Journal of Mining and Environment (JME) Vol. 17, No. 5, 2026, 1781-1800

Shahrood University of Technology

Journal of Mining and Environment (JME)

Journal homepage: www.jme.shahroodut.ac.ir

Iranian Society of Mining Engineering (IRSME)

Determination of Geochemical Probability Index Map Using Harmony Search and Cocoa Search Algorithms in Dehno Area, Central Iran

Fatemeh Vesmoridi, and Feridon Ghadimi*

Department of Mining Engineering, Arak University of Technology, Arak, Iran

Article Info

Received 28 July 2025

Received in Revised form 19 August 2025

Accepted 6 September 2025

Published online 6 September 2025

DOI: 10.22044/jme.2025.16587.3243

Keywords

Stream sediments

Stepwise factor analysis

Cocoa search and Harmony search algorithms

Khomain Dehno

Abstract

A total of 400 stream sediment samples were analyzed for 13 elements, and stepwise factor analysis was employed to generate geochemical maps indicative of mineralization. This method was utilized to develop a Geochemical Mineralization Probabilistic Index (GMPI) through a novel approach that produces geochemical evidence maps derived from stream sediment data. The study comprised a three-stage factor analysis of geochemical data collected from the Khomain Dehno region. The first factor included Zn, Pb, As, and Cd, accounting for 41.63% of the variance. The second factor comprised Mn, Mo, and Zr, explaining 21.86% of the variance, while the third factor consisted of Fe, Cu, and Ti, representing 7.79% of the variance. The cumulative variance explained by these three factors was 81%. Furthermore, a novel intelligent methodology, termed Relevant Vector Regression (RVR), enhanced with Cocoa Search (CS) and Harmony Search (HS) algorithms, is proposed for the prediction of the GMPI. The HS and CS algorithms were integrated with the RVR model to optimize its hyperparameters. In these models, Zn, Pb, As, and Cd served as input variables, while the GMPI was designated as the output variable. The performance of the predictive models was evaluated using Mean Squared Error (MSE) and the Coefficient of Determination (R^2). The results indicated that the RVR model optimized with the HS algorithm exhibits superior performance, achieving an R^2 value of 0.99256 and an MSE of 0.0031455. These findings underscore the efficacy of the proposed approach for accurate GMPI estimation.

1. Introduction

Mineral Potential Mapping (MPM) is a comprehensive, multi-stage process involving the creation of evidence maps where features indicative of the presence of mineral deposits are extracted, weighted, combined, and ultimately used to rank promising target areas for further exploration [1-3]. Within MPM, two primary approaches are employed for assigning evidence weights and integrating various evidence maps: knowledge-based methods and data-driven methods [4-6].

The integration of geochemical data from stream sediments with other types of mineral exploration data in knowledge-based MPM poses significant challenges, primarily due to the need for detailed analysis of multi-element geochemical anomalies as indicators of potential mineralization [7]. Analysis of stream sediment samples can reveal various geochemical relationships, some of which may serve as surface geochemical signatures of underlying mineral deposits. A key challenge in stream sediment geochemistry lies in identifying anomalous multi-element signatures that are indicative of mineralization [8, 9]. In this context, multivariate analysis methods are invaluable, as they enable the assessment of the relative importance of various combinations of geochemical variables [10].

The prediction of the Geochemical Mineralization Probability Index (GMPI) offers a valuable approach by generating mappable weights that can be incorporated as evidence maps in MPM. The application of stepwise factor analysis enhances the identification of anomalous

Corresponding author: ghadimi@arakut.ac.ir (F. Ghadimi)

geochemical signatures, improves detection of geochemical anomalies, and increases the proportion of explained variance in the dataset [11]. By employing GMPI, a novel data-driven fuzzification technique has been proposed to (a) effectively assign weights to geochemical anomaly classes derived from stream sediments and (b) improve the predictive accuracy of mineral potential maps, thereby enhancing exploration success [12, 13]. Furthermore, GMPI can be utilized to evaluate the predictive potential of each geochemical sample from stream sediments for identifying the type of reservoir located upstream of the sample site. It can also be efficiently integrated into knowledge-based MPM as an exploratory data analysis tool to generate evidence-weighted maps for under-explored regions [14-16].

Several studies have applied multivariate techniques to analyze geochemical exploration data [17–21]. Among these, factor analysis is widely employed and has proven particularly effective for interpreting geochemical data derived from stream sediments [22–24]. The primary objective of factor analysis is to explain variation within a multivariate dataset using a minimal number of factors and to uncover latent patterns in the data [25]. This makes factor analysis ideally suited for exploring inherent variability in geochemical datasets containing multiple elements and explains its frequent use as an exploratory data analysis tool.

While artificial neural networks (ANNs) are commonly applied to predict GMPI, challenges arise in determining optimal network architectures and managing randomness during model training [26–30]. Moreover, ANNs suffer from limitations such as a low learning rate [31]. In contrast, Support Vector Regression (SVR) is a deterministic global method that offers improved consistency. However, SVR requires tuning multiple parameters (such as penalty weight C and sensitivity parameter ε) and the careful selection of an appropriate kernel function. Relevant Vector Regression (RVR) is emerging as a promising alternative to SVR. Unlike SVR, RVR does not restrict the choice of kernel functions to those satisfying Mercer's theorem. Additionally, RVR requires tuning only the kernel width (σ), simplifying implementation. Its sparse modeling capability enables RVR to generate more generalized results with significantly reduced computational time compared to SVR.

A limitation in current research is the insufficient application of optimization algorithms such as Harmony Search (HS) and Cuckoo Search (CS) to aid in identifying geochemical anomalies. In the present study, an improved RVR model is proposed for the indirect estimation of GMPI. To optimize RVR parameters, the HS and CS optimization algorithms are employed, facilitating the selection of appropriate kernel parameters. The performance of these hybrid models (RVR-HS and RVR-CS) was evaluated using data available in the scientific literature. Notably, no prior research has been identified that applies these novel optimization methods for exploring anomalies in stream sediments. Thus, this study addresses this gap by implementing HS and CS algorithms to enhance the prediction of GMPI. One of the salient features of the article is the use and evaluation of two HS and CS algorithms to predict the Geochemical Mineralization Probability Index (GMPI) in the stream sediments of the Dehno Khomein mining area.

2. Geology

The Dehno-Khomein mineralization zone is located within the Sanandaj-Sirjan structural belt (Figure 1) [32]. Dehno granite of type S (sedimentary) is derived from the melting of slates, phyllites, and schists of lower Upper Triassic and Lower Jurassic age (Figure ۲). Before the Lower Cretaceous, it was replaced due to the Upper Jurassic orogenic phase and did not affect the Cretaceous units, the lowest of which is red sandstone. Since the granite does not have a mantle origin and the host rock is melted slate, phyllite, and low-grade schist, and it is not derived from melting of igneous rocks such as andesitic to basaltic volcanic rocks, it is therefore devoid of copper. After the emplacement of the granitic magma that formed the Dehno granitic intrusive body, its hydrothermal solutions were purely acidic and copper enrichment alteration was created around it. It should be noted that the melting of the oceanic crust of the Arabian plate subducting beneath Iran and the ascent of its mafic magma about 10 to 15 kilometers above the mantle within the crust, along with the heat produced, melted the aforementioned metamorphic rocks and generated the granitic magma. Mafic dioritic and gabbroic dikes in the Dehno area support this interpretation. Thus, magmatism in the region is hybrid or from two sources, mantle and crust and the contribution of mantle magma uplift to the surface is negligible. The granitic magma caused a weak thermal metamorphic halo in the surrounding sandstones and created weakly metamorphosed rocks such as albite and cordierite hornfels with a limited radius

of influence. Additionally, differentiation of the granitic magma produced tourmaline-bearing pegmatitic dikes and very light-colored granitic masses called leucogranite in Dehno. The granite mass in the region forms a batholith extending from the Oligocene to Azna. The Dehno intrusive body is similar to the pistachio granodiorite found in Sarband County. Also, its andalusite-bearing diorites are similar to andalusite hornfels or metamorphosed diorite like those in Malmir village, Hendudar city. The temperature of this dioritic body, sometimes gabbroic and mantle-derived, ranged from 800 to 900°C, which caused melting of Upper Triassic and Lower Jurassic slates and phyllites. In Reza Abad, the intrusive body occurs as stocks or boss-like spherical bodies connected beneath to the Dehno batholith or the Oligocene-Azna batholith [33].

Regarding Pb-Zn- mineralization in slates and phyllites, it is necessary to mention the Pb-Zn mineralization in Kazem Abad village (Khomein) east of Dehno village, and Golzard village above Dehno. There is a possibility of Pb-Zn presence in downstream slates and phyllites of the Dehno valley and in the alluvium downstream of the lead-zinc mine of Dehno and within several-meter-thick silicified veins. The source of Pb and Zn in high-grade slates and phyllites and low-grade sandstones has two genetic ages: syngenetic within primary shale and sandstone (springs) and epigenetic due to Pb and Zn leaching by hydrothermal fluids from shales into faults and fractures parallel to bedding, depositing sphalerite, galena, and occasionally chalcopyrite. It appears that the copper in the Dehno mine inside granites results from hydrothermal transport and leaching of chalcopyrite copper from slates and phyllites into the granites, seen as malachite or oxide forms, which is insignificant.

The oldest rock formations in this area consist of dark-colored shales and slates dating to the Triassic-Jurassic period, interspersed with carbonate layers. Jurassic units have experienced intrusions of granite and granodiorite (Figure 2). In the northwestern part of the zone, hornfels is present in contact zones adjacent to intermediate intrusive rocks. The Cretaceous sedimentary sequence begins with basal sandstones and conglomerates that grade upward into calcareous units. Limestones, along with associated slate and phyllite, serve as the host rocks for Zn and Pb deposits found in Dehno. Overlying these units are marl deposits of Cretaceous age. Copper (Cu) mineralization occurs near the intrusive bodies, manifesting as silica-iron oxide veins and malachite mineralization. Evidence of historic mining activities, including tunnels and exploratory trenches, is apparent in the region. The mineralization primarily comprises oxidized Pb and Zn carbonates, such as anglesite and smithsonite. These mineral deposits are typically concentrated within small fractures, cracks, and faults, occurring predominantly in a stratiform framework.

