Shahrood University of Technology

Journal of Mining and Environment (JME)

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

Iranian Society of Mining Engineering (IRSME)

Machine learning-based simulation of borehole grade identical twins from geophysical attributes: Comparative study of LR, GB, RF, and SVM in Kahang, Iran

Hassanreza Ghasemi Tabar[1], Sajjad Talesh Hosseini[2], Andisheh Alimoradi[2*], Mahdi Fathi[3], and Maryam Sahafzadeh[4]

1. Department of Mining, Petroleum and Geophysics, Shahrood University of Technology,Shahrood, Iran
2. Faculty Member, Department of Mining and Petroleum Engineering, Imam Khomeini International University, Qazvin, Iran
3. Senior Exploration Engineer, Kavoshgaran Consulting Engineers, Tehran, Iran
4. Senior Mining Engineer, SRK Consulting, , Vancouver, Canada

Article Info

Received 6 November 2024
Received in Revised form 30 April 2025
Accepted 13 May 2025
Published online 13 May 2025

DOI: 10.22044/jme.2025.15310.2936

Keywords

Linear regression
Gradient boosting
Random forest
Support vector machine
Identical twin

Abstract

Estimating ore grades during the exploration phase is often time-consuming and costly due to the need for extensive drilling. Geophysical surveys, as the last indirect exploration method before drilling, offer valuable insights into subsurface mineralization. This study introduces a novel approach for simulating “identical twins” of borehole copper grade values using geophysical attributes derived from the geoelectrical method in the Kahang porphyry copper deposit, central Iran. By treating the simulated values as digital twins of actual borehole grades, we employed four machine learning algorithms—Linear Regression (LR), Gradient Boosting (GB), Random Forest (RF), and Support Vector Machine (SVM)—to model the complex relationships between geophysical inputs and copper grades. After data preprocessing with Principal Component Analysis (PCA), a refined dataset was used to train, test, and validate each model. The results demonstrate that GB yielded the highest predictive accuracy, generating grade estimates closely aligned with actual values. This identical twin modeling approach highlights the potential of machine learning to enhance early-stage mineral exploration by reducing dependence on costly drilling.

1. Introduction

Generally, porphyry type reserves are the most important source of valuable metal elements such as copper and molybdenum, along with other important elements such as gold and silver in the world. According to available statistics, about 50 to 60 percent copper and more than 95 percent of global molybdenum production are obtained from these reserves. Most of these deposits have been expanded in metallogenic states known as porphyry copper belts. The porphyry copper belt of Iran, which is part of a large global copper belt with a length of four thousand kilometers, coincides with the northwest-southeast end with magmatic arc of Sahand Bazman or Orumiyeh Dokhtar. Due to the fact that the distribution of mineral content depends on many factors, the effects of many of these factors are not known and are not considered in the common mathematical model. Therefore, in each modeling, for the distribution of the ore grade, there is a simplification and assumptions about spatial variations. Distance is the only factor that is considered while other factors such as geological structure, formation environment, storage shape, ore type and degree of mineralization are also effective. Fortunately, all of these factors have an influence on geophysical data. The attractiveness of the intelligent techniques is due to the fact that they provide dynamic and nonlinear systems that are capable of learning. These techniques do not require assumptions about the factors affecting spatial variations around a borehole.

Corresponding author: alimoradi@eng.ikiu.ac.ir (A.Alimoradi)

As the historical trend, artificial neural network was used to estimate the storage of a deposit in central Iran by Shahabifar. A total of 57 wells are considered. To evaluate the designed network, four well's data were completely selected as validation data and the rest of the specimens were considered for train and test process. After training the neural network, the validation rate was about 73%. Tahmasebiand Khashkarani used neuro-fuzzy algorithm to estimate the copper grade in Kerman [1]. According to the results, the network has been able to accurately detect the spatial pattern between inputs and outputs and accurately evaluate the copper grade.

Karimi has used the multilayer perceptron neural network to evaluate the gold grade in Zarshouran gold mine [2]. Ghasemi Tabar and his colleagues have performed several algorithms, including Random Forest and Gradient Boosting on the northern anomaly of Choghart under the Python programming language [3].All of these works focused on the relation estimation between drill holes' grade values. Due to the availability of geophysical data and the existence of the relationship between geophysical data and bore hole values, the other scientists tried to find these relationships. Yuan and colleagues optimized the inversion of three models of geophysical data with the help of Particle Swarm Optimization and Ant Colony Optimization [4]. Alimoradi and his colleagues used a backpropagation network with 4 hidden layers to model a dyke with the inversion of the magnetometric data [5]. In another research with the same theme, with the help of a perceptron artificial neural network with two hidden layers, the modeling of the electro-seismic data of wells has been done [6]. FitzGerald tried to establish a relationship between airborne geophysical data and identify subsurface structures of the earth by using an artificial intelligence feature extraction technique [7]. As the most related work, Alimoradi and his colleagues evaluated the silver grade of the Zarshuran gold mine using drill spatial data, data from the Induced Polarization (IP) geophysical approach, and the cuckoo search machine learning algorithm. The findings indicated that grade values can be accurately approximated using geophysical data, particularly in locations without drilling data [8].

The purpose of this research study is to estimate the copper values using geophysical attributes and borehole coordinates. According to the limitation in number of layers, as well as the activation function limit and algorithm processing time in regular neural networks, it is necessary to use the other algorithms with high capabilities such as fast processing time, various activation functions, low layer or single-layer and high accuracy. Huang and his colleagues suggested linear regression (LR) to overcome these weaknesses [9]. Wang and his colleagues, has been studying the thickness of the ore with the help of seismic properties and random forest algorithm [10]. This has been a great deal of economic and time efficiency. The various functions of activation, fast processing time and the power of data analysis, which leads to generalization, were the reasons for choosing this method. Regarding the advantages of these algorithms and also two other algorithms (gradient boosting and support vector machine), the copper grade estimation of this research work has been investigated using LR, GB, RF and SVM algorithms.

2.Materials and Method

Kahang deposit is located in the central Iran, near Isfahan historical city. Figure 1 shows the location of Kahang deposit on the map as the red star. The area is completely mountainous. The satellite image of the area is shown in Figure 2.

The main petrological units of the area are: light brown to cream quartz monzo-diorite to quartz diorite rocks, the combination of quartz diorite with a weathered light brown and fresh light gray Eocene volcanic rocks, dacite lavas that can be seen in the form of single outcrops in the east of the area,dike, dome and thick dacite-rhyodacite lavas in light gray to pink color, different types of pyroxene-trachy-andesite, quartz-trachy-andesite and trachy-andesite. In addition to these features, the argillic and advanced argillic alterations in the area illustrates that the region is susceptible for porphyry deposit formation. After geological investigations done by the geology team, geophysical investigations were recommended to be performed. The main geophysical approach to check the target in porphyry type deposits is geo electrical method.

