-RESEARCH PAPER-

International Journal of Mining and Geo-Engineering

IJMGE

IJMGE 60-1 (2026) 43-55

DOI: 10.22059/IJMGE.2026.401315.595299

Prediction of Au grade in Carlin type using pathfinder elements by GMDH and MCMC in Zarshuran deposit

Feridon Ghadimi [a] *

[a] *Mining Department, Arak university of Technology, Arak, Iran.*

Article History:

Received: 27 August 2025.

Revised: 22 October 2025.

Accepted: 02 March 2026.

ABSTRACT

Pathfinder elements play a crucial role in the exploration of concealed and deep-seated mineral deposits. Their significance is particularly pronounced in the context of epithermal gold (Au) deposits, where their presence may serve as an indicator of nearby gold mineralization. Among these pathfinder elements, arsenic (As) and antimony (Sb) are considered the most critical for the exploration of epithermal Au systems. This study investigated Au Carlin type in the Zarshuran to highlight the utility of pathfinder elements in gold estimation. The analysis was conducted using the concentrations of 35 elements measured across 108 samples. The mineralization characteristics of the Zarshuran deposit exhibit notable similarities to those of epithermal gold deposits hosted in sedimentary rocks (Carlin-type), thus presenting a suitable exploration model for the northern Takab region. Selection of pathfinder elements was carried out through factor analysis, which revealed a strong positive correlation among Au, As, Cd, Pb, Sb, and Zn. Two predictive approaches were employed to estimate gold content: the Group Method of Data Handling (GMDH) neural network, and the Monte Carlo Markov Chain (MCMC) simulation. Neural network techniques, such as GMDH, are particularly well-suited for modeling datasets with both linear and nonlinear characteristics. In these models, As, Cd, Pb, Sb, and Zn were used as input variables. The predictive performance of the models was assessed using the coefficient of determination (R^2). The GMDH neural network achieved a superior performance with an R^2 value of 0.9483, outperforming the MCMC simulation. Based on these findings, the GMDH neural network is recommended as a robust and reliable method for predicting Au mineralization in other prospective exploration areas.

Keywords: *Monte Carlo Markov chain simulation, GMDH neural network, Au prediction, pathfinder element, Au-Zarshuran.*

1. Introduction

Pathfinder elements serve as critical indicators in mineral exploration due to their close association with key representative elements and their widespread distribution in the form of primary geochemical halos. These elements frequently undergo chemical decomposition, rendering them valuable for detecting concealed and deeply buried mineral deposits. Such hidden and deep deposits, located beneath the Earth's surface, pose significant challenges for identification and exploitation. Consequently, their discovery often necessitates the application of specialized geological exploration techniques and advanced geophysical methods. Approaches such as geochemical analysis, deep seismic surveys, and even sophisticated machine learning models are increasingly employed to accurately delineate these deposits [1–4]. In the context of Carlin-type deposits, prominent pathfinder elements include mercury (Hg), arsenic (As), and antimony (Sb) [5, 6]. Epithermal gold (Au) deposits, particularly those of the Carlin type, have attracted considerable interest from mining enterprises due to their relatively large reserves, suitability for open-pit mining, and straightforward mineralogical characteristics [7, 8].

The Zarshuran gold (Au) deposit is a prime example of Carlin-type mineralization and has been a key focus of exploration over the past two decades. Subsurface geochemical studies demonstrate a zonal elemental distribution within these deposits: Au, arsenic (As), and mercury (Hg) are enriched in the deposit zone; silver (Ag) and zinc (Zn) predominate in the sub-deposit zone; and copper (Cu) is concentrated in the supra-deposit zone [9]. The mineralization features of the Zarshuran deposit make it a valuable model for regional exploration in northern Takab [10–12]. Multivariate statistical methods, such as factor analysis, are particularly effective tools for selecting pathfinder elements from geochemical datasets [13-15]. Elements such as Au and As are commonly identified as critical pathfinder indicators in gold exploration. Factor analysis reduces complex geochemical data into interpretable factors, thereby aiding in the selection of elements most relevant for predicting Au mineralization.

In recent years, machine learning techniques, particularly those based on neural networks have exhibited significant potential in mineral exploration, owing to their ability to model both linear and nonlinear data patterns [16–18]. These methods are especially valuable when the relationships between dependent and independent variables are complex or poorly defined. The Adaptive Artificial Intelligence (AAI) algorithm has been applied in the exploration of metal deposits [19]. In addition, Artificial Neural Networks (ANNs) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) have been employed to predict pyrite oxidation in coal tailings [20, 21]. A recent advancement includes the development of an Explainable Artificial Intelligence (XAI) model designed to predict gold (Au) mineralization, providing enhanced interpretability of the decision-making process [22]. Moreover, techniques such as wavelet neural networks and Monte Carlo simulations have also been utilized in metal ore exploration [23]. An

* Corresponding author. *E-mail address:* ghadimi@arakut.ac.ir (F. Ghadimi).

Improved Artificial Neural Network (IANN) model has been successfully implemented to predict copper (Cu) and gold (Au) mineralization in epithermal gold deposits [24].

The Group Method of Data Handling (GMDH) is a self-organizing algorithm used for modeling and predicting complex systems. Although it is considered an early form of neural network, GMDH differs from conventional neural networks in both structure and learning methodology. GMDH networks automatically determine the optimal model structure and complexity during the training process [25]. The network is constructed incrementally, growing layer by layer, with neurons generated by combining pairs of neurons from the previous layer. These neurons typically employ polynomial transfer functions, often quadratic polynomials. At each layer, the best-performing neurons are selected based on prediction error assessed on validation data. This external criterion helps prevent overfitting and allows the network to be pruned to an appropriate size and complexity. The GMDH algorithm has been successfully applied in various fields, including time series prediction (economic, environmental, and financial data), system identification and modeling, signal processing, forecasting, and nonlinear regression problems.

Monte Carlo Markov Chain (MCMC) simulation is a robust computational method extensively applied in Bayesian statistics, machine learning, physics, and related disciplines. Monte Carlo methods are a class of algorithms that utilize repeated random sampling to obtain numerical results [26]. Specifically, Monte Carlo processes are stochastic in nature, where the subsequent state depends only on the current state, independent of the full history of previous states. MCMC integrates these concepts to generate samples from probability distributions that are otherwise difficult or impossible to sample directly [27]. The applications of MCMC include Bayesian parameter estimation, uncertainty quantification, and sampling from complex or high-dimensional distributions.

In this study, factor analysis was first used to identify key pathfinder elements for Au exploration. The selected pathfinder elements then served as input variables for two predictive models: the Group Method of Data Handling (GMDH) Neural Network and Monte Carlo Markov Chain (MCMC) simulation. These two models were presented, tested, and compared to assess their effectiveness in predicting Au grades. By combining statistical and neural network approaches, this study provides a robust framework to improve Au exploration in epithermal and Carlin type deposits.

2. Materials and methods

2.1. Location and geology of the deposit area

The Zarshuran Au deposit is located in West Azerbaijan Province, Iran, approximately 50 km north of Takab. The deposit is hosted within the Zarshuran shale unit, a stratigraphic sequence composed of limestone, black shale, dolomite, and calcareous sandstone (Figure 1) [28]. Mineralization is controlled by both stratigraphic and tectonic factors. The shale-carbonate unit serves as a favorable host rock (stratigraphic control), while east–west trending faults represent the key structural control. Notably, Au mineralization is concentrated at the intersections of these fault systems.

The ore mineral assemblage includes orpiment, realgar, stibnite, cinnabar, barite, and pyrite. Surrounding the mineralized zones, hydrothermal alteration features include silicification, argillic alteration, dolomitization, and oxidation, primarily affecting the black shales and limestones. Mineralization at the Zarshuran Au deposit occurred in three distinct stages:

(1) Carbonate Removal and Silica Replacement: The initial stage involved the leaching of carbonate components from the host rock and their replacement by silica, marking the onset of hydrothermal activity.

(2) Massive Silica Formation with Argillic Alteration: This represents the most significant mineralizing phase, characterized by extensive silicification and accompanying argillic alteration. This stage played a critical role in the development of the ore body.

(3) Veining and Arsenic Mineral Deposition: The final stage is defined by the formation of mineralized veins and the deposition of arsenic-bearing minerals, primarily orpiment and realgar, along fault zones and fracture networks.

This sequential mineralization pattern is characteristic of Carlin-type Au deposits hosted in sedimentary rocks. At Zarshuran, mineralization was further enhanced by the intrusion of hydrothermal fluids into the Zarshuran unit. These fluids, operating under reducing conditions and within a pressurized water system, facilitated the transport and deposition of ore minerals, particularly within the upper stratigraphic levels of the host sequence.

The Zarshuran gold deposit (155 t Au, average grade: 2.63 g/t), NW Iran, provides a new paradigm for understanding the multicomponent ore-forming processes and metallogeny of gold during the evolution of hydrothermal fluids. It is characterized by auriferous quartz veins and gold coexisting with disseminated Fe-As-S sulfide minerals that are hosted in a sequence of Early Cambrian metasedimentary rocks. The mineralizing fluid system can be described as carbonic-aqueous with low to moderate salinity (3.2–15.1 wt% NaCl equiv.) and medium temperature of 285 to 317 ◦C (early ore-stage) and 255 to 290 ◦C (late ore-stage), which suggests that phase separation was responsible for gold precipitation during late ore-stage As-Hg-Sb sulfide veins [29]. The δ18Ofluid ranges from − 7.8 to +4.2 ‰, and the δ D values for fluid inclusions in mineral range from − 105 to − 65 ‰, suggesting involvement of meteoric water during late- to post ore-stages. Our results indicate that the Zarshuran is a distal disseminated gold deposit formed during southward subduction of the Proto-Tethys oceanic lithosphere beneath the northern margin of Gondwanan terranes through the early Cambrian.

