dissimilarities among such relations. The data is then clustered. Applying the RBFC method to a set of color images verified that the said method outperformed the aforementioned traditional methods insofar as clustering accuracy and processing time were involved. The suggested method also improves reliability by removing randomness and reliably producing color clusters in the same sequence throughout multiple runs. Moreover, the RBFC approach has shown remarkable performance in large datasets like medical and satellite images. The findings highlight that the RBFC approach is an efficient and effective clustering method for complex data scenarios and image processing applications.

Keywords: Color image segmentation, Relationship between features clustering, Multi features data, Color image clustering

1. Introduction

In recent years, digital imaging technology has seen very fast growth, which has produced a great deal of image data. Due to this, effective and efficient clustering algorithms are required to analyze image data. Clustering algorithms are used to categorize similar data items into one or more clusters. The effectiveness and efficiency of different clustering algorithms on a variety of applications, such as object segmentation and color classification, are quite significant as they influence medical imaging diagnosis and navigation of autonomous vehicles. Choosing the right color space is very important

Received 4 March 2025; revised 9 June 2025; accepted 11 June 2025.
Available online 29 September 2025

* Corresponding author.
E-mail addresses: hisham.alameen@uomosul.edu.iq (H. Y. A. Alameen), ali.karabash@hku.edu.tr, ali_karabash2016@yahoo.com (A. Karah Bash).

https://doi.org/10.52866/2788-7421.1321
2788-7421/© 2025 The Author(s). This is an open-access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

Nomenclature

RBFC	Relationship Between Features Clustering	**SA**	Segmentation Accuracy
FC-means	Fuzzy C-means	**RGB**	Red-Green-Blue
TV	Template Vector	**RV**	Relationship Vector
SM	Shrinking Matrix	**ED**	Euclidean Distance
MD	Manhattan Distance	**JSON**	JavaScript Object Notation
HSV	Hue, Saturation, Value	**HDM**	Hamming Distance Measure
TM	template matrix	**DRCS**	Data Reduction and Coding Stage
TVCS	Template Vector Creation Stage	**MS**	Matching Stage
LOERGAN	Log of the Exponent Rule Generative Adversarial Network		

for color image clustering. Each color space has its own characteristics, which affect the performance and efficiency of the clustering algorithm involving image segmentation color. A study by Zhou et al. [21] proved the efficiency of the color LAB color space in the accurate segmentation of color images.

Experts in color image clustering have used cluster analyses for studying different applications in that field over many years. For instance, Zhang and Chen [20] conducted a critical review of images of ornamental surfaces to investigate visual perception associated with custom wardrobe finishes and identified design factors through color and texture data. To get over this limitation, robust distance measures have been devised as a long-term solution that is less influenced by noise and outliers. Moreover, clustering algorithms commonly use standard distance measures like Euclidean distance for their straightforwardness and computational ease [3]. Clustering algorithms are not sufficiently accurate in capturing the authentic structure of the dataset due to their exhibiting high outlier and noise sensitivity. It has been suggested that robust distance measures can provide a fruitful mechanism to enhance the inertia of the outcome against outliers and noise.

When robust statistical techniques are used, clustering algorithms become better able to preserve complex features in colored images and discover significant structures. The recent development of clustering algorithms has resulted in the combination of graph-based multi-view clustering techniques with customized models that can be used with diverse data. Hidden clusters in complex data are revealed through the use of dissimilarity matrices and objective weighting techniques. Moreover, advanced self-adaptive fusion approaches that utilize geometric distances have been shown to work very well for uncertainty management and combating the effects of noise and outliers.

As much progress has been made, there is still a demand for clustering algorithms that can effectively handle multi-attribute information, apart from proving high accuracy and robustness to randomness and outlier. B-spline curves are flexible enough to fit the curve used in modeling anything from field roads to almost anything else [21]. According to He and Zhang [9], using FC-means clustering image segmentation introduced the notion of partial membership to clusters which is better than other approaches like K-means. The outcome of this research has helped in enhancing the optimization of the performance and efficiency of color image clustering algorithms [15].

Although K-means, K-medoids, and other similar clustering algorithms have been widely used, these are still susceptible to various drawbacks [17]. As an illustration, Zhang proposed an adaptation of K-means that involves online sampling and is capable of enhancing the approximation ratio for outlier-rich datasets [20]. Similarly, Ditton et al. [5] proposed the Hamming Distance Measure (HDM) technique for image clustering.

After that, Tang et al. [2] employed new methods using robust distance-based measures to cluster multi-attribute data in color images. Image color clustering uses region-based methods and robust fuzzy clustering methodology. It is an attempt to improve the robustness of clustering when the presence of noise and outliers is present, without losing the features of images.

Additionally, Qi et al. [16] proposed new techniques for time-series data clustering based on an annealing machine to generate highly comparable clusters and ensure the robustness of outliers. D'Urso and Vitale [4] introduced a new technique for image partitioning into dictionaries encoding the structure of JavaScript Object Notation (JSON) characterized by key-value pairs. Recent developments introduced a clustering approach in image inpainting systems via the Log of the Exponent Rule Generative Adversarial Network (LOERGAN) algorithm, enhancing the efficacy as well as quality of image repair [13, 16, 25, 26].

Dong et al. [6] boosted multi-view clustering by integrating metric and graph learning approaches.

Table 1. Key insights from reviewed studies on color image clustering methods.

Reference	Case Study	Used Method	Inputs	Outputs
Zhou et al. [21]	Color image segmentation	K-means clustering	RGB values	Image segmentation with improved color differentiation
Zhang & Chen [20]	Custom wardrobe surface analysis	Multi-view clustering	Color moments, texture data	Analysis of ornamental surfaces using visual perception
D'Urso et al. [3]	Medical image processing	Euclidean distance-based clustering	Pixel intensity values	Medical image segmentation with high accuracy
He & Zhang [9]	Remote sensing image classification	FC-means	Satellite image data	Land cover classification and segmentation
Singh & Verma [17]	High-dimensional image clustering	K-medoids clustering	Multispectral images	Clustered data with reduced sensitivity to noise
Li et al. [12]	Object recognition in images	Graph-based clustering	Feature vectors	Accuracy enhancement in object identification
Kuwil et al. [11]	Large-scale data clustering	Gravity center-based method	Image datasets	Reduced computational complexity in clustering
Dong et al. [6]	Multi-view clustering	Metric & graph learning	Self-expressiveness property	High structural awareness in image data
Xiang [12]	Image dataset classification	Loss function optimization	Feature representations	Improved purity and fairness in clustering results
Niemelä & Kärkkäinen [14]	Multi-cluster learning	Attention-based clustering	Cluster correlation data	Reinforced distant positive samples for better separation
Qi et al. [16]	Time-series clustering	Annealing machine-based clustering	Temporal image data	Balanced classification with improved resilience to outliers
D'Urso & Vitale [4]	Image segmentation using metadata	JSON-based clustering	Key-value pairs in image dictionaries	Efficient segmentation with metadata-driven analysis
Liu et al. [13]	Clustering-focused image inpainting	LOERGAN (GAN-based)	Image patches	High-quality inpainting with generative modeling
Yang et al. [19]	Social network-based clustering	Graph-based techniques	Social graph data	Improved clustering accuracy for social image data
Ikotun et al. [10]	High-dimensional data clustering	Distance-based clustering	Multi-attribute datasets	Optimal cluster number determination
Gou & Zeng [7]	Big data clustering	Spark-based clustering	Large-scale image data	Scalable clustering with improved efficiency
Tang et al. (2024)	Robust clustering of noisy images	Region-based & fuzzy clustering	Image pixel intensity	Increased robustness to noise and outliers
Yadav & Sharma [18]	Student admissions interest mapping	K-means association rules	Enrollment datasets	100% accuracy in identifying key influencing factors
Poullis [15]	Real-world image clustering	Advanced multi-feature clustering	Image attributes	Optimized clustering for diverse image environments
Current Study	Color image clustering	Relationship Between Features (RBF)method	RGB image datasets (BSDS500 dataset)	High-speed, accurate, and non-random clustering results

Further, Xiang [12] designed a new loss function to maintain purity consistency within clusters and guarantee fairness across the cluster models [12, 27]. Niemelä and Kärkkäinen [14] also improved the related work with a recent attention mechanism for calculating the correlations of clusters. Table 1 highlights the contribution of diverse research to developing the design of different clustering methods.

Recent advancements in clustering have seen the rise of deep-learning-driven models, such as convolutional autoencoders and graph convolutional networks, which aim to learn compact representations and structure-aware relationships within data. These approaches have shown promising results, particularly in high-dimensional or abstract image data. However, they often demand extensive labeled datasets, high computational resources, and careful tuning. The current study focuses on a simpler alternative with lower computational cost, aiming for broader accessibility in applied image processing contexts [22–24, 28].

Building on these developments, this study introduces a novel statistical clustering method for color image analysis, emphasizing high feature correlations. The main aim of this article is to avoid the limitations of traditional clustering methods like K-means, K-medoids, and FC-means by utilizing a matrix-based methodology instead of conventional distance measures. The proposed method is quite efficient in both computation and clustering and is

extremely effective for color image processing. Also, the study provides a new process for studying the relationships between various features with an organized encoding scheme for producing similar and dependable clusters.

The key contributions of this research are summarized as follows:

1. The provided clustering technique relies on using pre-computed template matching (TM), enabling quick clustering of color images. This enables one to achieve quicker outcomes in monitoring systems, i.e., target recognition and tracking.
2. The precision and controlling capability for the number of clusters make the current method universal for diverse applications like image processing and pattern recognition.
3. The process guarantees stability and reliability by doing away with randomness, rendering it highly suitable for error-prone applications such as military technology.
4. It is scalable to large datasets, such as large color image datasets, medical images, satellite images, and easily deals with high-dimensional data of up to four or five dimensions without any loss of performance.