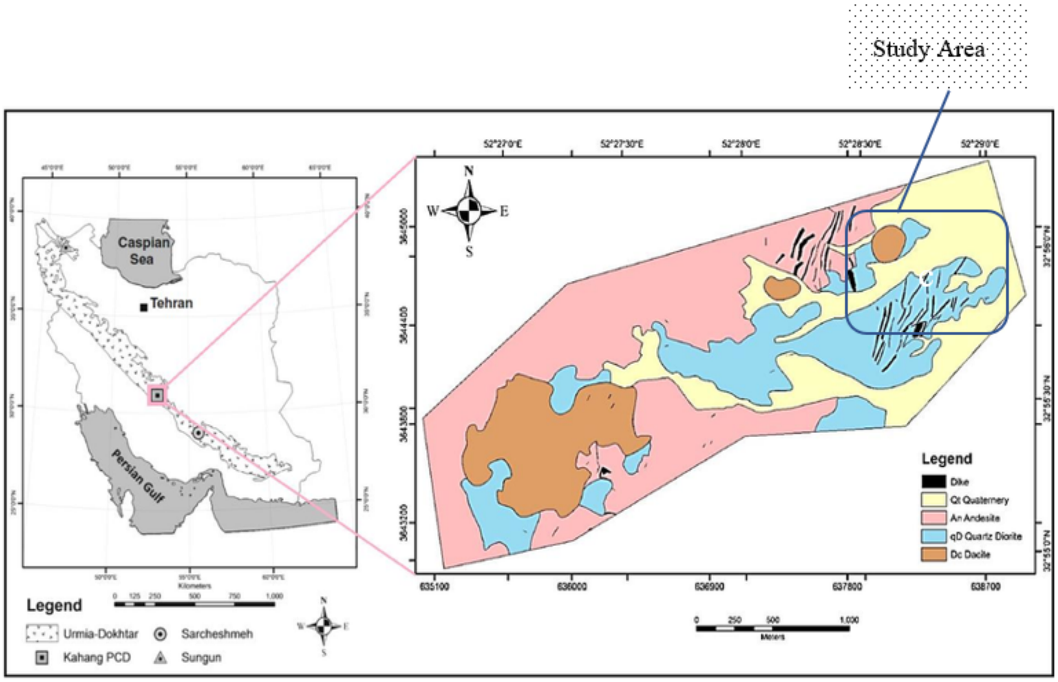

Figure 1. Geological map of Kahang porphyry copper deposit

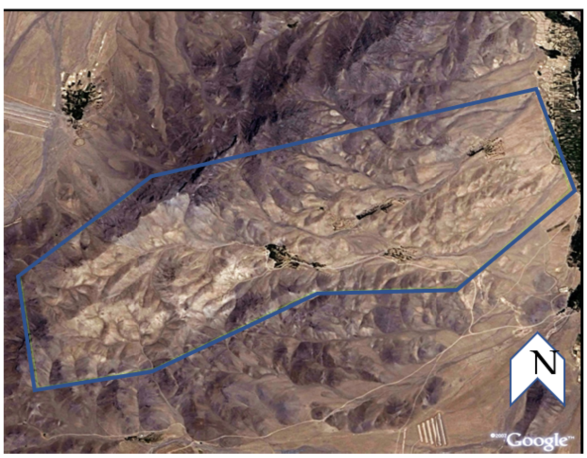

Figure 2. Satellite image of the area (Source: Google Earth)

The copper ore in the Kahang study area is not represented as a simple homogeneous horizontal layer. Instead, it is found within a complex geological matrix, where faults and fractures influence the mineralization. These geological complexities can significantly impact the effectiveness of the Vertical Electrical Sounding (VES) technique, which is typically more successful in areas with uniform subsurface layers. Given the heterogeneous nature of the subsurface geology at Kahang, the interpretation of VES data must consider these complexities to accurately assess the distribution of copper mineralization. Therefore, our study incorporates additional geophysical methods to complement the VES results and provide a more comprehensive understanding of the subsurface characteristics.

2.1. Geophysical Study

The basis of the geoelectrical test is the determination of the electrical resistivity of the earth. Materials have different electrical resistivity and can be defined with resistance test methods. According to Ohm's law, when an electric current I passes through a conductor with resistivity R, the relationship between the created potential (V) is V = RI and vice versa. If the potential difference V is applied between two conductors, then the current I passes through. For different materials, they vary in terms of potential. The type of geoelectrical method used in Kahang area is the method referred to as Vertical Electrical Sounding or VES.

At the deposit site, eight profiles with a distance of 100 meters and parallel to the northeast-southwest direction have been designed and surveyed, with 81 electrical soundings. The location of the VES boreholes is shown in Figure 3.

After geophysical survey, process and interpret, a grid of exploratory drillings was defined. Among drilled, sampled and analyzed boreholes,5boreholeswere drilled very closely to the vertical electrical soundings. Figure 4, shows the location of the VES and those 4 drill holes in the study area. We selected these5boreholes with the 5VESs beside them to define a data matrix as the raw data for intelligent inversion. The information that is obtained and put in the matrix are borehole coordinates, Cu grade, potential value(V), electrical current(I), geometric coefficient (K)and resistivity (P). The total 91 data set was finalized in which all variables were valid. Table 1 shows 23 rows of the data set from surface to the depth of 375 m as a sample. To prevent interference and lack of proper network identification, the principal component analysis (PCA) is used to improve network detection.

Figure 3. Position of the vertical electrical soundings (Source: Google Earth)

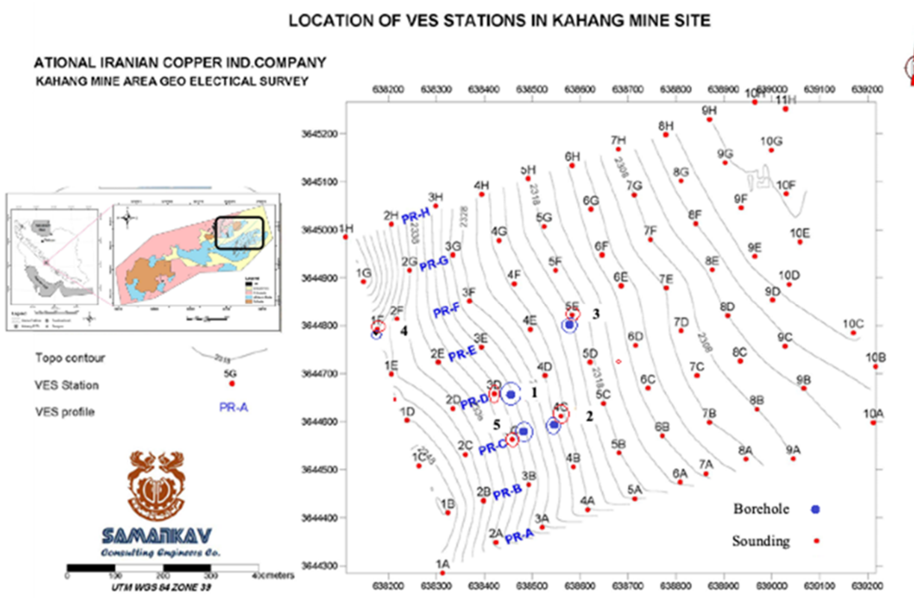

Figure 4. The location of the final selected boreholes along with the location of vertical electrical soundings

In this work, we utilized a dataset collected from the Kahang copper deposit. Each sample represents a unique data point that includes various geophysical attributes and corresponding copper grade values. The dataset was subjected to rigorous Quality Assurance/Quality Control (QA/QC) protocols to ensure the reliability and accuracy of the results. This included duplicate sampling, standard reference materials, and blank samples to monitor potential contamination and analytical errors.

The analytical methods employed for determining the copper grades were based on inductively coupled plasma mass spectrometry

(ICP-MS), which is recognized for its sensitivity and precision in trace element analysis. This robust dataset, with its stringent QA/QC measures and advanced analytical techniques, provides a solid foundation for applying machine learning algorithms to predict copper grades based on geophysical attributes. The insights gained from this dataset not only enhance our understanding of the Kahang deposit but also contribute to the broader field of mineral exploration.