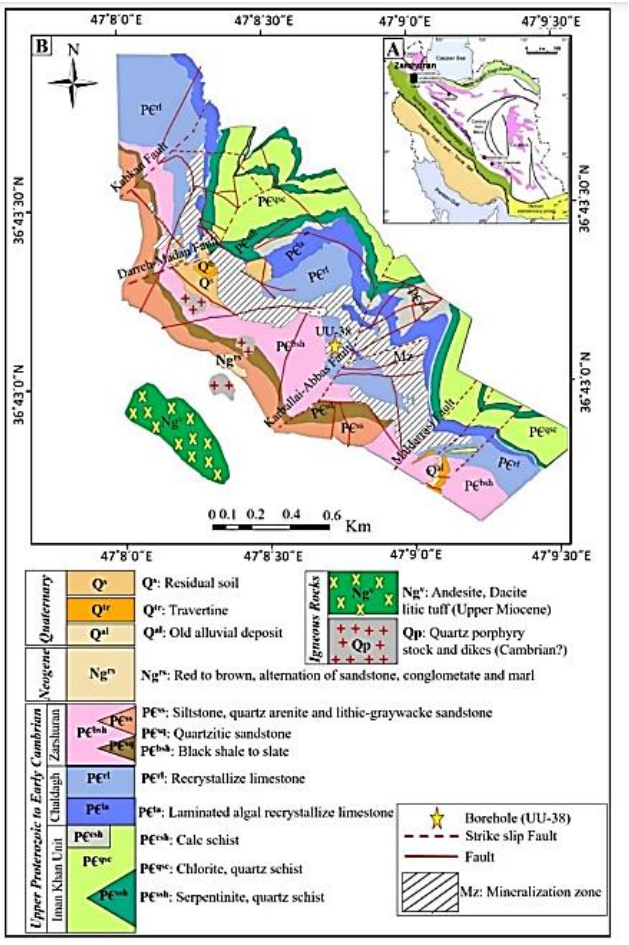

Figure 1. A: Structural map of Iran and location of the Zarshuran deposit, B: Geological map of the Zarshuran deposit [28].

2.2. Research method

A total of 108 rock samples were collected from rock outcrops in the Zarshuran area. The samples were powdered and sent to the reliable Zar-Azma laboratory for analysis of 35 elements, following the standards of the measuring device. Elemental concentrations were determined using four-acid digestion (HF, HCl, $HClO_4$, and HNO_3), followed by analysis with an ICP-MS device. For determining the Au concentration, the Fire Assay method was used. In this process, 30 grams of the sample were melted at 1100 °C, separating Au from the slag and allowing it to be absorbed by a Pb amalgam. During the cupellation stage, Pb was removed, and Au was extracted from the Ag amalgam. Finally, the Ag amalgam was dissolved using aqua regia, and the Au concentration was determined via ICP. The laboratory error percentage was controlled using duplicate elements, with an error margin below 10% for all elements and below 1% for Au. The data were normalized and standardized for multivariate statistical analysis using the isometric log-ratio (ilr) transformation (Equation 1) [30].

The isometric log-ratio (ILR) transformation was selected to preprocess the compositional geochemical data due to its effectiveness in addressing the inherent constraints and challenges associated with compositional datasets. Specifically, compositional data represent parts of a whole and thus reside in a simplex space rather than in conventional Euclidean space. Compared to other transformations, the Box-Cox transformation is useful for stabilizing variance and promoting normality; however, it does not specifically account for the constant-sum constraint characteristic of compositional data. Applying the Box-Cox transformation to compositional data may introduce spurious correlations as a result of closure effects. The centered log-ratio (CLR) transformation, by contrast, maps compositional data into an unconstrained space through the logarithm of each component divided by the geometric mean. However, the CLR transformation results in components that are perfectly collinear, as their values sum to zero. This characteristic can lead to challenges in multivariate analyses and may violate the assumptions underlying certain statistical and machine learning models. The ILR transformation addresses these limitations by generating orthonormal coordinates in real Euclidean space, ensuring that the transformed variables are linearly independent and thus more suitable for standard analytical techniques. This orthogonality improves both model interpretability and numerical stability. Accordingly, the ILR transformation was chosen as the most appropriate method for processing geochemical compositional data, enabling accurate and unbiased modeling while avoiding the statistical issues associated with alternative transformations. Pathfinder elements were identified through correlation coefficient analysis and factor analysis, which revealed significant relationships among the elements and their associations with gold (Au) mineralization.

$$ilr(x) = \sqrt{\frac{1}{2}} ln(\frac{x_i}{x_j}) \quad (1)$$

In the above equation, $\mathbf{x}$ represents the initial value of the data, and $\mathbf{x_j}$ denotes the geometric mean of the data. Two neural network methods were employed to estimate Au concentrations.

2.3. Neural network model

The Group Method of Data Handling (GMDH) neural network was first introduced by Ivakhnenko [31]. GMDH comprises a family of self-organizing algorithms utilized in mathematical modeling and machine learning. It autonomously determines both the structure and parameters of models based on empirical data, often employing polynomial functions to construct feedforward networks of optimal complexity. Renowned for its capability to capture complex, nonlinear relationships, GMDH's inductive approach enables it to identify patterns without relying on strong a priori assumptions. This method is widely applied in the prediction and optimization of metal deposit potential in mineral exploration [32]. To estimate the output (n), it is necessary to train the GMDH neural network using input values (M), which represent the experimental data. The relationship between inputs $X = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{in})$ and outputs $(\hat{y}_i)$ is mathematically expressed in Equation 2 [33].

$$\hat{y}_i = \hat{f}(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}) i = 1,2, \dots, M \quad (2)$$

The neural network must minimize the sum of squared errors (SSE) between the estimated and measured values to achieve accurate predictions (Equation 3).

$$\sum_{i=1}^{M} \left[\hat{f}(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}) - y_i\right]^2 \rightarrow min \quad (3)$$

Equation 2 defines the model function: how input data X are transformed into output predictions$\hat{y}_i$. Equation 3 is the loss function (SSE) that quantifies the error between predicted and actual outputs. The goal during training is to find f that minimizes the SSE.

The relationship between input and output data is obtained according to Equation 4.

$$y = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} a_{ijk} x_i x_j x_k + \dots \quad (4)$$

The quadratic form and two variables are according to Equation 5. The first layer of a GMDH neural network creates neurons for all possible pairs of input variables. Each neuron in the first layer is typically a quadratic polynomial of the form Equation 5

$$\hat{y} = G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \quad (5)$$

a_i Is obtained based on the regression method in Equation 5.

All neurons in the GMDH neural network consist of n input variables, so according to Equation 6 we have

$$\binom{n}{2} = \frac{n(n-1)}{2} \quad (6)$$

This is the binomial coefficient, often read as "n choose 2", and it represents the number of unique pairs of input variables from a total of n.

Neurons in the second layer are constructed according to Equation 7. Second-layer neurons are built from the outputs of the first layer neurons.

Here, x_{ip} and x_{iq} represent outputs of neurons from the previous layer, not original inputs. y_i is the output of the ith neuron in this layer. Each neuron still uses a pair of inputs and applies a polynomial function like before.

$$\{(y_i, x_{ip}, x_{iq}) | i = 1,2, \dots, M \& p, q \in 1,2, \dots, M\} \quad (7)$$

A variant of the function stated in Equation 3 is used for each M triple row.

These equations are described as Equation 8.

$$Aa = Y \quad (8)$$

where A is the equation's unknown coefficients vector.

$$a = \{a_0, a_1, a_2, a_3, a_4, a_5\} \quad (9)$$

$$Y = \{y_1, y_2, y_3, \dots, y_M\}^T \quad (10)$$

The values are in the observation vector (Equation 11, 12).

$$A = \begin{bmatrix} 1 x_{1p} x_{1q} x_{1p} x_{1q} x_{1p}^2 x_{1q}^2 \\ 1 x_{2p} x_{2q} x_{2p} x_{2q} x_{2p}^2 x_{2q}^2 \\ \dots \dots \dots \dots \dots \dots \\ 1 x_{Mp} x_{Mq} x_{Mp} x_{Mq} x_{Mp}^2 x_{Mq}^2 \end{bmatrix} \quad (11)$$

and,

$$a = (A^T A)^{-1} A^T Y \quad (12)$$

The structural model of the GMDH neural network is presented in Figure 2.

In this research, the characteristics of 108 field samples were carefully analyzed. The input dataset includes the elements As, Cd, Pb, Sb, and Zn, while the output dataset corresponds to Au concentrations.

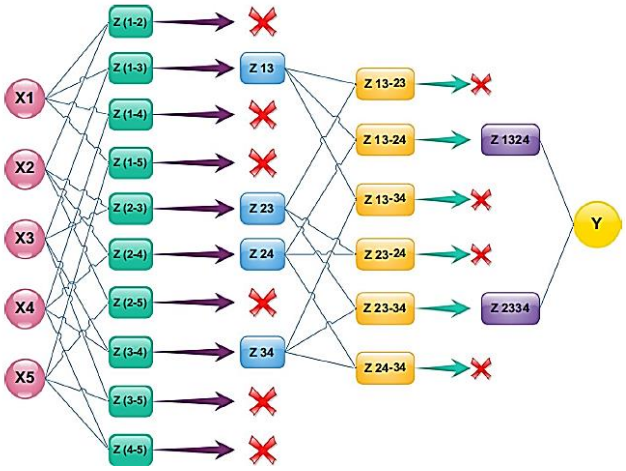

Figure 2. Structural model of GMDH neural network.