Moreover, after the end of the Jurassic and beginning of the Cretaceous, the Cretaceous sea advanced over the metamorphic rocks of Triassic and Jurassic (slates and phyllites) and also over the granitic intrusive body. The Lower Cretaceous in Dehno initially formed with coastal red sandstones and then massive continental shelf limestones with minor regression forming shales, marls, and lagoonal marly limestones. There are no Middle and Upper Cretaceous stratigraphic units in Dehno. During deposition of the limestones, old normal faults existed on the seafloor and lagoon, through which hot springs discharged mineralized fluids with Pb- Zn, sometimes Cu and Fe, precipitating sphalerite, galena, pyrite, and occasionally chalcopyrite during deposition and early diagenesis. Two mineralization horizons formed in the Lower Cretaceous: one at the contact of massive limestone with marl, and another within marl limestone of sulfide type, the latter being weak and thin (2 to 3 meters). Sulfide mineralization, especially sphalerite, may occur in massive limestone beneath the marl. After orogeny and exposure of limestone at the surface, meteoric waters entered sulfides via faults, transforming them into hemimorphite, smithsonite, anglesite, and cerussite, observed in the Dehno mine. The presence of barite, calcite, dolomite, and normal faults indicates an Irish-type mineralization in the area. Since the Mississippi Valley-type (MVT) mineralization contains barite and lacks normal faults but has thrust or reverse faults, it can be stated that the Pb-Zn types of Malayer-Esfahan can be either Irish or MVT. In the Khomein region, deposits in the middle of Malayer-Esfahan in limestones are mostly Irish type, while in slates and phyllites are Sedox type.

3. Material and methods

3.1. Sampling

A regional stream map was initially prepared to identify precise sampling locations. Sampling points were systematically selected to ensure a uniform and homogeneous sampling density across the Dehno-Khomein exploration area, with an

average density of approximately 60 samples per km² for stream sediments (Figure 3). Collected samples were crushed and ground to a particle size of 200 mesh. After thorough homogenization, approximately 20 grams of each sample were sent to an accredited laboratory for elemental analysis using Inductively Coupled Plasma (ICP) techniques. Elemental concentrations were determined following a four-acid digestion procedure (HF, HCl, $HClO_4$, and HNO_3), and subsequent measurement was performed via Inductively Coupled Plasma Mass Spectrometry (ICP-MS). In total, 400 samples were analyzed for 13 elements.

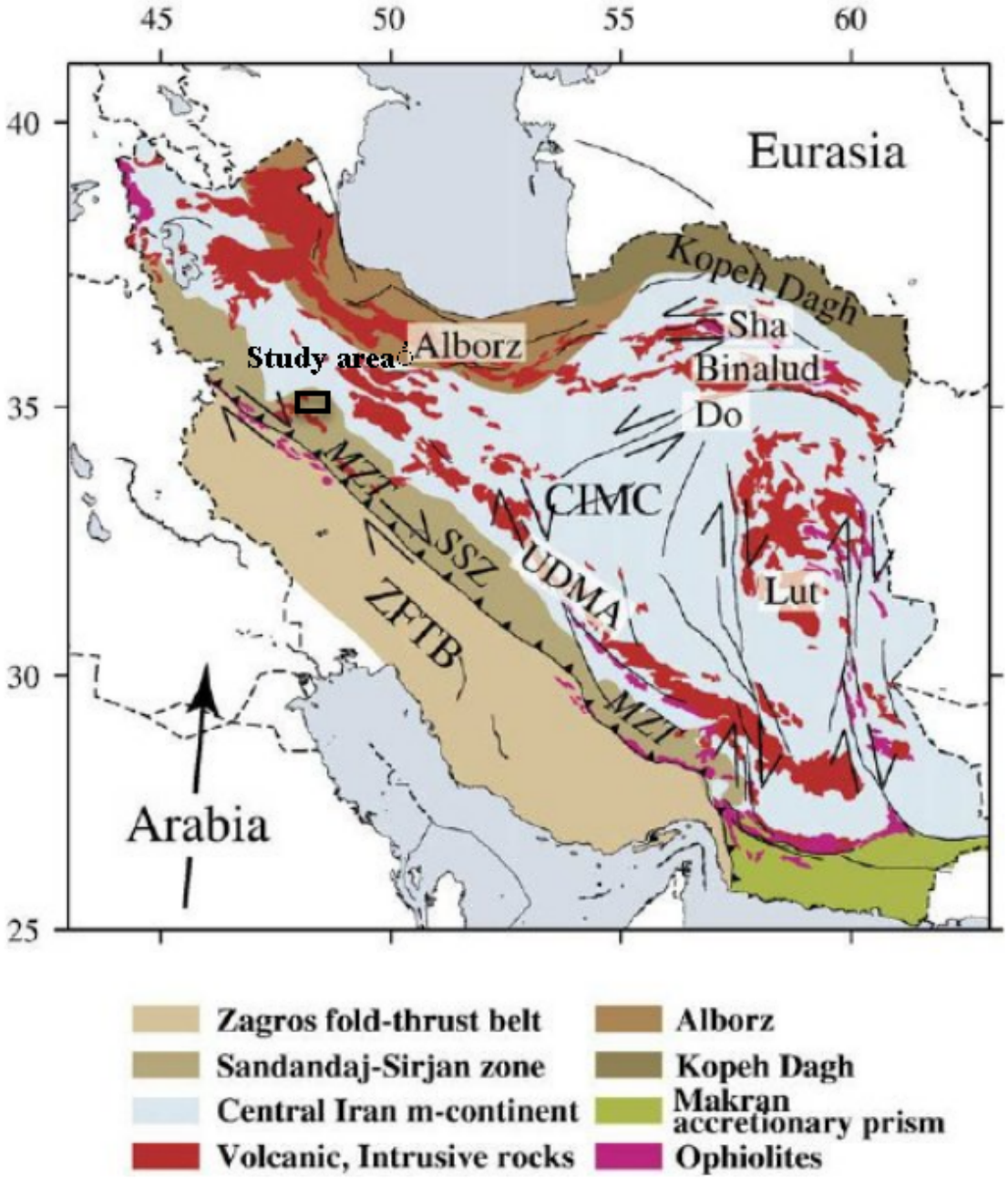

Figure 1. Simplified geological map of Iran showing the main tectonic regions. ZFTB: Zagros fold-thrust belt; SSZ: Sanandaj-Sirjan zone; MZT: Main Zagros thrust; UDMA: Urumieh- Dokhtar magmatic assemblage; CIMC: Central Iran micro-continent. Do is the Doruneh fault and Sha is the Shahrud fault [33]

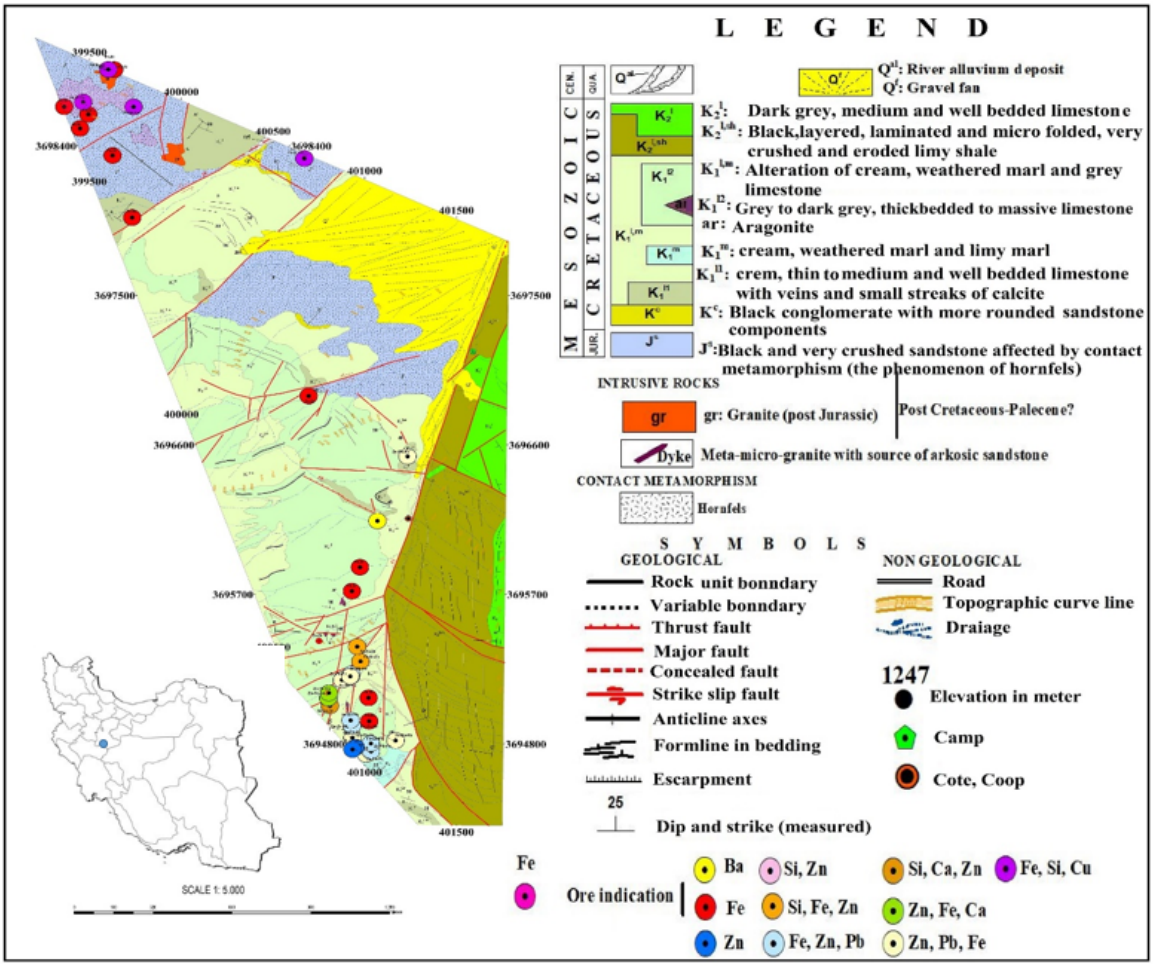

Figure 2. Geological map and profile of Zn and Pb deposits in Khomain Dehno

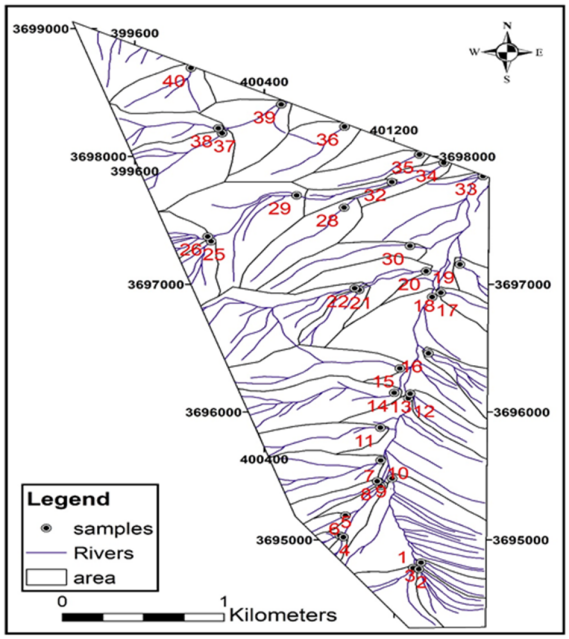

Figure 3. Location map of major stream sediment sampling points

3.2. Factor analysis

Stepwise factor analysis was employed in this study to generate geochemical scatter maps indicative of mineralization [34–38]. Since geochemical data typically deviate from a normal distribution, the central logarithmic ratio (CLR) transformation (Equation 1) was applied to normalize and standardize the data. The Varimax rotation method was subsequently used to simplify the factor structure and enhance interpretability.