Table1.Sample data from the total data set

	X	Y	Z	K	V	I	P	Grade (%)
1	638557	3644608	1.5	12.6	39.0	6.3	77.7	0.0000
2	638557	3644608	2.5	37.7	28.0	14.1	65.6	0.0000
3	638557	3644608	3.5	75.4	17.0	14.8	86.6	0.0194
4	638557	3644608	5.0	155.4	7.0	14.8	73.5	0.0228
5	638557	3644608	5.0	58.9	17.0	14.0	71.5	0.0228
6	638557	3644608	7.5	137.4	14.0	27.1	71.0	0.0343
7	638557	3644608	10.0	347.3	8.0	27.0	73.3	0.0765
8	638557	3644608	15.0	56.1	3.0	26.2	62.2	0.0320
9	638557	3644608	25.0	1566.0	2.3	59.0	61.0	0.0222
10	638557	3644608	25.0	377.0	11.0	60.4	71.2	0.0222
11	638557	3644608	35.0	3073.0	1.2	59.1	65.1	0.0291
12	638557	3644608	35.0	754.0	6.0	59.5	76.3	0.0291
13	638557	3644608	50.0	1554.0	1.2	22.6	82.6	0.0262
14	638557	3644608	75.0	3517.0	1.2	46.3	91.1	0.0968
15	638557	3644608	100.0	6264.0	0.5	37.9	82.7	0.0643
16	638557	3644608	100.0	1507.0	2.8	38.1	110.7	0.0643
17	638557	3644608	150.0	14114.0	0.5	77.5	92.0	0.0092
18	638557	3644608	150.0	3470.0	1.3	38.2	118.0	0.0092
19	638557	3644608	200.0	25104.0	0.5	96.1	130.6	0.0972
20	638557	3644608	200.0	6217.0	2.3	95.8	149.3	0.0972
21	638557	3644608	250.0	9750.0	2.1	125.0	163.8	0.1347
22	638557	3644608	300.0	3073.0	6.4	155.0	178.4	0.0063
23	638557	3644608	375.0	22015.0	1.3	155.0	184.6	0.1467

2.2. Preprocessing

The preliminary statistical analysis of the data utilized is shown in Table 2. According to the measurable characteristics such as the scattering of the data utilized in this table appears that most data and the scattering within the input data are related to Cu grades less than 0.14%, and within the data with grades higher than this grade, the number and scattering of the data The accessible ones are less. In this way, it can be predicted that less precision will be achieved within the data with a higher Cu grade. Also, the information in Table 3 shows the correlation between the input data and the Cu grade, which as shown is the foremost correlation on the spatial data.

Table 2. Statistical parameters of the studied data

	X	Y	Z	K	I	V	P	Cu (%)
mean	638426	3644640	86.81	4276.51	8.62	46.77	112.23	0.13
std	178.156	66.1024	95.35	6455.8	26.09	37.98	48.28	0.21
min	638019	3644578	1.5	12.6	0.4	6.3	54.77	0
25%	638420	3644608	15	377	1.2	15.75	73.38	0.03
50%	638454	3644620	35	1554	2.3	36	104.5	0.07
75%	638557	3644654	150	3517	6	81.2	129.6	0.14
max	638582	3644788	375	25104	187	155	264.7	1.14

Table 3. Correlations (percent) in inputs and Cu of the studied data

	X	Y	Z	K	I	V	P
Cu (%)	-0.26	0.26	0.02	0.07	-0.12	0.07	0.07

Multiple data analysis has a fundamental role in data science. If there are variables in each dataset, each variable can have multiple dimensions. Given that it is difficult to comprehend multiple-dimensional space, the Principal Component Analysis (PCA)reduces the dimensions of all observations based on the combined index and classification of similar observations. In this method, variables in a multimode space are summed up to a set of unconnected components, each of which is a linear combination of the main variables. The unconnected or less depended components are called core components of the PCA, derived from the special covariance matrices or the correlation matrix of the main variables. In general, the main application of the principal component analysis method is to reduce the number of variables and find the structure of the relationship between them.

The criteria used in this study to find the necessary components are:

- Coordinates of the data as X, Y, and Z which are necessary as the input data.
- Even K as the geometric coefficient will affect the geoelectrical parameters, but it should be considered as input data since different K values can give similar geoelectrical parameters.
- For other 3 geophysical variables, it is obvious that they are dependent on each other. The less dependent variable should be selected as the fifth input data.

Table4, shows the dependency of the variables to each other. According to the results of this table, it can be concluded that V (field potential) should be selected as the final input variable.

Table 4. PCA results

PC	1st Factor	2nd Factor	3rd Factor	4th Factor	5th Factor	6th Factor	7th Factor
1	0.00	0.00	0.00	0.00	0.00	0.99	0.07
2	0.00	0.00	0.00	0.00	0.00	-0.07	0.99
3	0.59	-0.31	0.00	-0.05	0.73	0.00	0.00
4	0.52	0.78	0.33	0.00	-0.03	0.00	0.00
5	-0.15	-0.19	0.55	0.77	0.00	0.00	0.00
6	0.56	-0.31	-0.28	0.38	-0.59	0.00	0.00
7	0.20	-0.39	0.70	-0.49	-0.25	0.00	0.00

Figure 5 shows the scattering plot of K, P,I and V. As it is shown in these plots, I and P have a regular trend by increasing the depth, but V has the unknown distribution like K.

Finally, and according to the PCA results, the input data were reduced to X, Y, Z, K and V and copper grade was considered as the output. Table5illustrates the final input and output databased on table 1.

Table5. Sample of the final data set which should be used in machine learning algorithm

	X	Y	Z	K	V	Grade (%)
1	638557	3644608	1.5	12.6	39.0	0.0000
2	638557	3644608	2.5	37.7	28.0	0.0000
3	638557	3644608	3.5	75.4	17.0	0.0194
4	638557	3644608	5.0	155.4	7.0	0.0228
5	638557	3644608	5.0	58.9	17.0	0.0228
6	638557	3644608	7.5	137.4	14.0	0.0343
7	638557	3644608	10.0	347.3	8.0	0.0765
8	638557	3644608	15.0	56.1	3.0	0.0320
9	638557	3644608	25.0	1566.0	2.3	0.0222
10	638557	3644608	25.0	377.0	11.0	0.0222
11	638557	3644608	35.0	3073.0	1.2	0.0291
12	638557	3644608	35.0	754.0	6.0	0.0291
13	638557	3644608	50.0	1554.0	1.2	0.0262
14	638557	3644608	75.0	3517.0	1.2	0.0968
15	638557	3644608	100.0	6264.0	0.5	0.0643
16	638557	3644608	100.0	1507.0	2.8	0.0643
17	638557	3644608	150.0	14114.0	0.5	0.0092
18	638557	3644608	150.0	3470.0	1.3	0.0092
19	638557	3644608	200.0	25104.0	0.5	0.0972
20	638557	3644608	200.0	6217.0	2.3	0.0972
21	638557	3644608	250.0	9750.0	2.1	0.1347
22	638557	3644608	300.0	3073.0	6.4	0.0063
23	638557	3644608	375.0	22015.0	1.3	0.1467

2.3. Grade estimation using machine learning

Artificial Neural Networks (ANNs) are inspired by human brain function and artificial processor units. These models are based on the assumption that the human brain can be learnt by the neural units (neurons).

In the present study, finalized dataset from the PCA processing will be used in four machine learning algorithms (Linear Regression, Gradient Boosting, Random Forest and Support Vector Machine) to find the relationship between geoelectrical attributes and the real values of copper grade. The method of work is to randomly divide the data set into train (70%), test (20%) and validation (10%)sets. These sets should be normalized before using in each algorithm.

