RESEARCH PAPER
Contrib. Sci. & Tech Eng, 2025, 2(3)
Homepage: https://cste.journals.umz.ac.ir/
DOI:10.22080/cste.2025.29024.1032

Optimizing Traditional Clustering Methods Using Meta-Heuristic Algorithms for Joint Set Identification in Copper Mines

Ali Rasouli [1], Akbar Esmaeilzadeh [2]*, Reza Mikaeil [2], Atalou, Solat [1]

[1] *Department of Mining Engineering, Ah.C., Islamic Azad University, Ahar, Iran*

[2] *Department of Mining Engineering, Environment Faculty, Urmia University of Technology, Urmia, Iran.*

Article Info
Received 10 April 2025
Accepted 10 June 2025
Available online 11 September 2025

Keywords:
Joint Set;
Clustering Method;
K-Means and C-Means Methods;
Harmony Search Algorithm;
PSO Algorithm;
Sungun Copper Mine.

Abstract:
The assessment of clustering methods is crucial for identifying the primary characteristics of joints in mining and rock engineering. Orientation is commonly used to characterize the deformation patterns and mechanical properties of rock formations. This study introduces an enhanced clustering method by integrating the Harmony Search (HS) and Particle Swarm Optimization (PSO) algorithms to classify joint sets based on orientation parameters—namely, dip and dip direction—in the Sungun copper mine. First, joint characteristics were clustered using K-means and Fuzzy C-means (FCM) techniques. The elbow method was applied to determine the optimal number of clusters, resulting in a four-cluster classification. Subsequently, both K-means and FCM were optimized using HS and PSO algorithms, and the joint data were evaluated based on three clustering quality criteria: the Davies-Bouldin index (DBI), the Calinski-Harabasz index (CHI), and the Silhouette coefficient (SI). The results showed that the FCM-PSO method achieved the highest ranking, yielding a DBI of 0.80, a CHI of 348.47, and an SI of 0.565. In contrast, integrating the HS algorithm with K-means and FCM did not improve clustering performance as expected. Furthermore, the K-means-PSO method performed worse than the FCM clustering approach, ranking third overall. Based on these findings, the FCM-PSO method, by effectively optimizing cluster centers, provides a reliable approach for classifying joint sets. The proposed method can be effectively applied in rock mass behavior analysis for large-scale open-pit mines such as the Sungun copper mine. The FCM-PSO method achieved the best results with DBI=0.80, CHI=348.47, and Silhouette=0.565.

***Corresponding Author:** a.esmailzadeh@uut.ac.ir

Supplementary information: Supplementary information for this article is available at https://cste.journals.umz.ac.ir/

Please cite this paper as: Rasouli, A. , Esmaeilzadeh, A. , Mikaeil, R. , & Atalou, S. (2025). Optimizing Traditional Clustering Methods Using Metaheuristic Algorithms for Joint Set Identification in Copper Mines. Contributions of Science and Technology for Engineering, 2(3), 57-72. doi:10.22080/cste.2025.29024.1032.

1. Introduction

Rock joints play a crucial role in determining the stability of rock masses. The primary characteristics of joints, which include orientation and spacing, are key factors in assessing rock deformation and strength [1]. Due to long-term stochastic geological processes, the spatial distribution of joints tends to exhibit a complex and random pattern [2]. Clustering analysis allows for the classification of these randomly distributed joints into dominant groups within a defined geological domain. By analyzing joint characteristics, engineers can evaluate rock mass stability. Therefore, clustering analysis of rock joints is critical in rock engineering stability assessments. Historically, two principal approaches have been adopted to identify joint sets. The first involves clustering joints based solely on orientation, while the second considers additional joint features alongside orientation. The dip and dip direction significantly affect the stability of dam foundations, tunnel linings, and slope benches in mining operations. Thus, the application of clustering techniques that incorporate multiple joint characteristics is essential and particularly suited for engineering applications. Traditional grouping of joints often relies on graphical methods such as stereographic pole plots, rose diagrams, and equal density plots [3, 4]. While intuitive, these methods are highly sensitive to subjective bias and the experience of the analyst, limiting their ability to reliably classify widely dispersed joints. To overcome these limitations, automated clustering techniques have been developed to objectively identify joint sets. Clustering analysis is a robust statistical method for resolving classification problems. Its objective is to identify intra-group similarities and inter-group differences. This method is widely utilized across various disciplines, including data mining, pattern recognition, information retrieval, microbiological analysis, and machine learning [5]. Shanley and Mahtab [6] were the first to apply clustering techniques for rock joint classification.

*ISSN **3060-6578***

Since then, numerous clustering approaches have been developed, including K-means [7, 8], Fuzzy C-means (FCM) [9, 10], Spectral Clustering [11, 12], and Affinity Propagation (AP) [13]. Both K-means and FCM are sensitive to the initial placement of cluster centres and prone to converging at local optima. To address this challenge, researchers have proposed various solutions to improve the initialization of cluster centres [3, 14-16]. These studies provided valuable insights into clustering joints based on orientation and related properties. In recent years, several researchers have employed soft computing and metaheuristic algorithms—such as the Artificial Bee Colony (ABC), Mutative Scale Chaos Optimization (MSCO), Quantum Particle Swarm Optimization (QPSO), Differential Evolution (DE), and Bacterial Foraging Optimization Algorithm (BFOA)—to address the challenges in determining the mechanical and hydraulic characteristics of rock masses [4, 7, 17-22]. Others have utilized Support Vector Machines (SVM), Genetic Algorithms (GA), and Grey Wolf Optimizer (GWO) for clustering or classification tasks in complex geological settings [1, 21-23]. Additionally, Ruan et al. [24] applied the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm for joint set identification, achieving better accuracy than FCM by effectively filtering out outliers. Wang et al. [25] employed FCM to improve prediction accuracy in machine learning models such as Support Vector Regression (SVR) and Random Forests. Yong et al. [26] integrated the Neutrosophic Genetic Algorithm (NGA) with K-means clustering to enhance efficiency and practical applicability. Zarean and Poormirzaee [27] demonstrated the effectiveness of PSO in solving complex geotechnical problems by accurately identifying joint orientations.

Although these efforts have advanced the field, key limitations remain. Specifically, many existing models lack robustness in heterogeneous field conditions and struggle with overlapping joint features. Furthermore, few studies have provided multi-index benchmarking under practical mining scenarios.

This study addresses these gaps by integrating PSO and HS algorithms with K-means and FCM to improve clustering reliability and performance. The methodology is validated using three well-established metrics: the DBI, CHI, and SI. The proposed approach is applied to real joint orientation data from the Sungun copper mine, offering a practical and optimized solution for joint set classification in rock engineering. The structure of this paper is as follows: Section 2 outlines the data, methodology, and steps of the proposed metaheuristic-based clustering framework. Section 3 presents a comparative evaluation using the three clustering indices. Section 4 concludes with a summary of findings.

2. Materials and Methods

2.1. Geological Study Area

The Sungun copper mine is located in East Azerbaijan Province, 130 kilometres northeast of Tabriz and 30 kilometres north of Varzeghan, at 46° 43' E and 38° 42' N (Figure 1).

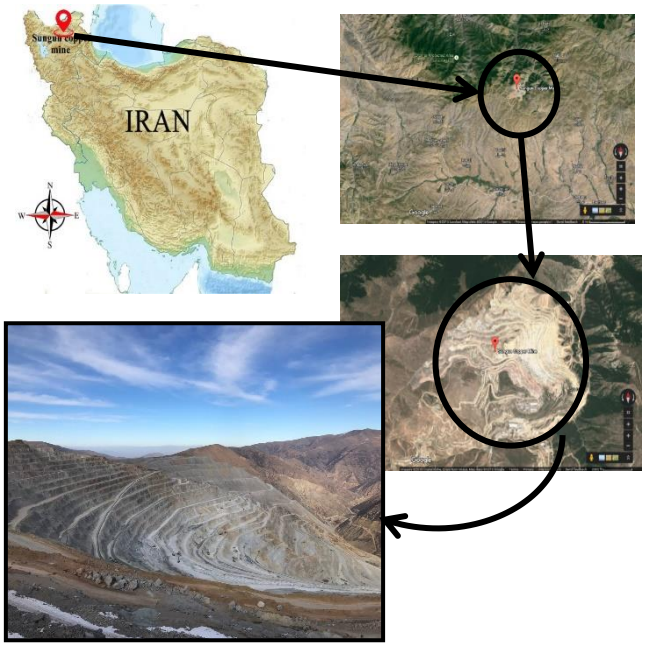

Figure 1. Location of Sungun copper mine

This mine is located in a mountainous region with an average elevation of 2000 meters above sea level. It is situated in the northwest of Iran within the global copper belt. The Sungun copper deposit is positioned in the Urumieh-Dokhtar Magmatic Arc (UDMA), which forms part of the Himalayan-Alpine metallogenic belt. This monzonitic-type deposit lies within a tectonic belt related to a subduction zone along the continental margin during the Tertiary period. The orebody, delineated within the Tafadili exploration network, spans approximately one square kilometre. High-grade veins are concentrated in the semi-deep portion of the peripheral monzonite. This classifies it as a peripheral-type copper deposit. The surrounding rock formations include Upper Cretaceous limestone units and andesitic-latitic volcanic sequences (Figure 2).

The oldest rocks in the area include a 500-meter-thick sequence of limestone interbedded with shale, and a 1500-meter-thick sequence of lava flows. Calc-alkaline andesitic dikes intrude the calc-alkaline and tuffaceous rocks [28].