The term CLR can mean different things depending on the field for example:

In microbiome or compositional data analysis, CLR usually means Centered Log-Ratio transformation, which is used to handle compositional data appropriately.

In machine learning, CLR can refer to Cyclic Learning Rate method.

Or it might be something else entirely depending on the study.

If it talking about compositional data, it typically uses the CLR method because:

It transforms compositional data (which are proportions summing to 1) into real space, allowing standard statistical analyses.

It addresses the issue of relative scale and the constant sum constraint in compositional data.

It helps avoid spurious correlations caused by the compositional nature of the data.

$$CLR(X) = Ln\ \frac{X}{X_j} \quad (1)$$

In equation 1, X represents the original data value, while X_j denotes the geometric mean of the dataset. To generate factor maps, it is necessary to calculate the factor score of each variable corresponding to the factor of interest. Given that factor scores varied across samples, the data were transformed using a fuzzy membership function (Equation 2). This transformation converted the factor scores into the Geochemical Mineralization Probability Index (GMPI), which ranges from 0 to 1 and was subsequently utilized within ArcGIS software for spatial analysis [4, 39]. The Fuzzy Factor Scores (FFS) were calculated as fuzzy weights based on percentile thresholds (minimum, 25%, 50%, 75%, 84%, 97.5%, and 99.5%) for each sample. All calculations were performed using Statistica software.

$$\text{GMPI} = \frac{e^{FS}}{1 + e^{FS}} \quad (2)$$

Where are:

GMPI- fuzzy weight

FS- factor score of each sample.

3.3. Relevance vector regression (RVR)

The Relevant Vector Regression (RVR) method, introduced by Tipping [40], is a specialized form of the Gaussian process. Unlike Support Vector Regression (SVR), RVR effectively quantifies uncertainty in the output estimates. Additionally, RVR exhibits better generalization and dispersion properties compared to SVR, which reduces the complexity of online predictions. Importantly, RVR does not require tuning of the error/maximum margin trade-off parameter (C), thereby decreasing computational time. Furthermore, the kernel function in RVR is not constrained to satisfy Mercer's condition, providing greater flexibility in model design. These advantages have contributed to the growing popularity of RVR, which has been successfully applied to various regression estimation problems [41, 42].

In the RVR approach, the system is assumed to be multi-input-single-output. Given N input vectors and N corresponding scaled targets $\{x_n, t_n\}_{n=1}^{N}$, the output $t = (t_1, \dots, t_N)^T$ can be expressed as a sum of an approximation $Vector = \left(y(x_{1,\dots,y(x_N)})\right)^T$. The objectives are obtained from the model with incremental noise:

$$t_n = y\ (x_n, w) + e \quad (3)$$

Where w is the weight vector and e is the random noise. The function y(x) is defined as follows:

$$y\ (x, w) = \sum_{i=1}^{N} w_i K(x, x_i) + w_0 = \sum_{i=1}^{N} w_i\ \emptyset(x) \quad (4)$$

$\Phi(x)$ is given as $\emptyset(x) = ([1, \mathrm{K}\ (x, x_1), K\ (x, x_2, \dots, K|(x, x_N)]$

The targets can be given as $p(t_n|x_n) = N(t|y(x_n), \delta^2)$.

$$p(t|w, \delta^2) = \frac{1}{2\pi\delta^2} exp\left\{-\frac{1}{2\pi\delta^2}\|t - \emptyset(x)w\|\right\} \quad (5)$$

Where $=(w_0, w_1, \dots, w_N) =,\ t = (t_{10}, t_2, \dots, t_N)$

And Φ is the $N\ (N + 1)$ design matrix

In this approach, Relevant Vector Regression (RVR) adopts a Bayesian perspective and constrains w and δ^2 by defining a prior probability distribution over the weights:

$$p(w|\alpha_i) = \prod_{i=1}^{N} N\left(w_i|0, \alpha_i^{-1}\right) = \frac{1}{2\pi^{(N+1)/2}} \prod_{i=1}^{N} \alpha_1^{1/2}\ exp\left(-\frac{a\alpha_{iw_i^2}}{2}\right) \quad (6)$$

$$p(\alpha) = \prod_{i=1}^{N} gamma\ (\alpha_i|a, b) \quad (7)$$

$$p(\beta) = gamma(\beta|a, b) \quad (8)$$

Where b=δ^2, a is an N+1 hyper-parameter, and gamma $(\alpha|a,b)$ is defined as equation 9.

$$1\Gamma(a)^{-1}b^a\alpha^{a-1}e^{-b\alpha}\Gamma(a)=\int_0^\infty t^{a-1}\,e^{-t}dt \qquad (9)$$

Also the posterior over weights can be considered through the Bayesian rule:

$$p(w|t,\alpha,\delta^2)=\frac{p(t|w,\delta^2)p(w|\alpha)}{p(t|\alpha,\delta^2)}=\frac{1}{2\pi^{(N+1)/2}}|\sum|^{-1/2}exp\{\left\{-\frac{1}{2}(w-\mu)^t\sum^{-1}(w-\mu)\right\} \qquad (10)$$

Where the posterior covariance and mean are defined as follow:

$$\sum=(\delta^{-2}2\emptyset^T2\emptyset+A)^{-1} \qquad (11)$$

$$\mu=(\delta^{-2}\sum\emptyset^Tt \qquad (12)$$

Where A=diag $(\alpha_1,\alpha_2,\dots,\alpha_N$. The likelihood distribution over the training targets is given by Tipping [43].

$$p(t|\,\alpha,\delta^2)=\int p(t|w,\delta^2)p(w|\alpha)dw=(2\pi)^{-N/2}\,|C|^{-1/2}exp\left\{\frac{1}{2}t^TC^{-1}t\right\} \qquad (13)$$

Where the covariance is given by C=$\delta^2I+\emptyset A^{-1}\emptyset^T$. A detailed explanation of the RVR approach can be found in Tipping [43].

3.4. Harmony search algorithm:

The Harmony Search (HS) algorithm is a metaheuristic technique inspired by the improvisation process of musicians [44-45]. Unlike traditional optimization methods, it does not require initial values for the decision variables. Instead, it employs a random search guided by a harmony memory, incorporating parameters such as the harmony consideration rate and the pitch adjustment rate [46]. The HS algorithm is relatively straightforward to implement and demands the tuning of only a few parameters. The algorithm can be described as follows:

Step 1: The optimization problem is defined as follows:

Minimize f(x) with respect to $x_{iL}\leq xi\leq x_{iu}$ (i=1, 2… N), where x_{iu} and x_{iL} representing the upper and lower bounds of the decision variables, respectively. The parameters of the Harmony Search (HS) algorithm are also specified at this stage. These parameters include the harmony memory size (HMS), which denotes the number of solution vectors in the harmony memory, the pitch adjustment rate (PAR), the harmony memory consideration rate (HMCR), the bandwidth (bw), and the number of improvisations (K), which serves as the stopping criterion. The value of K corresponds to the total number of function evaluations.

Step 2: The Harmony Search (HS) algorithm is initialized within the Harmony Memory (HM).

Harmony Memory refers to a memory structure where all solution vectors (sets of decision variables) are stored. The initial Harmony Memory is generated randomly within the desired region $[x_{iL}, x_{iu}]$ (i=1, 2 … N). This process is carried out according to the following equation 14.

$$x_i^j=x_{iL}+rand\,()\times(x_{iu}-x_{iL})\;\; j=1,2,\dots,HMS \qquad (14)$$

Where rand () is a random number from a uniform distribution of [0, 1].

Step 3: The harmony, x′= (x′$_1$, x′$_2$… x′$_n$), is generated based on three mechanisms:

random selection, memory consideration, and suboptimal adjustment [47-49]. In the random selection mechanism, the value of each decision variable in the harmony vector is chosen randomly from the range of possible values with a probability of (1−HMCR). The Harmony Memory Consideration Rate (HMCR), a value between 0 and 1, represents the probability of selecting a value from the historical values stored in the

Harmony Memory (HM), while (1−HMCR) denotes the probability of randomly selecting a value from the range of possible values [50].

$$x_i' = x_i' \in \{x_i^1, x_i^2, \dots, x_i^{HMS}\} \text{ with probability HMCR}$$

$$x_i' = x_i' \in x_i \text{ with probability } (1 - \text{HMCR}) \quad (15)$$

The value of each decision variable, obtained through memory consideration, is evaluated to determine whether pitch adjustment should be applied. If a decision is made to apply a pitch adjustment to x_i' with a probability equal to the Pitch Adjustment Rate (PAR), x_i' is replaced by the value $x_i' \pm u\,(-1, +1) \times bw$ represents the desired interval bandwidth for the continuous design variabls and $u\,(-1, +1)$ is a uniform distribution between -1 and 1. The value (1−PAR) determines the probability of no change being made. Therefore, pitch adjustment is applied to each variable as follows:

$$x_i' = x_i' \pm u\,(-1, +1) \times bw \text{ with probability HMCR} \quad (16)$$

$$x_i' = x_i' \text{ with probability HMCR} \times (1 - \text{PAR})$$

Step 4: Training the SVR Model and Evaluating the Fit

For this purpose, the entire dataset is randomly divided into two independent and non-overlapping subsets: a test set and a training set. The training set is used for the training process and the selection of optimal parameters, while the test set is employed to assess the predictive robustness and the overall capability of the model.

