Shahrood University of Technology

Journal of Mining and Environment (JME)

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

Iranian Society of Mining Engineering (IRSME)

Copper Ore Grade Prediction using Machine Learning Techniques in a Copper Deposit

Jairo Marquina-Araujo[1]*, Marco Cotrina-Teatino[1], José Mamani-Quispe[2], Eduardo Noriega-Vidal[1], Juan Vega-Gonzalez[3], and Juan Cruz-Galvez[3]

1. Department of Mining Engineering, Faculty of Engineering, National University of Trujillo, Trujillo, Peru
2. Department of Chemical Engineering, Faculty of Engineering, National University of the Altiplano of Puno, Puno, Perú
3. Department of Metallurgical Engineering, Faculty of Engineering, National University of Trujillo, Trujillo, Perú

Article Info

Received 6 January 2024
Received in Revised form 23 January 2023
Accepted 27 January 2024
Published online 27 January 2024

DOI: 10.22044/jme.2024.14032.2617

Keywords

Multi-layer perceptron artificial neural network
Random forests
Extreme gradient boosting
Support vector regression

Abstract

The objective of this research work to employ machine learning techniques including Multilayer Perceptron Artificial Neural Networks (ANN-MLP), Random Forests (RFs), Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR) to predict copper ore grades in a copper deposit located in Peru. The models were developed using 5654 composites containing available geological information (rock type), as well as the locations of the samples (east, north, and altitude) and secondary ore grade (Mo) obtained from drilling wells. The data was divided into 10% (565 composites) for testing, 10% (565 composites) for validation, and 80% (4523 composites) for training. The evaluation metrics included SSE (Sum of Squared Errors), RMSE (Root Mean Squared Error), NMSE (Normalized Mean Squared Error), and R^2 (Coefficient of Determination). The XGBoost model could predict the ore grade with an SSE of 15.67, RMSE = 0.17, NMSE = 0.34, and R^2 = 0.66, the RFs model with an SSE of 16.40, RMSE = 0.17, NMSE = 0.36, and R^2 = 0.65, the SVR model with an SSE of 19.94, RMSE = 0.19, NMSE = 0.43, and R^2 = 0.57, and the ANN-MLP model with an SSE = 21.00, RMSE = 0.19, NMSE = 0.46, and R^2 = 0.55. In conclusion, the XGBoost model was the most effective in predicting copper ore grades.

1. Introduction

Accurate prediction of mineral grades holds critical importance in the estimation of mineral resources, and plays a decisive role in multiple mining operations. These include the control of mineral grade, underground operations, optimization of open-pit mines, and the comprehensive planning and design of the mine [1]. Precise estimation of mineral grade is fundamental in the economic valuation of mining projects, influencing capital allocation, sustainability, depletion rates, and the overall feasibility of mining operations. However, estimating mineral grade presents considerable challenges, given the complex and multi-faceted nature of mineral deposition processes [2-4].

Traditionally, geometrical and geo-statistical methods have dominated the field of mineral resource estimation. Among these, Kriging stands out as a renowned estimation method in the mining industry, recognized for its accuracy in mineral resource estimation [2]. This method represents an ideal spatial regression technique, designed for regional or local block grade estimation through a linear combination of available data, with the aim of minimizing estimation error [5, 6]. Variations of Kriging such as Simple Kriging (SK) [7], Indicator Kriging (IK) [8], and Ordinary Kriging (OK) [9] have been widely applied in mineral resource estimation. Ordinary kriging, also known as the "best linear unbiased estimator," has become the most widespread technique in this area. This method allows for the estimation of values at unsampled locations within a region of interest, using data from the region and a variogram model

Corresponding author: jairomarquina@unitru.edu.pe (J, Marquina-Araujo)

interpreted from the available data, thereby, minimizing the expected error between the estimated and actual grades [10, 11]. It has been observed that the spatial distribution of estimates made through Kriging tends to exhibit an excessively smoothed character, leading to an overestimation of low-grade values and an underestimation of high-grade values [12, 13]. To address this issue, Deutsch and Journel introduced the technique of Sequential Gaussian Simulation (SGS), thus offering a solution to the smoothing effect inherent in kriging.