2.4. MCMC simulation

In this study, the parameter vector was initially derived from the dataset. The conventional approach assumes that the variables are known, although their precise values remain uncertain and must be estimated. Classical likelihood-based methods typically provide a point estimate of the parameter of interest. In contrast, the Bayesian framework models variables as probability distributions, allowing for a probabilistic representation of uncertainty. Bayesian inference offers several advantages over classical statistical methods in the context of mineral exploration:

(1) Uncertainty Quantification: Rather than producing a single deterministic prediction, the Bayesian model generates a distribution of possible outcomes or confidence intervals, which is essential in experimental settings or when dealing with noisy data.

(2) Regularization: Bayesian methods inherently guard against overfitting, often resulting in improved model generalization

(3) Model Comparison: Bayesian approaches provide principled mechanisms for model comparison and selection based on model evidence or marginal likelihood.

The Bayesian method operates on the premise that the experimenter begins with prior knowledge or preconceived beliefs about the process, which are subsequently updated as new data become available [34]. In the Bayesian framework, data assimilation is typically performed using Markov Chain Monte Carlo (MCMC) simulations.

The posterior density, which often involves high-dimensional integrals, is estimated through Monte Carlo simulation techniques. This process entails constructing a Markov chain with a stationary distribution that approximates the posterior probabilities. Brooks et al. [35] illustrated the application of Bayesian models using the MCMC algorithm, emphasizing its effectiveness in managing complex datasets. In this study, Bayesian analyses were conducted using WinBUGS, a Windows-based interactive software designed to facilitate Bayesian modeling and MCMC simulations. WinBUGS provides robust tools for parameter estimation and uncertainty quantification, making it well-suited for the tasks undertaken in this research.

WinBUGS, which stands for Windows Bayesian inference Using Gibbs sampling, is a specialized software tool developed to perform Bayesian statistical modeling. It employs Markov Chain Monte Carlo (MCMC) methods, particularly Gibbs sampling, to generate samples from complex posterior distributions that are often intractable through analytical computation. How Does WinBUGS Help in Bayesian Analysis?

(1) Model Specification: WinBUGS allows users to define Bayesian models using the flexible BUGS language, which supports the specification of hierarchical models, prior distributions, likelihood functions, and other model components.

(2) Automated Sampling: The software automatically executes MCMC simulations to approximate the posterior distributions of model parameters without requiring manual tuning.

(3) Parameter Estimation: Following the completion of MCMC simulations, WinBUGS provides estimates of posterior means, medians, credible intervals, and other relevant summary statistics.

(4) Uncertainty Quantification: By sampling from the full posterior distribution, WinBUGS delivers a comprehensive probabilistic description of the uncertainty associated with parameter estimates and predictions.

Why is this useful for the present study? When working with experimental data and Group Method of Data Handling (GMDH) neural networks, Bayesian approaches such as those implemented in WinBUGS enable the quantification of uncertainty in parameter estimates, rather than producing only point predictions. Additionally, prior knowledge about parameters can be incorporated into the analysis, thereby enhancing model robustness. WinBUGS simplifies the often complex computations inherent in Bayesian methods, allowing researchers to focus on model development and interpretation.

2.5. Model performance evaluation

To evaluate the performance of the GMDH neural network and MCMC simulation models, the root mean squared error (RMSE) and the coefficient of determination (R^2) were selected as accuracy metrics [36]. The mathematical definitions of RMSE and R^2 are provided in Equations 13, and 14.

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(t_k - \hat{t}_k)^2} \tag{13}$$

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(t_k - \hat{t}_k)^2}{\sum_{k=1}^{n} t_k^2 - \frac{\sum_{i=1}^{n} t^2{}_k}{n}} \tag{14}$$

where $\hat{t}_k$ and t_k are estimated and measured values, respectively, and n is the sample size.

3. Results and discussion

3.1. Univariate statistical analyses

Statistical parameters such as mean, median, minimum, maximum, skewness, and kurtosis of the elements are presented in Table 1.

3.2. Multivariate statistical investigations

Among the available multivariate statistical techniques Principal Component Analysis (PCA), Factor Analysis (FA), and Cluster Analysis (CA) factor analysis was selected for this study. Unlike PCA, factor analysis incorporates a rotation criterion that enhances interpretability, and unlike cluster analysis, it does not rely on a distance-based separation criterion. Factor analysis is designed to identify specific underlying relationships among a set of variables that may initially appear unrelated, within the framework of a hypothetical model. A primary objective of factor analysis is to reduce the dimensionality of the dataset while preserving its essential structure. The method operates under the assumption that latent patterns or linear relationships exist among the observed variables, which can be explained by a smaller set of unobserved variables known as factors. These factors are extracted to represent the shared variance among the original variables. By applying factor analysis, the complexity of the dataset is reduced from a large number of correlated variables to a smaller number of uncorrelated factors. These factors exhibit two important properties: (a) they account for a substantial proportion of the total variance in the dataset, and (b) they are mutually uncorrelated, having been derived as linear combinations of the original variables [37, 38]. In this study, five factors were retained based on eigenvalues greater than one, following Kaiser's criterion [39] (see Table 2).

Kaiser's criterion, which recommends retaining factors with

Table 1. Summary of Zarshuran deposit data statistics (all elements in ppm and Au in ppb).

Variables	Mean	Median	Min.	Max.	Std.dev.	Skewness	Kurtosis
Ag	1.79	0.26	2.88	8.50	1.61	0.93	0.93
Al	40055	36104	29863	265	0.26	-1.37	-1.37
As	2291	2.15	1.68	10975	0.15	2.15	0.22
Ba	435	333	345	1109	1.01	-0.20	-0.20
Be	0.91	0.70	0.63	2.30	0.69	-0.79	-0.79
Bi	0.38	0.38	0.00	0.40.	0.00	0.00	0.00
Ca	93077	83076	74959	255640	0.39	-1.11	-1.11
Cd	76.54	23.15	98.16	366	1.66	2.10	2.10
Ce	36.94	32.00	25.02	110	0.54	-0.74	-0.74
Co	10.66	10	7.57	33	0.55	-0.27	-0.27
Cr	83	76	55	247	1.23	1.95	1.95
Cu	36	28	29	89	0.80	-0.67	-0.67
Fe	25820	23103	14287	73521	0.35	-0.35	-0.35
K	14642	15130	10507	33245	0.18	-1.35	-1.35
La	19	16	13	63	0.61	-0.41	-0.41
Li	24	19	20	75	0.84	-0/33	-0.33
Mg	11397	5301	11719	33759	0.86	-0.88	-0.88
Mn	631	625	485	2094	0.44	-0/68	-0.68
Mo	0.97	0.96	0.10	1.30	1.08	1.74	1.74
Na	2604	2144	1894	6301	0.94	-0.20	-0.20
Ni	53	45	42	187	1.52	2.31	2.31
P	379	363	235	1103	0.42	-0.63	-0.63
Pb	664	79	975	2675	1.29	0.01	0.01
S	2282	1540	2318	7877	1.28	0.70	0.70
Sb	56	11	77	207	1.21	0.23	-0.23
Sc	7	6	5	17	0.30	-1.35	-1.35
Sr	6	6	3	15	1.50	1.99	1.99
Th	12	11	6.80	37	2.00	3.96	3.96
Ti	2358	2340	1779	7633	0.40	-0.67	-0.67
U	3.03	3.05	0.48	4.20	-0.15	-0.09	-0.09
V	65	59	48	167	0.47	-1.12	-1.12
Y	11	11	6	24	0.37	-0.69	-0.69
Yb	1.41	1.40	0.55	2.70	-0.0۲	-0.42	-0.42
Zn	820	318	1109	4100	1.73	2.17	2.17
Zr	12	6	13	62	1.93	3.06	3.06
Au	121	36	220	1458	3.52	5.28	15.28

eigenvalues greater than 1 during Principal Component Analysis (PCA) or Factor Analysis (FA), is widely employed in geochemical studies to simplify complex datasets while preserving meaningful variance. The underlying rationale of this criterion is that an eigenvalue exceeding 1 signifies that the corresponding factor explains more variance than a single standardized original variable, each of which contributes a variance of 1 by definition. By retaining only these factors, the analysis ensures that each extracted component offers substantial explanatory power, effectively reducing the dimensionality of the dataset without sacrificing critical geochemical information. In geochemical datasets, variables frequently exhibit strong correlations arising from shared mineralogical or geochemical processes. Kaiser's criterion facilitates the identification of dominant geochemical signatures or elemental associations that account for the majority of the variance within the dataset. This approach not only aids in data interpretation but also informs subsequent modeling efforts and exploration decision-making.

Table 2. Variance of factors obtained from factor analysis in the Zarshuran deposit.