Step 5: **Update the Harmony Memory**

After generating a new harmony vector x^{new}, the harmony memory is updated. If the fitness value of the improvised harmony vector $x^{new} = (x_1^{new}, x_2^{new}, \dots, x_N^{new})$ is better than the fitness value of the worst harmony in the memory, the worst harmony in the Harmony Memory (HM) is replaced by x^{new}, making it a new member of the harmony memory.

Step 6: End of the process

Steps 3 to 5 are repeated until the stopping criterion (maximum number of improvisations K) is met. In this paper, the kernel factor of the Gaussian kernel function RBF $\left(K_{RBF}\left(x_i, x_j = exp\left(-\frac{\|x_i - x_j\|^2}{\gamma^2}\right)\right)\right)$ is selected by the HS algorithm.

3.5. Cocoa search algorithm

A family of birds known as cuckoos is recognized for their unique lifestyle and aggressive breeding strategy [51-53]. To describe the Cuckoo Search (CS) algorithm, the following three idealized rules are used:

a) Each cuckoo lays one egg at a time and deposits it in a randomly selected nest;

b) The best nests, which contain high-quality eggs, are passed on to the next generation;

c) The number of available host nests is fixed, and an egg laid by a cuckoo is discovered by the host bird with a probability within the interval [0, 1]. A more detailed description of the CS algorithm can be found in Rajabioun [55]. Additionally, Figure 4 illustrates the flowchart of the CS algorithm. In this study, the kernel parameter of the Gaussian kernel function RBF $\left(K_{RBF}\left(x_i, x_j = exp\left(-\frac{\|x_i - x_j\|^2}{\gamma^2}\right)\right)\right)$ is selected by the CS algorithm.

3.6. RVR Optimized by HS and CS Algorithms

In the Relevant Vector Regression (RVR) method, the Harmony Search (HS) and Cuckoo Search (CS) algorithms serve as optimizers for tuning the RVR hyperparameters. Typically, RVR is integrated with these algorithms such that the predicted outputs from RVR are used to evaluate the fitting function. The optimal hyperparameter values are determined after a predefined maximum number of iterations. In this study, the objective function is defined by the Root Mean Square Error (RMSE), where a lower RMSE indicates higher estimation accuracy. The optimization process of the RVR hyperparameters using the HS and CS algorithms is illustrated in Figure 5.

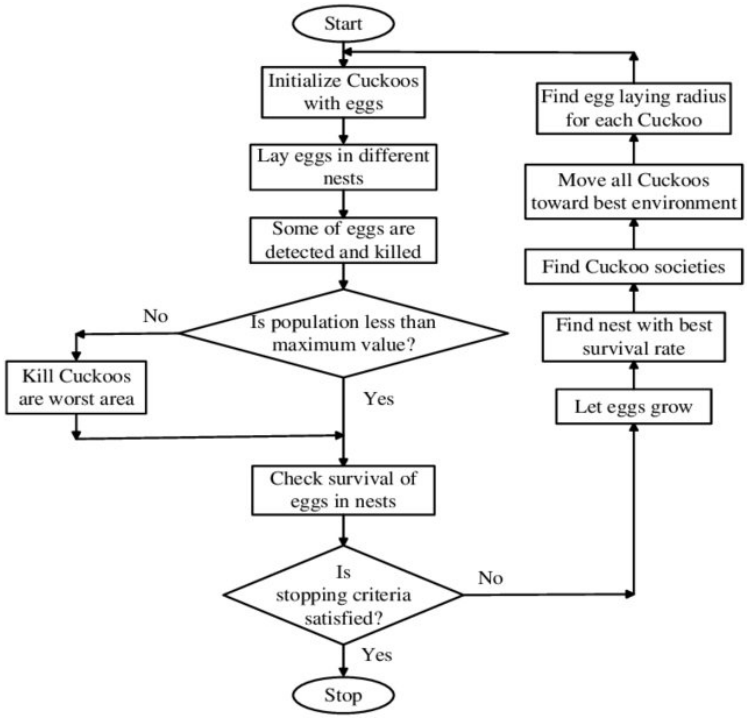

Figure 4. A flow chart of the CS algorithm [54]

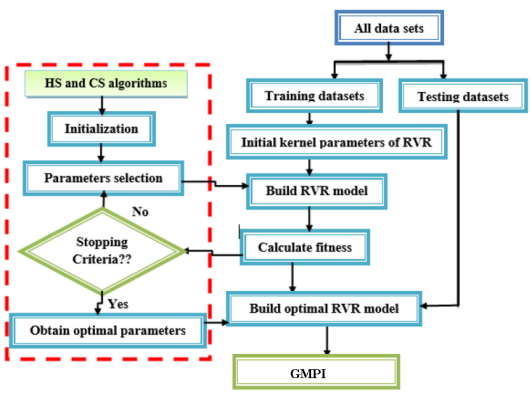

Figure 5. A flowchart of the RVR-CS and RVR-HS models for forecasting GMPI

4. Forecasting GMPI using RVR-HS and RVR-CS models

To predict the Geochemical Mineralization Probability Index (GMPI), it is essential to determine all relevant parameters, as the RVR-HS and RVR-CS models rely solely on the provided data and do not have prior knowledge of the target variable. The following sections detail the input and output parameters, as well as the procedure for predicting GMPI using the RVR-HS and RVR-CS models.

4.1. Database information

The primary objective of this study is to apply the aforementioned methodology to the prediction of the Geochemical Mineralization Probability Index (GMPI). The dataset employed to establish the relationship between input and output variables was collected from stream sediment samples. The complete dataset consists of 400 cases, of which 320 data points (80%) were used for model training, while the remaining 80 data points (20%) were reserved for evaluating model performance. The partial dataset includes concentrations of Zn, Pb, As, Cd, along with the GMPI values.

4.2. Performance Criterion

To assess accuracy, the difference between the model output and the actual output is considered the error, which is expressed in two ways: the Mean Squared Error (MSE) and the squared correlation coefficient (R^2) [55]. Let t_k be the actual value, $\hat{t}_k$ the predicted value for the kth observation, and n the number of observations. The MSE and R^2 are defined as follows:

$$\mathrm{MSE} = \frac{1}{n}\sum_{k=1}^{n}(t_k - \hat{t}_k)^2 \tag{16}$$

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(t_k - \hat{t}_k)^2}{\sum_{k=1}^{n} t_k^2 - \frac{\sum_{i=1}^{n} \hat{t}_k{}^2}{n}} \tag{17}$$

4.3. Algorithm Configuration

In the proposed RVR-HS and RVR-CS methods, fine-tuning of multiple parameters is necessary. For the Cuckoo Search (CS) algorithm, the maximum number of iterations is set to 50, the number of nests to 8, and the population size (number of cuckoos) to 25. The alien population detection rate is fixed at 25%, the coordination memory consideration rate at 0.4, the step adjustment rate at 0.1, the number of new coordinations at 15, and the ferret width reduction ratio at 0.99. To optimize the performance of the Relevant Vector Regression (RVR) model, parameters are adjusted differently during each iteration, with the best-performing configuration selected for this study. Similarly, in the Harmony Search (HS) algorithm, the maximum number of iterations is set to 50, the number of eggs/solutions to 0.75, and the beta value to 3.2.

5. Results and Discussion

5.1. Stepwise factor analysis

A three-stage factor analysis was conducted on the geochemical data of stream sediments from Khomain Dehno. In the first stage, four factors with eigenvalues greater than 1 were identified (Table 1). For this study, a factor loading threshold of 0.7 was applied to separate variables. When it performs factor analysis to select elements (variables), they usually consider several key criteria to ensure meaningful and reliable results. While the exact criteria can vary depending on the study, here are the common ones:

- Correlation among variables: Variables should be reasonably correlated with each other to justify factor analysis. Often, a correlation matrix or Bartlett's test of sphericity is used to check if the variables are sufficiently correlated.
- Sample size and data suitability: The dataset should have enough observations relative to the number of variables (commonly a ratio like 5–10 observations per variable. The Kaiser measure is often reported to assess sampling adequacy.
- Communality values: Variables should have adequate communalities (variance explained by the factors), usually above 0.3 or 0.4, meaning they share enough variance with others.
- Factor loadings: Elements are selected based on their loadings (correlation between the variable and the factor). Typically, loadings above a threshold (e.g., 0.4 or 0.5) are retained.
- Interpretability and theoretical justification: Factors and selected variables should make sense conceptually and theoretically.
- Eigenvalues and variance explained: Factors with eigenvalues greater than 1 (Kaiser Criterion) are usually retained and the cumulative variance explained by the factors is also considered [56].

Elements such as Ba and Ag, which exhibited factor loadings below this threshold, were excluded from the analysis. Subsequently, the next stage of factor analysis was performed on the remaining variables.

In the second stage, three factors were identified, and Sb was excluded from the analysis due to its factor loading being below the 0.7 threshold (Table 2). In the third stage, three factors were again identified (Table 3, Figure 6). The first factor comprised Zn, Pb, As, and Cd and its variance% is 41.63%: the second factor included Mn, Mo, and Zr and its variance% is 21.86% and the third factor consisted of Fe, Cu, and Ti and its variance is 7.79%. The cumulative variance of three factors is 81%. As no additional elements could be assigned to any of the existing factors at this stage, the factor analysis was concluded. Performing factor analysis in a stepwise manner enhances the likelihood of successful exploration.

The first factor is of greater importance compared to the other two factors. Stream sediments containing Zn and Pb play a crucial role in delineating the extent of Zn-Pb mineralization in the area. Accordingly, the stepwise factor analysis employed for anomaly detection in stream sediments have proven valuable in defining the extent of Zn and Pb anomalies. These methods have also helped in selecting the most appropriate approach for identifying multi-element geochemical anomalies in the region. The second factor highlights the abundance of Zn and Pb, which corresponds to Zn-Pb mineralization in the Cretaceous limestone units. These minerals predominantly occur in the form of oxide minerals such as smithsonite and cerussite**.** In some Zn-Pb oxide mines in the region, Pb sulfide occurs as galena at the surface, although Zn is more dominant. In the second factor**,** Mn, Mo, and Zr are associated with granite and are strongly influenced by faulting. Manganese mineralization, occurring as pyrolusite and iron oxides, is scattered along fault zones. The third factor, including Fe, Cu, and Ti, is associated with granite-rich areas. The faulted granitic intrusive body hosts Cu oxide mineralization, which is characterized by malachite and chalcopyrite mineralization. Significant iron

oxide mineralization is observed in the form of veins and weathered rocks stained with limonite and goethite, particularly in areas where tunnels have been excavated into the granite. Additionally, a half-meter-thick vein of malachite was identified within the granite inside one of the tunnels. Ti mineralization likely originates from rutile present in the granite and stream sediments.