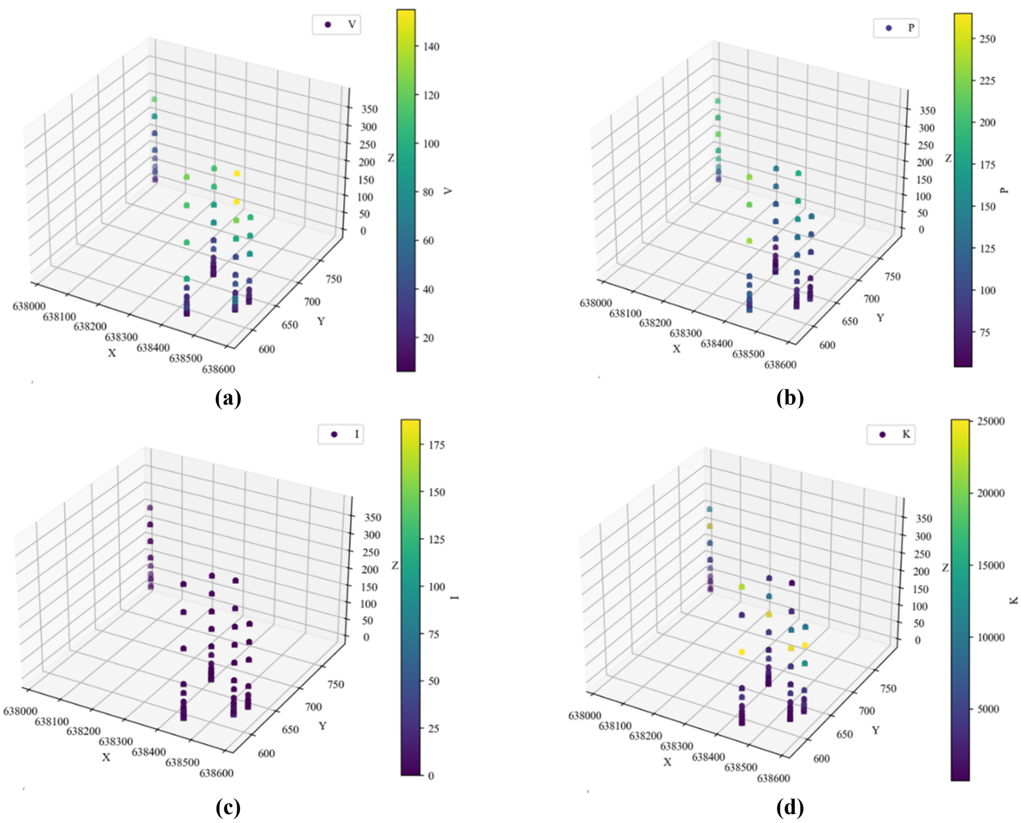

Figure 5. Scattering plot of the data (Source: Python 3.6)

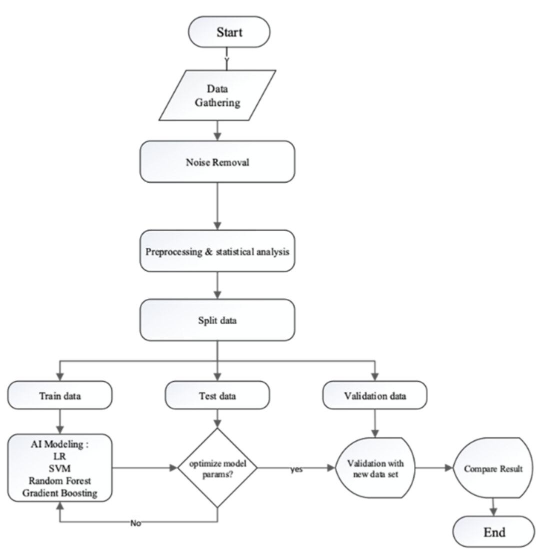

Figure 6. Flow chart of the modeling

Further, the results of each algorithm will be shown and discussed. Finally, the best algorithm for this study will be defined.

2.4. Machine learning algorithms

In this work, we employed four machine learning algorithms: Linear Regression (LR), Gradient Boosting (GB), Random Forest (RF), and Support Vector Machine (SVM) for VES data analysis. The selection of these algorithms is grounded in their diverse methodologies and proven effectiveness in managing complex and high-dimensional datasets commonly encountered in geophysical applications.

Linear Regression (LR) serves as a foundational linear model that is particularly suitable for binary classification tasks, offering interpretability and simplicity in understanding the relationship between input features and the target variable. However, its linearity may limit its performance in capturing more intricate patterns within the data.

On the other hand, Gradient Boosting (GB) and Random Forest (RF) are advanced tree-based ensemble methods known for their robustness in modeling nonlinear relationships and interactions among features. GB builds models sequentially, optimizing for errors made by previous models, which allows it to achieve high predictive accuracy. Conversely, RF constructs multiple decision trees and aggregates their predictions, providing resilience against overfitting and improving generalization on unseen data.

Support Vector Machine (SVM) is recognized for its capability to classify data by finding the optimal hyperplane that separates different classes, even in high-dimensional spaces. Its strength lies in handling nonlinear decision boundaries through the use of kernel functions, making it particularly effective for complex classification tasks. By comparing these algorithms, we aim to identify their respective strengths and weaknesses in the context of VES data analysis. This comparative approach not only enhances our understanding of how different features influence performance but also allows us to select the most appropriate algorithm for specific geophysical scenarios. Ultimately, our findings will contribute to more accurate interpretations of subsurface characteristics, facilitating better decision-making in mineral exploration.

2.4.1. Linear regression

Linear regression is a supervised statistical algorithm which looks for the best relationship between some dependent and independent variables in a linear manner. The functions are based on back propagation algorithms and almost this method is suitable for the problems with considerable input data. Figure 7 illustrates the function of the linear regression algorithm.

2.4.2. Gradient boosting

Gradient boosting is a powerful algorithm based on decision tree (DT) function which is capable in noise data controlling and finding the nonlinear model between data [27]. This algorithm looks for the weak points in the model and try to boost them in order to get the best result. It builds the model in a stepwise manner like other booster methods and generalizes the variable function of the decision tree by allowing arbitrary optimization. The Gradient Boosting is an integrated high performance and stable algorithm [15]. The training process and progress of the enhanced gradient algorithm based on the error function and repetition of the training process are shown in Figure 8.