Following the emplacement of the quartz monzonite mass, magmatic activity resulted in the injection of multiple dyke systems. Recent lithological studies indicate three major intrusive pulses in the region. The first pulse is associated with the Sungun-bearing quartz monzonite, which hosts porphyry Cu-Mo mineralization. This was later intruded by two younger quartz-bearing units: urite-granodiorite and barren diorite. Post-mineralization dykes, mostly trending NW-SE and dipping SW (occasionally N-S), have crosscut the main mass.

Figure 2. The Urumieh-Dokhtar volcanic belt hosts large porphyry copper deposits, including the Sungun copper deposit, as illustrated in the structural map of Iran

These are divided into four dyke generations: (a) First-generation (DK1): Derived from late-stage quartz diorite to granodiorite. subdivided into DK1a, DK1b, and DK1c. (b) (DK2): Gabbro-diorite compositions, observed outside the main pit. (c) (DK3): Mainly composed of diorite. (d) (DK4): Latitic to trachyandesitic rocks from the Pliocene-aged Chaldaghi dome. These dykes exhibit considerable age variations and were excluded from the present study [29].

The Chaldaghi subvolcanic dome intrudes the SE margin of the Sungun porphyry mass. The latest magmatic activity stems from the Calderi Dash Dibi volcano to the SW. Pyroclastic deposits and lava flows from this eruption overlie the porphyry and later-stage dykes [30].

A variety of sulfide and oxide minerals formed during mineralization. Copper sulfides include chalcopyrite, bornite, chalcocite, and covellite. Other sulfides present are pyrite, molybdenite, galena, sphalerite, marcasite, and pyrrhotite. Precious metals such as gold and silver exist alongside oxides like ilmenite, rutile, magnetite, and goethite. The Sungun deposit shows three mineralization zones: oxide, supergene, and hypogene (Figure 3).

The hypogene zone varies from 0 to 200 meters in thickness (average: 80 m). The supergene zone ranges between 0 and 212 meters (average: 100 m). The entire hypogene section is 320 to 500 meters thick, averaging 340 meters. Sulfur mineralization occurs in supergene and hypogene zones. Due to rough topography and the proximity of the supergene upper surface to groundwater, this zone shows an uneven top surface. Main alteration zones: potassic, propylitic, phyllic, and argillic. Surface and shallow zones show strong alteration, with white, cream, red, and brown rock colors indicating weathering.

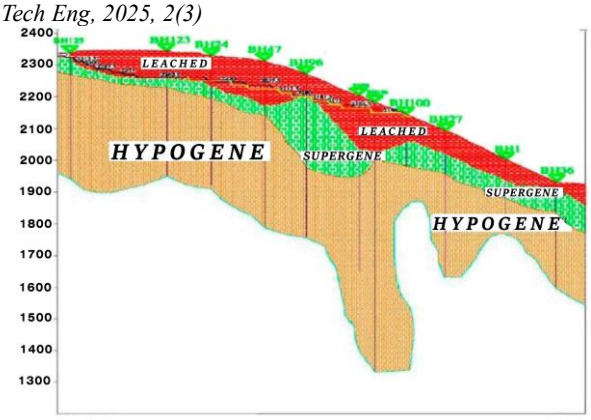

Figure 3.Distribution of subsidence, supergene, and hypogene zones in a vertical section of the Sungun copper deposit [31].

The color of the Sungun porphyry varies with alteration: Greenish-gray to olive in propylitic, light gray to cream and white in argillic, light gray to dark in phyllic, Dark gray with pink feldspar in potassic. Color index:10 to 35 leucocratic natures. Primary minerals: plagioclase (40–45%), potassium feldspar (30–35%), amphibole (5–10%), biotite (5–10%), and quartz (5–10%) [28].

The type of host rocks plays a key role in applying fuzzy clustering for joint identification in Sungun. Dominant rocks: granodiorite, diorite, and quartz diorite. These are suitable for geotechnical studies due to favorable physical and mechanical properties. Their igneous nature ensures high stability, hardness, and data reliability. Due to extensive fracturing and mineralization, they are ideal for optimized fuzzy clustering.

Granodiorite: High hardness; ideal for joint classification. Diorite and quartz diorite: Aid in fuzzy analysis due to dense mineralogy and structural resistance.

2.2. Methodology

Although up to 10 characteristics of joints can be recorded in the field, only two key attributes, dip and dip direction, are commonly used to identify joint sets. In this study, data were collected from 19 benches in the Sungun copper mine using a compass clinometer, which is a well-established method for structural geological surveys due to its high accuracy and direct field applicability. Unlike remote sensing techniques like LiDAR or photogrammetry—which require post-processing and complex calibration—the clinometer method provides real-time, high-resolution measurements with minimal instrumental bias. This approach resulted in the collection of 376 joint sets (Figure 4), with each joint characterized by its dip and dip direction. (Figure 5) presents the box plots of the collected data. These boxplots indicate that some outlier values are present in the initial data set, particularly within the dip data. To ensure accurate modelling and prevent deviations, it is necessary to eliminate these outliers from the database. To systematically identify these outliers, the Z-score method was employed, which standardizes data points and flags those exceeding a predefined threshold (typically $|Z| > 3$) as outliers.

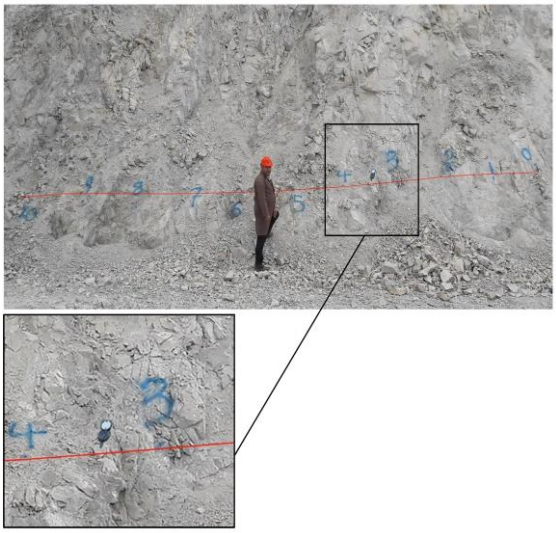

Figure 4. Joint surface mapping using a compass clinometer in the Sungun copper mine

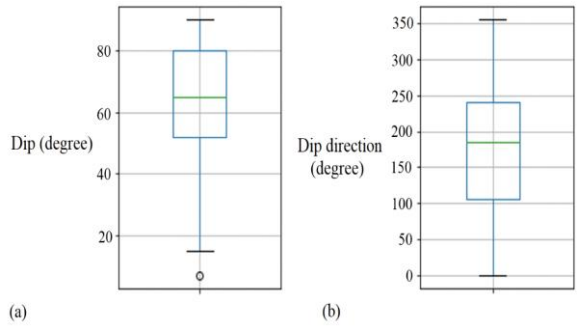

Figure 5. Box plot of collected joint surface data: (a) Dip, (b) Dip Direction

The Z-score statistic was applied to eliminate outlier data. The Z-score, also known as the standard score, is a statistical measure that quantifies how far a data point deviates from the mean of a dataset in terms of standard deviations. This method helps determine whether a given data point falls within the normal range or qualifies as an outlier compared to the rest of the dataset, as defined by the following equation [32]:

$$z = \frac{x-\mu}{\sigma} \quad (1)$$

where z Standardized value (Z-score), x Raw data value, μ Mean of the dataset, and σ Standard deviation of the dataset. Data points exceeding three standard deviations from the mean in any given column were identified as outliers and subsequently removed. After filtering outliers, a total of 375 joint set data points remained for analysis. The measured parameters are presented in Table 1.

When analysing the orientation of a joint mathematically, it is commonly represented as a pole using the equal-angle projection in the upper hemisphere Figure 6. The spatial orientation is defined by the dip direction (α), ranging from 0° to 360°, and the dip angle (β), ranging from 0° to 90°. In a Cartesian coordinate system, a unit normal vector e_i represents the orientation of a joint. This orientation is typically characterized by its direction cosines, denoted as $e_i = (x_i, y_i, z_i)$ [33]. Figure 7 presents a stereographic plot of the joint sets in the Sungun copper mine.

$$x_i = \cos\alpha \sin\beta$$
$$y_i = \sin\alpha \sin\beta \quad (2)$$
$$z_i = \cos\beta$$

Table 1. Statistical characteristics of input parameters

Joint features	Unit	Datasets	Values			
			Maximum	Minimum	Average	Standard deviation
Dip	Degree	375	90	20	65.13	17.82
Dip direction	Degree	375	355	0	177.56	91.76

2.3. Optimal Number of Clusters

A crucial aspect of the data clustering process is determining the optimal number of clusters, which serves as a key input for the clustering algorithm. This study employed the Elbow method to identify the appropriate number of clusters. The Elbow method is a widely used and effective approach for determining the optimal number of clusters by analysing the percentage of variance explained as a function of the number of clusters. This technique suggests selecting the number of clusters until the addition of another cluster no longer results in meaningful improvements in data modelling. By plotting the percentage of variance explained by clusters against the number of clusters, it becomes evident that the initial clusters contribute significantly to explaining variance. However, beyond a certain point, the marginal benefit of adding additional clusters diminishes. The elbow criterion is not always easily identifiable [34]. The percentage of variance is calculated as the ratio of between-group variance to total variance, commonly referred to as the F-test, which is computed as follows [35]:

$$F = \frac{\sum_{i=1}^{k}(n_i(\bar{Y}_i-\bar{Y}))/(K-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left((\bar{Y}_{ij}-\bar{Y}_i)\right)/(N-K)} \quad (3)$$

where i represents the index of the group, j is the index of a data point within group i, n_i is the number of data points in group i, $\bar{Y_i}$ represents the mean of group i, $\bar{Y}$ is the global mean of the dataset, $\bar{Y_{ij}}$ is the value of the j^{th} data point in group i, K is the total number of clusters, and N is the total number of data points. If the ratio of between-group

variability to within-group variability is substantial, this index will yield a high value.