Factor	Eigenvalue	Cumulative Eigenvalue	Variance%	Cumulative variance%
1	13.65	13.65	39.02	39.02
2	10	23.66	28.58	67.6
3	2.87	26.53	8.21	75.82
4	1.61	28.15	4.62	80.44
5	1.12	29.28	3.22	83.67

Among the five selected factors, the first factor accounted for the largest proportion of variance, explaining 39.02% of the total variability. This was followed by the second factor, which explained 28.58%, and the third factor, contributing 8.21% of the overall variance. Examination of the variables associated with each factor revealed that the second factor comprised the elements As, Cd, Pb, Sb, Zn, and Au (Table 3), underscoring the strong interrelationships among these elements and their collective influence on the observed geochemical variability

Gold (Au) is well-known for its heterogeneous and often nuggety distribution in geological samples, which poses significant challenges for exploration and modeling. This spatial variability implies that small-scale sampling may not adequately represent the true grade distribution, potentially introducing bias and issues of representativeness in data-driven models. To address these challenges in our modeling approach, several strategies were implemented:

(1) Incorporation of Pathfinder Elements: Rather than relying solely on direct Au measurements, which can be sparse or highly variable, the model incorporates concentrations of associated pathfinder elements (As, Cd, Pb, Sb and Zn). These elements tend to display more consistent spatial patterns linked to Au mineralization, thereby providing a more stable and representative proxy for predicting Au grades across heterogeneous datasets.

(2) Use of Advanced Neural Network Modeling (GMDH): The Group Method of Data Handling (GMDH) neural network excels at capturing nonlinear and complex relationships within multivariate data. Its self-organizing structure helps the model adapt to subtle patterns and irregular distributions inherent in heterogeneous Au data, improving the robustness of predictions even when direct Au values are scattered or patchy.

(3) Data Partitioning and Validation: To ensure that the model did

Table 3. Factor loading of variables in factor analysis.

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Ag	-0.39	0.52	0.20	-0.22	0.23
Al	0.96	-0.01	0.08	-0.01	0.12
As	-0.21	**0.92**	-0.05	0.02	0.01
Ba	0.26	0.57	0.08	-0.05	0.67
Be	0.81	-0.16	0.20	0.14	0.41
Ca	-0.30	-0.09	0.79	0.01	-0.40
Cd	-0.21	**0.92**	-0.05	0.02	-0.02
Ce	0.95	-0.20	-0.05	0.01	0.06
Co	0.84	0.16	-0.08	0.33	0.16
Cr	0.66	0.28	0.35	0.18	0.25
Cu	0.35	0.65	0.13	-0.12	0.50
Fe	0.87	0.08	0.18	0/07	0.24
K	0.95	0.02	0.08	-0.04	0.05
La	0.95	-0.16	-0.02	0.01	0.06
Li	0.32	0.58	0.45	-0/11	0.44
Mg	0.06	-0.56	-0.31	0/33	-0.31
Mn	0.69	-0.36	-0.32	0.29	0.05
Mo	-0.01	0.44	0.31	-0.01	0.11
Na	0.26	0.51	0.12	-0.04	0.73
Ni	0.42	0.10	0.07	0.76	0.34
P	0.93	-0.12	-0.08	-0.13	0.12
Pb	-0.17	**0.87**	0.05	-0.18	0.22
S	0.30	0.23	0.09	-0.66	0.38
Sb	-0.08	**0.93**	-0.01	-0.17	0.05
Sc	0.96	-0.18	-0.09	0.01	0/04
Sr	-0.22	0.45	-0.42	0.01	0.09
Th	0.12	0.28	0.17	0.12	0.87
Ti	0.94	0.15	0.09	-0.12	0.05
U	0.44	-0.07	-0.68	0.01	-0.06
V	0.97	-0.04	0.11	-0.03	0.11
Y	0.83	-0.17	-0.41	0.05	-0.01
Yb	0.70	-0/29	-0.55	0.05	-0.19
Zn	0.15	**0.83**	0.09	0.27	0.15
Zr	0.54	0.43	-0.19	0.12	-0.37
Au	-0.17	**0.79**	0.01	0.21	0.20

not overfit localized anomalies or non-representative samples, the dataset was carefully partitioned into training and testing subsets. This approach allowed the model to learn from a broad spectrum of data variability while validating its predictive performance on independent data, thereby enhancing confidence in its generalizability.

(4) Preprocessing and Outlier Management: Prior to modeling, exploratory data analysis and correlation assessments helped identify and manage extreme values or outliers that could distort model training. This step ensured that anomalously high or low Au grades did not unduly bias the model, improving the representatively of the input data.

(5) Multi-Element Integration and Dimensionality Reduction: By integrating multiple pathfinder elements and selecting the most relevant variables, the model reduced noise and focused on key geochemical signals linked to Au mineralization, thereby mitigating the impact of heterogeneity in any single variable.

Through the integration of these combined approaches, our modeling effectively mitigates the challenges posed by the heterogeneous distribution of Au data, resulting in more reliable grade estimation and improved anomaly detection. This, in turn, enhances the practical applicability of the model for guiding exploration decisions, despite the inherent spatial variability of Au in geological settings. In similar studies, the low-intensity nature of Au anomalies and their heterogeneous distribution within geological samples have posed significant challenges for sampling and sample preparation processes [6]. In multi-element geochemical exploration programs, the specific analytical and procedural requirements for detecting precious metals such as Au are often underemphasized. As a result, the data collected for these elements frequently suffer from reduced reliability and accuracy, undermining the confidence in anomalies identified from such datasets [40]. Consequently, Au anomalies derived from inadequately processed data are often regarded as unreliable. Nevertheless, when interpreted correctly, multi-element geochemical datasets can provide valuable insights into the presence of Au mineralization [41, 42]. Pathfinder elements such as As, Cd, Pb, Sb, and Zn commonly occur in association with Au and can serve as effective proxies for its detection, depending on the mineralization style. In the Zarshuran region, the application of neural network models to identify Au geochemical anomalies and estimate Au concentrations has demonstrated the utility of advanced computational techniques in overcoming these limitations and improving anomaly detection accuracy [43].

This study demonstrated that gold (Au) grades can be reliably estimated based on the concentrations of key pathfinder elements. Furthermore, geochemical anomalies can be effectively identified by analyzing the predicted Au grade values generated by neural network models. However, upon detection of such anomalies, it is crucial to validate the model predictions against direct analytical measurements of Au to ensure accuracy. Integrating these model-derived estimates with complementary geological and geochemical evidence enhances both the precision of anomaly identification and the reliability of Au content estimation. This combined approach strengthens exploration decision-making by reducing uncertainty and improving target delineation. In geochemical exploration, anomalies detected through pathfinder elements serve as indirect indicators of potential gold mineralization. Elements such as As, Sb, Pb, Cd, and Zn frequently exhibit spatial or geochemical associations with gold, particularly in specific deposit types, such as epithermal or orogenic systems.

Due to gold's often low concentrations and highly heterogeneous distribution, relying solely on direct Au analyses can sometimes result

in missed targets or underestimation of mineralized zones. While direct Au analyses provide definitive measurements of gold content in samples, offering conclusive evidence of mineralization where detected, they are limited by factors such as the nugget effect, low detection limits, and sampling bias. These limitations can prevent a full characterization of the spatial extent of gold anomalies. Consequently, the role of pathfinder-based anomaly detection is to identify areas with elevated geochemical potential even when Au concentrations are low or below detection limits. This approach guides follow-up sampling and drilling efforts and extends exploration coverage beyond regions where direct gold detection is either impractical or cost-prohibitive. Conversely, direct Au analyses are essential for confirming the presence and concentration of gold and play a critical role in validating model predictions and interpreting anomalies. In practice, combining pathfinder-based methods with direct Au measurements enables broader and more sensitive anomaly detection, ground-trothing, and accurate resource estimation. Together, these complementary approaches contribute to a more comprehensive and reliable exploration strategy.

3.3. Modeling

3.3.1. Experimental database

The purpose of this paper is to estimate Au grade using two advanced computational approaches: the GMDH neural network and MCMC simulation methods. For this study, 106 data points were selected to establish the relationships between the input and output parameter sets. The input dataset includes pathfinder elements (As, Cd, Pb, Sb, Zn), while the output dataset corresponds to Au concentrations.

In this study, the dataset was divided into an 80/20 split, with 80% of the data used for training and 20% reserved for testing. This split is a commonly accepted practice in machine learning because it provides a sufficiently large portion of data for the model to learn underlying patterns while keeping aside enough samples to evaluate the model's performance on unseen data. The 80/20 ratio strikes a practical balance:

(1) Adequate training data ensures the model can effectively learn complex relationships.

(2) Sufficient test data allows for a meaningful assessment of the model's predictive accuracy and generalization capability.

However, the fixed 80/20 split has inherent limitations, particularly with relatively small datasets such as ours (108 data points), where the selection of test samples can significantly influence evaluation outcomes. To mitigate this issue, more robust validation techniques, such as k-fold cross-validation, can be employed. In k-fold cross-validation, the dataset is partitioned into k subsets (folds).

The model is then trained and evaluated k times, each iteration using a different fold as the test set while the remaining folds serve as the training set. This method reduces variability associated with any single train-test split, provides a more reliable estimate of model performance, and helps to prevent bias in performance evaluation, which is especially important when working with smaller datasets.

While k-fold cross-validation offers enhanced robustness, it is computationally more intensive. For practical and computational efficiency, the 80/20 split remains a valid initial approach, with cross-validation recommended for future work or more critical model assessments [44]. Table 4 provides a statistical summary of the input and output datasets. The probability distribution functions (PDFs) of the variables (As, Cd, Pb, Sb, Zn and Au) used in the model are shown in Figure 3. These PDFs help visualize how the values of each variable are spread or distributed. By looking at these distributions, you can better understand the nature of the data whether it's skewed, symmetric, concentrated around certain values, or spread out. Examination of figures a to f showed that the distribution of each variable (As, Cd, Pb, Sb, Zn and Au) is right-skewed.