As a final point, the Relationship Between Features Clustering (RBFC) algorithm introduced in the research can be implemented to carry out image segmentation of a color image. By using the formulation that is based on a matrix, the solution enhances accuracy, performance, and stability and thus ensures reproducible and systematic cluster results. The advancements in image processing that this study will afford, through the utilization of both in practical applications, as well as theoretical research in computer vision and pattern recognition.

The paper structure is organized as follows: Section 1 provides an introduction and literature review of the research topic. Section 2 outlines the methodology of this study. Section 3 provides a comprehensive analysis of the results and discusses the findings. Finally, Section 4 concludes this study.

2. Materials and methods

The main objective of the present study is to develop an improved statistical clustering model known as the Relationship Between Features Clustering (RBFC) method for clustering color images. The research also investigates the performance and efficiency of RBFC with regard to K-means, K-medoids, and FC-means. To analyze the effectiveness of the clustering techniques, the Berkeley Segmentation Dataset (BSDS500) is exploited [1].

2.1. The K-means clustering method

The K-means clustering algorithm is an unsupervised ML algorithm that divides data into k clusters. The objective of the algorithm is to optimize positions of centroids to minimize variance within the cluster [18]. It uses a mathematical objective function to minimize the distance of data points from cluster centroid (smaller is better).

$$\min_{C} \sum_{i=1}^{k} \sum_{x_j \in C_i} ||x_j - \mu_i||^2 \tag{1}$$

where C_i is i-th cluster, x_j is a data point, μ_i is the centroid of a cluster C_i.

In recent years, significant advances have led to the development of numerous types of K-means, including causal clustering, in which aggregating unspecified counterfactual functions estimates the effectiveness of heterogeneous processes. This major aspect has transformed subgroup analysis in complex data sets. It is also proposed that association rules in K-means be integrated for accuracy improvement in specific scenarios, i.e., student admissions analysis. Furthermore, the adaptive K-means clustering approach was introduced to solve class imbalance issues in ML. Dynamically, the approaches identify the optimal numbers of clusters for various datasets, thus enhancing the classification performance. These advances reflect the variety and complexity of K-means clustering approaches within various fields.

2.2. K-medoids clustering method

The K-medoids is a robust clustering algorithm that is better at grouping data points into clusters, thereby reducing outliers in big datasets. In contrast to the K-means algorithm, the K-medoids choose real data points to serve as the centers of clusters, or medoids. Therefore, this algorithm is more resistant to noise and outliers [8]. The main aim of the objective function in K-medoids is to reduce the sum of the differences between data points and their respective medoids, provided mathematically as:

$$\min_{M} \sum_{i=1}^{k} \sum_{x_j \in C_i} d\left(x_j, m_i\right) \tag{2}$$

where M is the set of m_i medoids, C_i represents i-th cluster, x_j refers to the data points within the cluster C_i, and $d(x_j, m_i)$ is the dissimilarity (typically measured by a distance metric) between data point x_j and medoid m_i.

The k-medoids algorithm was demonstrated to attain over 90.6% accuracy in the detection of power consumption outliers, mitigating the impacts of transient outliers. Similarly, K-medoids was employed in clustering education indicator data for Indonesian provinces, with a demonstration of its outlier handling and insightful contributions to policy-making. Additionally, the method was utilized in examining human development indices, which identified significant regional inequalities in development. Briefly, K-medoids is also recognized for being simple to implement and robust across a range of applications, such as customer support clustering and digital transformation analysis.

2.3. Fuzzy C-means (FC-means) clustering method

Fuzzy C-means (FC-means) is a method of clustering used in data analysis that creates two or more clusters from one piece of data. Unlike the traditional clustering methods like K-means and K-medoids, which are exclusively assigned to one cluster, this method incorporates the notion of fuzziness, where each data point belongs to more than one cluster. The main goal of the FC-means algorithm is to reduce the objective function as follows:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m}||x_i - c_j||^2 \tag{3}$$

where J_m is the objective function to be minimized, N is the number of data points, C is the number of clusters, u_{ij} is the membership degree of the data point x_i in cluster j, m is a fuzziness parameter, typically $m > 1$, x_i is a data point, and c_i is the center of the cluster j. FC-means' capability to deal with overlapping clusters enhances its flexibility concerning standard hard clustering algorithms. Furthermore, it offers a more meaningful representation of data in ambiguous cluster boundaries.

This algorithm is more complex than standard hard clustering algorithms like K-means and K-medoids. It may also fall into local minima, producing suboptimal clustering results. The main benefit of FC-means is that it can deal with situations where points in the color data can be part of multiple clusters, making it more adaptable in setups with discrete or smooth color changeovers.

2.4. Proposed novel relationship between features clustering (RBFC) method

The clustering of high-dimensional data sets, i.e., color images, needs effective dimensionality reduction while simultaneously retaining the intrinsic relationships among data points—a critical aspect of the subject. The current research suggests an advanced clustering approach, titled Relationship Between Features Clustering (RBFC), after its fundamental working principle, to resolve these problems. The RBFC class of models employs a novel matrix representation to represent and reduce the interdependence among various attributes in the data. In contrast to conventional clustering methods, which classify color image data by similarity in feature values, the RBFC method here classifies data by examining interactions between several features within an individual object.

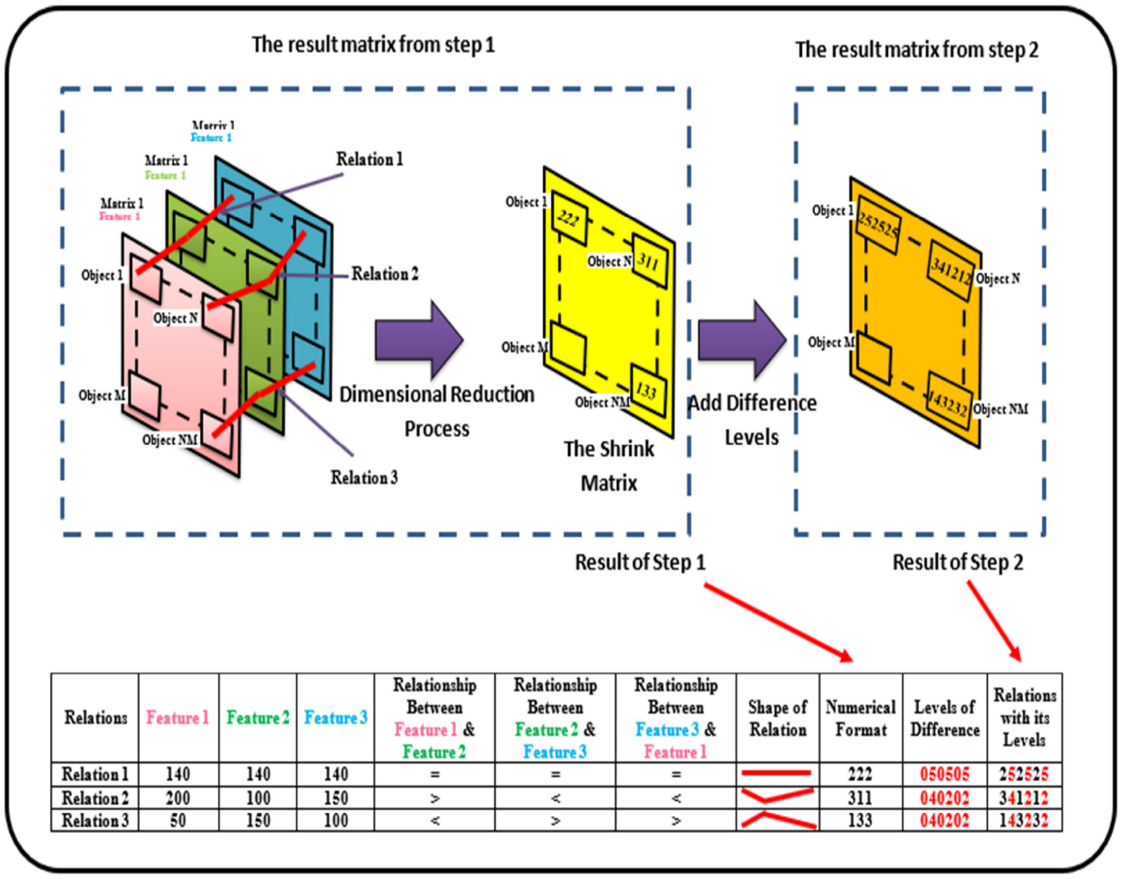
The suggested clustering technique uses three fundamental procedures: data reduction and coding, building a template matrix (TM) based on the first-phase results, and template matching with a color image data matrix. These three phases in the existing clustering technique are elaborated as follows:

A. Data Reduction and Coding Stage (DRCS)

The primary aim of the Data Reduction and Coding Stage (DRCS) of the proposed clustering method is to generate a new two-dimensional matrix from the three RGB image matrices, referred to as the relation or shrinkage matrix (SM). This step involves the dimensionality reduction of the data obtained from the color images. The SM preserves the relationships between various features of the same object. In this context, these features are compared using mathematical relationships such as greater than, less than, and equal to ($>$, $<$, $=$), respectively. Next, these mathematical relationships are encoded numerically (Labeling), where the "greater than" relation is assigned the number 3, "less than" is assigned the number 1, and "equal to" is assigned the number 2. As a result, the data from the RGB images is transformed from a more complex structure into a simpler form for clustering. This is achieved by ordering the elements of the three matrices from left to right, followed by the labeling process. Ultimately, the SM is converted into the relationship vector (RV).

The leveling process is then applied to compute the difference values between features, where the numerical values are divided according to the color property, which spans a range from 0 to 255. This division is performed using a factor of 28, producing a set of new gradient levels. The choice of 28 as the divisor is not arbitrary; it is based on a calculated selection that results in 9 distinct levels, ranging from 0 to 9, in alignment with the following mathematical equation:

$$levels = \frac{\max value}{28} = 9.1 \approx 9\ Levels \tag{4}$$



Relations	Feature 1	Feature 2	Feature 3	Relationship Between Feature 1 & Feature 2	Relationship Between Feature 2 & Feature 3	Relationship Between Feature 3 & Feature 1	Shape of Relation	Numerical Format	Levels of Difference	Relations with its Levels
Relation 1	140	140	140	=	=	=		222	050505	252525
Relation 2	200	100	150	>	<	<		311	040202	341212
Relation 3	50	150	100	<	>	>		133	040202	143232

Fig. 1. The procedure of RGB image reduction and feature coding stage of RBFC method.