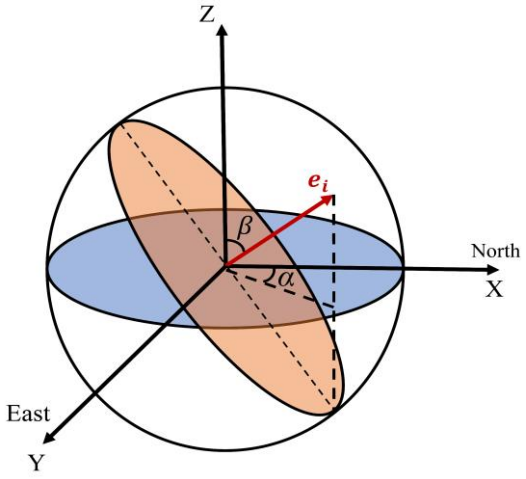

Figure 6. Representation of vector data in spherical coordinates

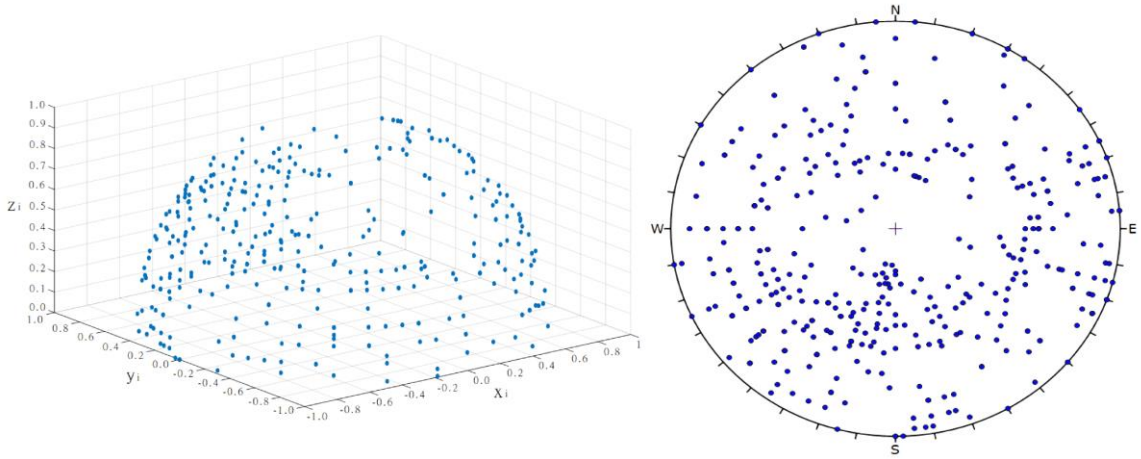

Figure 7. Scatter of joint sets data a: Spherical space b: Two-dimensional space equivalent to the upper hemisphere projection
Scatter of joint sets data: (a) Spherical coordinates, (b) Two-dimensional equivalent of upper hemisphere projection

2.4. K-means clustering method

The K-means clustering algorithm, first introduced by MacQueen in 1967 [36], is one of the most widely used methods for data clustering. In this method, if X = {X1, X2, ..., Xn} represents a dataset and V = {V1, V2, ..., Vc}is a set of cluster centres, then each data point is assigned to the nearest cluster centre at any given time. Here, n represents the number of data points, and c is the number of clusters. The objective of the K-means clustering method is to minimize the objective function j(v), which is defined as follows:

$$j(v) = \sum_{i=1}^{c}\sum_{j=1}^{c}\left\|x_{ij} - v_j\right\|^2 \quad (4)$$

where x_{ij} denotes the value of the i^{th} data point with respect to the j^{th} cluster, n is the total number of data points, c is the total number of clusters, v_j is the centroid of the j^{th} cluster, and $\|x_{ij} - v_j\|^2$ is the squared Euclidean distance between data point x_{ij} and centroid v_j.The centroid of each cluster is determined using the following equation:

$$v_i = \frac{1}{c_i}\sum_{j=1}^{c_i} x_{ij}\ ,\ i = 1,2,\dots,c \quad (5)$$

where v_i is the centroid of the i^{th} cluster, x_{ij} represents the j^{th}data point in cluster i, c_i is the number of data points in cluster i, and c is the total number of clusters. The K-means clustering algorithm partitions the dataset into K clusters by initially assigning data points to clusters in a randomized manner. In each iteration, the algorithm computes the Euclidean distance between each data point and the cluster centroid, subsequently reassigning the point to the nearest cluster. This iterative process continues until no further changes occur in cluster assignments. Figure 8 presents the flowchart of the K-means clustering method.

2.5. Fuzzy C-means (FCM)

Fuzzy C-means (FCM) is a data clustering algorithm that partitions a dataset into c clusters, where each data point is assigned a specific membership grade for each cluster. This method was initially introduced by Dunn [37] and later refined by Bezdek et al. [38]. The primary objective of FCM is to classify a collection of n data points into c distinct

clusters. The dataset is represented by a matrix A, defined as $A = \{a_1, a_2, \dots, a_n\}$,where each data point consists of m features, expressed as $a_i = \{a_i^1, a_i^2, \dots, a_i^m\}$. Equation (6) defines the objective function that guides the clustering process. The goal of this function is to minimize the Euclidean distance between each data point and its corresponding cluster centroid while simultaneously maximizing the Euclidean distance between different cluster centroids to enhance cluster separability.

$$j(u, v) = \sum_{k=1}^{n} \sum_{i=1}^{c} (v_{ik})^2 (d_{ik}^2) \quad (6)$$

where: v_{ik} is the degree of membership of the k^{th} data point (pattern vector a_k) in the i^{th} cluster, n is the total number of data points, c is the total number of clusters, $j(u, c)$ is the total clustering cost (to be minimized) and d_{ik}^2 is the squared Euclidean distance between data point a_k and cluster centroid c_i, which is mathematically defined as:

$$d_{ik}^2 = (a_k - c_i)^T (a_k - c_i) \quad (7)$$

a_k is the k^{th}data vector (pattern) in the dataset, c_i is the centroid of the i^{th}cluster, d_{ik}^2 represents the squared Euclidean distance in a multidimensional feature space and In Equation 7, the superscript (T) denotes the transpose operator, which converts a column vector into a row vector. This operation is used to compute the squared Euclidean distance through the inner product of the difference vector and its transpose. Equations 8 and 9 are used to iteratively update the cluster centers and the membership matrix in the FCM algorithm. Specifically, Equation 8 calculates the centroid $c_i$of the i^{th} cluster by taking a weighted average of all data points, where the weights are the squared membership degrees. Equation 9 updates the membership values for each data point and cluster in the next iteration step$(r + 1)$, based on the relative distances between the data point and all cluster centers:

$$C_i = \frac{\sum_{k=1}^{n} (v_{ik})^2 . ak}{\sum_{k=1}^{n} (v_{ik})^2 . ak} \quad (8)$$

$$v_{ik}^{(r+1)} = \left[\sum_{j=1}^{c} \left(\frac{d_{ik}^r}{d_{jk}^r} \right)^2 \right]^{-1} \quad (9)$$

Figure 9 presents the flowchart of the FCM method. The iterative process continues until the improvement in comparison to the previous iteration falls below a predefined threshold ε , as shown in Equation 10. Here, the variable r represents the iteration step, and the superscript$(r + 1)$ refers to the values computed at the next iteration.