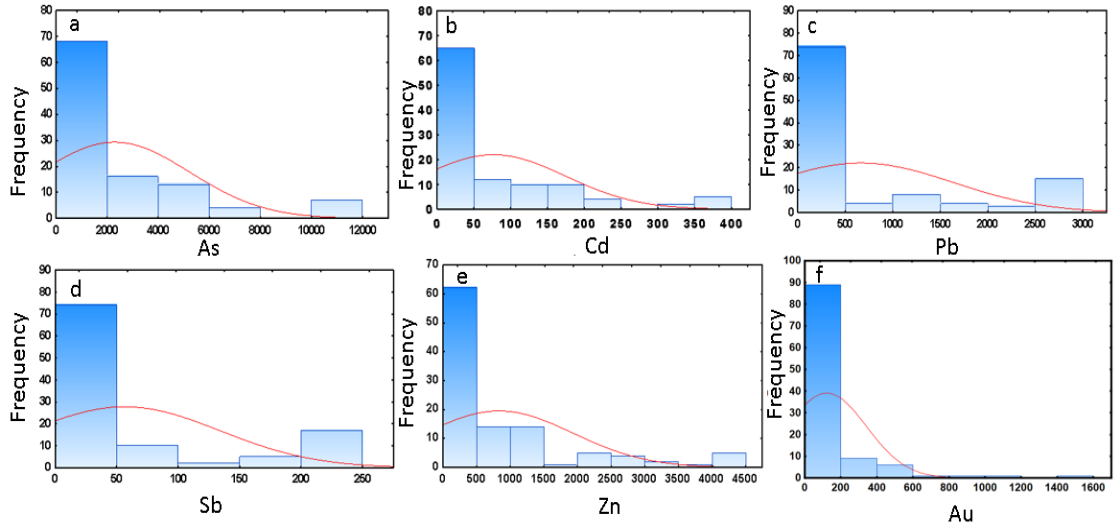

Figure 3. Continuous probability distribution of variables in models.

Table 4. Statistics description of inputs and output datasets.

Parameters		Mean	Minimum	Maximum	Standard Deviation
Inputs	As	2291	29	10975	2938
	Cd	76	5	100	98
	Pb	664	7	2675	975
	Sb	56	0.95	207	77
	Zn	820	0.75	4100	1109
Output	Au	121	4	1458	220

3.3.2. GMDH neural network model

In our results, the error values observed for the training dataset are generally lower than those for the test dataset. This discrepancy is characteristic of neural network modeling and illustrates the fundamental concept of model generalization. During the training phase, the model learns patterns directly from the input data, enabling it to closely fit the training examples and often resulting in reduced error values. However, the true measure of a model's effectiveness lies in its capacity to generalize that is, to accurately predict outcomes on new,

unseen data, as represented by the test dataset. The marginally higher errors observed in the test set indicate that, although the model performs well on familiar data, it encounters greater challenges when predicting novel samples. This difference is anticipated, as the test data may contain variations or complexities not fully captured during training. Crucially, the relatively small increase in error between the training and test datasets suggests that the model does not suffer from overfitting; rather, it has successfully learned meaningful patterns that extend beyond the original training data.

Eighty percent of the total 108 data points were allocated for model training, while the remaining 20% were reserved for evaluating the model's predictive accuracy. The error metrics generated by the GMDH neural network are summarized in Table 5 and visualized in Figure 4, which also includes a comparison between actual and predicted values. The relationship between the GMDH-estimated Au grades and the corresponding measured values for both the training and testing datasets is presented in Figure 5, further illustrating the model's performance across both phases. These results demonstrate a strong alignment between the predicted and observed Au grades, thereby confirming the reliability and robustness of the GMDH neural network in this application. When comparing the prediction results between the training and test datasets both derived from the same overall dataset, it is observed that the fit between predicted and measured values appears stronger in the test dataset than in the training dataset. This seemingly counterintuitive outcome is primarily attributed to the difference in dataset size and complexity. The training dataset, being substantially larger, encompasses a broader range of variability, making the task of fitting a predictive model more challenging. As a result, the model must account for greater heterogeneity within the training data. In contrast, the smaller test dataset typically spans a narrower range of values, which can lead to a closer match between predicted and observed values simply due to reduced variability. Therefore, the improved fit observed in the test dataset should be interpreted with caution. It does not necessarily indicate superior model generalization but may instead reflect the inherent differences in data distribution and complexity between the training and test sets [45]. Furthermore, as shown in the comparison of predicted and measured values, the error curves for the GMDH neural network model in the test dataset do not completely overlap, indicating some residual variance. Figure 5 illustrates the estimated versus measured Au values for both the training and test datasets using the GMDH model. Notably, the coefficient of determination (R^2) exceeds 0.9 for both datasets, indicating a high level of predictive accuracy across training and testing phases.

Beyond comparing the errors between the predicted and measured values in both the training and test datasets, the root mean square error (RMSE) and the coefficient of determination (R^2) are also utilized to assess prediction performance. As presented in Table 5, when the error indicators are computed for the dataset, the results of the GMDH neural network model for the training dataset are generally higher than those for the test dataset. A potential reason for this is that the model is primarily trained using the training dataset, with the test dataset reserved for prediction evaluation. Since the test dataset is relatively small, the calculation results on the training dataset (R^2 = 0.9949, RMSE = 0.0607) are better than those on the test dataset (R^2 = 0.9483, RMSE = 0.1379). This approach ensures strong model performance while maintaining the interpretability of the nonlinear component [46, 47].

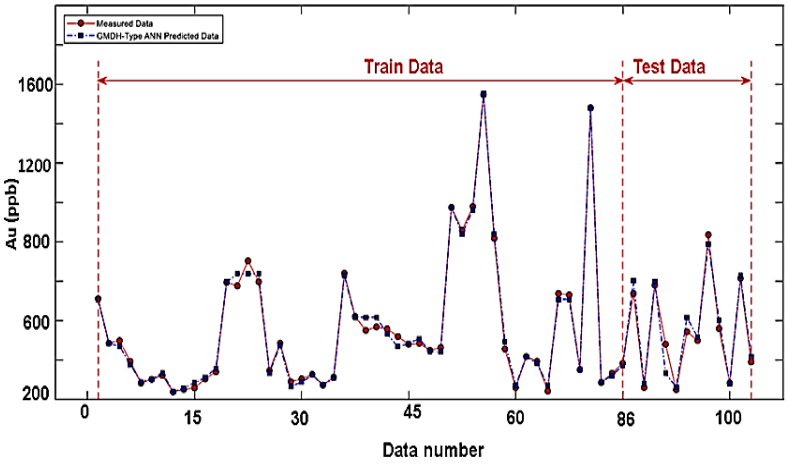

Figure 4. Comparison of estimated and measured values using the GMDH neural network model.

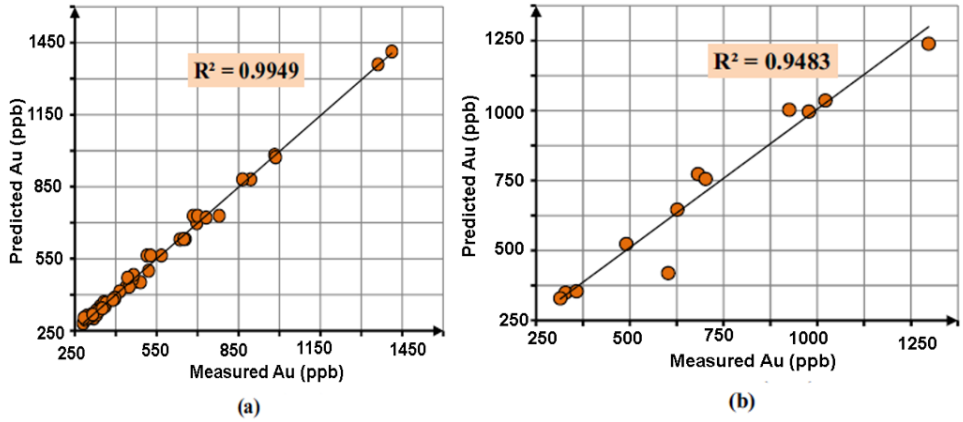

Figure 5. Estimated and measured values of Au for a) training dataset, b) testing dataset in GMDH.

Table 5. Error estimation using the GMDH neural network model.

Training		Testing	
RMSE	R^2	RMSE	R^2
0.0607	0.9949	0.1379	0.9483

K-fold cross-validation is a statistical technique used to more reliably assess the performance of a predictive model, particularly when working with limited data. In this approach, the entire dataset is divided into k equally sized subsets, or folds. The model is then trained on k-1 folds and tested on the remaining fold. This process is repeated k times, with each fold serving exactly once as the test set. After all k iterations, performance metrics such as accuracy, R^2, or error are averaged across all folds to provide a final, more stable estimate of the model's performance. This method ensures that every sample is used for both training and validation, maximizing the utility of the available data. Because the results are averaged over multiple splits, the performance estimate is less sensitive to the way the data is partitioned. This leads to a more reliable evaluation and helps determine whether a model's performance is consistent across different subsets of the data [48]. We are addressing a binary classification problem using a very small dataset consisting of only 108 examples. The dataset is imbalanced, with 80% of the samples belonging to the majority class and 20% to the minority class. To evaluate model performance, the k-fold cross-validation is employed (Figure 6). At each iteration, the model's performance is monitored not only on the validation fold, but also on the training samples. When the performance metric (e.g., accuracy, F1-score, or another suitable measure) is consistently higher on the training data compared to the validation or test data, this may indicate that the model is overfitting. Overfitting occurs when the model learns the training data too well, including noise or patterns that do not generalize to unseen data.