Table 2. The numeric representation of the relationships between features for three pixels.

Pixels	R	G	B	Relationship between R-G	Relationship between G-B	Relationship between B-R	Numerical Codes
First	200	30	30	R > G	G = B	B < R	321
Second	200	60	60	R > G	G = B	B < R	321
Third	50	150	50	R < G	G > B	B = R	132

where the maximum value is 255. Finally, the number of digits required to represent each color point (pixel) is reduced from nine digits to six. The mathematical relationship encoding and leveling processes significantly accelerate the subsequent processing stages and decrease the comparator time, while preserving the key distinctions between various features. The data reduction and feature coding stages are illustrated in Fig. 1.

To illustrate the digital representation of relationships among three features, consider a color image with three pixels, each represented by red, green, and blue colors, denoted as R, G, and B, respectively. These three colors take on different values across the three pixels, as detailed in Table 2.

Notably, despite the distinct G and B values in the first and second pixels, both share the same numeric representation (321). This observation becomes clearer when examining the color values of these pixels within the image. The two points exhibit a subtle gradient difference in red, highlighting the classification that is based on the overall properties of the object. The leveling process plays a crucial role in addressing this issue. This suggests that classification is grounded in the intrinsic characteristics of the object itself.

To solve the challenge of classification relying solely on the intrinsic characteristics of the object, the proposed clustering model discretizes the differences between the values of two features into levels ranging from 0 to 9. The resulting value is then appended to the numerical symbol representing the relationship between the two properties, as shown in Fig. 2. For instance, Fig. 2 illustrates the leveling process applied to the first pixel as presented in Table 1, resulting in the code (362116). When the same process is applied

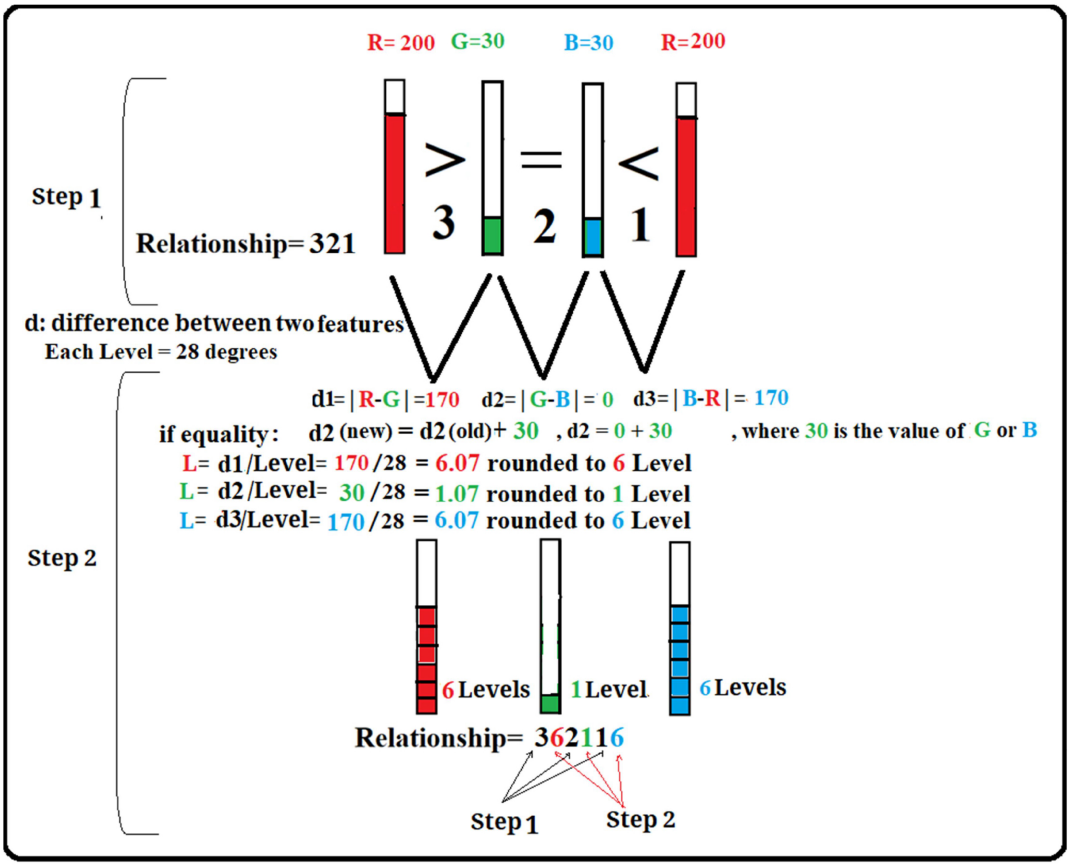


Fig. 2. The leveling process of the relationship between the feature properties of the data.

to the second pixel, it generates the code (362216), highlighting a distinction between the two codes. This effectively resolves the issue of slight differences in the color gradient between these pixels.

B. Template Vector Creation Stage (TVCS)

Once the SM is completed, the RBFC method generates the Template Vector (TV). This one-dimensional vector is derived from the SM and contains unique numbers extracted from the code numbers within the SM. Each element in the TV corresponds to a distinct cluster. To illustrate this concept, consider the relation vector RV = [311, 331, 113, 113, 113, 222, 222, 311, 321, 321, 321, 113]. In this case, the TV would be TV = [113, 222, 311, 321, 331].

C. Matching Stage (MS)

The final step of the RBFC clustering method is the Matching Stage (MS), where the actual clustering takes place. In this stage, the Template Vector (TV) is used to classify data points into distinct clusters by comparing the Relation Vector (RV) with the TV. During the matching process, each data point is evaluated against the TV to determine the cluster it belongs to. This is accomplished by taking intersections of the RV and the TV and then grouping items in the RV that correspond to items in the TV into clusters. Consider the RV and TV vectors in the example above. Five clusters are obtained by taking the intersection of two vectors' intersection operation (RV ∩ TV). For example, Cluster1 = [0, 0, 113, 113, 113, 0, 0, 0, 0, 0, 113], and similar patterns are observed for other clusters. Fig. 3 depicts clearly the procedure. The MS generates a set of well-defined clusters, each representing a set of data points with comparable properties determined by the relationships recorded in previous steps. This matching process ensures that data points are accurately grouped according to their encoded connections, increasing classification performance. As a result, the RBFC approach efficiently reduces the time and computer resources required to cluster an RGB image.

D. Cluster Number Determination

In the proposed clustering approach, TV is constructed to encompass all possible relationships between the RGB components of a color pixel. Each pixel, composed of three values (R, G, and B), can express various internal relationships when encoded numerically. For instance, a pixel might yield a label

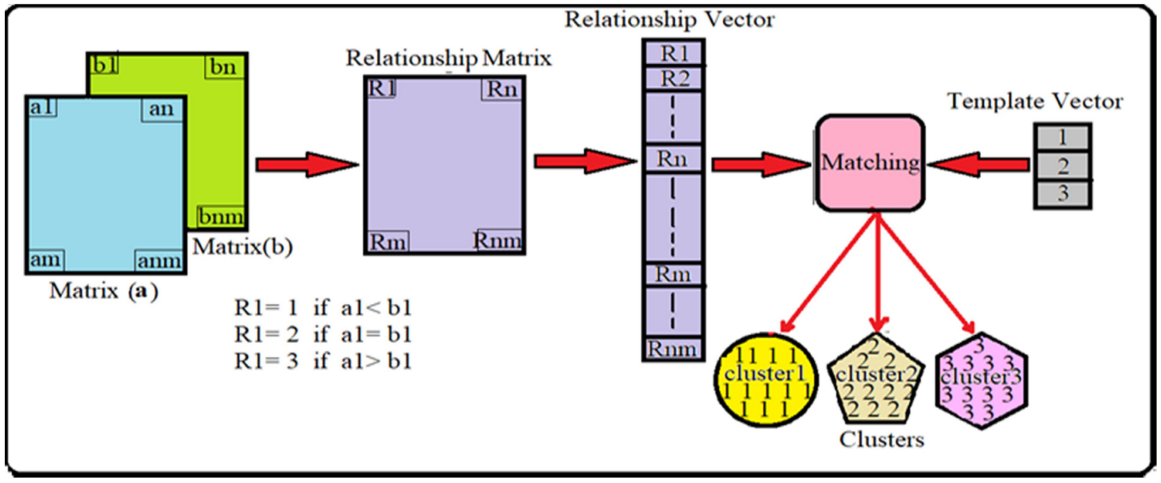


Fig. 3. Depicts the cluster creation process using the suggested RBFC clustering approach.

such as 152535 or 342414 depending on how the relationships between its color channels are interpreted. Considering all such permutations, the static template is initially designed to hold up to 64 unique patterns, each representing a distinct cluster. Once a color image is fully encoded into these label formats, every resulting label is compared against the entries in the template. A match identifies the cluster to which that pixel belongs. Depending on the content of the image, not all 64 template elements may be utilized — for example, an image with 20 distinct color clusters will only match 20 out of the 64 predefined template entries, leaving the rest unused.

To enhance flexibility and better adapt to images with varying levels of complexity, the algorithm also includes an optional dynamic template construction step. This routine analyzes the image after it has been encoded, samples one instance of each unique label, and constructs a template directly from the data. As a result, the template becomes adaptive in size — it may contain 10, 20, or even more cluster definitions depending on the diversity within the image. The total number of detected clusters is then reported, and users can specify whether they prefer to visualize all clusters or focus only on those that are dominant. In general, the static template approach yields faster execution since the reference structure is preloaded. However, the dynamic version offers improved adaptability for scenarios where cluster counts are not known in advance or where user-defined clustering granularity is required.