Table 1. Factor loadings of the first step for the stream sediments geochemical data in Khomain Dehno

Variables	Factor1	Factor2	Factor3	Factor4
Zn	**0.88**	0.11	**-0.29**	0.03
Pb	**0.89**	-0.02	**0.05**	0.08
As	**0.81**	0.31	**0.29**	0.06
Ba	0.61	0.62	-0.17	-0.05
Cd	**0.83**	0.05	**0.20**	0.13
Fe	0.03	**0.75**	0.48	0.14
Mn	0.40	**0.78**	-0.06	0.21
Mo	0.25	**0.77**	0.10	-0.03
Cu	0.14	0.23	**0.91**	0.03
Ag	0.55	-0.01	0.37	-0.31
Sb	0.10	0.08	0.05	**0.93**
Ti	-0.04	**0.73**	0.59	0.06
Zr	-0.16	**0.86**	0.14	-0.01
Eigenvalue	5.38	2.84	1.31	1.01
Variance%	41.63	21.86	10.07	7.79
Cumulative variance%	41.63	63.22	73.29	81

Table 2. Factor loadings of the second step for the geochemical data of stream sediments

Variables	Factor1	Factor2	Factor3
Zn	**0.85**	0.28	-0.25
Pb	**0.93**	0.02	0.02
As	**0.80**	0.29	0.33
Cd	**0.84**	0.06	0.16
Fe	0.03	0.51	**0.77**
Mn	0.26	۰,۷۷	0.24
Mo	0.17	**0.77**	0.31
Cu	0.22	-0.07	0.91
Sb	0.15	0.39	-0.06
Ti	-0.02	0.40	0.86
Zr	-0.24	**0.74**	0.44
Variance%	44.77	21.72	16.35
Cumulative variance%	44.77	66.49	82.84

Table 3. Factor loadings of the third step for the geochemical data of stream sediments

Variables	Factor1	Factor2	Factor3
Zn	**0.88**	0.17	-0.28
Pb	**0.92**	-0.03	0.06
As	**0.82**	0.30	0.29
Cd	**0.83**	0.01	0.21
Fe	0.05	0.57	**0.71**
Mn	0.42	**0.78**	0.09
Mo	0.23	**0.83**	0.12
Cu	0.18	0.09	**0.93**
Ti	-0.02	0.58	**0.75**
Zr	-0.19	**0.85**	0.25
Variance%	46.19	22.72	16.42
Cumulative variance%	46.19	68.91	85.33

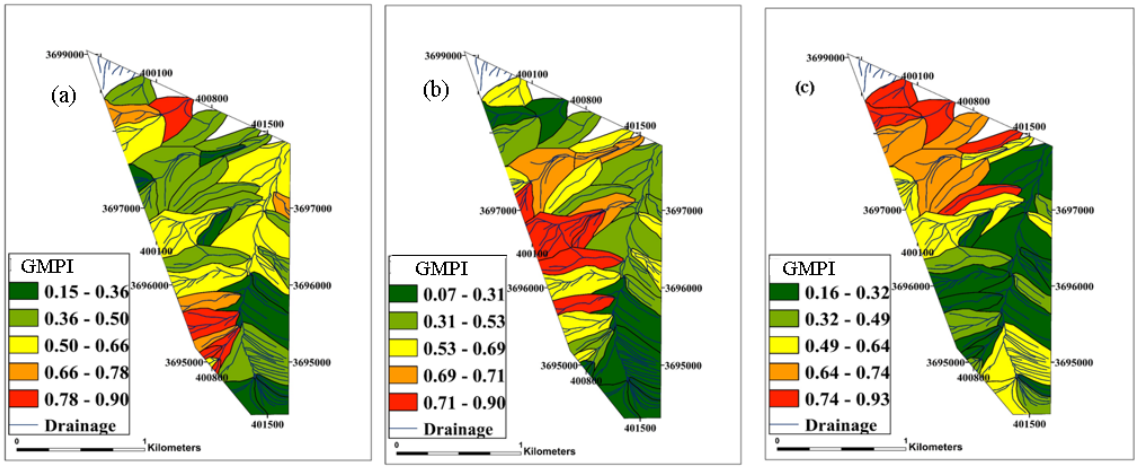

Figure 6. Scatter map of GMPI (a: Factor 1; b: Factor 2; c: Factor 3)

5.2. Forecasting Model

In this study, the RVR-HS and RVR-CS models were developed in the MATLAB environment to predict the Geochemical Mineralization Probability Index (GMPI) based on the available data. The complete dataset, consisting of 400 data points, was randomly divided into two subsets: 80% for training the models and 20% for testing to evaluate their reliability. Considering that Zn and Pb mineralization in the region occurs predominantly in local limestones, the second factor comprising Zn, Pb, As, and Cd was selected as the basis for GMPI prediction. Accordingly, these elements served as input variables, while GMPI was designated as the output variable in the models.

In data-driven system modeling, several preprocessing steps are typically performed prior to analysis to eliminate outliers, address missing data, and remove inappropriate entries. These steps ensure that the raw data extracted from the database are suitable for modeling. To facilitate the training process and improve prediction accuracy, all data samples were normalized using a linear mapping function to scale values within the interval [0, 1]. Following model development, the relationship between the estimated GMPI values generated by the RVR-HS and RVR-CS models and the measured values during both the training and testing phases is illustrated in Figures 7 through 10. As shown in these figures, the RVR-HS model results exhibit strong agreement with the actual data, indicating good predictive accuracy.

5.3. Results of Cocoa Search (CS) and Harmony Search (HS) algorithms for GMPI Prediction

The values of the coefficient of determination (R^2) and Mean Squared Error (MSE) indices for the prediction models developed using the Cuckoo Search (CS) and Harmony Search (HS) algorithms during both the training and testing phases are presented in Tables 4 and 5. The measured and predicted GMPI values for the training and testing datasets generated by the CS and HS algorithms are illustrated in Figures 9 through 11, respectively. According to Table 4, the nonlinear relationships predicted by the CS and HS algorithms demonstrate high accuracy in estimating GMPI. Furthermore, the predicted GMPI values closely match the measured values across all 400 data samples, with 320 samples allocated for training and 80 for testing. This indicates that the developed models exhibit very low error and effectively predict actual GMPI values. Therefore, the relationships derived from these algorithms can be reliably applied to estimate GMPI in other case studies.In this study, the Geochemical Mineralization Probability Index (GMPI) from stream sediments was predicted using the CS and HS algorithms, selected for their simplicity and efficiency, which reduce computational time [30, 57]. Both algorithms are designed to model nonlinear relationships between input variables (X) and output variables (Y). Their advantages include: (1) the ability to predict Y based on X, (2) testing hypotheses about the relationship between X and Y, and (3) adaptability in predicting spatial patterns of Y [23].

The R^2 measure quantifies the proportion of variance in the dependent variable explained by the

model, with values ranging from 0 to 1. Higher R^2 values generally indicate better model fit; however, excessively high R^2 values may result in overfitting, especially in complex models, thereby reducing generalizability. Nonlinear models are often preferred when variable relationships are inherently nonlinear or when the data structure demands greater model complexity. These models tend to be easier to interpret and are well-suited for predicting complex biological and geochemical systems [26]. Additionally, nonlinear models can be effectively combined with Relevant Vector Regression (RVR), wherein residuals from the nonlinear model are further modeled using RVR. This hybrid approach enhances predictive performance while maintaining interpretability [28, 58].

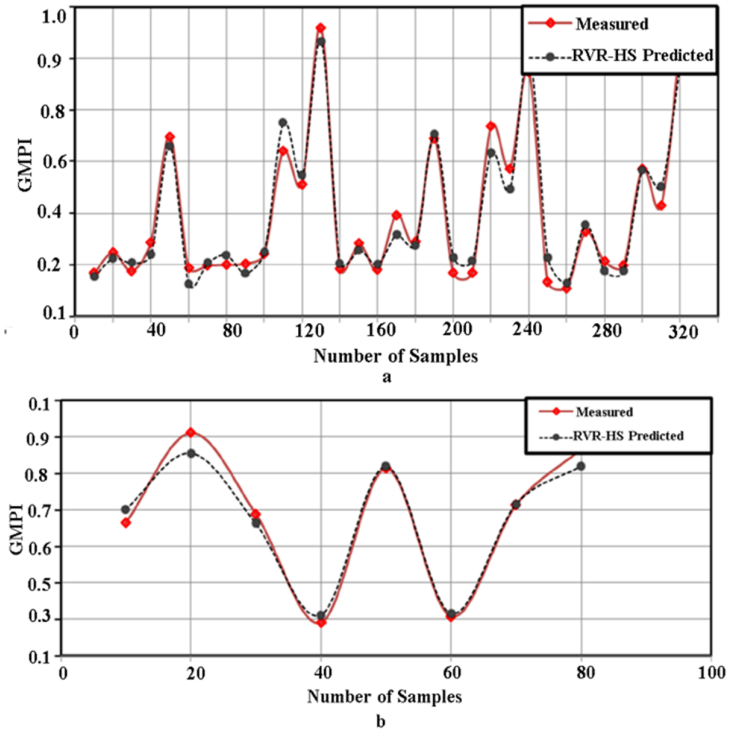

Figure7. Correlation between the measured and estimated using the RVR-HS model for a) training datasets b) testing datasets

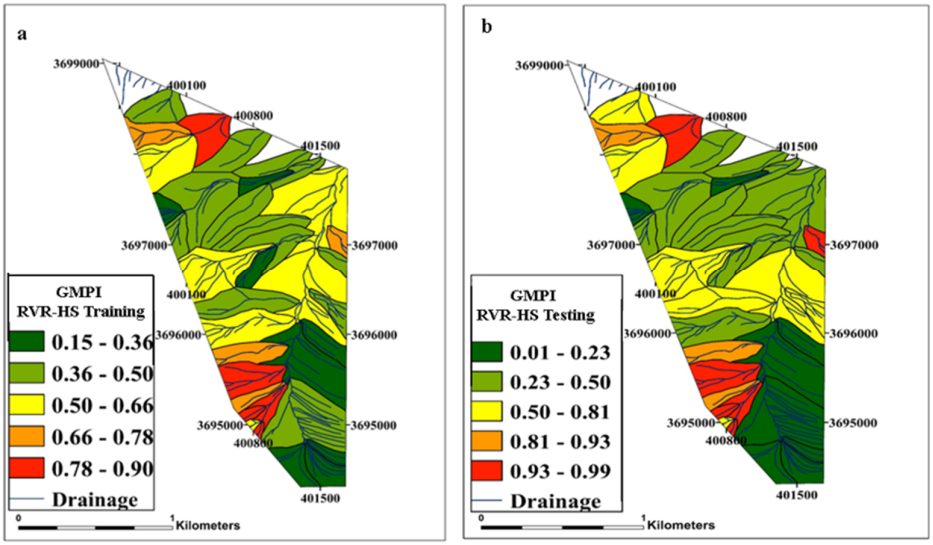

Figure 8. The measured and estimated GMPI using the RVR-HS model for a) training datasets b) testing datasets.