In studies on mineral grade prediction using machine learning techniques, Batlile et al. [2] implemented an Artificial Neural Network (ANN) model that integrated variables such as lithology, alteration, east-west orientation, north-south orientation, altitude, dip, and skewness to predict copper grade. The results indicated that the proposed ANN model outperformed traditional machine learning methods, achieving a coefficient of determination (R^2) of 0.584, correlation coefficient (R) of 0.765, Mean Absolute Error (MAE) of 0.0018, Mean Square Error (MSE) of 0.0016, and Root Mean Square Error (RMSE) of 0.041. Emrah and Topal [14], on the other hand, developed a grade estimation technique that combined multi-layer feed-forward Neural Network (NN) models with the k-Nearest Neighbors (kNN) method, demonstrating its effectiveness with an Mean Absolute Error (MAE) of 0.507 and an R^2 of 0.528. In contrast, a traditional model based solely on the coordinates of the sample points recorded an MAE of 0.862 and an R^2 of 0.112.

Li et al. [15] introduced an innovative non-linear and adaptive method based on Wavelet Neural Networks (WNN). This approach, combining features of the wavelet transform with the advantages of artificial neural networks, offered a quick, robust, and promising alternative estimation technique to established methodologies in the field. Chatterjee et al. [16] explored the use of Genetic Algorithm (GA) techniques and k-means clustering for neural network modeling in the context of a lead-zinc deposit. They observed that the proposed method showed considerable efficacy in predicting zinc grades, although it did not represent significant improvements in the prediction of lead grades.

Guerra et al. [17] introduced an innovative application of hyperspectral systems in the mining industry, focusing on the characterization and quantification of minerals in drill cores. Specifically, they implemented conventional neural networks in conjunction with hyperspectral data to estimate copper concentration in drill cores from the Olympic Dam iron oxide copper-gold deposit. The qualitative and quantitative analyses of the results obtained underline the effectiveness of the proposed method. It achieved a Mean Squared Error (MSE) of 0.48 in estimating the percentage of copper along the drill cores, suggesting a notable capability of the method to determine the spatial distribution of the mineral grade. This technique is especially useful in identifying areas of interest for more detailed analysis. On the other hand, Nagpal et al. [18] adopted an artificial intelligence-based approach for mineral grade prediction, using a multi-layer perceptron neural network model and neural network regression models. These models were applied to predict mineral grade from petro-physical data collected through comprehensive geo-physical studies that captured 21 distinct properties of the mineral. These studies demonstrate the growing relevance of advanced data processing and modeling techniques in mineral resource estimation and highlight the potential of these methodologies to enhance accuracy and efficiency in mining exploration and exploitation.

Mostafaei and Ramazi [19] in their research work address the construction of a 3D model for induced polarization (IP) and resistivity (Rs) data of the Madan Bozorg copper mine in Iran. The objective was to quantify the uncertainties of these models using geo-statistical methods and borehole data. They included a methodology of designing and performing four geophysical profiles, processing data to obtain 2D sections of IP and Rs, and their subsequent use in building 3D block models. They obtained a high correlation between model anomalies and mineralization found in drill holes with a value of 95.1%, validating the construction of 3D models from 2D IP and Rs data with an acceptable average error level of 0.001125. Therefore, integrating geo-physical and geo-statistical methods to optimize drill hole locations, potentially reducing the amount needed for copper exploration, saving time and costs. Mostafaei and Ramazi [20] in their research work conducted at the Abassabad copper mine, located in the Miami-Sabzevar mineralization belt, northwest Iran, combined drilling data with induced polarization (IP) and electrical resistivity (Rs) data to estimate mineral resources, effectively reducing the number of drill holes from 20 to 7 at two locations and optimizing their location, resulting in cost savings. Results included copper estimates using regression,

where an average grade of 0.78% and 2.11 million tons of ore were estimated. Another method used was Multivariate Regression Analysis (MRA) with an average grade of 1.43% copper and 2.46 million tons of ore estimated, and with the cokriging method, an average grade of 0.92% copper and 1.85 million tons of copper ore estimated were obtained; this led to an overall summary of an average copper grade of 0.71% and an estimated total of 1.95 million tons. Comparison of these models with actual copper data revealed estimation errors of 8.2% for regression, 26.1% for MRA and 5.1% for cokriging, demonstrating the efficiency of this integrated methodology in mineral resource estimation.