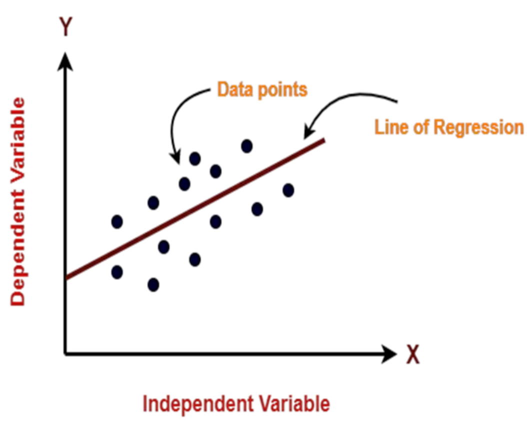

Figure 7. Linear regression function [11]

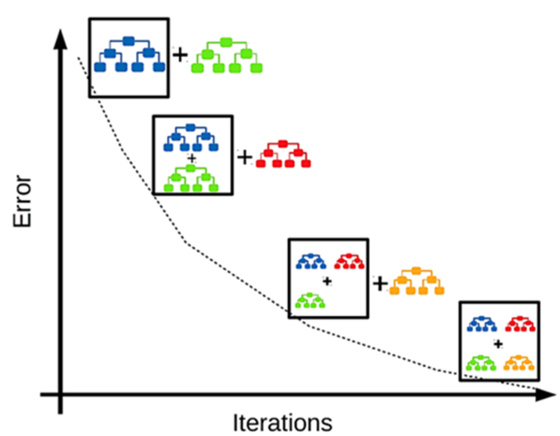

Figure 8. Gradient boosting training process [11]

2.4.3. Random forest

Random forest or Random decision forest is a hybrid learning method for classification and regression, which is based on a structure consisting of many decision trees, on the training time and the output of classes (classification), or for the average predictions of each tree [12]. Random Forest is suitable for decision trees that undergo pre-fitting in the training set. Although this algorithm is very user-friendly and has only two input parameters of the network, which are the number of trees and the number of variables of subsets, it is not highly sensitive to the value of these parameters [13]. On the other hand, the random forest algorithm (Figure 9) is a tree-based algorithm that uses the features of several decision trees to make decisions. In fact, this algorithm uses averaging to improve performance and control overfitting [14].

2.4.4. Support vector machine

Support vector machine is a supervised learning algorithm capable of providing generalization performance over a wide range of problems. This method is one of the machine learning methods which can perform the tasks of binary classification and regression estimation. This algorithm performs the classification by building an n-dimensional hyperplane that optimally divides the data into two or more separate categories, which maximizes the margin of separation between different categories. In other words, the super plane determines the separation of different categories in such a way that each category has the largest distance from others [16]. In Figure 10, the procedure of the support vector machine algorithm in differentiating groups of similar data has been shown.

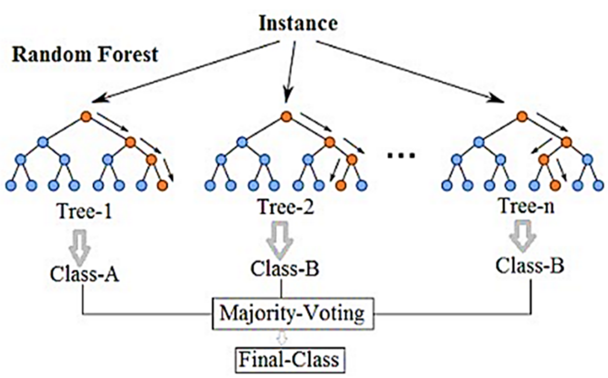

Figure 9. Random forest performance procedure [15]

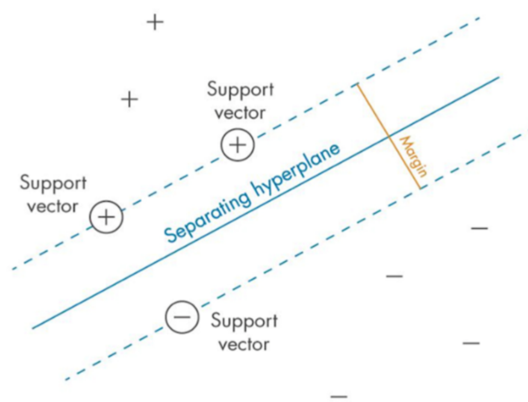

Figure 10. Support vector machine performance [17]

3. Results and Discussion

Figure 11 shows the results of training for each machine learning model. As it can be seen, gradient boosting has the best performance in training data set and all real and predicted values completely overlapped, with regression value about 1. After training, each model should be applied for the test data set to check the compatibility of models in predicting copper grades from input data which have not involved in the training procedure. By doing this stage, each model can be modified in training parameters to get the best test estimation values. Figure 12 shows the results of testing each network.

According to the test results, it is clear that all four algorithms have almost similar performance in predicting test data. The weak point of all models is to predict the value of sample with copper grade more than 1. Table 6 explains this problem. As it can be seen, more than 75 percent of the real grades are less than 0.04 %. Also checking whole data set illustrates that less than 10 percent of the data have assay value more than 1 %. This will influence the model performance specifically for test and validation data. There are two scenarios for this issue: The first scenario is to keep these values and accept the model performance with mentioned problem in detecting these values. The second scenario is to eliminate these high values. Although models will get better fitting to the real data in this scenario, but the main problem is the elimination of near 10% of the data with values that affect the final reserve estimation model. Also this will reduce the number of data less than 91. Since this study is a research work to show the compatibility of machine learning in estimating grade values from geophysical attributes, we chose the first scenario, but will give some comments to enhance the model performance in high grade values in concluding remarks chapter.

Finally, the models applied to the validation data set which are unknown for the algorithms. This process shows the robustness of the models in predicting data from other geophysical surveys but in the same area. Figure 13 shows the results of applying models on validation data.

Table 6. Statistical parameters of the modeling data

	Cu (%)	Predicted Cu (%)
Mean	0.13	0.11605178
Std	0.21	0.170844185
Min	0	0
25%	0.035	0.04
50%	0.07	0.07
75%	0.14	0.14
max	1.14	1.14