$$\|J^{r+1} - J^r\| \leq \varepsilon \quad (10)$$

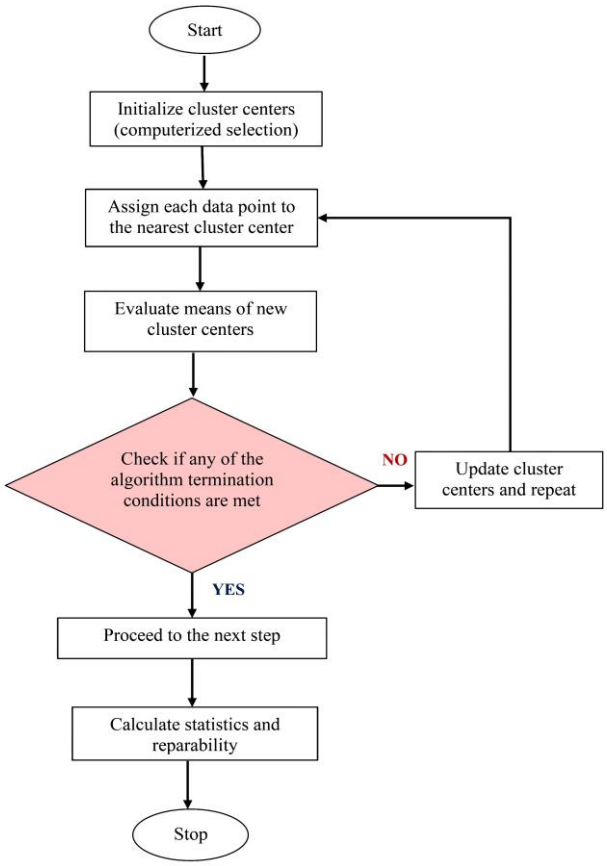

Figure 8. Flow chart of K-means clustering method

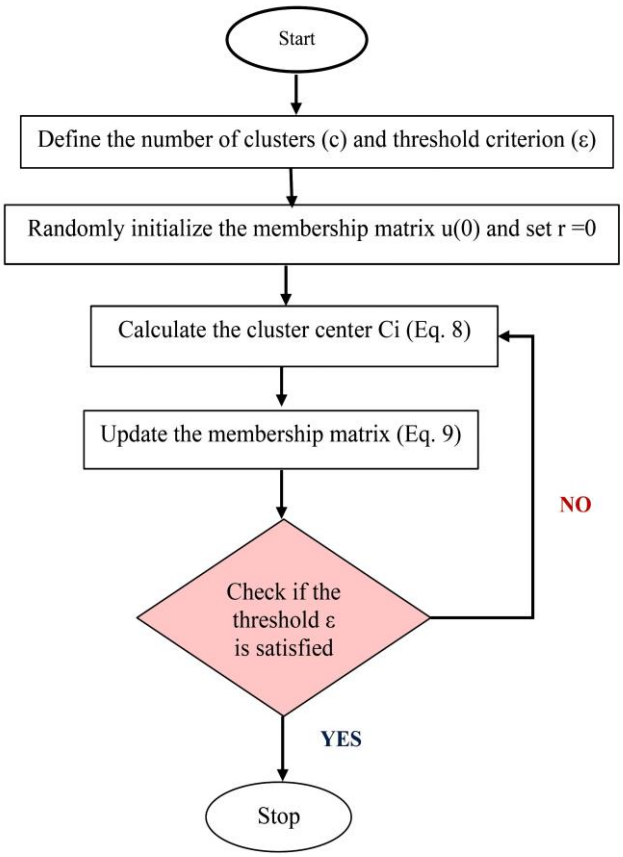

Figure 9. Flow chart of FCM clustering method

2-6. Harmony Search (HS) Algorithm

The HS algorithm, introduced by Geem et al. [39], is a metaheuristic optimization technique designed to handle complex and uncertain systems. The HS algorithm demonstrates strong performance in processing both linear and nonlinear data structures. The fundamental concept of HS optimization is inspired by the process of musical composition, where musicians create harmonious melodies by selecting and adjusting musical notes. Similar to this process, the HS algorithm iteratively refines solutions by selecting the most optimal parameters, considering a limited memory capacity [40].

Outline of the HS Algorithm Steps: Step 1: Initialization of Harmony Memory (HM) A set of initial solutions is generated randomly based on the problem constraints. For an n-dimensional optimization problem, the HM is initialized with a predefined size, referred to as Harmony Memory Size (HMS). This process follows Equation 11 [41]:

$$HM = \begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_n^1 \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{HMS} & x_2^{HMS} & \cdots & x_n^{HMS} \end{bmatrix} \quad (11)$$

Each potential solution is represented as $[x_1^i, x_2^i, \ldots, x_n^i]$, where $i = 1,2, \ldots, HMS$.

Step 2: Generating a New Harmony from the HM

Step 3: Updating the HM – The newly generated solution is evaluated. If it proves to be better than the worst solution currently stored in the HM, it replaces that solution, ensuring continuous improvement of the harmony memory.

Step 4: Iteration Process – Steps 2 and 3 are repeated iteratively until the termination criteria—based on accuracy and the number of iterations—are met. Figure 10 illustrates the flowchart of the HS algorithm.

2.7. Particle Swarm Optimization (PSO) Algorithm

The PSO algorithm, a well-known bio-inspired optimization technique, was first introduced by Kennedy and Eberhart in 1995 [42]. PSO is an efficient and straightforward optimization algorithm that iteratively searches for an optimal solution within a given solution space [42]. The core idea behind PSO is inspired by the social behaviour of bird flocks and fish schools, where individuals (particles) adapt their movements based on both their own experiences and the collective intelligence of the swarm. The steps of the PSO algorithm are as follows [43]:

Step 1: Initialization – A population of particles is randomly distributed in the search space; each is assigned an initial position and velocity. The objective function evaluates the initial fitness values of the particles.

Step 2: Particle Movement – The velocity of each particle is updated based on three factors: (1) its current velocity, (2) its distance to its best-known position, and (3) its distance to the best-known global position found by the swarm. The particle's new position is then determined based on the updated velocity.

Step 3: Evaluation and Update – Each particle's new position is evaluated using the objective function. If a better solution is found, the particle updates its personal best position and fitness value. If the particle's solution is better than any previously discovered solution, the global best position is updated.

Step 4: Termination – Steps 2 and 3 are repeated until a termination criterion is met, such as reaching the maximum number of iterations or achieving a predefined fitness threshold.

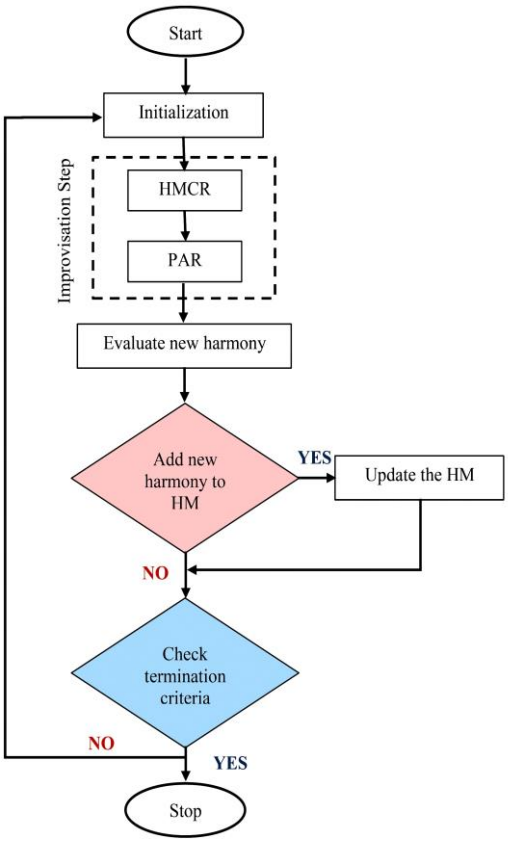

Figure 10. Flowchart of the HS

The PSO algorithm is widely recognized for its simplicity, ease of implementation, and minimal parameter tuning requirements. Several studies have demonstrated that multi-objective PSO offers significant advantages in terms of convergence speed, accuracy, and solution diversity [44, 45]. Figure 11 presents the flowchart of the PSO algorithm.

2.8. Criteria Evaluation

Assessing the performance of clustering methods is essential. The evaluation process requires the application of criteria that assess the effectiveness, accuracy, and precision of each proposed method. This paper evaluates each method based on three widely used criteria: Calinski-Harabasz (CHI), Davies-Bouldin (DBI), and Silhouette indices (SI). The CHI is defined as:

$$CHI = \frac{SS_B}{SS_W} \times \frac{(N-K)}{(K-1)} \quad \text{Calinski-Harabasz index (CHI)} \tag{12}$$

Here, SS_B represents the overall variance between clusters, SS_W denotes the total variance within clusters, k is the number of clusters, and N is the total number of datasets. The total between-cluster variance (SS_B) is defined as follows [46]:

$$SS_B = \sum_{i=1}^{k} n_i m_i - m^2 \tag{13}$$

where m_i indicates the center of the i^{th} uster, k represents the total number of clusters, m represents the overall mean of the dataset, and $|m_i - m|$ denotes the Euclidean distance between two vectors. The formula for calculating the total within-cluster variance (SS_W) is as follows:

$$SS_W = \sum_{i=1}^{k} \sum_{x \in c_i} |x - m_i|^2 \tag{14}$$

In this context, x represents a specific data point, k denotes the number of clusters, c_i refers to the i^{th} cluster, m_i represents the centroid of that cluster, and $|x - m_i|$ represents the Euclidean norm (Euclidean distance) between the data point x and the cluster centroid m_i. An optimal clustering is achieved when the between-cluster variance (SS_B) is maximized while the within-cluster variance (SS_W) is minimized. Consequently, a higher CHI ratio indicates better clustering performance, reflecting well-separated and compact clusters.

The DBI quantifies clustering quality by assessing the ratio of intra-cluster dispersion to inter-cluster separation. It is calculated as follows [47]:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{j\neq i}\{D_{i.j}\} \quad \text{Davies-Bouldin index (DBI)} \quad (15)$$

where $D_{i.j}$ is the ratio of within-cluster variance to between-cluster distance for the i^{th} and j^{th} clusters and is defined as follows:

$$D_{i.j} = \frac{(\overline{d_i}+\overline{d_j})}{d_{i.j}} \quad (16)$$

Here, $\overline{d_j}$ represents the mean distance between each data point in cluster j and its centroid, $\overline{d_i}$ represents the mean distance between each data point in cluster i and its centroid, and $d_{i.j}$ specifies the Euclidean distance between the centroids of clusters i and j. A lower DBI indicates a more optimal clustering solution.

The SI evaluates how similar each data point is to others within its cluster relative to points in different clusters and is calculated as follows [48]:

$$SI = \frac{b_i - a_i}{\max(a_i \cdot b_i)} \quad \text{Silhouette index (SI)} \quad (17)$$

The variable a_i represents the mean distance from the i^{th} point to all other points within the same cluster, while b_i denotes the smallest mean distance from the i^{th} point to any neighboring cluster. The silhouette score ranges from -1 to +1. A higher silhouette value indicates that a data point is well-clustered within its own group and minimally associated with neighbouring clusters. Clustering quality is considered more favourable when most points exhibit high silhouette values.