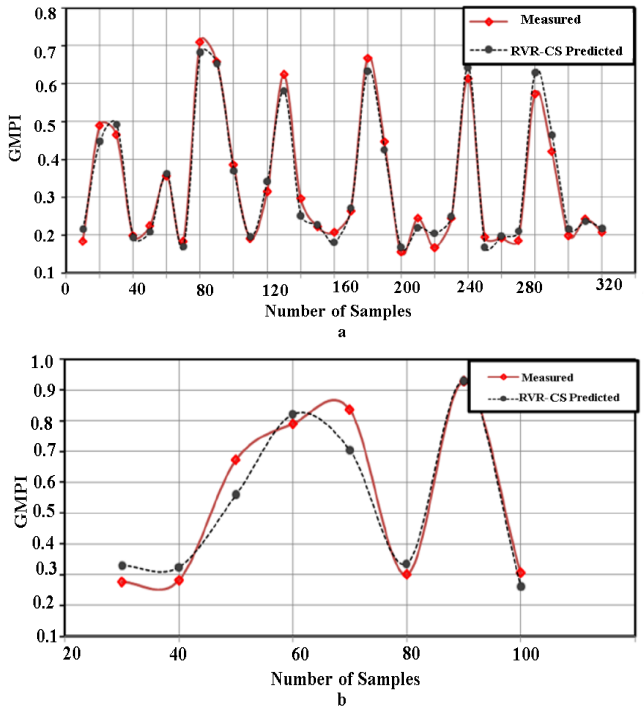

Figure 9. Correlation between the measured and estimated the RVR-CS model for a) training datasets b) testing datasets

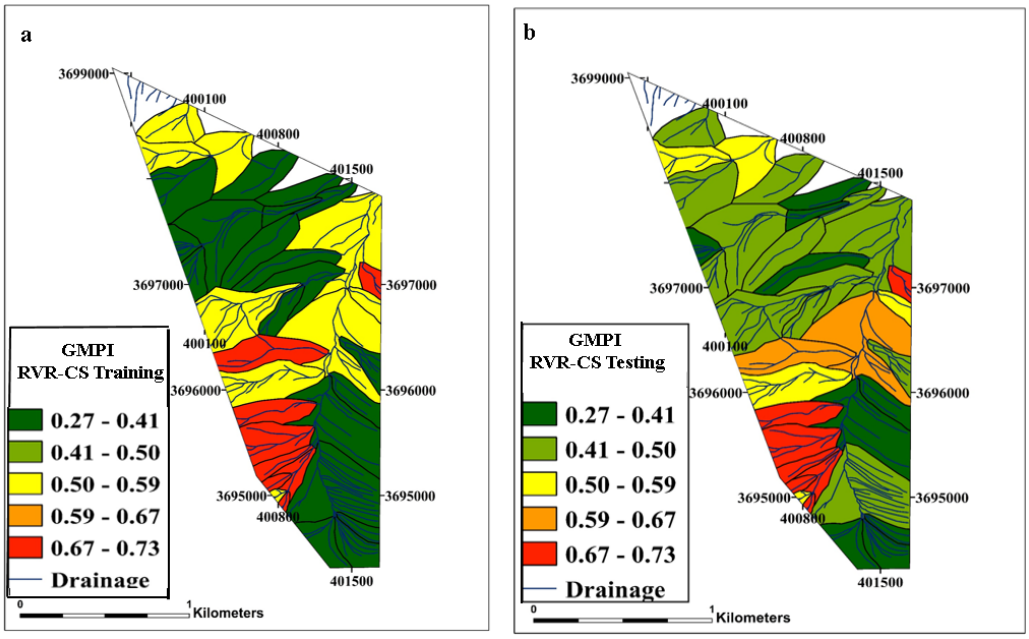

Figure 10. The measured and estimated GMPI using the RVR-CS model for a) training datasets b) testing datasets

The high R^2 values observed in this study for the geochemical parameters may reflect a nonlinear relationship but also raise the possibility of overfitting, suggesting that the model may be overly tailored to this specific dataset. While this may be advantageous for developing region-specific GMPI management strategies, it necessitates further validation. Future analyses might include alternative adjustment techniques or modeling approaches to address this concern. Moreover, the CS and HS methods are particularly advantageous for datasets characterized by nonlinear relationships, small sample sizes, and

low dimensionality, as supported by previous studies [19, 22].

The robustness of the CS and HS methods was evaluated using the R^2 metric. The geochemical variables Zn, Pb, As, and Cd, derived from stream sediments of Khomain Dehno, were incorporated as input variables in the modeling process. The RVR structure employed included multiple input layers corresponding to these geochemical variables and a single output layer for GMPI prediction. Performance analysis of the RVR-HS and RVR-CS models is summarized in Table 4. The RVR-HS model emerged as the most accurate predictive model, achieving an R^2 of 0.99256 and an MSE of 0.0031455. Furthermore, as shown in Figure 11 and Table 5, the RVR-HS model consistently exhibited lower MSE values compared to other models across nearly all cases, indicating superior predictive capability. It should be noted that this model was developed using a limited dataset and therefore may not be generalizable to the entire region; however, its predictive performance can potentially be improved with the inclusion of additional data.

The Relevant Vector Regression modeling framework demonstrated strong predictive performance in this study. Figures 7 and 8 illustrate the nonlinear models for GMPI estimation using the CS and HS algorithms, respectively. The results indicate that the HS-based methods provide particularly high accuracy, as reflected in the overall data relationships depicted in these figures.

Table 4. Performance analysis of the RVR-HS and RVRCS models for forecasting GMPI

	Description	MSE	R^2
RVR-HS model	Training	0.0023446	0.9886
	Testing	0.0031203	0.9903
RVR-CS model	Training	0.0023455	0.9821
	Testing	0.0038432	0.9542

Table 5. Comparing the performance of RVR-HS model in forecasting GMPI with different fractions of training and testing data

Training/Testing (%)	Model	MSE (Train)	MSE (Test)	R^2 (Train)	R^2 (Test)
90/10	RVR-HS	0.0024567	0.0044612	0.98032	0.99222
80/20	**RVR-HS**	**0.0022038**	**0.0031455**	**0.98194**	**0.99256**
70/30	RVR-HS	0.0022325	0.0057420	0.97645	0.98255
60/40	RVR-HS	0.0056345	0.0236790	0.95333	0.96344

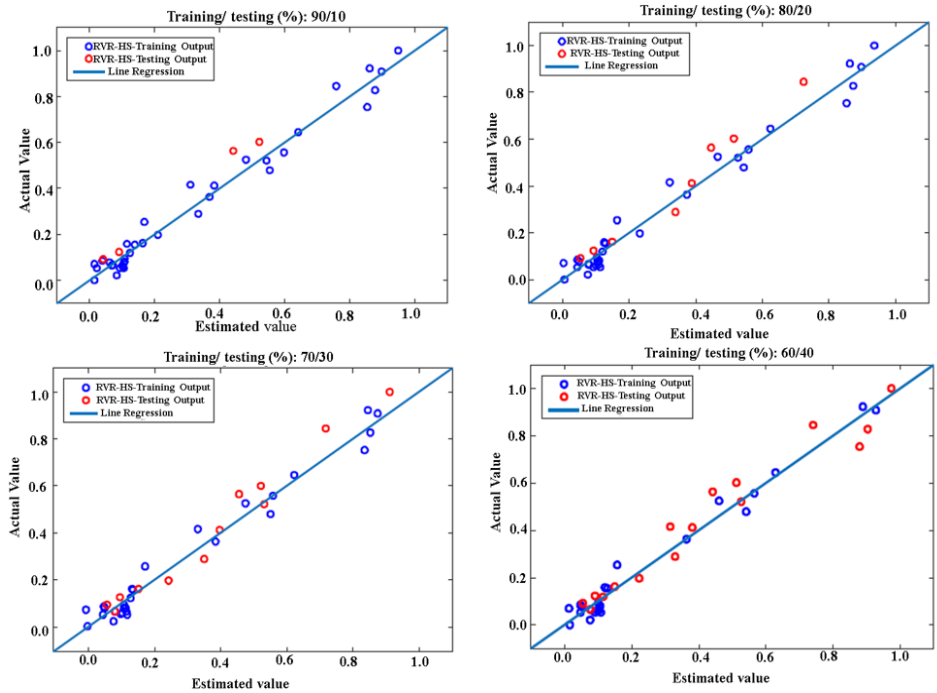

Figure 11. Comparing performance of the RVR-HS model with different fractions of training and testing data

5.4. Comparative Performance of GMPI Prediction Models

The results of this study demonstrate that Relevant Vector Regression (RVR) combined with the Harmony Search (HS) algorithm provides an accurate and reliable method for estimating Geochemical Mineralization Probability Index (GMPI) values. The R^2 values for the HS model ranged from 0.9886 to 0.9903, indicating excellent model fit. Moreover, the commonly used 80/20 split of training and testing data proved to be an effective approach for GMPI estimation using the RVR algorithm [59]. Notably, both the CS and HS models outperformed the standalone RVR model in predicting GMPI values. Among these, the GMPI values estimated by the HS model exhibited a high degree of agreement with observed values, demonstrating superior predictive performance (Tables 4 and 5).

The findings suggest that the Harmony Search model is more suitable than the Cuckoo Search (CS) model for GMPI prediction. The HS algorithm is recognized for its simplicity and computational efficiency, which have contributed to its success in various real-world applications [60]. This conclusion aligns with existing literature that supports the advantages of the HS algorithm in similar modeling tasks [50]. Additionally, the low Mean Squared Error (MSE) values associated with the HS model further confirm its superior predictive capability, as smaller MSE values correspond to better model performance [61].