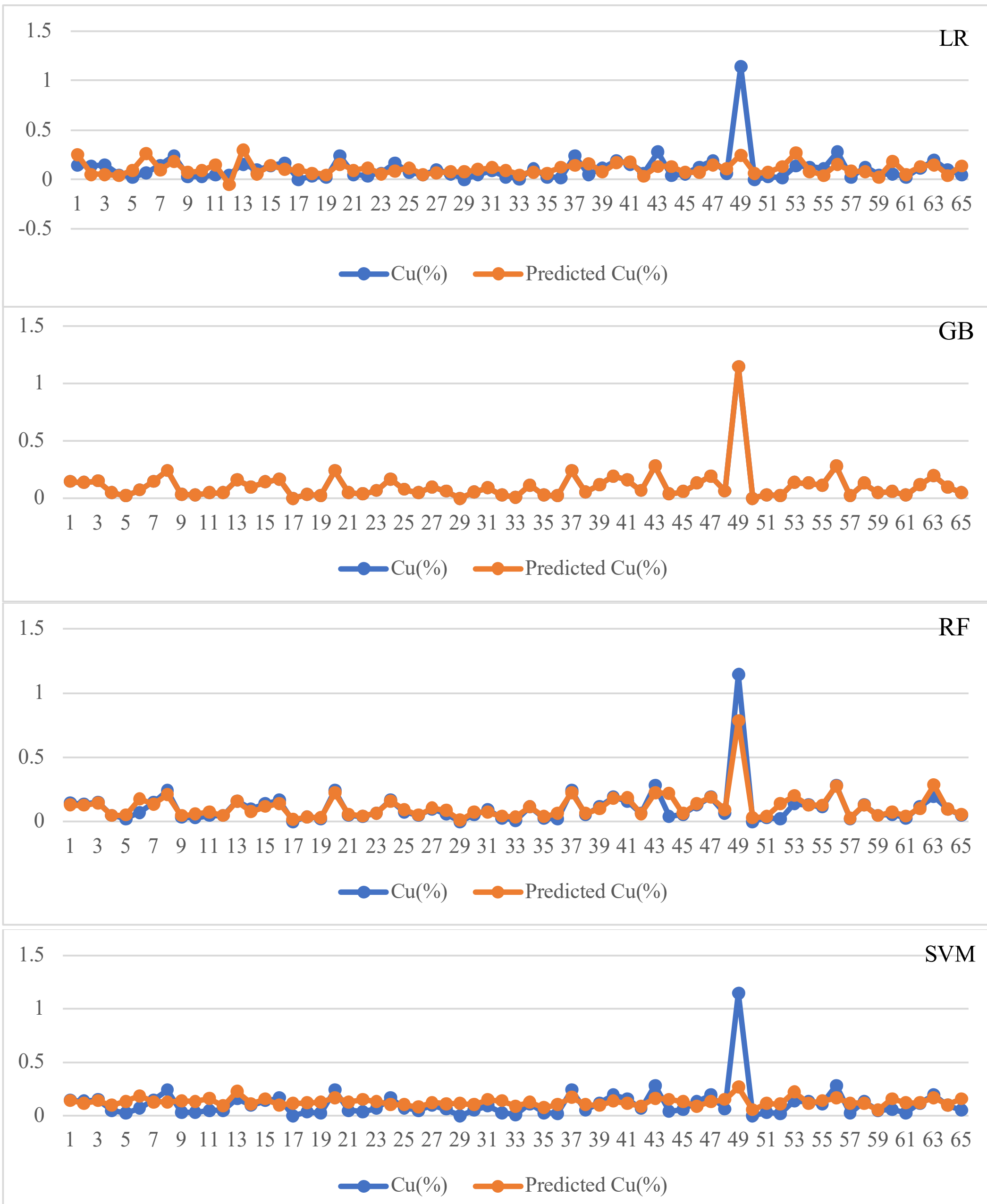

Figure 11. Results for training data set

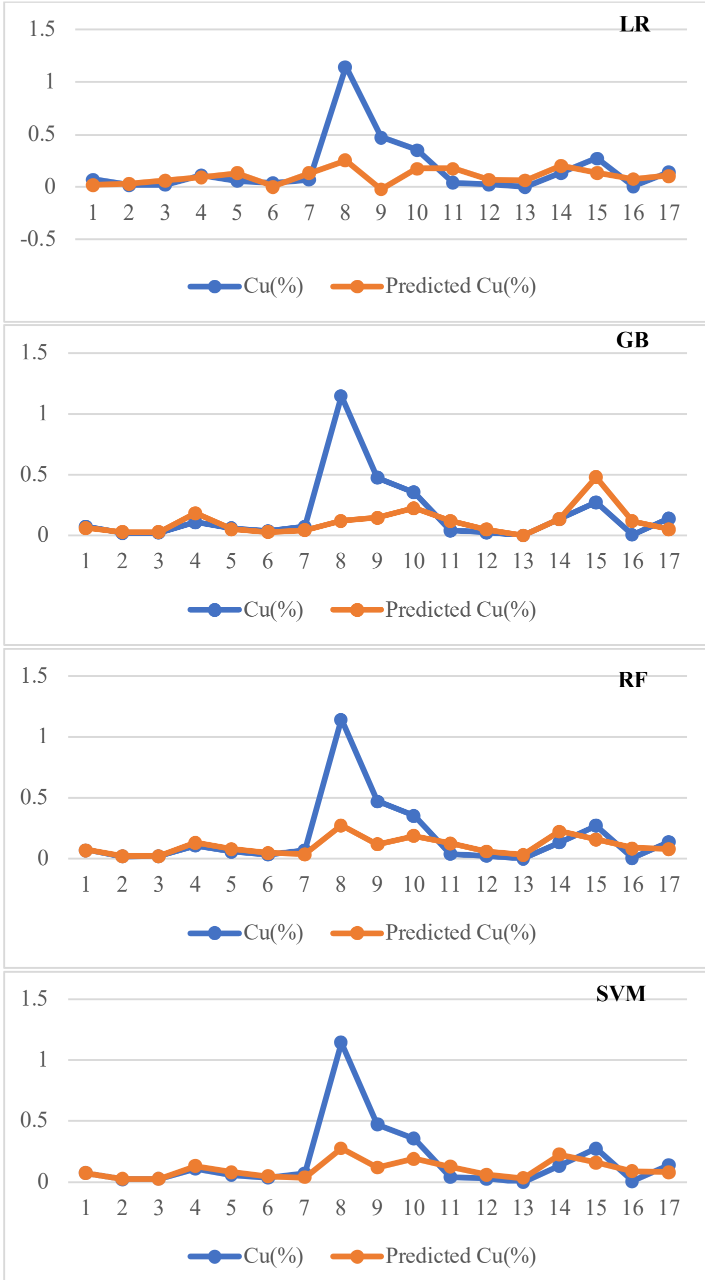

Figure 12. Results for testing data set

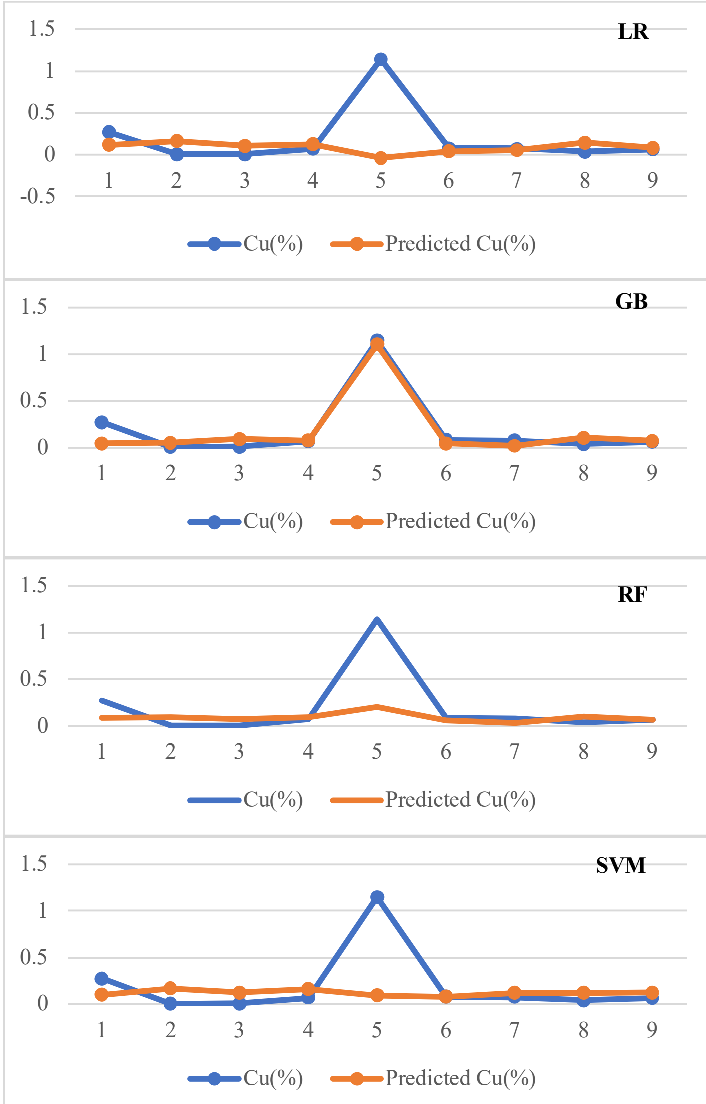

Figure 13. Results for validation data set

Table 7 shows a brief result for any model in the form of error and regression values. This table also confirm the graphical results shown in previous figures and illustrates that the best algorithm in this modeling is gradient boosting. Except the data with high grade values which are not enough in quantity to be completely known for models in this study, all grade values can be estimated accurately by this algorithm. Figure 14 also shows the scattering and spatial accuracy of the results predicted by the GB algorithm.