E. Sensitivity Level Control

The sensitivity level in the RBFC algorithm reflects its ability to distinguish variations between adjacent pixel values—essentially serving as an indicator of how pixel property differences are interpreted. In simpler terms, it gauges how finely the algorithm detects divergence among neighboring pixel attributes. One of the strengths of the RBFC approach lies in its capacity to modulate this sensitivity. Increasing the sensitivity leads to a higher number of clusters due to the algorithm's ability to recognize even slight pixel differences, while decreasing it results in more merging, thereby grouping neighboring pixels with similar properties into broader clusters.

To formalize this concept, a parameter termed Reduce of Sensitivity (RS) is introduced, allowing users to adjust clustering resolution. The sensitivity spectrum is divided into nine discrete levels, with each level determining the granularity of clustering. The classification of the Difference Between Features (DBF)—such as between the R and G components—is determined using the following equations:

$$RS = 1, 2, 3, \ldots..9 \tag{5}$$

$$Level = DBF/(RS * 28) \tag{6}$$

For instance, when $RS = 1$, the algorithm achieves the highest sensitivity, segmenting data into nine distinct clusters, each corresponding to a 28-point interval. As illustrated in Table 3 and Fig. 4, each pixel is treated as its own cluster due to the maximum sensitivity to small differences. Conversely, setting $RS = 2$ effectively doubles the interval size to 56 points (2×28), leading to merging of adjacent pixels and the formation of four broader clusters, as shown in Table 4 and Fig. 5. Similarly, an $RS = 3$ value reduces the sensitivity further, producing only three clusters from the same nine pixels, with each cluster now encompassing three neighboring values, as demonstrated in Table 5 and Fig. 6.

This adjustable mechanism demonstrates that the number of output clusters is directly dependent on the *RS* value, enabling users to tailor the clustering



Fig. 4. The nine clusters resulting from dividing the gradient ruler, in addition to a tenth cluster representing the white background of the ruler.

Table 3. Represents 9 values that were adopted for 9 adjacent pixels on the credit score. Each pixel represents one cluster.

Pixels	Clusters	RGB	RGBR	Label
P1	C1	15 43 43	15 43 43 15	1 1 2 1 3 1
P2	C2	15 71 71	15 71 71 15	1 2 2 2 3 2
P3	C3	15 100 100	15 100 100 15	1 3 2 3 3 3
P4	C4	15 125 125	15 125 125 15	1 4 2 4 3 4
P5	C5	15 160 160	15 160 160 15	1 5 2 5 3 5
P6	C6	15 188 188	15 188 188 15	1 6 2 6 3 6
P7	C7	15 216 216	15 216 216 15	1 7 2 7 3 7
P8	C8	15 244 244	15 244 244 15	1 8 2 8 3 8
P9	C9	15 255 255	15 255 255 15	1 9 2 9 3 9

resolution according to application needs—whether seeking finer segmentation or broader generalization.

F. Computational Complexity Analysis

The computational efficiency of the proposed RBFC method can be evaluated by analyzing its three primary stages: feature relationship encoding, template vector generation, and cluster matching. For an input image containing n pixels and assuming a fixed number of features (e.g., three for RGB), the encoding stage processes each pixel individually, resulting in a time complexity of approximately O(n). The TV is generated from a fixed-length relationship matrix, which is also processed linearly.

Table 4. Shows the division of the nine pixels into five clusters.

Pixels	Clusters	RGB	RGBR	Label
P1	C1	15 43 43	15 43 43 15	1 1 2 1 3 1
P2		15 71 71	15 71 71 15	1 1 2 1 3 1
P3	C2	15 100 100	15 100 100 15	1 2 2 2 3 2
P4		15 125 125	15 125 125 15	1 2 2 2 3 2
P5	C3	15 160 160	15 160 160 15	1 3 2 3 3 3
P6		15 188 188	15 188 188 15	1 3 2 3 3 3
P7	C4	15 216 216	15 216 216 15	1 4 2 4 3 4
P8		15 244 244	15 244 244 15	1 4 2 4 3 4
P9	C5	15 255 255	15 255 255 15	1 5 2 9 3 9

G. Cluster Synthesis Feature

The RBFC method provides a practical capability for synthesizing selective clusters. Following the initial execution of the algorithm, each identified cluster is presented in a predefined order. Users are given the flexibility to choose specific clusters of interest for reconstruction, effectively allowing for the reassembly of an image composed only of selected segments. This functionality proves particularly useful in scenarios where the aim is to suppress noise-related clusters while retaining meaningful regions of the image. By specifying the indices of relevant clusters in a subsequent operation, the method reconstructs a refined composite image containing only the desired components. Such an approach enables targeted analysis or enhancement by filtering out visually or semantically irrelevant data in complex color images.

3. Results and discussion

This study evaluates the performance of the cutting-edge RBFC algorithm for grouping RGB images. The performance and efficiency of the proposed technique were evaluated using colored disk images, medical, satellite, noisy images, and the Berkeley Segmentation Dataset (BSDS500). Fig. 7 shows a sample of images from the Berkeley segmentation dataset (BSD500) that contains 500 images of natural scenes with 481 × 321 pixels. The present study compares the Radial Basis Function of RBFC method with

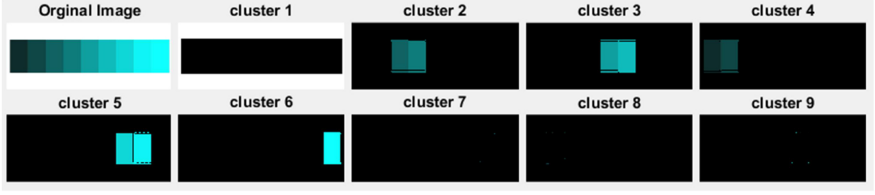


Fig. 5. The merging of every two adjacent pixels into a single cluster after reducing the sensitivity (RS = 2), resulting in five clusters in addition to the background cluster.



Fig. 6. The merging of every three adjacent pixels into a single cluster after reducing the sensitivity (RS = 3), resulting in three clusters in addition to the background cluster.

Table 5. The division of the nine pixels into three clusters.

Pixels	Clusters	RGB	RGBR	Label
P1	C1	15 43 43	15 43 43 15	1 1 2 1 3 1
P2		15 71 71	15 71 71 15	1 1 2 1 3 1
P3		15 100 100	15 100 100 15	1 1 2 1 3 1
P4	C2	15 125 125	15 125 125 15	1 2 2 2 3 2
P5		15 160 160	15 160 160 15	1 2 2 2 3 2
P6		15 188 188	15 188 188 15	1 2 2 2 3 2
P7	C3	15 216 216	15 216 216 15	1 3 2 3 3 3
P8		15 244 244	15 244 244 15	1 3 2 3 3 3
P9		15 255 255	15 255 255 15	1 3 2 3 3 3

three conventional clustering techniques. They are K-means, K-medoids, and FC-means on RGB images. The performance and effectiveness of clustering algorithms are gauged with the segmentation accuracy (SA) metric [12].

$$SA = \frac{\text{Number of correctly} - \text{classified pixels}}{\text{Total number of pixels}} \quad (7)$$

Moreover, the performance of the clustering method can also be effectively investigated through the purity issue, which assesses the extent to which a cluster is composed of data points from a single class. Purity measures the proportion of data points properly assigned to their respective clusters. The mathematical model of purity can be expressed as:

$$\text{purity} = \frac{1}{N}\sum_{k} \max_{j} \left|C_k \cap T_j\right| \quad (8)$$

where N is the total number of data points. k is the number of clusters. C_k is the set of points assigned to cluster i. T_j is the set of points belonging to class j. The $max_j |C_k \cap T_j|$ represents the size of the largest class in cluster i. The purity results are presented in Table 6.

The purity value ranges from 0 to 1. Its maximum value refers to perfect clustering, where each cluster includes one class only, and a value closer to 0 suggests poor clustering performance. Purity is a simple, intuitive evaluation metric, but it does not account for the number of clusters or their size and may favor a clustering solution with many small, homogeneous clusters.

Purity evaluation was performed on the first implementation of each of the clustering methods discussed in this study. The results of the Purity evaluation are presented in Table 6. For the RBFC method, it was applied to two cases: before image processing (RBFC1) and after image processing (RBFC2). The Purity results for the methods are as follows: K-means = 0.942704, K-medoids = 0.942952, FC-means = 0.941736, and RBFC1 = 0.9749. After preprocessing the color image, the results of RBFC2, when applied to a 12-color image with a size of 500 × 500, were as follows: Outlier Points = 0, Processing Time = 0.7321 seconds, and Purity of RBFC2 = 1.

3.1. Results of colored disk images

In this study, color image clustering was performed using 12-color and 24-color images. Figs. 8 and 9 present the results of applying the proposed RBFC model alongside three traditional methods—K-means, K-medoids, and FC-means—on a 12-color disk image across two execution stages. These figures



Fig. 7. An example of images from the Berkeley segmentation dataset (BSD500).

Table 6. Represents the results of the Purity evaluation.