The study also evaluated the performance of the RVR-HS model across different fractions of training and testing datasets, a common practice in the literature to assess model robustness [62-64]. The 80/20 training/testing split consistently demonstrated the effectiveness and reliability of the HS algorithm in accurately predicting GMPI values (Table 5, Figure 10). The lowest performance was observed with a 40/60 split, where the R^2 and MSE values were 0.96344 and 0.0236790, respectively (Table 5), indicating that reduced training data negatively impacts model accuracy.

6. Conclusions

The Geochemical Mineralization Probability Index (GMPI) is a critical parameter for identifying geochemical anomalies and assessing the suitability of geochemical data in various stream sediment applications. However, direct determination of GMPI values is often time-consuming and labor-intensive. As a result, indirect predictive methods are commonly employed to estimate GMPI. In this study, we propose a novel approach utilizing Relevant Vector Regression (RVR) optimized by Harmony Search (HS) and Cuckoo Search (CS) algorithms for GMPI prediction. The HS and CS algorithms serve as optimization tools to determine the optimal values of user-defined parameters within the RVR framework. The integration of these optimization techniques has significantly enhanced the predictive performance of the RVR model. The following results were obtained:

- The RVR-HS model, with R^2=0.99256 and MSE = 0.0031455, proves to be a reliable system modeling technique for predicting GMPI, offering excellent accuracy and robustness.
- The use of evolutionary algorithms significantly improves the speed and accuracy of determining the optimal values of the geochemical parameters.
- It is important to note that the presented model was developed based on a limited dataset and cannot be generalized to all stream sediment geochemical data. However, this model has the potential to be further developed with the addition of more data.

References

[1]. Ghasemzadeh, S., Maghsoudi, A., & Yousefi, M. (2021). Identifying porphyry-Cu geochemical footprints using local neighborhood statistics in Baft area, Iran. *Frontier Earth Sciences*, 15, 106–120.

[2]. Saremi M, Yousefi S, & Yousefi M. (2024). Combination of Geochemical and Structural Data to Determine Exploration Target of Copper Hydrothermal Deposits in Feizabad District. *Journal of Mining and Environment*, 15(3), 1089-1101.

[3]. Saremi, M., Maghsoudi, A., Ghezelbash, R., Yousefi, M., & Hezarkhani, A. (2025). Targeting of Porphyry Copper Mineralization Using a Continuous-based Logistic Function Approach in the Varzaghan District, North of Urumieh-Dokhtar Magmatic Arc. *Journal of Mining and Environment,* 16 (4), 1417-1435.

[4]. Seifpanahi Shabani K, & Vaezian A. (2017). Magnetic Nano mineral and acid mine drainage interaction: An experimental study. *Journal of Mining and Environment*, 8(3), 447-453.

[5]. Yousefi, M., Kamkar-Rouhani, A., & Carranza, E.J.M. (2012). Geochemical mineralization probability index (GMPI): a new approach to generate enhanced stream sediment geochemical evidential map for increasing probability of success in mineral potential mapping, *Journal of Geochemical Exploration*, 115, 24-35.

[6].Yousefi, M., Carranza, E.J.M., Kreuzer, O.P., Nykänen, V., Hronsky, J.M. A., & Mihalasky, M.J. (2021). Data analysis methods for prospectivity modelling as applied to mineral exploration targeting: state-of-the-art and outlook. *Journal of Geochemical Exploration*, 229, 106839.

[7].Seyedrahimi-Niaraq. M., & Mahdiyanfar, H. (2025). Analysis of closed and open numerical systems of geochemical data in spatial statistics environment in order to separate anomalous areas. *Science Report*, 1, 15(1), 21254.

[8].Wang, J., Zhou, Y., & Xiao, F. (2020). Identification of multi-element geochemical anomalies using unsupervised machine learning algorithms: a case study from Ag–Pb–Zn deposits in north-western Zhejiang, China. *Applied Geochemistry*, 120, 104679.

[9]. Mostafaei, K., Kianpour, M.N., Yousefi, M., & Saleki, M. (2025). Improving the results of the fractal model of geochemical Mineralization Probability Index Using the Gray Wolf Algorithm on the Stream Sediments Data of Sarduiyeh-Baft Area. *Journal of Mining and Environment*, 16(4), 1165-1178.

[10]. Grunsky E., Greenacre M., & Kjarsgaard B. (2024). GeoCoDA: Recognizing and validating structural processes in geochemical data. A workflow on compositional data analysis in lithogeochemistry. *Applied Computer Geoscience*, 22, 100149.

[11].Ghadimi F., Ghomi M., & Aref Sedigh M. (2016). Identification of Ti- anomaly in stream sediment geochemistry using of stepwise factor analysis and multifractal model in Delijan district, Iran. *International Journal of Mining and Geo-Engineering* 50(1), 77-95.

[12]. Keykha Hoseinpoor, M., & Aryafar, A. (2014). The Use of Robust Factor Analysis of Compositional Geochemical Data for the Recognition of the Target Area in Khusf 1:100000 Sheet, South Khorasan, Iran. *International Journal of Mining and Geo-Engineering*, 48 (2), 191-199.

[13]. Shahrestani, S., Mokhtari, A.R., & Alipour-Asll, M. (2019). Assessment of Estimated Bedrock and Stream Sediment Geochemical Backgrounds in Catchment Basin Analysis. *National Resources Research*, 28, 1071–1087.

[14].Jodeiri Shokri, B., Dehghani, H., Shamsi, R., & Doulati Ardejani, F. (2020). Prediction of acid mine drainage generation potential of a copper mine tailings using gene expression programming-a case study. *Journal of Mining and Environment*, 11(4), 1127-1140.

[15]. Hadadi, F., Jodeiri Shokri, B., Zare Naghadehi, M., & Doulati Ardejani, F. (2021). Probabilistic prediction of acid mine drainage generation risk based on pyrite oxidation process in coal washery rejects-A Case Study. *Journal of Mining and Environment*, 12(1), 127-137.

[16]. Mahdiyanfar, H., & Seyedrahimi-Niaraq, M. (2022). Improvement of geochemical prospectivity mapping using power spectrum–area fractal modelling of the multi-element mineralization factor (SAF-MF). *Geochemistry: Exploration, Environment, Analysis*, 22.

[17]. Ayari, J., Barbieri, M., Barhoumi, B., Belkhiria, W., Braham, A., Dhaha, F., & Charef, A. (2022). A regional-scale geochemical survey of stream sediment samples in Nappe zone, northern Tunisia: implications for mineral exploration. *Journal of Geochemical Exploration*, 235, 106956.

[18]. Bigdeli, A., Maghsoudi, A., & Ghezelbash, R. (2022). Application of self-organizing map (SOM) and K-means clustering algorithms for portraying geochemical anomaly patterns in Moalleman district, NE Iran. *Journal of Geochemical Exploration*, 233, 106923.

[19]. Hajihosseinlou, M., Maghsoudi, A., & Ghezelbash, R. (2024). Intelligent mapping of geochemical anomalies: adaptation of DBSCAN and mean-shift clustering approaches. *Journal of Geochemical Exploration*, 258, 107393.

[20]. Saremi, M., Maghsoudi, A., Hajihosseinlou, M., & Ghezelbash, R. (2024). Detection of multivariate porphyry Cu signatures and anomaly enhancement: Incorporation of SFA, GMPI, and Grey Wolf Optimization. *Geochemistry*, 84 (4), 126207.

[21]. Hajihosseinlou, M., Maghsoudi, A., & Ghezelbash, R. (2023). A Novel Scheme for Mapping of MVT-Type Pb–Zn Prospectivity: Light GBM, a Highly Efficient Gradient Boosting Decision Tree Machine Learning Algorithm. *National Resources Research*, 32, 2417–2438.

[22]. Mirzabozorg, S.A.A.S., Abedi, M., & Yousefi, M. (2024). Enhancing training performance of convolutional neural network algorithm through an autoencoder-based unsupervised labeling framework for mineral exploration targeting. *Geochemistry*, 84 (4), 126197.

[23]. Mirzabozorg, S.A.A.S., & Abedi, M. (2023). Recognition of mineralization-related anomaly patterns through an autoencoder neural network for mineral exploration targeting. *Applied Geochemistry*, 158, 105807.

[24]. Saremi, M., S Agha Seyyed Mirzabozorg, S.A., Maghsoudi, A., Abedi, M., & Dehghan Niri, R. (2025). Cu-bearing signatures from multi-element geochemical data, a correct strategy to implement a Convolutional Autoencoder Algorithm. International Journal of Mining and Geo-Engineering, 59 (1), 1-10.

[25]. Esmaeiloghli, S., Lima, A., & Sadeghi, B. (2024). Lithium exploration targeting through robust variable selection and deep anomaly detection: an integrated application of sparse principal component analysis and stacked autoencoders. *Geochemistry*, 84 (4), 126111.

[26]. Xiong, Y., & Zuo, R. (2020). Recognizing multivariate geochemical anomalies for mineral exploration by combining deep learning and one-class support vector machine. *Computer and Geoscience* 140, 104484.

[27]. Ahsan, M., Celik, T., & Genc, B. (2021). A Comparative Study on Machine Learning Algorithms for Geochemical Prediction Using Sentinel-2 Reflectance Spectroscopy. *Journal of Mining and Environment*, 12 (4), 987-1001.

[28]. Pambudi, E.A., Badharudin, A.Y., & Wicaksono, A.P. (2021). Enhanced k-means by using grey wolf optimizer for brain MRI segmentation, ICTACT. *Journal of Soft Computer*, 11(3), 2353-2358.

[29]. Daviran, M., Ghezelbash, R., & Maghsoudi, A. (2024). GWOKM: a novel hybrid optimization algorithm for geochemical anomaly detection based on Grey wolf optimizer and K-means clustering. *Geochemistry*, 84 (1), 126036.

[30]. Mostafaei, K., Yousef, Kreuzer, O., & Kianpour, M.N. (2025). Simulation-based mineral prospectivity modeling and Gray Wolf optimization algorithm for delimiting exploration targets. Ore Geology Review, 177, 106458.

[31]. Lui, T.C.C., Gregory, D.D., Anderson, M., Lee, W., & Cowling, A.A. (2022). Applying machine learning methods to predict geology using soil sample geochemistry. *Applied Computer Geoscience*, 16, 100094.

[32]. Khavari, M., Ghadimi, F., & Mojeddifar, S. (2020). Application of finite difference method in multifractal techniques to identify Pb–Zn anomalies: a case study from Khomeini, Iran. *Mining, Metallurgy & Exploration*, 37, 1765-1777.