Jafrasteh et al. [21] explored the application of machine learning methods such as neural networks, random forests, and Gaussian processes in estimating copper grade in mineral deposits. They compared the performance of these methods with established geo-statistical techniques including ordinary kriging and indicator kriging. Their findings revealed that Gaussian processes, particularly those adapted with a symmetric standardization of spatial inputs, and an anisotropic kernel resulted in significantly more accurate predictions. Furthermore, they highlighted that including information about rock type in the predictor variable set significantly improves the outcomes. On the other hand, Hekmatnejad et al. [22] demonstrated the effectiveness of disjunctive kriging in predicting mineral grades. In two case studies, they found that disjunctive kriging outperformed ordinary kriging in scenarios involving gold grades with heavy-tailed distributions. In another case, related to copper grades with a moderately skewed distribution, disjunctive kriging showed comparable accuracy to ordinary kriging but with less conditional bias. Finally, Goswami et al. [23] analyzed the effectiveness of General Regression Neural Networks (GRNN) and Support Vector Regression (SVR) in improving the prediction of mineral quality estimates in Western India. They took three-dimensional geographical coordinates and four underlying lithological units as input factors, while the output factors included CaO, Al_2O_3, Fe_2O_3, and SiO_2. To validate the results of their comparative analysis between these models, they employed Ordinary Kriging (OK). The results indicated that the GRNN model significantly outperformed the SVR model in terms of accuracy and reliability of the estimates.

The central objective of this study is to introduce an innovative approach for predicting mineral grade, utilizing machine learning techniques such as the advanced Artificial Neural Network (ANN) model [2] [24,25], Random Forests (RFs) [26-27], Support Vector Regression (SVR) [28-29], and Extreme Gradient Boosting (XGBoost) [30]. These models integrate key spatial information such as east, north, and altitude coordinates, along with the secondary interest grade (in this case, silver grade) and specific geological data (rock types) as inputs. The output variable of the model is the copper ore grade. The development of this paper is structured as what follows. To facilitate a comprehensive understanding, Section 2 details the dataset information and methodology employed. Section 3 presents the results obtained from the proposed artificial intelligence models. Finally, Section 4 concludes the study, summarizing the main findings and contributions.

2. Materials and methods

In this section, we first present the background and observations on the dataset of drill holes. Then we introduce the machine learning techniques and their functioning for estimating mineral grade.

2.1. Data preparation

The database belongs to a copper deposit in Peru. Most of the studied area is occupied by magnetite rocks, granodiorites, dacite porphyries, calcareous sediments, and associated volcanics. Due to a confidentiality agreement, the authors are not permitted to refer to the name of the deposit or any other mineralization that might expose the deposit and/or the mining company. The studied area contains copper (%) and molybdenum (%) grade values from 185 drillings. The average distance between drill holes is 30 meters. The average length of the drillings is about 480 meters. The mineral is extracted from five different lithologies, which for this study were coded, where Rock 1 is Magnetite Skarn, Rock 2 is Granodiorite, Rock 3 is Dacite Porphyry, Rock 4 is Calcareous Sediments, and Rock 5 is Catalina Volcanics. The raw data from the drillings were composed based on lithology, resulting in a total of 5654 drillings (Figure 1).