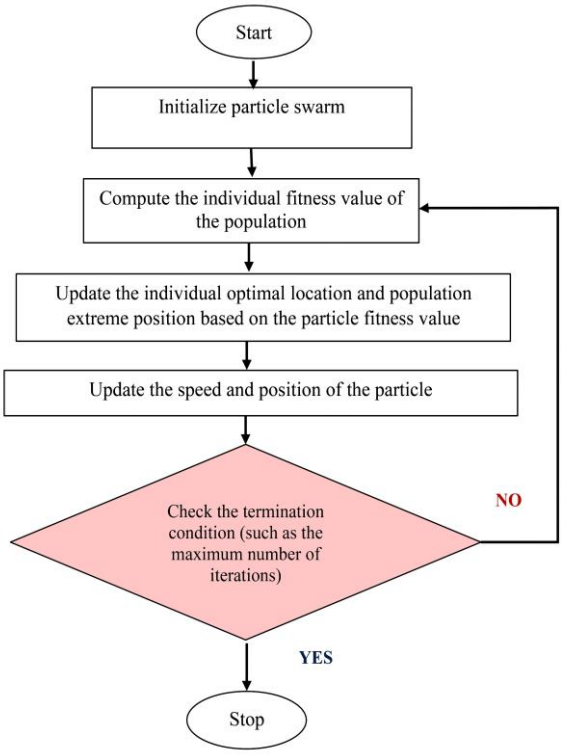

Figure 11. Flowchart of the PSO algorithm

3. Results

3-1. Determining the optimal number of clusters

As illustrated in Figure 12, the first step in the clustering process involved segmenting the dataset into clusters ranging from 2 to 7 and computing the elbow index for each scenario. Based on the methodology described in the previous section, the optimal number of clusters for the Sungun copper mine dataset was determined to be four. The graph initially shows a steep decline, with a distinct elbow shape observed at four clusters. Beyond this point, the rate of change diminishes significantly, and the curve stabilizes, becoming nearly parallel to the X-axis.

3.2. Preparation of Clustering Model based on K-means and FCM algorithm

Given the unique characteristics of the clustering process, its quality can be assessed using multiple evaluation methods. By analyzing the spatial distribution of the data, it is possible to determine clustering effectiveness and the accuracy of data assignment under specific conditions, considering the geometric structure and spatial positioning of the data. However, visual quality assessment methods are often insufficient due to the large volume of data and their intricate spatial distribution. To address these limitations, specific indicators and criteria have been developed to evaluate clustering performance. In this study, the values of the three aforementioned clustering quality criteria were computed to assess the efficiency and accuracy of each method. After determining the optimal number of clusters for the rock joint dataset, K-means and FCM methods were applied to classify the data. Figure 13 illustrates the

clustering of joints based on the K-means and FCM algorithms, demonstrating that both methods partition the data into four distinct classes. Table 2 presents the evaluation criteria values for data clustering using these two methods.

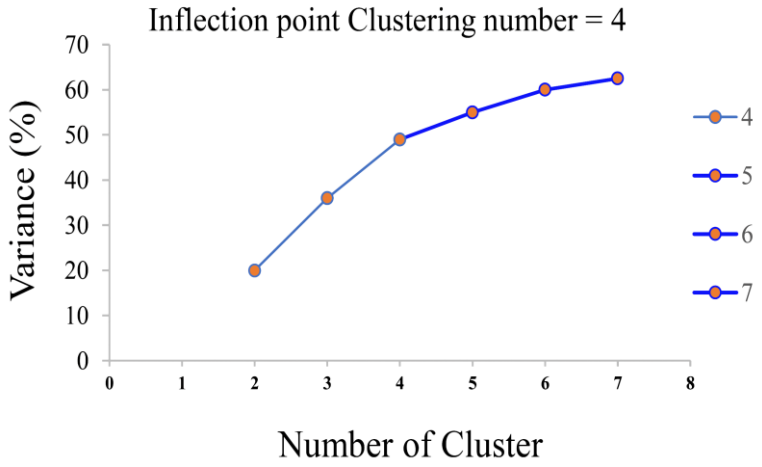

Figure 12. Data variance percentage versus the number of clusters

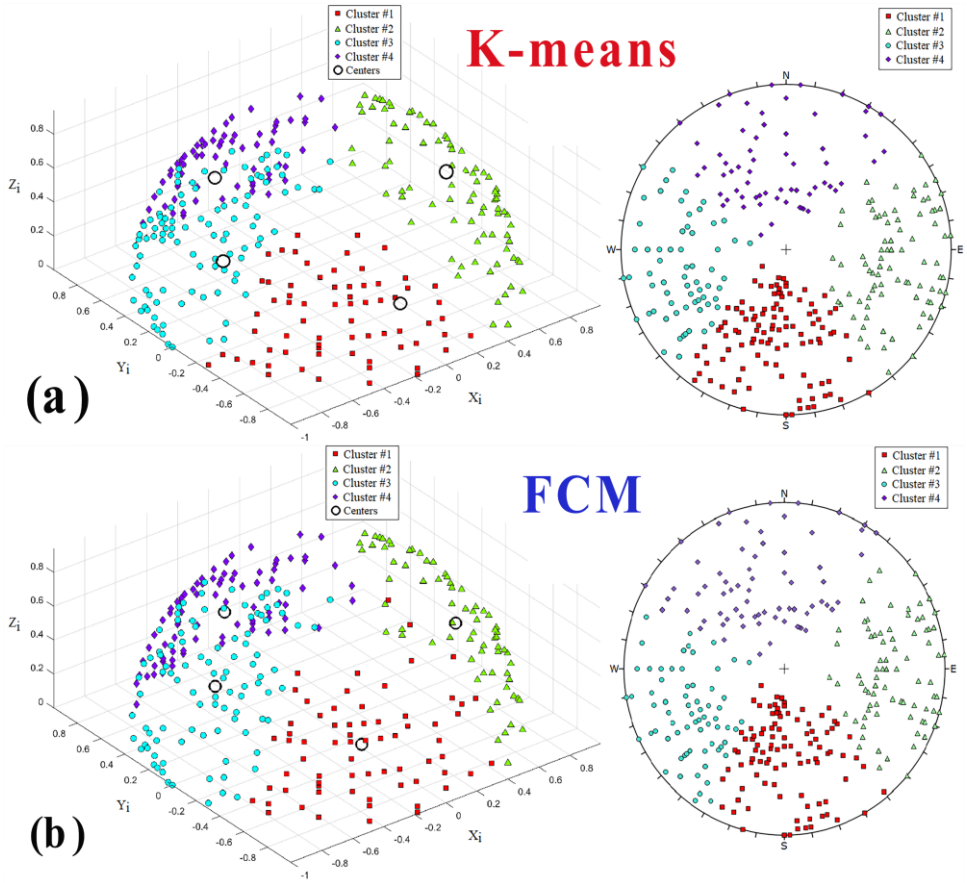

Figure 13. Stereographic Projection of Clustering Results (a) K-means Method, (b) FCM Method

Table 2. Values of Calinski-Harabasz, Davies-Bouldin, and Silhouette Criteria for Data Clustering

Methods	Values of the criteria of data		
	Calinski-Harabasz	Davies-Bouldin	Silhouette
K-means	312.77	0.88	0.541
FCM	346.05	0.82	0.565

The Davies-Bouldin criterion is widely recognized as a key metric for evaluating the quality of clustering methods, where lower values indicate superior clustering performance. In this context, the FCM method, with a value

of 0.82, outperforms the K-means method, which has a value of 0.88. The Calinski-Harabasz criterion serves as the second evaluation metric for clustering performance, favouring higher values as indicators of better clustering quality. Based on this measure, the FCM method (346.05) achieves more effective clustering compared to the K-means method (312.77). The silhouette criterion is the third metric used to assess clustering methods for rock joint data extracted from Sungun Mine. According to this criterion, the FCM method, with a silhouette value of 0.565, demonstrates the highest clustering performance.

3.3. Optimization of K-means and FCM clustering using HS and PSO algorithms

This study explores the integration of HS and PSO with FCM and K-means as unsupervised clustering techniques to analyze rock joint datasets from various case studies. To ensure optimal performance, it is crucial to define the control parameters of each algorithm. These parameters significantly impact the optimization results and the convergence speed of the algorithms. In many cases, there are no predefined equations or explicit rules for setting these parameters. Instead, expert judgment and dataset-specific adjustments, often determined through trial and error, are relied upon. For the HS algorithm, the key parameters include: Harmony Memory Size (HMS), Maximum Number of Iterations (MaxIt), Harmony Memory Consideration Rate (HMCR), Pitch Adjustment Rate (PAR), For the PSO algorithm, the essential parameters are: Inertia Weight (W), Swarm Size (PS), Inertia Weight Damping Rate (WDR), Maximum Number of Iterations (MaxIt), Following the initial analysis, the control parameters for the HS algorithm were set to a maximum of 1000 iterations and a Pitch Adjustment Rate (PAR) of 0.1, indicating that neighboring values were selected with a 10% probability. Furthermore, based on previous research, the typical range for HMS is 40 to 100. In this study, a value of 50 was deliberately chosen through a trial-and-error approach. Table (3) presents the parameter values used in the HS and PSO algorithms.

Table 3. Parameter values of the HS and PSO algorithms

Meta-heuristic algorithm	Hyperparameters	Values
HS	HMS	50
	MaxIt	1000
	HMCR	0.2
	PAR	0.1
	W	1
PSO	PS	50
	WDR	0.99
	MaxIt	500

Figures 14 to 15 and 16 illustrate the iteration-based optimization process and the stereographic plot of the clustering results for all methods, respectively. Figure 14 shows that the clustering process using HS-FCM and HS-K-means reached their optimal costs after 188 and 416 iterations, with final values of 167 and 169, respectively. Similarly, Figure 15 indicates that the PSO-K-means and PSO-FCM methods achieved their optimal costs after 48 and 33 iterations, with corresponding values of 156 and 160. These figures clearly demonstrate that the PSO algorithm converges faster than the HS algorithm, as indicated by the lower computational cost. It is important to note that the best value is a dimensionless number. Additionally, if the difference between the outcomes of two consecutive iterations falls below the minimum acceptable precision, the optimization process stabilizes. According to Figure (16), all clustering techniques successfully classified the rock joint datasets from the case studies. Furthermore, Table 4 and Figure 17 present a comparative analysis of the clustering methods based on the three previously discussed evaluation criteria.