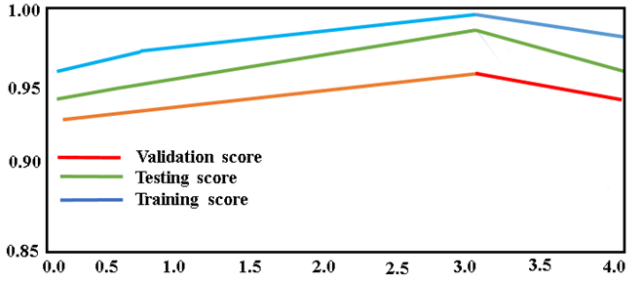

Figure 6. K-fold cross-validation for validation score, testing score and training score.

3.3.3. MCMC simulation model

In the MCMC model, the dataset comprising 108 geochemical samples was divided into two subsets: 80% of the data were allocated for model training, while the remaining 20% were reserved as a validation set to independently evaluate the model's predictive performance. This partitioning ensured that model accuracy could be assessed on data unseen during training. Prior to modeling, an initial correlation analysis was conducted to examine the relationships between the independent variables As, Cd, Pb, Sb, and Zn and Au grades. This exploratory analysis aimed to identify geochemical associations and potential pathfinder elements relevant to gold mineralization. Variables exhibiting the strongest correlations with Au were selected as candidate predictors, guiding the choice of input features for the MCMC algorithm. Thus, the correlation analysis served a dual purpose: it reduced dimensionality by excluding weakly related variables and improved model interpretability by highlighting key geochemical interactions pertinent to Au mineralization (see Equations 15 to 22).

Model#1

$$Au = a_1(As) + a_2(Cd) + a_3(Pb) + a_4(Sb) + a_5(Zn) + a_6 \quad (15)$$

Model#2

$$Au = a_1(As)^{b1} + a_2(Cd)^{b2} + a_3(Pb)^{b3} + a_4(Sb)^{b4} + a_5(Zn)^{b5} + a_6 \quad (16)$$

Model #3

$$Au = \frac{a_1(As)^{b1}+a_2(Cd)^{b2}+a_3(Sb)^{b3}+a_6}{a_4(Pb)^{b4}+a_5(Zn)^{b5}+a_7} \quad (17)$$

Model#4

$$Au = \frac{a_1(As)^{b1}+a_2(Cd)^{b2}+a_3(Sb)^{b3}+a_6}{a_4(Pb)^{b4}+a_5(As)^{b5}+a_7} \quad (18)$$

Model #5

$$Au = \frac{a_1(As)^{b1}+a_2(Cd)^{b2}+a_3(Zn)^{b3}+a_6}{a_4(Pb)^{b4}+a_5(Sb)^{b5}+a_7} \quad (19)$$

Model #6

$$Au = \frac{a_1(Pb)^{b1}+a_2(Cd)^{b2}+a_3(Zn)^{b3}+a_6}{a_4(As)^{b4}+a_5(Sb)^{b5}+a_7} \quad (20)$$

Model #7

$$Au = \frac{a_1(As^{b1}.Cd^{b2})+a_2+exp(Sb)+a_3}{a_4\exp(Zn)Pb^{b3}+a_5} \quad (21)$$

Model #8

$$Au = \frac{a_1(As)^{b1}+a_2(Cd)^{b2}+a_3(Pb)^{b3}++a_4}{a_5\exp(Sb.Zn)+a_6} \quad (22)$$

In this study, the undefined variables within the candidate models were treated as random variables. The primary objective was to identify the most suitable model for comparing Au data utilizing a Bayesian framework. Model parameters were estimated through the Bayesian Markov Chain Monte Carlo (MCMC) method, implemented via the WinBUGS software. For the stochastic nodes in the Bayesian framework, lognormal or normal distributions were assigned to the grades of As, Cd, Pb, Sb, and Zn, respectively. The models were formulated using the WinBUGS language, and a trial-and-error approach was employed to optimize the modeling parameters. For Model #4, the mean values of the uncertain parameters (b_1, b_2, b_3, b_4, b_5) and (a_1, a_2, a_3, a_4, a_7) were determined as follows:

$b_1 = 6.056$, $b_2 = -59.27$, $b_3 = 5.524$, $b_4 = 26.56$, $b_5 = -12.06$

$a_1 = 36.77$, $a_2 = 0.07215$, $a_3 = -0.05011$, $a_4 = -0.1842$, $a_7 = 0.08812$, $a_7 = 14.25$

These parameter values were employed to estimate Au grades with optimal accuracy. Both the training and testing datasets were utilized to identify the most appropriate model and to evaluate its predictive performance. The results are summarized in Table 6 and illustrated in Figure 7. Representative predictive distributions of the GMDH neural network model architecture, utilizing various inference methods, are shown in Figure 5. These distributions indicate that comparable results were obtained for both estimated and measured values [49]. In the example presented, the estimated and measured values exhibit similar distributions with closely aligned mean values and relatively low uncertainty estimates. The training dataset yields a narrower distribution and more precise predictions, as anticipated. Predictive accuracies, as measured by the performance metrics listed in Table 5, remain consistent across the different datasets used for training and testing (Table 5) as well as the resulting model (Table 6).

Overall, the predictive performance is notably consistent across the various datasets for both training and testing. Models generally exhibit superior performance on datasets #3 and #4 compared to the others, with dataset #4 outperforming dataset #3 across most metrics, including R^2 and RMSE. In general, the models achieve their best results on datasets #2 through #5, with R^2 values exceeding 0.7. This trend is observed for both Bayesian and non-probabilistic models. When comparing different model architectures, the performance difference between the top-performing models (#4 and #5) across all datasets is minimal. Additionally, when evaluating various inference methods, the GMDH neural network model demonstrates comparable predictive

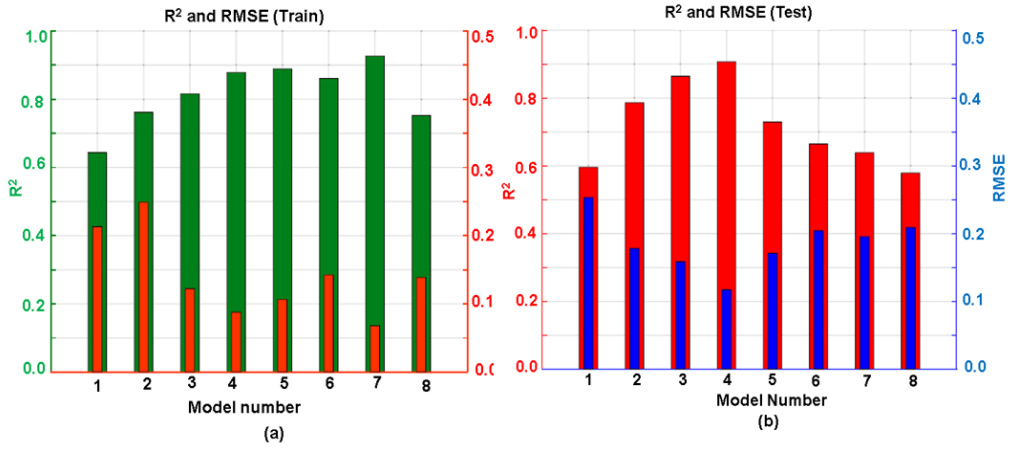

Figure 7. Outcome model for a) training and b) testing.

accuracy to its probabilistic counterparts in terms of RMSE and R^2 across all datasets. Notably, models #1 and #8 exhibit significantly poorer performance, with R^2 values below 0.6 and RMSE values exceeding 0.2 relative to the other models, indicating that these models are unsuitable for predicting gold grades.

Table 6. Outcome model for testing and training datasets.

Model No.	Training		Testing	
	RMSE	R^2	RMSE	R^2
#1	0.213199	0.6439	0.253214	0.5957
#2	0.249071	0.7616	0.178118	0.7858
#3	0.122254	0.8159	0.158516	0.8644
#4	0.087882	0.8785	0.117455	0.9069
#5	0.106348	0.8889	0.171069	0.7291
#6	0.142462	0.8610	0.204343	0.6646
#7	0.068133	0.9266	0.195556	0.6387
#8	0.138635	0.7521	0.209141	0.5787

The results identified Model #4 as the best-performing model for predicting Au grade. Based on evaluations of both training and testing datasets, Model #4 demonstrated superior accuracy and reliability, whereas Model #8 was found to be the least effective for Au grade prediction. The observation that the GMDH (Group Method of Data Handling) model outperforms MCMC Bayesian approaches in predictive accuracy and anomaly detection carries significant implications, particularly within the context of geochemical exploration and broader practical applications. While MCMC Bayesian methods are highly regarded for their robust uncertainty quantification an essential feature in early-stage exploration characterized by limited data and geological uncertainty they are often computationally demanding. Moreover, these methods typically require expert input for defining prior distributions and performing convergence diagnostics, which can constrain their practical utility in rapid, iterative field-based workflows.