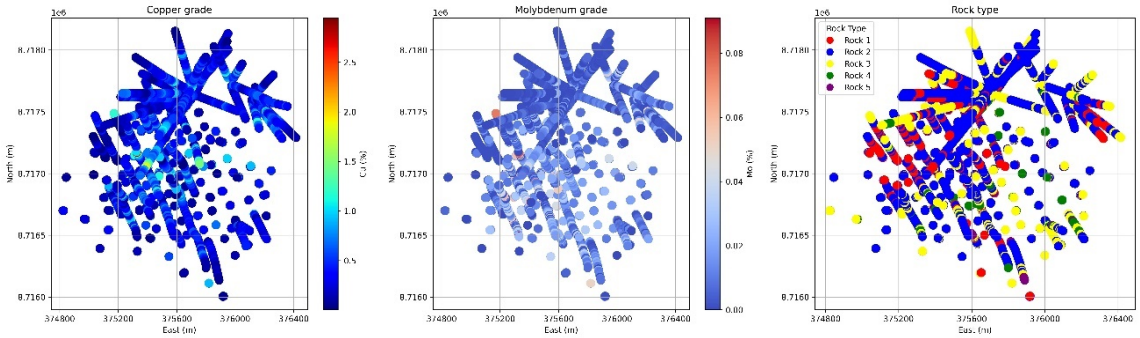

Figure 1. Distribution of drill holes.

Based on the statistical analysis of the composite data, it was shown that the average copper grade value was 0.430%, with a variance of 0.084 and a standard deviation (Std) of 0.290. This variance and standard deviation indicate a low distribution compared to the mean. Table 1 shows the statistical parameters of the data.

Table 1. Descriptive statistics of the data set.

Statistic	East (m)	North (m)	Elevation (m)	Rock type	Molybdenum (%)	Copper (%)
Count	5654	5654	5654	5654	5654	5654
Mean	375606	8717016	4474	2.163	0.014	0.430
Std	307	394	170	0.783	0.015	0.290
Min	374821	8716003	4050	1.000	0.000	0.002
25%	375393	8716738	4340	2.000	0.003	0.227
50%	375602	8716996	4463	2.000	0.010	0.378
75%	375825	8717272	4607	3.000	0.021	0.578
Max	376415	8718153	4902	5.000	0.091	2.949
Variance	94394	154875	28743	0.613	0.0002	0.084

Figure 2 presents a detailed histogram of the copper ore grades, highlighting a range of values from a minimum (Y_{min}) of 0.002% to a maximum (Y_{max}) of 2.949%, with an average of 0.430%. This histogram reveals that the distribution of copper grades is positively skewed, and represents significant variability, suggesting the presence of extreme values within the dataset. This observation is indicative of the heterogeneity of the deposit, with areas that exhibit high concentrations of copper, while others show significantly lower grades.

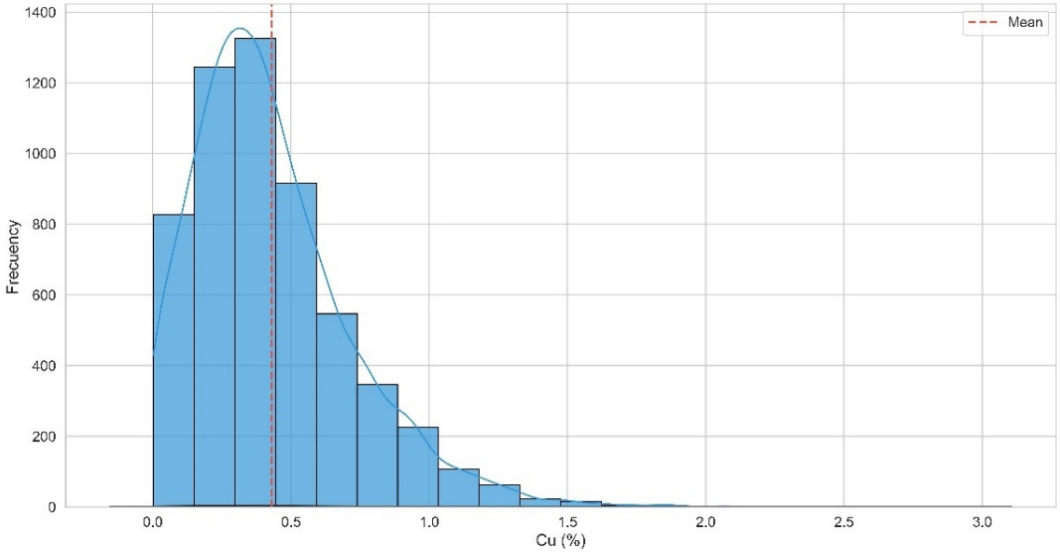

Figure 2. Histogram of the distribution of copper ore grades.