[33].Kaviani, A., Denis Hatzfel, D., Paul, A., Tatar, M., & Priestley, K. (2009). Shear-wave splitting, lithospheric anisotropy, and mantle deformation beneath the Arabia–Eurasia collision zone in Iran. *Earth and Planetary Science Letters*, 286, 371–378.

[34]. Afzal, P., Misagh Mirzaei, M., Yousefi, M., Adib, A., Khalajmasoumi, M., Zarifi, Z., Foster, P., & Bijan Yasreb, A. (2016). Delineation of geochemical anomalies based on stream sediment data utilizing fractal modeling and staged factor analysis. *Journal of Africa Earth Science*, 119, 139-149.

[35]. Afzal, P., Yusefi, M., Mirzaie, M., Ghadiri-Sufi, E., Ghasemzadeh, S.L., & Saein, D.(2019). Delineation of podiform-type chromite mineralization using geochemical mineralization prospectivity index and staged factor analysis in Balvard area (SE Iran). Journal of Mining and Environment, 10(3), 705-715.

[36]. Aryafar, A., Yousefi , S., Khosravi, V., & Khorashadi, M. (2020). Using stepwise factor analysis (SFA) and geochemical mineralization probability index (GMPI) in order to intensify the geochemical anomalies associated with vein-type copper mineralization in Kardgan 1: 100000 sheet, East of Iran. *Journal of Mining Engineering*, 15 (47), 1-13.

[37]. Tipping, M.E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1, 211-244.

[38]. Qin, Y., & Wang, F. (2011). Tunneling-induced ground surface settlement prediction based on relevance vector machine. In: 2011 International Conference on Electric Technology and Civil Engineering (ICETCE).IEEE, pp. 925-927.

[39]. Gholami, R., Moradzadeh, A., Maleki, S., Amiri, S., & Hanachi, J. (2014). Applications of artificial intelligence methods in prediction of permeability in hydrocarbon reservoirs. *Journal of Petrology Science Engineering*, 122, 643-656.

[40]. Lou, J., Jiang, Y., Shen, Q., & Wang, R. (2018). Failure prediction by relevance vector regression with improved quantum-inspired gravitational search. *Journal of Network and Computer Application*, 103,171-177.

[41]. Koohzadi, F., Afzal, P., Jahani, D., & Pourkermani, M. (2021). Geochemical exploration for Li in regional scale utilizing Staged Factor Analysis (SFA) and Spectrum-Area (S-A) fractal model in north central Iran. *Iranian Journal of Earth Sciences*, 13, 299-307.

[42]. Saadati, H., Afzal, P., Torshian, H., & Solgi, A. (2020). Geochemical exploration for Li using Geochemical Mapping Prospectivity Index (GMPI), fractal and Stage Factor Analysis (SFA) in NE Iran. *Geochemistry: Exploration, Environment, Analysis*, 20, 461-472.

[43]. Tipping, M.E. (2000). The relevance vector machine. In: Advances in neural information processing systems. pp 652-658.

[44].Geem, Z.W. (2009). Music-inspired harmony search algorithm: theory and applications, Vol 191.Springer Verlag.

[45]. Mikaeil, R., Ozcelik, Y., Ataei, M., & Shaffiee Haghshenas, S. (2019). Application of harmony search algorithm to evaluate performance of diamond wire saw. *Journal of Mining and Environment*, 10(1), 27-36.

[46].Fattahi, H. (2020). A new method for forecasting uniaxial compressive strength of weak rocks. *Journal of Mining and Environment*, 11(2), 505-515.

[47].Geem, Z.W., Kim, J.H., & Loganathan, G. (2001). A new heuristic optimization algorithm: harmony search. Simulation. 76 (2), 60-68.

[48]. Moh'd Alia, O., & Mandava, R. (2011). The variants of the harmony search algorithm: an overview. *Artificial Intelligence Reviews*, 36 (1), 49-68.

[49]. Yuan, X., Zhao, J., Yang, Y., Wang, Y. (2014). Hybrid parallel chaos optimization algorithm with

harmony search algorithm. *Applied Soft. Computer*, 17, 12-22.

[50]. Jaberipour, M., & Khorram, E. (2010). Two improved harmony search algorithms for solving engineering optimization problems. Communications in Nonlinear Sci. and Numerical Simulat. 15 (11), 3316-3331.

[51]. Yang, X.S., & Deb, S. (2010). Engineering optimisation by cuckoo search. *International Journal of Mathmatical Model Numbering Optimum*, 1 (4), 330-343.

[52]. Valian, E., Mohanna, S., & Tavakoli, S. (2011). Improved cuckoo search algorithm for feedforward neural network training. *International Journal of Artificial Intelligence and Applications*, 2 (3), 36-43. https://www.researchgate.net/publication/266074649

[53]. Yildiz, A.R. (2013). Cuckoo search algorithm for the selection of optimal machining parameters in milling operations. *International Journal of Advance Manufacture Technology*, 64 (1-4), 55-61.

[54]. Rajabioun, R. (2011). Cuckoo optimization algorithm. *Applied Soft Computer*, 11 (8), 5508-5518.

[55]. Fattahi, H., & Moradi, A. (2018) A new approach for estimation of the rock mass deformation modulus: a rock engineering systems-based model. *Bulletin Engineering Geology Environment*, 77 (1), 363-374.

[56]. Kaiser, H.F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3), 187-200.

[57]. Wang, J., Ouyang, H., Zhang, C. Li, S., & Xiang, J. (2023). A novel intelligent global harmony search algorithm based on improved search stability strategy. *Science Report,* 13, 7705.

[58]. Boucher, T.F., Ozanne, M.V., Carmosino, M.L., Dyar, M.D., Mahadevan, S., Breves, S.E., Lepore, K.H., & Clegg, S.M. (2015). A study of machine learning regression methods for major elemental analysis of rocks using laser-induced breakdown spectroscopy. Spectrochimica Acta Part B: Atomic Spectroscopy, 107 1-10.

[59]. Liu, J., Cheng, Q., & Wang, J. (2015). Identification of geochemical factors in regression to mineralization endogenous variables using structural equation modeling. *Journal of Geochemical Exploration*,150, 125-136.

[60]. Nasir, M., Sadollah, A., Grzegorzewski, P., Yoon, J.H., & Geem, Z.W. (2021). Harmony Search Algorithm and Fuzzy Logic Theory: An Extensive Review from Theory to Applications. *Mathematics* 9(21), 2665.

[61]. Jacobson, S.H. (1994). Optimal mean squared error analysis of the harmonic gradient estimators. *Journal of Optimum Theory Application*, 80, 573–590.

[62]. Sabbaghi, H. (2024). Recognition of multi-element geochemical anomalies related to Pb–Zn mineralization with applying graded support vector machine in the Varcheh district of Iran. Model. *Earth System Environment*, 10, 7159–7172.

[63]. Karim, M., Behrang, K., & Jeff, B. (2025). Automatic variogram inference using pre-trained Convolutional Neural Networks. *Applied Computer Geoscienc*, 25. 100219.

[64]. Ahmed, S., Ghosh, K.K., Singh, P.K., Geem, Z.W., & Sarkar, R. (2020). Hybrid of Harmony Search Algorithm and Ring Theory-Based Evolutionary Algorithm for Feature Selection. IEEE Access, 8, 102629-102645.

انجمن مهندسی معدن ایران
نشریه مهندسی معدن و محیط زیست
نشانی نشریه: www.jme.shahroodut.ac.ir
دانشگاه صنعتی شاهرود

تعیین نقشه شاخص احتمال ژئوشیمیایی با استفاده از الگوریتم های هارمونی و کوکو در رسوبات آبراهه ای دهنو خمین، ایران

فاطمه ویسمریدی[۱] و فریدون قدیمی[۲]

۱. دانشکده علوم زمین، مهندسی معدن، دانشگاه صنعتی اراک، ایران
۲. دانشکده علوم زمین، مهندسی معدن، دانشگاه صنعتی اراک، ایران

اطلاعات مقاله

تاریخ ارسال: ۲۰۲۵/۰۷/۲۸
تاریخ داوری: ۲۰۲۵/۰۸/۱۹
تاریخ پذیرش: ۲۰۲۵/۰۹/۰۶

DOI: 10.22044/jme.2025.16587.3243

کلمات کلیدی

رسوبات آبراهه‌ای
تحلیل عاملی مرحله‌ای
رگرسیون بردار وابسته
الگوریتم هارمونی و کوکو
دهنو خمین

چکیده

چهارصد نمونه رسوب آبراهه‌ای از منطقه دهنو خمین جمع آوری و در هر نمونه ۱۳ عنصر تعیین شد. تحلیل عاملی مرحله‌ای با تولید نقشه های ژئوشیمیایی همراه بوده که معرف کانی زایی است. رسوبات آبراهه‌ای برای تعیین شاخص احتمال کانی‌زایی ژئوشیمیایی (GMPI) استفاده شد. در این تحقیق طی تحلیل عاملی مرحله‌ای از رسوبات آبراهه‌ای، سه مرحله تعیین شد. عامل اول شامل عناصر روی، سرب، آرسنیک و کادمیوم دارای ۴۱/۶۳ درصد واریانس بود. عامل دوم با واریانس ۲۱/۸۶ درصد شامل عناصر منگنز، مولیبدن و زیرکانیم و عامل سوم با واریانس ۷/۷۹ درصد شامل عناصر آهن، مس و تیتانیم بود. واریانس تجمعی در این سه تحلیل عاملی ۸۱ درصد تعیین شد. از مدل‌های هوش مصنوعی نظیر رگرسیون بردار وابسته (RVR) و الگوریتم‌های هارمونی و کوکو برای پیش بینی شاخص احتمال کانی‌زایی ژئوشیمیایی استفاده شد. در این مدل‌ها عناصر روی، سرب، آرسنیک و کادمیوم به عنوان داده‌های ورودی و داده‌های شاخص احتمال کانی‌زایی ژئوشیمیایی به عنوان داده‌های خروجی انتخاب شدند. مدل‌های پیش‌بینی با استفاده از روش‌های خطای مربع میانگین (MSE) و ضریب تعیین (R^2) صحت سنجی شدند. مدل رگرسیون بردار وابسته (RVR) و مدل الگوریتم هارمونی با توجه به خطای مربع میانگین (MSE) برابر ۰/۰۰۳۱۴۵۵ و ضریب تعیین (R^2) برابر ۰/۹۹۲۵۶بهترین مدل برای تخمین شاخص احتمال کانی‌زایی ژئوشیمیایی (GMPI) تعیین شدند.

نویسنده مسئول مکاتبات: ghadimi@arakut.ac.ir (ف. قدیمی)