Table7. Brief results of four metaheuristic models

Model		Mean absolute error	Mean squared error	Median absolute error
Linear regression	**Train**	**0.07**	**0.01**	**0.05**
	Test	**0.12**	**0.05**	**0.06**
	Valid	**0.18**	**0.12**	**0.09**
SVM	Train	0.07	0.01	0.06
	Test	0.14	0.05	0.08
	Valid	0.17	0.10	0.08
Random forest	Train	0.02	0.00	0.01
	Test	0.10	0.04	0.03
	Valid	0.14	0.08	0.05
Gradient boosting regressor	**Train**	**0.00**	**0.00**	**0.00**
	Test	0.11	**0.06**	**0.02**
	Valid	**0.05**	**0.01**	**0.04**

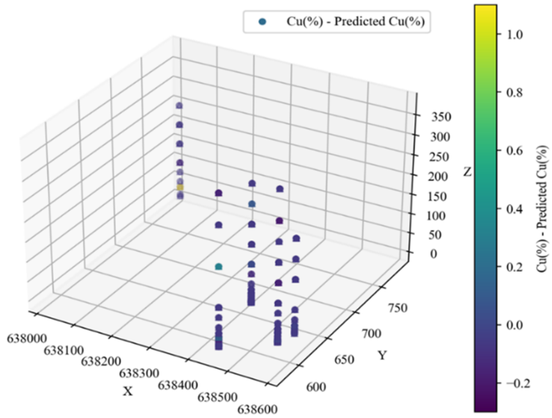

Figure 14. Result of GB algorithm (Source: Python 3.6)

4. Conclusions

Exploration drilling to get direct samples and real grade values of the ore of interest exerts some difficulties such as increasing time and cost in performing reserve estimation. Geophysical data, as the last indirect data in any exploration project, have invaluable mathematical based information about the deposit. In this paper, we successfully implemented and tested four machine learning computational agent (LR, GB, RF and SVM) to consider the unknown nonlinear relationships between geophysical attributes and copper ore grade in our prediction problem. Our approach uses coordinates, geometric coefficient and field potential as input system variables. Intelligent machines seek the relationship between these input variables adaptively and strives to a desirable output which is, in our case, the real copper grade values obtained from the direct sampling after exploration core drilling.

We considered a newly explored deposit for testing our methodology. Kahang cooper deposit case showed that the gradient boosting machine learning algorithm could train itself very well with practically complete correlation between real copper grade values and the predicted ones (correlation coefficient R of almost one). The algorithm also exhibited a remarkable capability in estimating test and validation data, even though the error values increased a bit in test process for which we speculate the followings as possible reasons for this peculiarity:

- The number of data for this modeling was 91 data set which is low to give the best result.
- Among 91 data, less than 10 percent of the data have grade values more than 0.4 % which makes the training process of the machine some difficult.

The remedy would be obtaining more data and increase the number of valid data set (with logic distribution between high, medium and low grades), then augmenting the training of the algorithm with the new dataset. We speculate this would enhance the accuracy of the algorithm predictions for the test and validation data and even increase the generalization of the machine in the area of interest.

our findings from the Kahang copper deposit not only demonstrate the effectiveness of the Gradient Boosting algorithm in predicting copper grade values but also provide valuable insights for similar geophysical exploration projects. The strong correlation observed between the predicted and actual copper grades underscores the potential of using machine learning techniques to improve reserve estimation processes in various geological settings. While this study focuses on the Kahang deposit, the methodologies and insights derived can be applied to other mineral exploration sites with comparable geophysical attributes. Future research should aim to validate our approach across different deposits, which would further establish the robustness and adaptability of our machine learning framework in diverse geological environments. By expanding the dataset and exploring additional case studies, we can enhance the generalizability of our findings and contribute to more efficient exploration practices in the mining industry.

Acknowledgment

Supporting for the information and data of this project by Kavoshgaran Consulting Engineers (KCE) is appreciated. The comments received from and the enlightening discussions with our anonymous reviewers improved the paper are appreciated.

Reference

[1]. Tahmasebi, P., & Khashkarani, A. (2009). A method for optimization of the neural network for estimating grade using the information system of porphyry copper system of Saunajil-Ahar. Year 21, No. 81, 31–36.

[2]. Karimi, A. (2016). Determination of geophysical indicators of IP/RS relating to precise values and estimates in Zarshuran gold mine using artificial neural network technique. Master's thesis, Faculty of Engineering, Imam Khomeini International University.

[3]. Ghasemi Tabar, H. R., Alimoradi, A., Hemmati Ahooi, H. R., Fathi, M., & Sarookhani, M. (2024). Intelligent borehole simulation with Python

programming.*Journal of Mining and Environment, 15 (2)*.

[4]. Yuan, S., Wang, S., & Tian, N. (2009). Swarm intelligence optimization and its application in geophysical data inversion. *Applied Geophysics, 6 (2), 166–174*.

[5]. Alimoradi, A., Angorani, S., Ebrahimzadeh, M., & Shariat Panahi, M. (2011). Magnetic inverse modelling of a dike using the artificial neural network approach.*Near Surface Geophysics, 9 , 339–347.*

[6]. Ardjmandpour, N., Pain, C., Singer, J., Saunders, J., Aristodemou, E., & Carter, J. (2011). Artificial neural network forward modelling and inversion of electrokinetic logging data.*Geophysical Prospecting, 59 , 721–748.*

[7]. FitzGerald, D. (2019). Artificial intelligence techniques to the interpretation of geophysical measurements.*ASEG Extended Abstracts, 1 , 1–5*.

[8]. Alimoradi, A., Maleki, B., Karimi, A., Sahafzadeh, M., & Abbasi, S. (2020). Integrating geophysical attributes with new cuckoo search machine-learning algorithm to estimate silver grade values–case study: Zarshouran gold mine.*Journal of Mining and Environment, 11 (3), 865–879.*

[9]. Huang, G., Yu, Z., & Kheong Sew, C. (2006). Extreme learning machine: Theory and applications.*Neurocomputing, 70 , 489–501*.

[10]. Wang, X., Li, Y., Chen, T., Yan, Q., & Ma, L. (2017). Quantitative thickness prediction of tectonically deformed coal using extreme learning machine and principal component analysis: A case study. *Computers and Geosciences, 101, 38–47*.

[11]. Breiman, L. (2001). Random forests.Machine Learning, 45(1), 5–32.

[12]. Fukunaga, K., & Hostetler, L. D. (1975). The estimation of the gradient of a density function, with applications in pattern recognition.*IEEE Transactions on Information Theory, 21 (1), 32–40.*

[13]. Yizong, C. (1995). Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 17 (8), 790–799.*

[14]. Dorin, C., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis.IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5), 603–619.

[15]. Simorgh, M., Alimoradi, A., Hemmati Ahooi, H., Salsabili, M., Fathi, M., Ghasemi Tabar, H., & Rezakhani, P. (2023). Application of supervised machine learning inversion in the estimation of iron ore grade from geophysical data: Comparative investigation of GB, RF and SVM algorithms. *Journal of Geomine, 1 (3), 123–136.*https://doi.org/10.22077/jgm.2024.7358.1018

[16]. Ashok Kumar, L. (2020). Proceedings of international conference on artificial intelligence, smart grid and smart city applications.*In Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*.