Execution Time	Clustering Methods	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
First	Kmeans1	14590	14587	75221	14547	14350	14670	14719	14499	28838	14675	14479	14265	560
Second	Kmeans2	14265	75475	14760	14501	14350	14675	14547	14575	14719	14740	14479	14590	14324
	Purity	**0.9427**												
First	Kmediods1	14572	14564	14741	14608	14506	28765	14560	75260	14350	14752	14477	526	14319
Second	Kmediods2	75493	14752	14547	14350	14572	14714	14675	14506	14262	14592	14741	14319	14477
	Purity	**0.9429**												
First	Fcmeans1	14462	14157	14440	14566	75801	14161	14687	14777	724	14191	14947	28572	14515
Second	Fcmeans2	14760	14157	14498	14161	963	14516	13761	14191	14551	28762	14776	14849	76055
	Purity	**0.9417**												
First	RBFC1	14733	13839	13750	14747	14815	13765	72388	13828	14725	13893	14753	13756	14733
	Purity	**0.9749**												
Second	RBFC2	15459	14215	15592	14106	15318	14094	72388	14148	15473	14271	15482	14112	15342
	Purity	**1**												

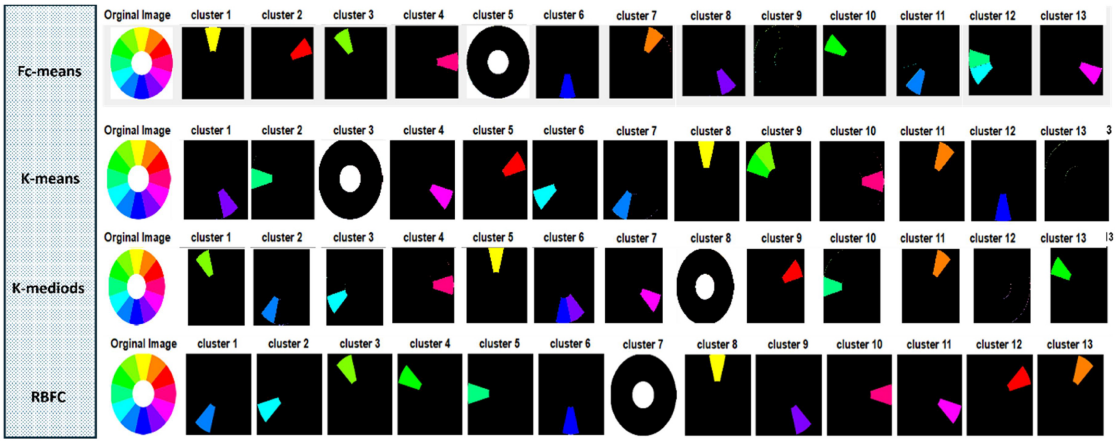


Fig. 8. Results of the four cluster methods in the first execution for the 12 color images.

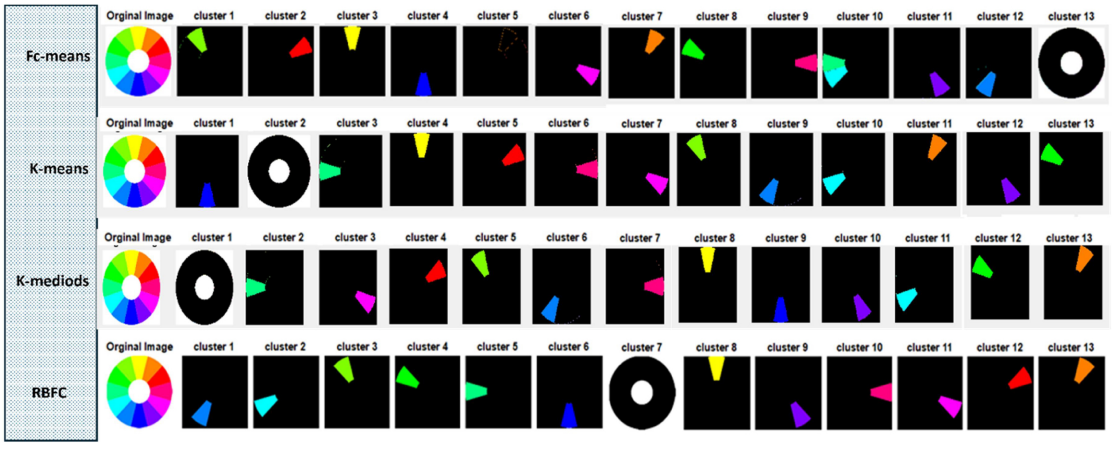


Fig. 9. Results of the four cluster methods in the second execution for the 12 color images.

highlight the stability of these methods in terms of maintaining the order and locations of the clusters, as well as showing the execution times across both runs.

The results reveal that the inherent randomness in the three classical methods negatively impacted the arrangement of clusters, leading to inconsistent displays between the first and second executions. In contrast, the RBFC model demonstrated stability by maintaining the same order and index in displaying 13 clusters, including 12 color segments and the background cluster, across both executions.

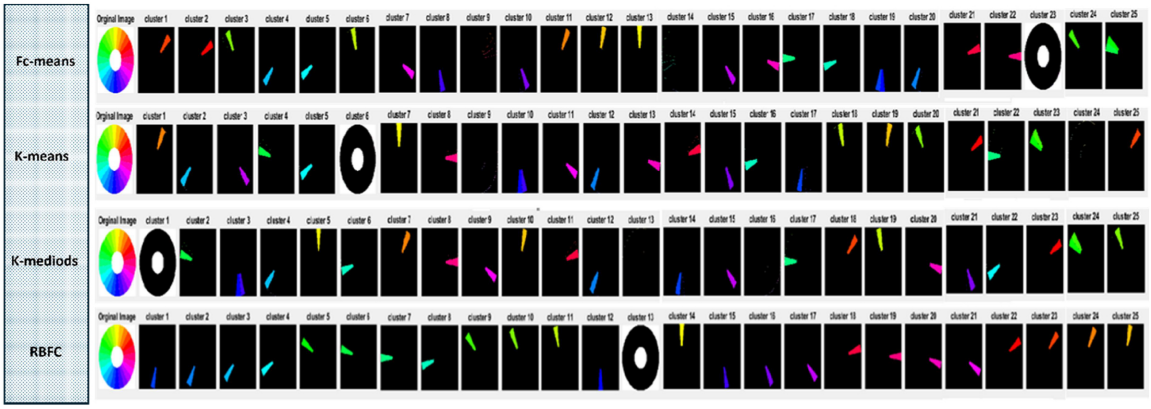


Fig. 10. Results of the four cluster methods for the 24 color images.

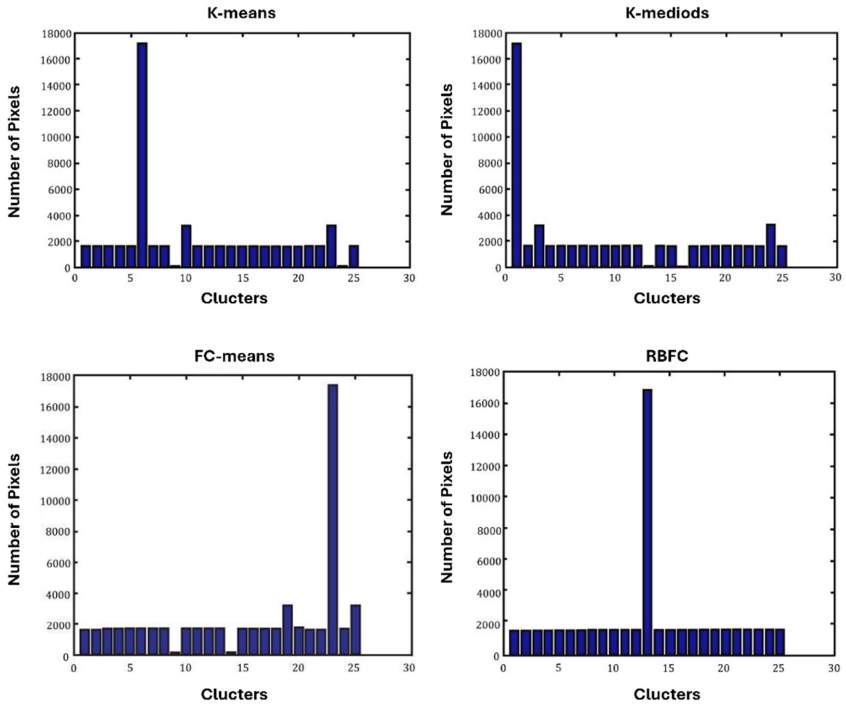


Fig. 11. Clusters-wise point distributions for the four clustering methods in the 24 color images.

A closer analysis of the FC-means method in the first execution, as shown in Fig. 10, reveals that the color segment does not appear in cluster 9 and instead merges with cluster 12. Similarly, in the K-means method, the color segment is missing from cluster 13 and instead appears combined with cluster 9 in the second execution, as depicted in Fig. 11. A comparable issue is observed with the K-medoids method.

Regarding execution time, the FC-means method required 3.822 and 3.807 seconds for the first and second executions, respectively. Although K-means performed more rapidly, taking 2.226 and 1.812 seconds for the respective executions, K-medoids took longer, requiring 4.711 and 3.901 seconds for both runs. In contrast, the proposed RBFC method demonstrated much shorter execution times, with the first and second runs taking 0.191 and 0.185 seconds, respectively.

To demonstrate the effectiveness of RBFC method compared to conventional clustering techniques, its performance was evaluated using a 24-color disk image, as shown in Fig. 10. Additionally, Fig. 11 illustrates the number of points within each cluster, with each column representing the number of points in a given cluster. The longest column corresponds to the white background of the image, identified as the twenty-fifth cluster. The execution times for the methods were recorded as follows: 7.980 seconds for

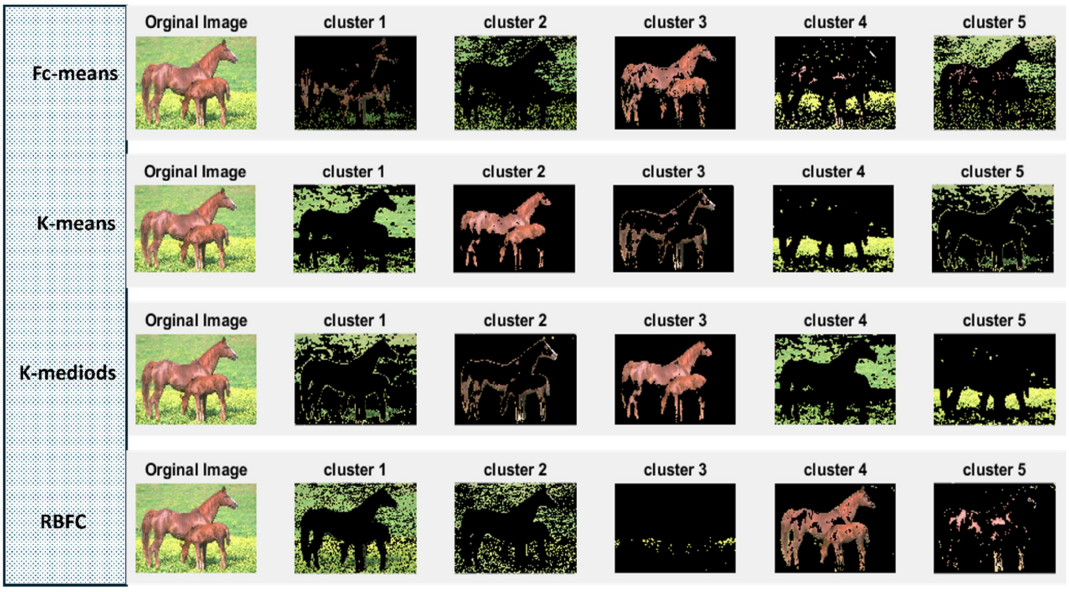


Fig. 12. Illustrates the findings of horses' image clustering. The first column of each row depicts the original image. The first row shows the FC-means method outcomes. The second row illustrates the K-means method outcomes. The third row shows the results of the K-medoids method. The fourth row depicts the proposed RBFC method findings.