To handle the different types of data and ranges present in the dataset (i.e. geographical coordinates, lithology, and mineral grade), the values of each feature were normalized based on the mean and standard deviation. This was done to scale the different values of the features to a common scale [31]. The value of a data point in each feature was recalculated by subtracting the population mean of a given feature from an individual point and then dividing the difference by the population standard deviation [32]. Each instance $x_{i,n}$ of the data is transformed into $x'_{i,n}$ in the following manner:

$$x'_{i,n} = \frac{x'_{i,n} - \mu_i}{\sigma_i} \quad (1)$$

where μ and σ denote the mean and standard deviation of the i-th feature, respectively [33] [31].

2.2. Multi-layer perceptron artificial neural network (ANN-MLP)

Artificial neural networks have been employed to address numerous geo-statistical problems due to their capability to capture nonlinear trends. A Multi-layer Perceptron (MLP) is a type of neural network that includes one or more hidden layers. MLPs can approximate any arbitrary function, given enough nodes in the hidden layers [34]. For an input vector x, the output of an MLP with a single hidden layer and H nodes in the hidden layer is given by:

$$\hat{y}(x) = \sum_{j=1}^{H} v_j f\left(w_j^T x + w_{bj}\right) + v_{b,} \quad (2)$$

where W = $(w_1, w_2, \ldots, w_H)$ is the weight matrix of the network between the nodes of the input layer and the hidden layer; $\{w_{bj}\}_{j=1}^{H}$ are the biases of the neurons in the hidden layer; $\{v_i\}_{j=1}^{H}$ are the weights of the connections between the hidden layer and the output node; v_b is the bias weight of the output; and f is a non-linear activation function. In this study, the hyperbolic tangent function is used as the activation function:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3)$$

Other activation functions such as sigmoid, probit, and rectified linear functions are also commonly used [35]. After testing various configurations of the neural network, the best results were achieved with a minimal error rate using a neural network architecture consisting of an input and an output. The neural network architecture comprised an input layer with 6 input neurons and 5 hidden layers, two with 224 neurons each, one with 320 neurons, another with 96 and 32 neurons each, and an output layer with 1 neuron (Figure 3). The model was constructed using the Python programming language version 3.11.7 and the deep learning library TensorFlow 2.15.0.

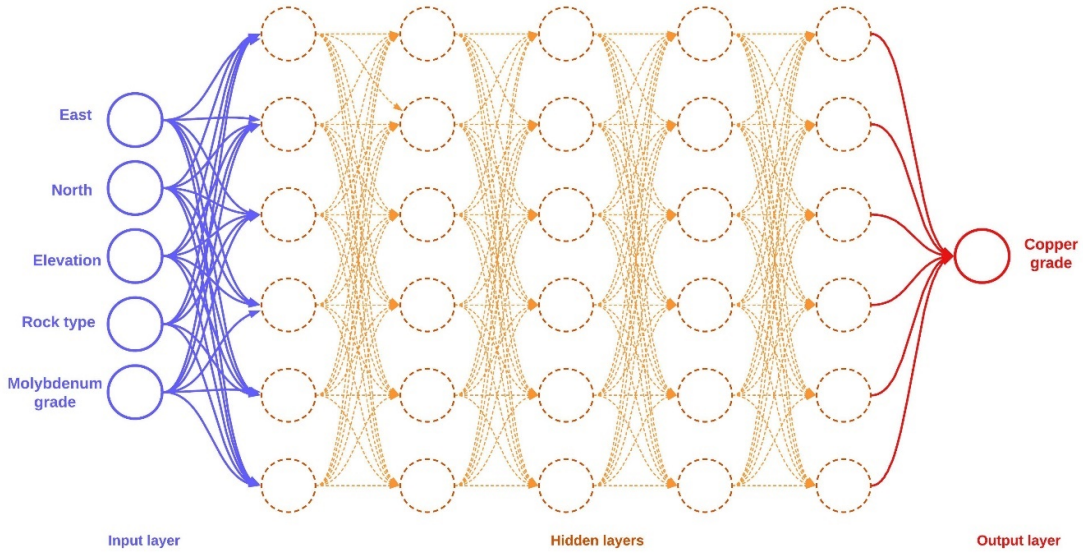

Figure 3. Structure of the Multilayer Perceptron Artificial Neural Network (ANN-MLP).