[17]. Liu, H., Wen, S., Li, W., Xu, C., & Hu, C. (2009). Study on identification of oil/gas and water zones in geological logging based on support-vector machine. *Fuzzy Information and Engineering, 2 , AISC 62, 849–857.*

[18]. Automation, P., & Khashkarani, A. (2011). The use of neural-fuzzy-genetic networks to estimate characteristics in porphyry copper deposit in Zar-Kerman Valley.*Volume 6, No. 12 , 1–9.*

[19]. Azadi, M., & Mirmohammadi, M. (2014).*Geometric-Genetic and Mineralogical Classification of Veinlets and Breccias in Kahang Porphyry Copper Deposit, Northern East Isfahan, Iran*. https://www.researchgate.net/publication/259010665

[20]. Embaby, A., Ismael, A., Faisal, A. Ali, H., Farag, A., Mousa, B.G., Gomaa, S., Elwageeh, M. (2023). An approach based on Machine Learning Algorithms, Geostatistical *Technique, and GIS analysis to estimate phosphate ore grade at the Abu Tartur Mine, WesternDesert, Egypt. Mining of Mineral Deposits, Volume 17, Issue 1, pp.* 108 – 119,.

[21]. Fathi, M., Alimoradi, A., & Hemmati Ahooi, H. (2021). Optimizing the extreme learning machine algorithm using particle swarm optimization to estimate iron ore grade.*Journal of Mining and Environment, 12 (2), 397–411*.

[22]. Hassani Pak, A. A. (2001).Exploratory data analysis. Tehran: Tehran University Press.

[23]. Ismael, A., Embaby, A., Faisal A. Ali, H. A. Farag, B. G., Gomaa, S., Elwageeh, M. (2024). Prediction of Iron Ore Grade using Artificial Neural Network, Computational Method, and Geostatistical Technique at El-Gezera Area, Western Desert, Egypt. *Journal of Mining and Environment (JME), Volume 15, Issue 3, May 2024, Pages 889-905,*

[24]. Jolliffe, I. T. (1986). Principal component analysis and factor analysis.*In Principal Component Analysis (Chapter 7). Springer, New York.*

[25]. Kalagari, A. S. (1992).Principles of geophysical exploration.*Radish Publishing.*

[26]. Kia, M. (2011).Neural networks in MATLAB. *Kian Publishing.*

[27]. Li, X., Xie, Y., Guo, Q., & Li, L. (2010). Adaptive ore grade estimation method for the mineral deposit evaluation. Mathematical and Computer Modelling, 52(11–12), 1947–1956. https://doi.org/10.1016/j.mcm.2010.04.018

[28]. Luo, X., Yang, X., Chang, X., & Zhang, C. (2015). Prediction of hidden dangers in mine production using

timeliness managing extreme learning machine for cloud services. *IEEE Computer Society*. https://doi.org/10.1109

[29]. Mahdi Farh, S., & Kamali, G. (2013). Estimation of iron grade in anomalous of the red hill of Sangan iron ore using three comparative neural fuzzy algorithms. No. 5, 1–10.

[30] .Mannhaj, B. M. (2014). Fundamentals of neural networks. *Amirkabir University of Technology.*

[31]. Narrosi, G. H. (2016). Exploratory geophysics. Tehran University Press.

[32]. Shahibi Far, M. (2004). Estimation of reservoir deposit with artificial neural network technique. *Iranian Mining Engineering Conference*, 12-14 Bahman, Tarbiat Modarres University, Iran.

[33]. Shirani Bidabadi, B. (1996).Multi-stage transformation petrology of skarns in Isfahan pressure zone. Ph.D. thesis, Tabriz University.

[34]. Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. Chemometrics and Intelligent Laboratory Systems.

[35]. Zahidi, M. Geological map of Kashan. Geological Survey of the Country.

[36]. Zooseri, M., Amami, N. H., Hariun, A., & Amani, H. M. (2007). Petrography and fault of the alteration zones of the volcanic volcanic aggregate complex. *Third Conference on Applied and Environmental Geology , Dec. 20, Tehran, Iran.*

انجمن مهندسی معدن ایران

نشریه مهندسی معدن و محیط زیست

نشانی نشریه: www.jme.shahroodut.ac.ir

دانشگاه صنعتی شاهرود

شبیه‌سازی مبتنی بر یادگیری ماشین با استفاده از روش نمونه‌های همسان برای مدل سازی عیار گمانه‌های اکتشافی با استفاده از داده‌های ژئوفیزیکی: مطالعه مقایسه‌ای روش‌های رگرسیون خطی، گرادیان تقویت شده، جنگل تصادفی و ماشین بردار پشتیبان در منطقه کهنگ، ایران

حسن رضا قاسمی تبار[1]، سجاد تالش حسینی[2]، اندیشه علیمرادی[2*]، مهدی فتحی[3]، مریم صحاف زاده[4]

۱. دانشجوی دکتری، دانشکده معدن، نفت و ژئوفیزیک، دانشگاه صنعتی شاهرود، ایران

۲. عضو هیئت علمی، دانشکده معدن و نفت، دانشگاه بین‌المللی امام خمینی، ایران

۳. مهندس ارشد اکتشاف، مهندسین مشاور کاوشگران، ایران

۴. مهندس ارشد معدن، شرکت مشاوره SRK، کانادا

اطلاعات مقاله

تاریخ ارسال:

تاریخ داوری:

تاریخ پذیرش:

DOI: 10.22044/jme.2025.15310.2936

کلمات کلیدی

رگرسیون خطی

نگرادیان تقویت شده

جنگل تصادفی

ماشین بردار پشتیبان

دوقلوهای همسان

چکیده

برآورد عیار کانسنگ در مرحله اکتشاف مواد معدنی به‌دلیل نیاز به حفاری‌های گسترده معمولاً زمان‌بر و پرهزینه است. روش‌های ژئوفیزیکی به‌عنوان آخرین روش اکتشاف غیرمستقیم پیش از حفاری، اطلاعات ارزشمندی درباره کانی‌زایی زیرسطحی ارائه می‌دهند. در این مطالعه، رویکردی نوین با استفاده از شبیه‌سازی «دوقلوهای همسان» برای مدل سازی مقادیر عیار مس در گمانه های اکتشافی با استفاده از ویژگی‌های ژئوفیزیکی بدست آمده از روش ژئوالکتریکی در معدن مس پورفیری کهنگ واقع در مرکز ایران معرفی شده است. با در نظر گرفتن مقادیر شبیه‌سازی‌شده به‌عنوان دوقلوهای دیجیتال عیار واقعی گمانه ها، چهار الگوریتم یادگیری ماشین شامل رگرسیون خطی (LR)، گرادیان تقویت شده (GB)، جنگل تصادفی (RF) و ماشین بردار پشتیبان (SVM) برای مدل‌سازی روابط پیچیده بین ورودی‌های ژئوفیزیکی و عیار مس به‌کار گرفته شدند. پس از پیش‌پردازش داده‌ها با تحلیل مؤلفه‌های اصلی (PCA)، مجموعه داده بهینه‌شده برای آموزش، آزمون و اعتبارسنجی هر مدل استفاده گردید. نتایج نشان داد که الگوریتم گرادیان تقویت شده بالاترین دقت پیش‌بینی را داشته و برآوردهای عیار آن به مقادیر واقعی بسیار نزدیک بود. این رویکرد مدل‌سازی دوقلوهای همسان، پتانسیل روش های یادگیری ماشین را در بهبود اکتشافات معدنی، در مراحل اولیه و کاهش وابستگی به حفاری‌های پرهزینه برجسته می‌سازد.

نویسنده مسئول مکاتبات: *alimoradi@eng.ikiu.ac.ir* (ا. علیمرادی)