FC-means, 6.365 seconds for K-means, 10.549 seconds for K-medoids, and 0.330 seconds for RBFC.

3.2. Results of real-world color images

To further evaluate the clustering performance, three real-world images—horses, flowers, and a landscape—were selected from the Berkeley Segmentation Dataset (BSDS500). It is important to note that the results are based solely on image color information, without considering the spatial relationships between pixels, and the segment number K was set according to the ground truth. Fig. 12 shows the clustering results for the horse image using the four clustering methods. The execution times for RBFC, K-means, K-medoids, and FC-means were 0.404, 3.535, 6.268, and 4.298 seconds, respectively.

The results indicated that the RBFC method delivered the best performance, accurately distinguishing the horse (object) from the background. It was also the fastest method. The RBFC method effectively reduced randomness and produced clear, well-separated clusters while maintaining a strong separation between the object and the background. In contrast, the traditional methods showed varying degrees of performance. The K-means method exhibited moderate accuracy in clustering the horse and the background, but it was affected by some randomness, resulting in less distinct clusters compared to the RBFC. The K-medoids method was the slowest and struggled to separate the horse from the background, leading to disorganized clusters due to random clustering behavior. Although the FC-means method produced more flexible clusters, it did not achieve the same level of clarity in separating the object from the background as the RBFC method. Some interference and noise were observed, negatively affecting its accuracy.

Figs. 13 and 14 show the clustering results for the flower and landscape images, respectively. For Fig. 13, the execution times for RBFC, K-means, K-medoids, and FC-means were 0.447, 3.293, 8.709, and 5.514 seconds, respectively. In Fig. 14, the execution times for the four methods were 0.449, 2.236, 6.846, and 4.736 seconds, respectively. Again, the proposed RBFC clustering method outperformed the other three methods in clustering color images.

Finally, the findings indicate that, based on clustering quality and implementation time across the three images (horse, flower, and landscape), the four clustering methods in this study can be ranked from best to worst as follows: RBFC, FC-means, K-means, and K-medoids.

3.3. Analysis of clustering under salt-and-pepper noise distortion

Figs. 15 and 16 present the results of applying two distinct clustering techniques—RBFC and FC-means—to an image artificially corrupted with salt-and-pepper noise, a commonly used model to simulate high-frequency pixel-level corruption.

In Fig. 15, the application of the proposed RBFC method leads to the extraction of 11 clusters from

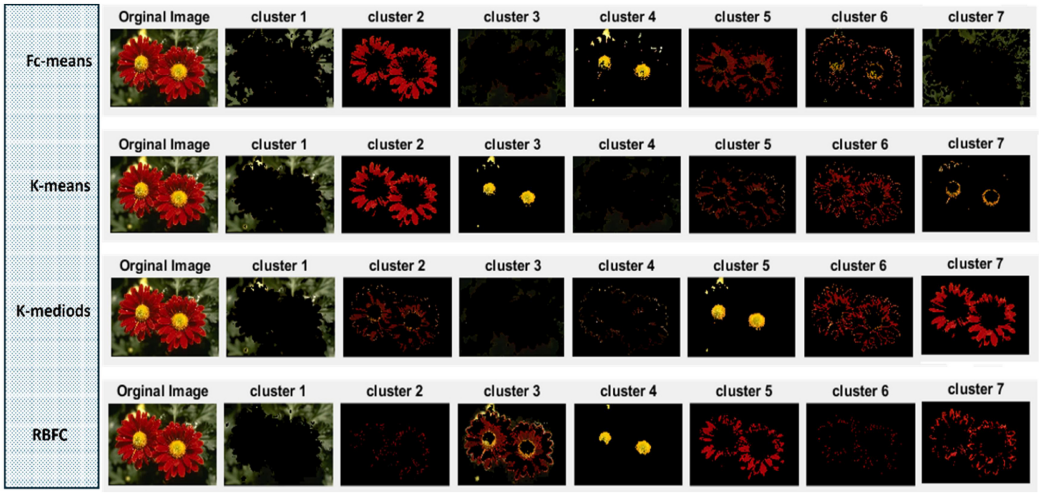


Fig. 13. Flowers image clustering results. The first column of each row illustrates the original image. The first row shows the FC-means method findings. The second row depicts the K-means findings. The third row presents the K-medoids findings. The fourth row depicts the proposed RBFC method outcomes.

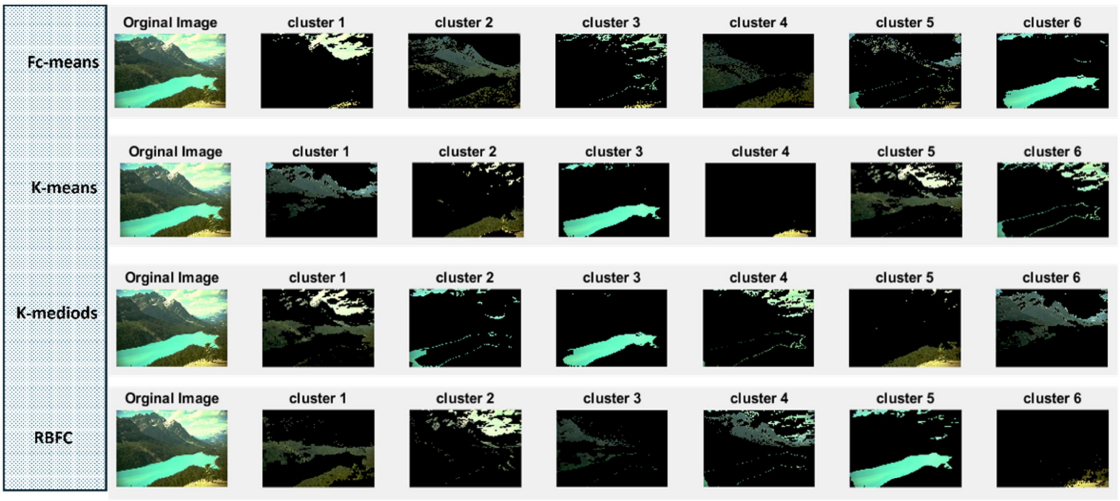


Fig. 14. Clustering outcomes of Landscape image. The first column of each row shows the original image. The first row depicts the FC-means method results. The second row displays the K-means method outcomes. The third row presents the findings of the K-medoids method. The fourth row shows the result of the proposed RBFC method.

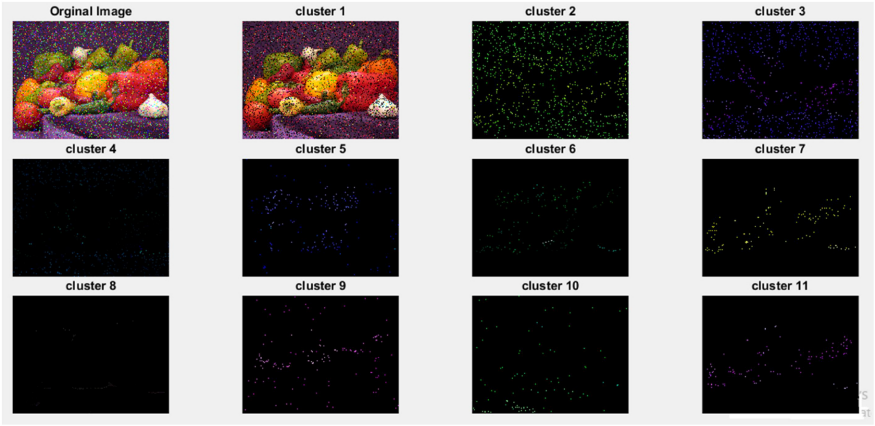


Fig. 15. Clusters resulting from applying RBFC method to a noisy image distorted by salt and pepper noise.

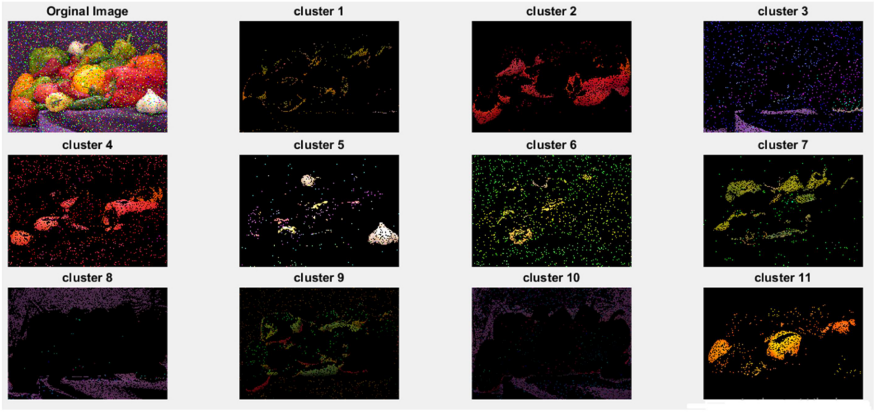


Fig. 16. Clusters resulting from applying FC-means clustering to a noisy image distorted by salt and pepper noise.

the noisy image. Despite the considerable distortion, the RBFC method maintains a strong capacity for distinguishing the true structural components of the image from the noise. Notably, the primary object in the original image is preserved within the first few clusters, and the visual coherence of dominant features is sustained. The sparse distribution of noise across separate clusters demonstrates the model's robustness in isolating disruptive pixel values from the actual content, rather than blending the two.