In contrast, the GMDH model offers a more efficient and scalable alternative, enabling quicker data processing and model refinement without extensive prior knowledge or complex diagnostic procedures. This advantage facilitates more agile decision-making in exploration campaigns, where timely identification of mineralization anomalies is critical. Consequently, the superior predictive performance of the GMDH model suggests that it may serve as a valuable complement or alternative to Bayesian methods, particularly in scenarios demanding high throughput and adaptability. This finding underscores the potential for integrating data-driven machine learning techniques into exploration workflows to enhance predictive accuracy while maintaining operational efficiency [50].

By contrast, GMDH offers automatic structure optimization (selecting relevant variables and model complexity), lower computational overhead, and faster model deployment without sacrificing accuracy. This makes GMDH especially appealing for operational teams needing fast turnaround between data collection and decision-making, such as in adaptive drilling campaigns or time-sensitive remote explorations. Because GMDH is data-driven and self-organizing, it adapts well to complex, nonlinear, and multi-dimensional relationships in geochemical datasets, varied geological settings without reconfiguration, and multi-element datasets with unknown or weakly understood element–mineralization relationships. This contrasts with MCMC, which, while flexible in theory may require recalibration or domain-specific priors to remain reliable across different terrains or deposit types.

The superior performance of the GMDH model highlights its strong potential for integration into real-time decision-support systems tailored for exploration teams for instance, field-deployable software platforms. Beyond mineral exploration, GMDH may also be effectively applied to a range of other resource prediction challenges, such as groundwater prospectively mapping, environmental contamination assessment, and agricultural soil optimization. Its data-driven design and user-friendly implementation make it particularly well-suited for use by non-experts, thereby enabling field geologists and technicians to leverage advanced predictive modeling through simplified, intuitive interfaces.

4. Conclusions

For predicting gold (Au) grades, two models MCMC and GMDH are also valuable tools for quality control of laboratory analyses and for comparing results across different laboratories. When data is incomplete or geochemical maps cannot be produced, these models offer reliable predictions and act as guides for the next phases of exploration. In practical mineral exploration, Markov Chain Monte Carlo (MCMC) Bayesian methods not only provide predictions but also quantify the uncertainties associated with those predictions. These uncertainties are crucial for informing risk-based decision-making in later stages of exploration. For example:

(1) Drill Target Prioritization: Areas with high predicted potential but also high uncertainty might be deprioritized in favor of moderately prospective zones with lower uncertainty ensuring more reliable returns on drilling investment.

(2) Resource Allocation: Knowing the confidence intervals around model outputs allows exploration teams to allocate budgets and effort more efficiently, focusing on areas where the probability of success is both high and well-constrained.

(3) Strategic Planning: In frontier or data-sparse regions, predictive uncertainty can help identify where additional sampling or geophysical surveys would most reduce risk, leading to smarter iterative exploration strategies.

In essence, MCMC-driven uncertainty estimates transform exploration from a purely predictive exercise into a probabilistic risk management process, increasing both scientific rigor and economic efficiency.

In this study, a Bayesian inference-based approach was implemented using WinBUGS software to predict gold (Au) grades and identify the best-performing model. The input variables for the models included the grades of arsenic (As), cadmium (Cd), lead (Pb), antimony (Sb), and zinc (Zn), which were selected based on factor analysis. Among the tested models, Model #4 was identified as the best prediction model according to its R^2 and RMSE performance metrics. The GMDH neural network model outperformed the MCMC simulation model, providing more accurate results for both training and testing datasets.

From a practical exploration standpoint, the GMDH neural network is preferred because it automatically selects the most relevant input variables and effectively models complex, nonlinear relationships between geochemical data and mineralization patterns. This adaptability is crucial in real-world situations where geological data are often noisy, incomplete, and highly variable. In regions with heterogeneous mineral deposits such as low-intensity gold anomalies distributed irregularly precise modeling is essential to distinguish true signals from background noise. GMDH's ability to capture nonlinear patterns enhances the detection of subtle anomalies that simpler models might overlook.

When working with large datasets containing many elements, GMDH helps reduce dimensionality by selecting the most influential variables, thereby improving model robustness and reducing false positives. In areas with limited historical data, GMDH can build predictive models without relying on extensive prior assumptions, supporting decision-making in uncertain environments. In such cases, high model precision directly leads to better targeting of drilling sites, minimizing costly mistakes and increasing the likelihood of discovering economically viable mineral deposits. This makes GMDH a valuable tool for optimizing exploration efficiency and resource allocation.

References

[1]. Liu, H., Zhang, B., Wang, X., Wang, Q., Du, Y., Zhang, B., Cui, Y., Zhou, J., Liu, B., & Li, J. (2024) Exploration indication of hidden gold deposits using the fine-grained soil prospecting method and its nano-micron metal migration evidence in the alluvial soil covered area, Jiaodong. Journal of Geochemical Exploration, 263: 107514. https://doi.org/10.1016/j.gexplo.2024.107514

[2]. Ohta, A., Imai, N., Terashima, S., &Tachibana, S. (2005). Influence of surface geology and mineral deposits on the spatial distributions of elemental concentrations in the stream sediments of Hokkaido, Japan. Journal of Geochemical Exploration, 86: 86-103. https://doi.org/10.1016/j.gexplo.2005.04.002

[3]. Li, X., Li, L.H., Zhang, B.L., & Guo, Q.J. (2013). Hybrid self-adaptive learning-based particle swarm optimization and support vector regression model for grade estimation. Neurocomputing 118(1): 179-190. https://doi.org/10.1016/j.neucom.2013.03.002

[4]. Cheng, Q. (2007). Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. Ore Geology Review, 32: 314–324. https://doi.org/10.1016/j.oregeorev.2006.10.002

[5]. Samal, R.A., Mohanty, K.M., & Fifarek, H.R. (2008). The Backward elimination procedure for a predictive model of gold concentration. Journal of Geochemical Exploration, 97: 69–82. https://doi.org/10.1016/j.gexplo.2007.11.004

[6]. Tabatabaei, H.S., Roshani Rodsari, P., & Mokhtari, R.A. (2015). Predicting Potential Mineralization Using Surface Geochemical Data and Multiple Linear Regression Model in the Kuh Panj Porphyry Cu Mineralization (Iran). Arabian Journal Science Engineering, 40:163–170. https://doi.org/10.1007/s13369-014-1482-z

[7]. Nazarpour, A., Rostami Paydar, G., & Carranza, E.J.M. (2016). Stepwise regression for recognition of geochemical anomalies: A case study in Takab area, NW Iran. Journal of Geochemical Exploration, 168:150-162. https://doi.org/10.1016/j.gexplo.2016.07.003

[8]. Li, S., Xu, L., Wang, Z., Yang, C., Tan, L., Nie, R., Meng, M., Li, J., Zhang, B., & Liu, J. (2023) Application of tectono-geochemistry method for weak information extraction of Carlin-type gold deposits in Yunnan–Guizhou–Guangxi, SW China. Ore Geology Reviews, 163: 105813. https://doi.org/10.1016/j.oregeorev.2023.105813

[9]. Saljoughia, S., & Hezarkha, A. (2019). Identification of geochemical anomalies associated with Cu mineralization by applying spectrum-area multi-fractal and wavelet neural network methods in the Shahr-e-Babak mining area, Kerman, Iran. Journal of Mining and Environment, 10(1): 49-73. https://doi.org/10.22044/jme.2018.6949.1533

[10]. Asadi Harooni, H. (2000). The Zarshuran gold deposit model was applied in a mineral exploration GIS in Iran. PH.D. Thesis, Delft University, the Netherlands.

[11]. Aliyari, F, Afzal, P., & Sharif, A. (2017). Determination of geochemical anomalies and gold mineralized stages based on litho-geochemical data for Zarshuran Carlin-like gold deposit (NW Iran) utilizing multi-fractal modeling and stepwise factor analysis. Journal of Mining and Environment, 8(4): 593-610. https://doi.org/10.22044/jme.2017.5252.1340

[12]. Yousefi, T., Abedini, A., Aliyari, F., Calagari, A.A. (2019). Mineralogy and fluid inclusion investigations in the Zarshuran gold deposit, north of Takab, NW Iran. Iranian Journal of Crystallogrphy and Mineral, 27 (3): 537-550. https://doi.org/10.29252/ijcm.27.3.537

[13]. Rezaei, S., Lotfi, M., Afzal, P., Jafari, M.R., & Shamseddin Meigoony, M. (2015). Delineation of Cu prospects utilizing multifractal modeling and stepwise factor analysis in Noubaran 1:100,000 sheet, Center of Iran. Arabian Journal Geoscience, 8 (9): 7343–7357. https://doi.org/10.1007/s12517-014-1755-6

[14]. Seyedrahimi-Niaraq, M., & Mahdiyanfar, H. (2024) Fractal modeling of the Cu-Au mineralization principal component values by considering the rejection of multivariate outlier data. Iranian Society of Mining Engineering, 19(62):16-38. https://doi.org/10.22034/ijme.2024.2011288.1981

[15]. 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, https://doi.org/10.1144/geochem2022-015

[16]. Othman, A.A, Gloaguen, R. (2017). Integration of spectral, spatial and morphometric data into lithological mapping: A comparison of different Machine Learning Algorithms in the Kurdistan Region, NE Iraq. Journal Asian Earth Science, 146: 90-102. https://doi.org/10.1016/j.jseaes.2017.05.005

[17]. Zhang, H., Niu, F., Zhang, J., & Yu, X. (2022). Prediction of Three-Dimensional Fractal Dimension of Hematite Flocs Based on Particle Swarm Optimization Optimized Back Propagation Neural Network. Mining, Metallurgy & Exploration, 39: 2503–2515. https://doi.org/10.1007/s42461-022-00684-z