2.3. Random forests (RFs)

Random Forest (RFs) is an ensemble technique, where the outcomes of a collection of random decision trees are combined to obtain a global prediction [36, 37]. This method is applicable for both regression and classification of practical interest [38]. At each internal node of the tree, a binary decision is made based on a boolean test. For example, if the selected attribute for splitting is ordinal, the test involves determining whether the attribute's value is above a threshold. Instances for which the response to the test is true (i.e. the attribute value is above the specified threshold) are assigned to one of the child nodes. Those with a false response (i.e. the attribute value is below the threshold) are assigned to the other child node. In this way, the training data are divided into separate subsets. The division is made so that within each subset, the classes are better separated (in classification problems) or the prediction error is minimized (in regression). In random trees, the boolean test of a specific internal node is selected as the best split resulting from considering only a subset of attributes randomly chosen [39]. The tree grows until a new split does not lead to purer nodes or a specified pre-pruning criterion is met (for example, there are too few instances assigned to a node, or the maximum depth of the tree is reached). Each tree in the forest is constructed from an independent Bootstrap sample of the data, as in bagging [40].

Predictions will be made at the terminal nodes (leaves) of the tree based on the training instances that have been assigned to those nodes by the sequence of tests at the root node and the subsequent intermediate nodes that connect the root to the corresponding leaf. In regression, the prediction is the average value of the response variable over the training instances assigned to that leaf. In classification, the final prediction of the ensemble is obtained by majority vote. In regression, the outcome of an ensemble of size T is the mean of the predictions from the random trees in the ensemble:

$$\hat{y}(x) = \frac{1}{T}\sum_{t=1}^{T} \hat{y}^{(t)}(x) \quad (4)$$

where $\hat{y}^{(t)}(x)$ is the output of the t-th regression tree.

2.4. Extreme gradient boosting (XGBoost)

XGBoost is a parallel tree boosting system based on the gradient boosting method [41]. It utilizes a model comprising a set of Classification and Regression Trees (CART) [42]. While XGBoost appears similar to GBDT, it has some inherent features that differ from GBDT such as the second-order Taylor expansion and the embedded normalization function [43, 44]. The XGBoost model can be briefly explained as follows:

For a dataset $D = (x_i, y_i)(x_i \in R^m, y_i \in R, i = 1,2,\dots,n)$ containing n instances with m dimensions, and a model trained with G trees, the predictions are obtained by the following:

$$\hat{y}_i = \sum_{k=1}^{G} f(x_i), f_k \in F(i = 1,2,\dots,n) \quad (5)$$

where f is the hypothesis space, and $f(x)$ is a regression tree: $F = \{f(x) = w_{q(x)}\}(q: R^m \rightarrow \{1,2,\dots,T\}, w \in R^T)$ $q(x)$. Here, $q(x)$ is the leaf node, and w is the score of the leaf [45].

To construct an ideal model, it is necessary to minimize the objective function to find the optimal parameters. This can be divided into a loss function (L) and a model complexity function (Ω).

$$J = L + \Omega \quad (6)$$

$$L = \sum_{i=1}^{n} L(y_i - \hat{y}_i)^2 \quad (7)$$

$$\Omega = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \quad (8)$$

where γ and λ are parameters that prevent the model from overfitting. The objective function used during training is as follows:

$$J(f_t) = \sum_{i=1}^{n} (y_i - (\hat{y}^{(t-1)} + f_t(x_i)))^2 + \Omega(f_t)) \quad (7)$$

where $\hat{y}^{(t-1)}$ is the predicted value of the $t-1th$ model, and $f_t(x_i)$ is the new function added at the $t-1th$ time. The lower the value of J, the better the efficiency of the model.