Conversely, Fig. 16 depicts the outcome of using the FC-means algorithm on the same noise-affected image. While the algorithm successfully segments the image into several clusters, the resulting separation exhibits notable limitations. Although the clusters appear to capture portions of the object, there is considerable overlap between object features and noise artifacts. In particular, noise pixels—especially those from the salt component—are dispersed within the same clusters as valid image regions, making it difficult to differentiate between true object boundaries and noise intrusions. As a result, the visual integrity of the segmented output is compromised, reducing its effectiveness in scenarios that demand precise noise separation, such as preprocessing stages in medical or remote sensing image analysis.

These comparative observations underscore the superior performance of the RBFC method in handling noisy data environments. Its ability to maintain clarity and suppress randomness makes it especially suitable for real-world applications involving degraded visual input.

3.4. Comparative analysis of clustering results on medical and satellite imagery

To assess the robustness and performance of the proposed method, it was applied to complex image types—specifically, medical and satellite images, both of which are characterized by dense structural details and significant overlap among components. These image categories pose a considerable challenge for clustering algorithms due to their intricate texture and color distribution. Fig. 17 presents the original medical and satellite images selected for this evaluation, serving as representative test cases for high-complexity scenarios.

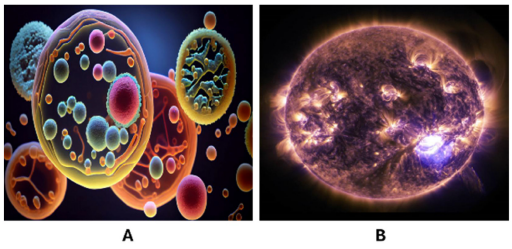


Fig. 17. The medical and satellite images used in this study. (A) The medical image. (B) The satellite image.

The visual performance of clustering algorithms, when applied to challenging real-world domains such as medical and satellite images, brings to light the practical limits and advantages of each method. These images are inherently complex, characterized by overlapping textures, subtle gradients, and intricate spectral variations—conditions that often expose the weaknesses of conventional clustering approaches. In the case of the medical image (Fig. 18), traditional methods like FC-means and K-means succeeded in processing the full-size image (1960 × 3430 pixels), yet at a significant computational cost. The FC-means algorithm required approximately 3000 seconds to complete the task, while K-means needed 1172 seconds. The K-medoids method, on the other hand, failed to process the original high-resolution image entirely, throwing an error that necessitated downscaling the image to 800 × 1000 pixels. Even

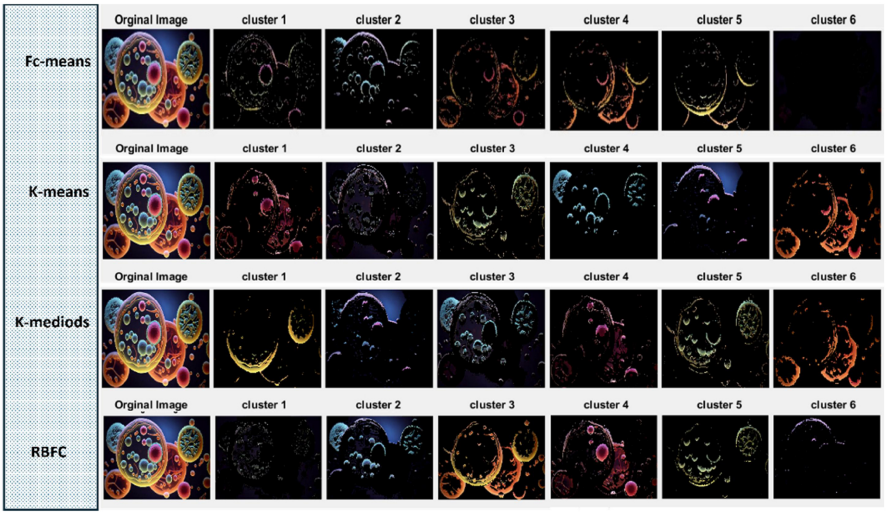


Fig. 18. Clustering outcomes of medical images. The first column of each row shows the original image. The first row depicts the FC-means method results. The second row displays the K-means method outcomes. The third row presents the findings of the K-medoids method. The fourth row shows the result of the proposed RBFC method.

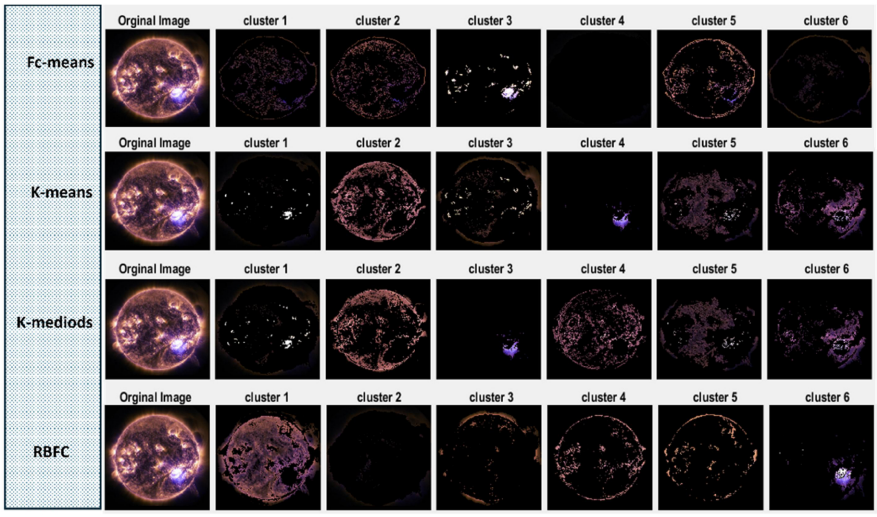


Fig. 19. Clustering outcomes of satellite image. The first column of each row shows the original image. The first row depicts the FC-means method results. The second row displays the K-means method outcomes. The third row presents the findings of the K-medoids method. The fourth row shows the result of the proposed RBFC method.

then, the execution time remained relatively high at 250 seconds. In contrast, the RBFC method managed to process the full-resolution medical image without failure, completing the task in just 83 seconds, demonstrating superior computational efficiency and robustness even under high data loads.

A similar pattern emerged in the satellite image evaluation (Fig. 19). With an original image resolution of 1280×1280 pixels, RBFC again outperformed all other methods, completing the clustering task in only 6 seconds. K-means followed with a significantly longer runtime of 77 seconds. The FC-means method required 90 seconds, while K-medoids once more failed to handle the original image size, forcing a reduction to 900×900 pixels and consuming 186 seconds. Beyond processing time, the quality of the output clusters further reinforced the advantage of the RBFC model. In both image types, RBFC maintained better structural coherence, cleaner segmentation boundaries, and reduced visual noise—critical qualities for medical diagnostics and satellite analysis alike.

Altogether, the empirical results not only confirm the theoretical strengths of the RBFC model in terms of clarity and adaptability but also highlight its practical feasibility for high-resolution, real-world

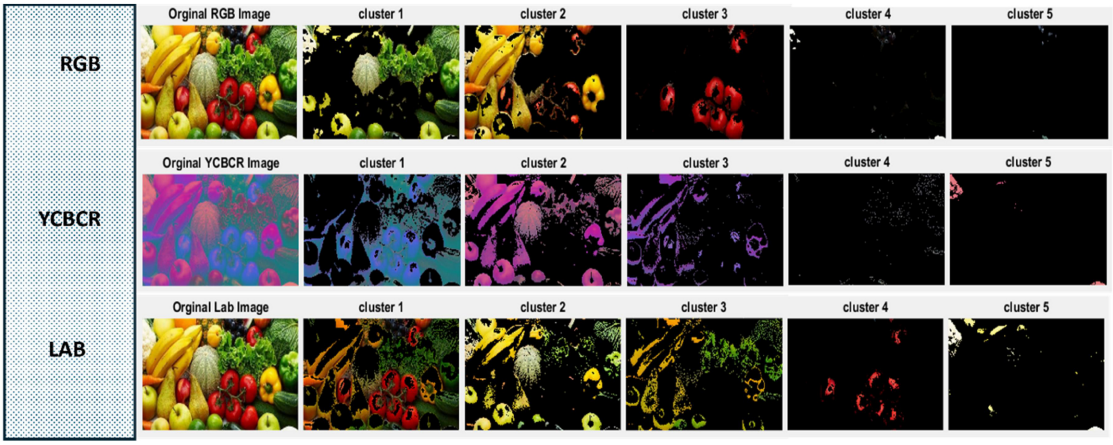


Fig. 20. The results of the RBFC method evaluation using RGB, YCBCR, and LAB images

applications. Its consistent handling of image complexity, combined with fast execution and resilience to scaling issues, positions RBFC as a powerful candidate for advanced image clustering across domains where both accuracy and speed are essential.

3.5. The RBFC method evaluation using YCBCR and LAB images

Fig. 20 presents the evaluation results of the RBFC method applied to RGB, YCBCR, and LAB image formats. The experimental outcomes demonstrate the RBFC model's robustness and adaptability across these image types. With its combination of high speed and resilience to scaling challenges, RBFC emerges as a strong candidate for advanced image clustering tasks in domains that demand both accuracy and efficiency.

3.6. Comparison between clustering methods of this study

When comparing the three conventional clustering methods with the proposed RBFC method, it is evident that FC-means offers flexibility and performs well in scenarios where color data points overlap across clusters. Despite its advantages, FC-means is more computationally demanding than K-means, making it less suitable for real-time applications but beneficial for more complex clustering tasks. Although the K-medoids method offers increased robustness in handling outliers, it comes with a significantly higher computational cost. The proposed RBFC method is recommended for applications that require a balance between speed and robustness, depending on whether the data relationships are linear or non-linear.