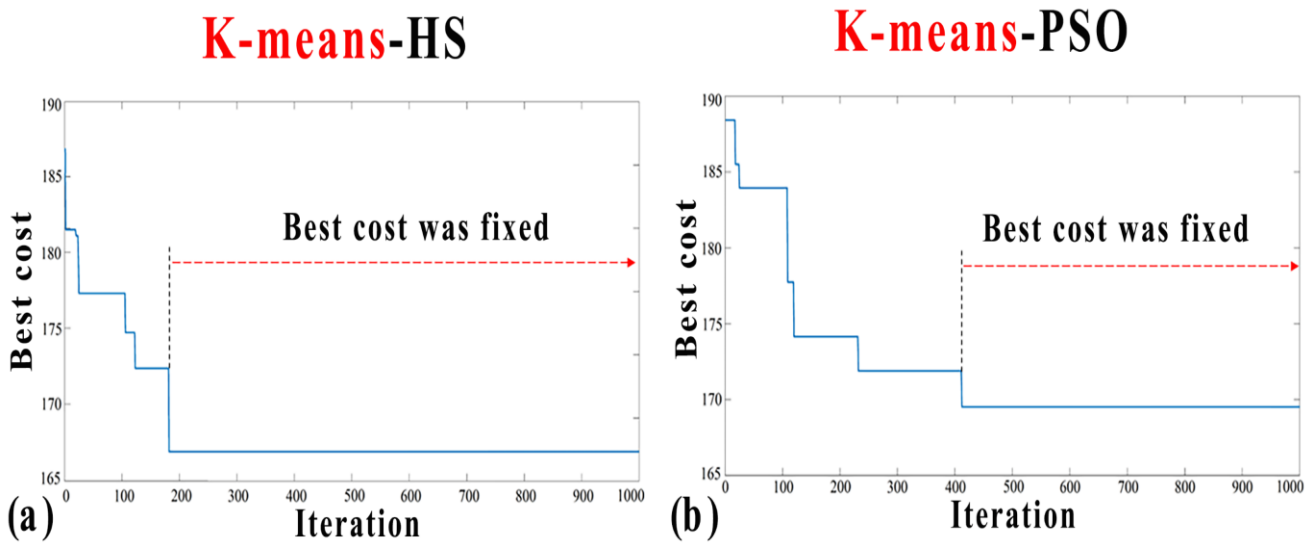

Figure 14. Best cost per iteration achieved by the algorithms:(a) K-means-HS (b) K-means-PSO

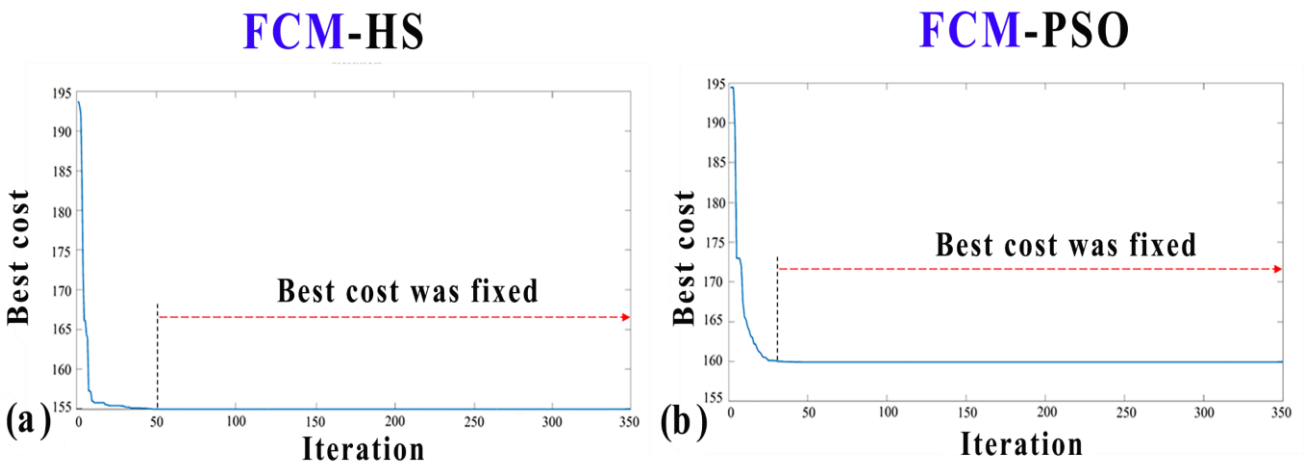

Figure 15. Best cost per iteration achieved by the algorithms:(a) FCM-HS (b) FCM-PSO

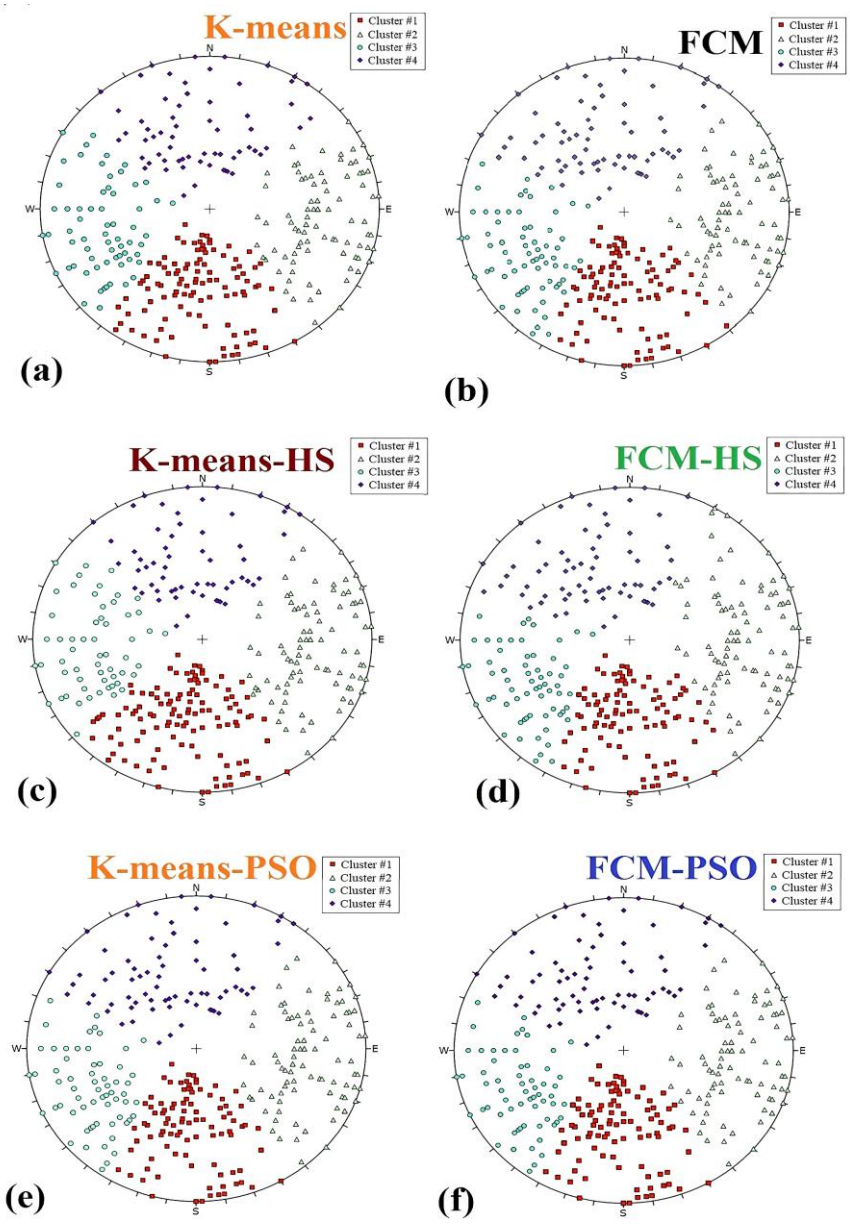

Figure 16. Comparison of clustering methods (a) K-means, (b) FCM, (c) K-means-HS, (d) FCM-HS, (e) K-means-PSO, (f) FCM-PSO

Table 4. Evaluation criteria for different clustering methods

Clustering method	\Evaluation criteria		
	Davies-Bouldin	Calinski-Harabasz	Silhouette
K-means	0.88	312.77	0.541
K-means-PSO	0.82	345.47	0.566
K-means-HS	0.84	341.69	0.560
FCM	0.82	346.05	0.565
FCM-PSO	0.80	348.47	0.576
FCM-HS	0.85	329.23	0.550