[18]. Nandi, B.P., Singh, G., Jain, A., & Tayal, D.K. (2024). Evolution of neural network to deep learning in prediction of air, water pollution and its Indian context. International Journal of Environment Science Technology, 21: 1021–1036. https://doi.org/10.1007/s13762-023-04911-y

[19]. Zuo, R., Cheng, Q., Xu, Y., Yang, F., Xiong, Y., Wang, X., & Kreuzer, O. (2024). Explainable artificial intelligence models for mineral prospectivity mapping. Sci. China Earth Science, 67: 2864–2875. https://doi.org/10.1007/s11430-024-1309-9

[20]. Jodeiri Shokri, B., Ramazi, H., Doulati Ardejani, F., & Sadeghi Amirshahidi, M.H. (2014). Prediction of pyrite oxidation in a coal washing waste pile applying artificial neural networks (ANNs) and adaptive neuro-fuzzy inference systems (ANFIS). Mine Water and the Environment, 33: 146-156. https://doi.org/10.1007/s10230-013-0247-3

[21]. Balogun, S., & Ogwueleka, T.C. (2023). Performance prediction for wastewater treatment plant effluent cod using artificial neural network. International Journal Environment Sciences Technology, 20: 12659–12668. https://doi.org/10.1007/s13762-023-04823-x

[22]. Pradhan, B., Jena, R., Talukdar, D., Mohanty, M., Sahu, B.K., Raul, A.K., &Abdul Maulud, K.N. (2022). A New Method to Evaluate Gold Mineralisation-Potential Mapping Using Deep Learning and an Explainable Artificial Intelligence (XAI) Model. Remote Sensing, 14: 4486. https://doi.org/10.3390/rs14184486

[23]. Chen, G., Huang, N., Wu, G., Luo, L., Wang, D., &Cheng, Q. (2022). Mineral prospectivity mapping based on wavelet neural network and Monte Carlo simulations in the Nanling W-Sn metallogenic province. Ore Geology Review, 143: 104765. https://doi.org/10.1016/j.oregeorev.2022.104765

[24]. Bazdar, H., & Imamalipour, A. (2024). Application of an improved artificial neural network model for prediction of Cu and Au concentration in the porphyry copper-epithermal gold deposits, Case study: Masjed Daghi, NW Iran. International Journal of Mining and Geo-Engineering, 58(4): 327-339. https://doi.org/10.22059/IJMGE.2024.376761.595167

[25]. Shen, C., Asante-Okyere, S., Yevenyo- Ziggah, Y., Wang, L., & Zhu, X. (2019) Group Method of Data Handling (GMDH) Lithology Identification Based on Wavelet Analysis and Dimensionality Reduction as Well Log Data Pre-Processing Techniques. Energies, 12(8): 1509; https://doi.org/10.3390/en12081509

[26]. Deng, S., Zhang, N., Kuang, B., Li, Y., & Sun, H. (2022) Bayesian Markov Chain Monte Carlo inversion of surface-based transient electromagnetic data. SN Applied. Sciences, 4:254 .https://doi.org/10.1007/s42452-022-05134-5

[27]. Liu, B. Yan Liang, Y. (2017) An introduction of Markov chain Monte Carlo method to geochemical inverse problems: Reading melting parameters from REE abundances in abyssal peridotites. Geochimica et Cosmochimica Acta, 203: 216-234. https://doi.org/10.1016/j.gca.2016.12.040

[28]. Mehrabi, B., Yardley, B.W.D., & Cann, J.R. (1999). Sediment-hosted disseminated gold mineralization at Zarshuran, NW Iran. Mineralium Deposita, 34: 673–696. https://doi.org/10.1007/s001260050227

[29]. Tale Fazel, E., Paˇsava, J., Wilke, F.D, H., Oroji, A., & Andronikova, I. (2023) Source of gold and ore-forming processes in the Zarshuran gold deposit, NW Iran: Insights from in situ elemental and sulfur isotopic compositions of pyrite, fluid inclusions, and O–H isotopes. Ore Geology Reviews, 156: 105382. https://doi.org/10.1016/j.oregeorev.2023.105382

[30]. Paravarzar, S., Maarefvand, P., Maghsoudi, A., & Afzal, P. (2015). Correlation between geological units and mineralized zones using fractal modeling in Zarshuran gold deposit (NW Iran). Arabian Journal of Geosciences, 8:3845–3854 https://doi.org/10.1007/s12517-014-1453-4

[31]. Aitchison, J. (1986). Statistical analysis of compositional data. UK: Chapman and Hall, London, 416 p

[32]. Ivakhnenko, A.G. (1971). Polynomial theory of complex systems IEEE Transactions on Systems, Man, and Cybernetics, 364-378.https://doi.org/ 10.1109/TSMC.1971.4308320

[33]. Oh, H.J.S., & Lee, S. (2010). Application of artificial neural network for gold–silver deposits potential mapping: a case study of Korea. National Resources Research, 19(2): 103-124. https://doi.org/10.1007/s11053-010-9112-2

[34]. Rigol-Sanchez, J.P., Chica-Olmo, M., & Abarca-Hernandez, F. (2003). The Artificial neural networks as a tool for mineral potential mapping with GIS. International Journal Remote Sensing, 24(5): 1151-1156. https://doi.org/10.1080/0143116021000031791

[35]. Gimenez, O., Bonner, S.J., King, Ruth., Parker, R.A., Brooks, S.P., Jamieson, L.E., Grosbois, V., Morgan B.J.T., & Len, T. (2009). WinBUGS for population ecologists: Bayesian modeling using Markov Chain Monte Carlo methods. In: Modeling demographic processes in marked populations. Springer, pp 883-915. https://hdl.handle.net/10023/677

[36]. Brook, S., Gelman, A., Jones, G., & Meng, X.L. (2011). Handbook of Markov chain Monte- Carlo. CRC Press

[37]. Li, H., Li, X., Yuan, F., Jowitt, S., Zhang, M., Zhou, J., Zhou, T., Li, X., Ge, C., & Wu, B. (2020). Convolutional neural network and transfer learning based mineral prospectivity modeling for geochemical exploration of Au mineralization within the Guandian-Zhangbaling area, Anhui province, China. Applied Geochemistry, 122: 104747. https://doi.org/10.1016/j.apgeochem.2020.104747

[38]. Reimann, C., & Filzmoser, P. (2000). Normal and lognormal data distribution in geochemistry: death of a myth, Consequences for the statistical treatment of geochemical and environmental data. Environmental Geology, 39: 1001–1014. https://doi.org/10.1007/s002549900081

[39]. Jolliffe, T. (2002). Principal component analysis. Springer Verlag, New York, 488 pp.

[40]. Kaiser, H.F. (1958). Varimax criterion for analytic rotation in factor analysis. Psychometrika, 23:187-200. https://doi.org/10.1007/BF02289233

[41]. Cheng, Q. (2012). Singularity theory and methods for mapping geochemical anomalies caused by buried sources and for predicting undiscovered mineral deposits in covered areas. Journal of Geochemical Exploration, 122: 55–70. https://doi.org/10.1016/j.gexplo.2012.07.007

[42]. Nazarpour, A., Omran, N.R., Paydar, G.R., Sadeghi, B., Matroud, F., & Mehrabi Nejad, A. (2015). Application of classical statistics, log-ratio transformation, and multifractal approaches to delineate geochemical anomalies in the Zarshuran gold district, NW Iran. Chemie der Erde Geochemistry, 75: 117-132. https://doi.org/10.1016/j.chemer.2014.11.002

[43]. Nazarpour, A. (2018). Application of C-A fractal model and exploratory data analysis (EDA) to delineate geochemical anomalies in the: Takab 1:25,000 geochemical sheet, NW Iran. Iranian Journal of Earth Science, 10: 173-180.

[44]. Misra, D., Samanta, B., & Bandopadhyay, S. (2007). Evaluation of artificial neural networks and kriging for the prediction of arsenic in Alaskan bedrock-derived stream sediments using gold concentration data. International Journal of Mining Reclamation and Environment, 21(4): 282-294. https://doi.org/10.1080/17480930701259294

[45]. Xiong, Y., & Zuo, R. (2020). Recognizing multivariate geochemical anomalies for mineral exploration by combining deep learning and one-class support vector machine. Computer Geoscience, 140: 104484. https://doi.org/10.1016/j.cageo.2020.104484

[46]. 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. https://doi.org/10.1016/j.sab.2015.02.003

[47]. 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. https://doi.org/10.21917/ijsc.2021.0336

[48]. Wong, T.T. (2015) Performance evaluation of classification algorithms by *k*-fold and leave-one-out cross validation. Pattern Recognition, 48(9): 2839-2846. https://doi.org/10.1016/j.patcog.2015.03.009

[49]. Mostafaei, K., Shahoo Maleki, S., Jodeiri Shokri, B., & Yousefi, M.(2023) Predicting gold grade by using support vector machine and neural network to generate an evidence layer for 3D prospectivity analysis. International Journal of Mining and Geo-Engineering, 57(4):435-444. https://doi.org/10.22059/IJMGE.2023.362951.595087

[50]. Moradi, M., Asghari, O., Norouzi, G.H., Riahi, M., & Sokooti, R. (2015). Joint Bayesian Stochastic Inversion of Well Logs and Seismic Data for Volumetric Uncertainty Analysis. International Journal of Mining and Geo-Engineering, 49(1): 131-142. https://doi.org/10.22059/IJMGE.2015.54636