While the RBFC method provides notable gains in terms of speed and clarity, especially for structured image data, it is important to acknowledge that deep learning and graph-based clustering models continue to evolve and remain strong in areas that require a deeper semantic understanding. These models rely on learned features and adaptive network structures, which help them uncover complex, abstract relationships—something particularly valuable when working with noisy inputs or unstructured data. Although the RBFC method was not developed to compete directly with such advanced architectures, it offers an alternative that is simpler, less resource-intensive, and more interpretable. In future work, we aim to explore how elements of the RBFC model can be integrated with or benchmarked against neural or graph-based clustering techniques to assess their combined potential in broader applications.

3.7. The limitations of this study

A. Fixed Number of Clusters:

The suggested technique relies on predetermined template vectors to calculate the number of clusters, which limits flexibility. The number of clusters in real-world applications is frequently unknown and varies greatly. Without major user effort or template revision, this approach may not automatically react to diverse clustering demands.

B. Handling of Highly Overlapping Clusters:

Although the technique demonstrates ability in the separation of objects and background, its behavior in the case of highly overlapping clusters (as in noisier or more complex images) has not been thoroughly studied. Techniques like FC means can be expected to give improved performance in overlapping data,

which indicates that the proposed technique may require modification in those cases.

C. Computational Scalability:

Although the RBFC technique outperforms standard approaches such as K-means, K-medoids, and FC-means regarding speed and execution time, the approach's scalability in real-time processing has not been completely tested.

D. Sensitivity to Parameter Tuning:

The robustness of the proposed RBFC clustering method may be affected by parameter tuning, e.g., the robust distance measure parameters and the templates for modeling attribute relationships.

3.8. Future directions of this study

1. **Scalability for large datasets:** Extension of the RBFC clustering algorithm for efficient processing of big data applications and large-scale image datasets.
2. **Real-time processing:** Enhancing the technique for real-time clustering in dynamic environments like video monitoring and surveillance systems.
3. **High-resolution image clustering:** Assessing the RBFC method's performance on extremely high-resolution images in domains such as remote sensing images.
4. **Integration with DL:** Enhancing feature extraction and clustering accuracy by integrating DL frameworks with the RBFC clustering approach.
5. **Adaptive and hybrid clustering models:** The RBFC approach can be modified or combined with other clustering approaches for increased flexibility and robustness.

4. Conclusion

The Relationship Between Features Clustering (RBFC) method is presented in this work as a cutting-edge, state-of-the-art clustering technique that exhibits exceptional clustering performance and computing efficiency on various real-world image datasets. Through the efficient use of a specialized matrix formulation that clearly models and optimizes the relationship between features, RBFC greatly improved accuracy and timings in contrast to the conventional K-means, K-medoids, and FC-means clustering methods. The results show that the RBFC method successfully separates objects from their backgrounds within complex visual scenes by creating distinct and uniform clusters with limited internal variation.

Furthermore, RBFC clustering's resilience and adaptability—two essential characteristics for real-world image segmentation and pattern recognition tasks—are highlighted by its clarity and efficacy, especially in difficult images like landscapes. In the end, the suggested RBFC clustering approach surpasses K-means, K-medoids, and FC-means in terms of efficiency and result quality, and it is the best model in this study because of its remarkable speed, accuracy, and capacity to manage intricate picture features with little unpredictability.

Funding

This research received no external funding.

Competing interests

Regarding this topic, no possible conflicts of interest were disclosed.

Availability of data and materials

Not applicable.

Acknowledgement

The authors extend their sincere gratitude to the Dean of the College of Computer Science and Mathematics, the Head of the Department of Computer Engineering, and the Head of the Department of Cybersecurity at the University of Mosul for their invaluable support and facilitation in granting us access to the college's resources and providing scientific guidance.

References

1. P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898–916, 2010.
2. Y. Chen, P. Tan, M. Li, H. Yin, and R. Tang, "K-means clustering method based on nearest-neighbor density matrix for customer electricity behavior analysis," *International Journal of Electrical Power & Energy Systems*, vol. 161, p. 110165, 2024.
3. P. D'Urso, L. De Giovanni, and V. Vitale, "A robust method for clustering football players with mixed attributes." *Annals of Operations Research*, vol. 325, no. 1, pp. 9–36, 2023.
4. P. D'Urso and V. Vitale, "A Kemeny distance-based robust fuzzy clustering for preference data," *Journal of Classification*, vol. 39, no. 3, pp. 600–647, 2022.
5. E. Ditton, A. Swinbourne, and T. Myers, "Selecting a clustering algorithm: A semi-automated hyperparameter tuning framework for effective persona development," *Array*, vol. 14, 100186, 2022.

6. Y. Dong, D. Fan, Q. Ma, and S. Ji, "An improved clustering method for multi-view images," Image and Graphics: 10th International Conference, ICIG 2019, Beijing, China, August 23–25, 2019, Proceedings, Part III, vol. 10, pp. 134–144, 2019.
7. Z. Gou and Y. Zeng, "Analysis of spatio-temporal behavior of research travelers based on spark big data computing model," *Procedia Computer Science*, vol. 228, pp. 494–501, 2023.
8. S. Harsono, E. Utami, and A. Yaqin, "The association rule methods and k-means clustering for optimization mapping of new students admission." *2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, pp. 1–6, 2024.
9. Z. He and Y. Zhang, "Weighted least-squares color digital halftoning," *2015 International Conference on Intelligent Systems Research and Mechatronics Engineering*, pp. 84–87, 2015.
10. A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Information Sciences*, vol. 622, pp. 178–210, 2023.
11. F.H. Kuwil, Ü. Atila, R. Abu-Issa, and F. Murtagh, "A novel data clustering algorithm based on gravity center methodology," *Expert Systems with Applications*, vol. 156, p. 113435, 2020.
12. X. Li, B. Kao, S. Luo, and M. Ester, "ROSC: Robust spectral clustering on multi-scale data," *Proceedings of the 2018 World Wide Web Conference*, pp. 157–166, 2018.
13. C. C. Liu, N. F. Greenwald, A. Kong, E. F. McCaffrey, K. X. Leow, D. Mrdjen, B. J. Cannon, J. L. Rumberger, S. R. Varra, and M. Angelo, "Robust phenotyping of highly multiplexed tissue imaging data using pixel-level clustering," *Nature Communications*, vol. 14, no. 1, p. 4618, 2023.
14. M. Niemelä and T. Kärkkäinen, "Improving clustering and cluster validation with missing data using distance estimation methods," In Computational Sciences and Artificial Intelligence in Industry (pp. 123–133). Springer, 2022.
15. C. Poullis, "Large-scale urban reconstruction with tensor clustering and global boundary refinement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 5, pp. 1132–1145, 2019.
16. Y.-F. Qi, Y.-H. Shao, C.-N. Li, and Y.-R. Guo, "Locally finite distance clustering with discriminative information," *Information Sciences*, vol. 623, pp. 607–632, 2023.
17. V. Singh and N. K. Verma, "Gene expression data analysis using feature weighted robust fuzzy-means clustering," *IEEE Transactions on NanoBioscience*, vol. 22, no. 1, pp. 99–105, 2022.
18. S. Yadav and S. Sharma, "Study Of existing methods & techniques of K-means clustering," *Educational Administration: Theory And Practice*, vol. 30, no. 4, pp. 1806–1813, 2024.
19. W. Yang, Z. Zhang, Y. Zhao, Y. Gu, L. Huang, and J. Zhao, "CABGSI: An efficient clustering algorithm based on structural information of graphs," *Journal of Radiation Research and Applied Sciences*, vol. 17, no. 3, p. 101040, 2024.
20. J. Zhang and Y. Chen, "Research on color and texture characteristics and visual perception of custom wardrobe finishes," *BioResources*, vol. 19, no. 3, 2024.
21. L. Zhou, Y. Zhang, D. Peng, and D. Wu, "Improved K-means clustering color segmentation for road perception," 2015.
22. M. A. Ahmed, S. M. Nafl, H. Baharin, and P. N. E. Nohuddin, "Prioritise five Tafseer translators using clustering technique for Surah Al-Baqarah," *Al-Iraqia Journal for Scientific Engineering Research*, vol. 3, no. 1, pp. 75–86, 2024.
23. A. A. Karah Bash and A. A. Khail, "Advanced multi-layer deep learning model for accurate estimation of heat transfer and flow designing parameters across diverse dataset configurations," *Engineering Applications of Artificial Intelligence*, vol. 151, p. 110723, 2025.
24. M. A. Mohammed and A. M. Ali, "Enhanced cancer subclassification using multi-omics clustering and quantum cat swarm optimization," *Iraqi Journal for Computer Science and Mathematics*, vol. 5, no. 3, p. 37, 2024.
25. A. Y. Yousif and B. A. Sarray, "Convex optimization techniques for high-dimensional data clustering analysis: A review," *Iraqi Journal for Computer Science and Mathematics*, vol. 5, no. 3, p. 29, 2024.
26. A. S. Hussain, M. S. Sulaiman, S. M. Hussein, E. A. Az-Zo'bi, and M. Tashtoush, "Advanced parameter estimation for the gompertz-makeham process: A comparative study of MMLE, PSO, CS, and bayesian methods," *Statistics, Optimization & Information Computing*, 2025.
27. H. R. Hussein, S. R. Hussein, A. S. Hussain, and M. A. Tashtoush, "Estimating parameters of software reliability growth models using artificial neural networks optimized by the artificial bee colony algorithm based on a novel NHPP," *Mathematical Modelling of Engineering Problems*, vol. 12, no. 1, 2025.
28. A. S. Hussain, K. B. Mahmood, I. M. Ibrahim, A. F. Jameel, S. Nawaz, and M. A. Tashtoush, "Parameters estimation of the Gompertz-Makeham process in non-homogeneous poisson processes: Using modified maximum likelihood estimation and artificial intelligence methods," 2025.