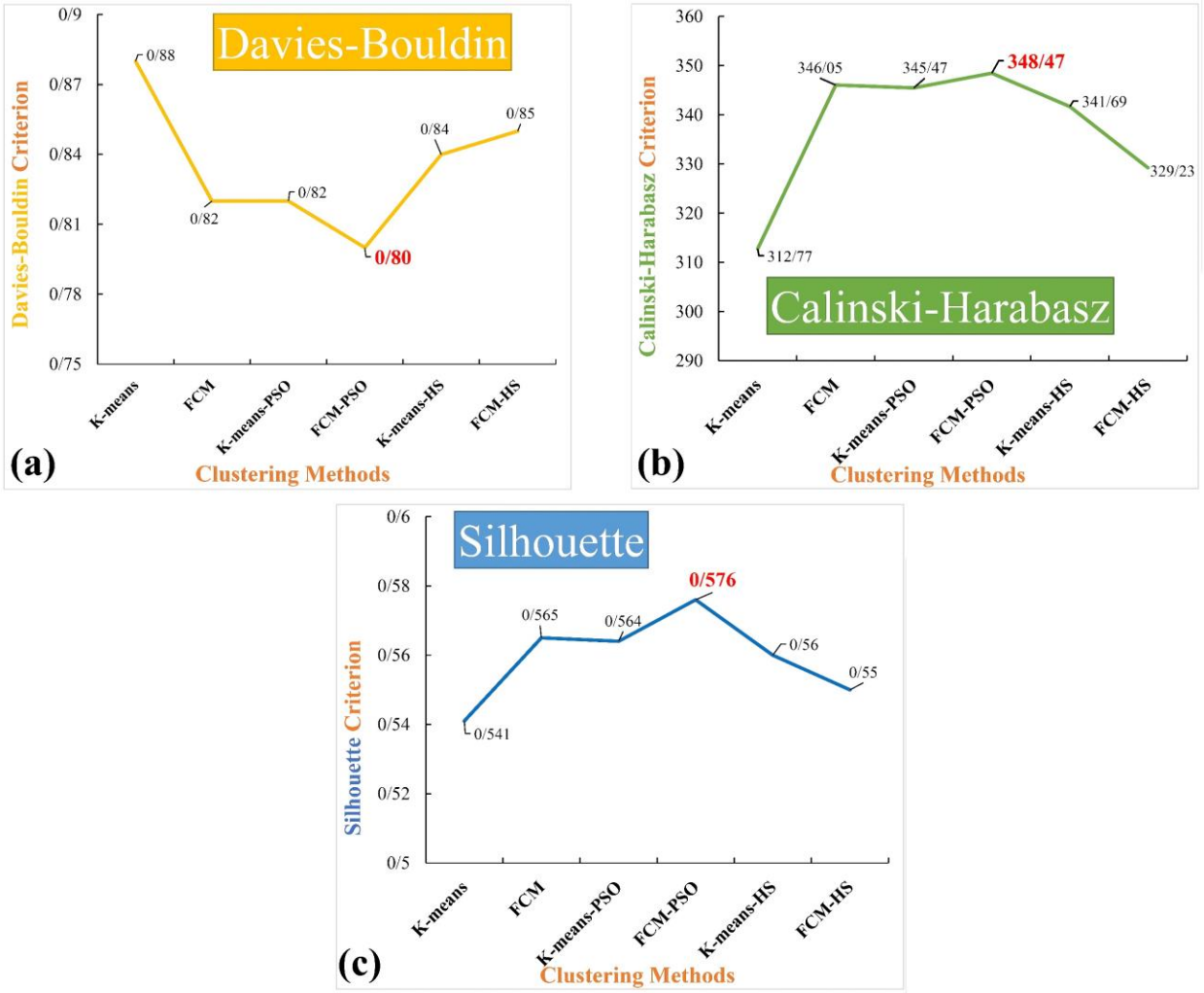

Figure 17. The results relate to the current research's evaluation criteria for clustering methods. a: Davies-Bouldin criterion b: Calinski-Harabasz criterion c: Silhouette criterion

4. Discussion

Based on the results presented in Table 4 and Figure 17, the FCM-PSO method has demonstrated superior performance in clustering joint-related data compared to other methods. According to the DBI, the FCM-PSO method achieves the lowest value (0.80), indicating its effectiveness in clustering. In contrast, the FCM-HS method does not show any improvement over the standard FCM method, as FCM alone attains a slightly better score of 0.82. Moreover, the PSO algorithm enhances the efficiency of the K-means method, reducing the DBI from 0.85 to 0.82.

For the Calinski-Harabasz criterion, which serves as the second evaluation metric, the FCM-PSO method ranks first with a value of 348.47, followed by the FCM (346.05) and K-means-PSO (345.47) methods. Similar to the previous metric, the FCM-HS method underperforms relative to the standard FCM approach. However, the K-means-PSO

method improves the clustering performance of the original K-means algorithm.

Analyzing the SI as the third evaluation criterion, the FCM-PSO method again outperforms all other methods with a value of 0.576. The K-means-PSO and FCM methods rank second and third with values of 0.566 and 0.565, respectively. In contrast, the K-means and FCM-HS methods exhibit lower silhouette scores, indicating suboptimal clustering performance. Overall, the FCM-PSO clustering method consistently demonstrates superior performance across all three evaluation criteria, making it the most effective approach for joint data classification in this study.

The findings of this study have significant practical implications for geotechnical and mining engineering, especially in complex geological settings like the Sungun porphyry copper deposit. Accurate delineation of joint sets facilitates more reliable evaluations of slope stability, tunnel orientation, and excavation strategies. The demonstrated superiority of the FCM-PSO algorithm provides a powerful and adaptable framework for capturing the inherent complexity and anisotropy of rock mass Joints. By enhancing the precision of joint classification, this method contributes not only to safer and more cost-effective mine design but also to improved ore recovery through optimized excavation and minimized damage to the surrounding rock mass. Furthermore, the algorithm's robustness in handling noisy, overlapping, and large-scale data makes it particularly well-suited for application in heterogeneous open-pit mining environments.

5. Conclusion

Rock joints significantly influence the mechanical behavior of rock masses, especially in open-pit mining operations during mineral extraction. Accurate identification and classification of rock joints—typically based on their slope and orientation characteristics—are essential for effective geotechnical analysis. In recent years, various clustering techniques have been employed to classify rock joints. However, meta-heuristic algorithms have shown superior capabilities in optimizing clustering tasks by identifying near-optimal solutions and reducing clustering errors.

This study focused on classifying joint orientation and slope data from 19 levels of the Sungun copper mine ramp using both K-means and FCM clustering methods. The optimal number of clusters was determined using the elbow method, resulting in a four-cluster solution. Both clustering techniques were subsequently enhanced by incorporating HS and PSO to improve centroid positioning and minimize intra-cluster distances. Clustering performance was quantitatively evaluated using three metrics: Davies-Bouldin index, Calinski-Harabasz index, and Silhouette coefficient.

Key findings of this study include:

- The FCM method consistently outperformed the standard K-means algorithm across all evaluation metrics. The integration of PSO into FCM (i.e., the FCM-PSO method) yielded the best performance, with the lowest Davies-Bouldin index (0.80), highest Calinski-Harabasz score (348.47), and highest Silhouette coefficient (0.57).

- The K-means-PSO method also improved upon the basic K-means performance, although it remained less effective than the FCM-based models.

- The FCM-HS method did not demonstrate substantial improvements, indicating that PSO may offer more effective optimization in this context.

Given its demonstrated effectiveness, the FCM-PSO method presents a powerful approach for classifying complex joint datasets in real-world mining applications. The ability to robustly handle noisy and overlapping geological features makes it particularly suitable for open-pit mining scenarios.

Future research directions may include:

- Conducting comparative analyses using other meta-heuristic optimization algorithms such as Grey Wolf Optimizer (GWO), Ant Colony Optimization (ACO), or Firefly Algorithm.

- Integrating deep learning methods with clustering algorithms to enhance performance on larger and more complex geological datasets.

- Applying the proposed hybrid framework to three-dimensional joint data for improved spatial characterization in underground mines.

6. References

[1] Hu, J., Zhou, T., Ma, S., Yang, D., Guo, M., & Huang, P. (2022). Rock mass classification prediction model using heuristic algorithms and support vector machines: a case study of Chambishi copper mine. Scientific Reports, *12*(1), 928. doi:10.1038/s41598-022-05027-y.

[2] Yang, B., Zhao, W., & Duan, Y. (2022). Critical damage threshold of brittle rock failure based on Renormalization Group theory. Geomechanics and Geophysics for Geo-Energy and Geo-Resources, 8(5), 135. doi:10.1007/s40948-022-00441-y.

[3] Cui, X., & Yan, E. chuan. (2020). Fuzzy C-Means Cluster Analysis Based on Variable Length String Genetic Algorithm for the Grouping of Rock Discontinuity Sets. KSCE Journal of Civil Engineering, 24(11), 3237–3246. doi:10.1007/s12205-020-2188-2.

[4] Xu, L. M., Chen, J. P., Wang, Q., & Zhou, F. J. (2013). Fuzzy C-means cluster analysis based on mutative scale chaos optimization algorithm for the grouping of discontinuity sets. Rock Mechanics and Rock Engineering, 46(1), 189–198. doi:10.1007/s00603-012-0244-z.

[5] Ye, J. (2017). A netting method for clustering-simplified neutrosophic information. Soft Computing, 21(24), 7571–7577. doi:10.1007/s00500-016-2310-z.

[6] Shanley, R. J., & Mahtab, M. A. (1976). Delineation and analysis of clusters in orientation data. Journal of the

International Association for Mathematical Geology, 8(1), 9–23. doi:10.1007/BF01039681.

[7] Tokhmechi, B., Memarian, H., Moshiri, B., Rasouli, V., & Noubari, H. A. (2011). Investigating the validity of conventional joint set clustering methods. Engineering Geology, 118(3–4), 75–81. doi:10.1016/j.enggeo.2011.01.002.

[8] Zhou, W., & Maerz, N. H. (2002). Implementation of multivariate clustering methods for characterizing discontinuities data from scanlines and oriented boreholes. Computers & Geosciences, 28(7), 827–839. doi:10.1016/S0098-3004(01)00111-X.

[9] Hammah, R. E., & Curran, J. H. (1999). On distance measures for the fuzzy K-means algorithm for joint data. Rock Mechanics and Rock Engineering, 32(1), 1–27. doi:10.1007/s006030050041.

[10] Hammah, R. E., & Curran, J. H. (2000). Validity Measures for the Fuzzy Cluster Analysis of Orientations. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12), 1467–1472. doi:10.1109/34.895981.

[11] Jimenez, R. (2008). Fuzzy spectral clustering for identification of rock discontinuity sets. Rock Mechanics and Rock Engineering, 41(6), 929–939. doi:10.1007/s00603-007-0155-6.

[12] Jimenez-Rodriguez, R., & Sitar, N. (2006). A spectral method for clustering of rock discontinuity sets. International Journal of Rock Mechanics and Mining Sciences, 43(7), 1052–1061. doi:10.1016/j.ijrmms.2006.02.003.

[13] Liu, J., Zhao, X. D., & Xu, Z. he. (2017). Identification of rock discontinuity sets based on a modified affinity propagation algorithm. International Journal of Rock Mechanics and Mining Sciences, 94, 32–42. doi:10.1016/j.ijrmms.2017.02.012.

[14] Li, Y., Wang, Q., Chen, J., Xu, L., & Song, S. (2015). K-means Algorithm Based on Particle Swarm Optimization for the Identification of Rock Discontinuity Sets. Rock Mechanics and Rock Engineering, 48(1), 375–385. doi:10.1007/s00603-014-0569-x.

[15] Ma, G. W., Xu, Z. H., Zhang, W., & Li, S. C. (2015). An enriched K-means clustering method for grouping fractures with meliorated initial centers. Arabian Journal of Geosciences, 8(4), 1881–1893. doi:10.1007/s12517-014-1379-x.

[16] Song, S., Wang, Q., Chen, J., Li, Y., Zhang, W., & Ruan, Y. (2017). Fuzzy C-means clustering analysis based on quantum particle swarm optimization algorithm for the grouping of rock discontinuity sets. KSCE Journal of Civil Engineering, 21(4), 1115–1122. doi:10.1007/s12205-016-1223-9.

[17] Song, T., Chen, J., Zhang, W., Xiang, L. J., & Yang, J. H. (2015). A method for multivariate parameter dominant partitioning of discontinuities of rock mass based on artificial bee colony algorithm. Rock Soil Mech 36: 861–868.

[18] Ding, Q., Huang, R., Wang, F., Chen, J., Wang, M., & Zhang, X. (2018). Multi-Parameter Dominant Grouping of Discontinuities in Rock Mass Using Improved ISODATA Algorithm. Mathematical Problems in Engineering, 2018(1), 5619404. doi:10.1155/2018/5619404.

[19] Shirazy, A., Hezarkhani, A., Shirazi, A., Khakmardan, S., & Rooki, R. (2021). K-Means Clustering and General Regression Neural Network Methods for Copper Mineralization probability in Char-Farsakh, Iran. Türkiye Jeoloji Bülteni / Geological Bulletin of Turkey, 65(1), 79–92. doi:10.25288/tjb.1010636.

[20] Mikaeil, R., Haghshenas, S. S., Haghshenas, S. S., & Ataei, M. (2018). Performance prediction of circular saw machine using imperialist competitive algorithm and fuzzy clustering technique. Neural Computing and Applications, 29(6), 283–292. doi:10.1007/s00521-016-2557-4.

[21] Esmaeilzadeh, A., & Shahriar, K. (2019). Optimized fuzzy cmeans – fuzzy covariance – fuzzy maximum likelihood estimation clustering method based on deferential evolutionary optimization algorithm for identification of rock mass discontinuities sets. Periodica Polytechnica Civil Engineering, 63(2), 674–686. doi:10.3311/PPci.13885.

[22] Zheng, S., Jiang, A. N., Yang, X. R., & Luo, G. C. (2020). A New Reliability Rock Mass Classification Method Based on Least Squares Support Vector Machine Optimized by Bacterial Foraging Optimization Algorithm. Advances in Civil Engineering, 2020(1), 3897215. doi:10.1155/2020/3897215.

[23] Ruan, Y., Chen, J., Fan, Z., Wang, T., Mu, J., Huo, R., Huang, W., Liu, W., Li, Y., & Sun, Y. (2023). Application of K-PSO Clustering Algorithm and Game Theory in Rock Mass Quality Evaluation of Maji Hydropower Station. Applied Sciences (Switzerland), 13(14), 8467. doi:10.3390/app13148467.

[24] Ruan, Y., Liu, W., Wang, T., Chen, J., Zhou, X., & Sun, Y. (2023). Dominant Partitioning of Discontinuities of Rock Masses Based on DBSCAN Algorithm. Applied Sciences (Switzerland), 13(15), 8917. doi:10.3390/app13158917.

[25] Wang, R., Ni, Y., Zhang, L., & Gao, B. (2025). Grouped machine learning methods for predicting rock mass parameters in a tunnel boring machine-driven tunnel based on fuzzy C-means clustering. Deep Underground Science and Engineering, 4(1), 55–71. doi:10.1002/dug2.12082.

[26] Yong, R., Wang, H., Ye, J., Du, S., & Luo, Z. (2024). Neutrosophic genetic algorithm and its application in clustering analysis of rock discontinuity sets. Expert Systems with Applications, 245, 122973. doi:10.1016/j.eswa.2023.122973.

[27] Zarean, A., & Poormirzaee, R. (2016). Joint inversion of ReMi dispersion curves and refraction travel times using particle swarm optimization algorithm. In Journal of Mining and Environment 7(1), 67–79.

[28] Hezarkhani, A., & Williams-Jones, A. E. (1998). Controls of alteration and mineralization in the Sungun porphyry copper deposit, Iran: evidence from fluid inclusions and stable isotopes. Economic Geology, 93(5), 651–670. doi:10.2113/gsecongeo.93.5.651.

[29] Calagari, A. A. (1997). Geochemical, stable isotope, noble gas and fluid inclusion studies of mineralization and alteration at Sungun porphyry copper deposit, East Azarbaidjan, Iran: Implications for genesis. PhD Thesis, The University of Manchester, Manchester, United Kingdom.

[30] Edwards, A. C. (Ed.). (2001). Mineral resource and ore reserve estimation - the AusIMM guide to good practice (monograph 23). Minerals Engineering, 14(9), II. doi:10.1016/s0892-6875(01)80033-9.

[31] Bagheryan, A. (2006). Copper concentration process in the Sungun copper complex (Internal report). National Iranian Copper Industries Company, Tabriz, Iran.

[32] Aggarwal, V., Gupta, V., Singh, P., Sharma, K., & Sharma, N. (2019). Detection of spatial outlier by using improved Z-score test. Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI 2019, 2019-April, 788–790. doi:10.1109/icoei.2019.8862582.

[33] Wong, L. N. Y., & Liu, G. (2010). An improved K-means clustering method for the automatic grouping of discontinuity sets. 44th US Rock Mechanics Symposium - 5th US/Canada Rock Mechanics Symposium, 10, 27 June, 2010, Salt Lake City, United States.

[34] Ketchen, D. J., & Shook, C. L. (1996). The application of cluster analysis in strategic management research: An analysis and critique. Strategic Management Journal, 17(6), 441–458. doi:10.1002/(sici)1097-0266(199606)17:6<441::aid-smj819>3.0.co;2-g.

[35] Adolfsson, A., Ackerman, M., & Brownstein, N. C. (2016). To cluster, or not to cluster: How to answer the estion. TKDD, 17, 1-9.

[36] MacQueen, J. (1967). Multivariate observations. Proceedings of the 5th Berkeley Symposium on Mathematical Statistic sand Probability, 1, 281-297, Volume 1: Statistics, University of California Press, Berkeley, United States.

[37] Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. Journal of Cybernetics, 4(1), 95–104. doi:10.1080/01969727408546059.

[38] Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences, 10(2–3), 191–203. doi:10.1016/0098-3004(84)90020-7.

[39] Geem, Z. W., Kim, J. H., & Loganathan, G. V. (2001). A New Heuristic Optimization Algorithm: Harmony Search. Simulation, 76(2), 60–68. doi:10.1177/003754970107600201.

[40] Wang, X., Gao, X.-Z., & Zenger, K. (2015). An Introduction to Harmony Search Optimization Method. Springer, Cham, Switzerland. doi:10.1007/978-3-319-08356-8.

[41] Haghshenas, S. S., Haghshenas, S. S., Geem, Z. W., Kim, T. H., Mikaeil, R., Pugliese, L., & Troncone, A. (2021). Application of harmony search algorithm to slope stability analysis. Land, 10(11). doi:10.3390/land10111250.

[42] Kennedy, J. and Eberhart, R. (1995) Particle swarm optimization. Proceedings of ICNN'95 - International Conference on Neural Networks, 4, 1942–1948. doi:10.1109/icnn.1995.488968.

[43] Kennedy, J. (1998). The behavior of particles. Evolutionary Programming VII. EP 1998. Lecture Notes in Computer Science, vol 1447. Springer, Berlin, Germany. doi:10.1007/BFb0040809.

[44] Shen, J., Wang, C., Wang, R., Huang, F., Fan, C., & Xu, L. (2015). A Band Selection Method for Hyperspectral Image Classification based on improved Particle Swarm Optimization. International Journal of Signal Processing, Image Processing and Pattern Recognition, 8(4), 325–338. doi:10.14257/ijsip.2015.8.4.28.

[45] Gao, H., Yang, Y., Zhang, X., Li, C., Yang, Q., & Wang, Y. (2019). Dimension reduction for hyperspectral remote sensor data based on multi-objective particle swarm optimization algorithm and game theory. Sensors (Switzerland), 19(6). doi:10.3390/s19061327.

[46] Caliñski, T., & Harabasz, J. (1974). A Dendrite Method Foe Cluster Analysis. Communications in Statistics, 3(1), 1–27. doi:10.1080/03610927408827101.

[47] Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1(2), 224–227. doi:10.1109/TPAMI.1979.4766909.

[48] Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65. doi:10.1016/0377-0427(87)90125-7.