2.5. Support vector machine regression (SVR)

Support Vector Regression (SVR) is a supervised learning method, and its performance depends on the training and testing dataset. The datasets must be partitioned for training and testing such that both have a similar distribution [28]. To examine the distribution of the dataset, t-statistics can be analyzed. The objective of SVR is to identify a function for which all training patterns or datasets can have a maximum deviation, ε, from

the target values, and at the same time, the flatness should be as high as possible. Generally, all training data points are located within the boundary $(-\varepsilon\ a + \varepsilon)$. SVR was used in this study, and is represented by the Equation [46]:

$$K(x, x') = exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \quad (8)$$

where $\|x - x'\|^2$ is recognized as the squared Euclidean distance between the two feature vectors, and sigma (σ) is the spread of the distribution used in the kernel function.

2.6. Model evaluation

The performance of the models was assessed using datasets with four error indices: Sum of Squared Errors (SSE), Root Mean Squared Error (RMSE), Normalized Mean Squared Error (NMSE), and Coefficient of Determination (R^2). All indices were determined from the predicted grade (represented by o) and the observed grades (represented by p) of the test samples. An SSE value closer to 0 indicates that the model has a lower random error component, making it more useful for prediction [28]. It is determined as follows:

$$SSE = \sum_i (o_i - p_i)^2 \quad (9)$$

The RMSE should ideally be zero for a perfect model. The RMSE of a model's prediction relative to the observed values is defined as the square root of the mean squared error [28]:

$$RMSE = \sqrt{\frac{1}{n_{ts}} \sum_i (o_i - p_i)^2} \quad (10)$$

The dimensionless form of RMSE can be represented by NMSE. Like RMSE, a lower NMSE indicates the development of a better prediction model. However, unlike RMSE, where a higher value indicates poor prediction; a higher NMSE does not imply the same [28]. It can be determined as:

$$NMSE = \frac{1}{n_{ts}} \frac{\sum_i (o_i - p_i)^2}{o_i p_i} \quad (11)$$

The R-squared value (known as the coefficient of determination) describes what portion of the variance between the two variables (observed and predicted values) is described by the legal fit. Models with an R^2 above 55% are considered satisfactory, less than 30% are suspect, and more than 75% are excellent [47]. This can be determined as:

$$R^2 = \frac{(\sum (o_i - \bar{o_i})(p_i - \bar{p_i}))^2}{\sum (o_i - \bar{o_i})^2 \sum (p_i - \bar{p_i})^2} \quad (12)$$

3. Results

In the research work, a rigorous data partitioning has been implemented, reserving 10% of the composites (565 composites) for testing, an additional 10% for validation, and the remaining 80% (4523 composites) dedicated to model training. This partitioning structure ensures the integrity of model validation, and reflects a commitment to obtaining robust and verifiable predictive estimates. Table 2 encapsulates the strategic configuration of hyperparameters for each employed method: from the Multilayer Perceptron Artificial Neural Network (ANN-MLP) architecture, which includes five hidden layers with a carefully calibrated distribution of neurons, to the Random Forests (RFs) configuration with 100 estimators and a maximum depth of 30, aiming to limit complexity and prevent overfitting. In the case of Extreme Gradient Boosting (XGBoost), parameters such as learning rate (Eta) and the number of estimators has been meticulously tuned to optimize the sequential tree building process. Additionally, the Support Vector Regression (SVR) model has been fine-tuned with a focus on the balance between margin and classification error, using an RBF kernel and a carefully selected parameter C.

Table 2. Hyper-parameters for machine learning methods.

Method	Parameter	Value	Description
Multi-layer Perceptron Artificial Neural Network (ANN-MLP)	Input layers	1	Single input layer.
	Hidden layers	5	Model's capacity to extract complex patterns.
	Neurons in the 1st hidden layer	224	Neurons per layer, determining model complexity.
	Neurons in the 2nd hidden layer	96	
	Neurons in the 3rd hidden layer	320	
	Neurons in the 4th hidden layer	224	
	Neurons in the 5th hidden layer	32	
	Output layers	1	Output layer for predictions.
	Activation function	Tanh	Allows modeling of nonlinear responses.
	Dropout	0.2	Reduces overfitting by randomly deactivating neurons.
	Optimizer	Adam	Efficient in weight adjustment.
	Loss function	Mean Squared Error	Measures squared difference between predictions and actual values.
Random Forests (RFs)	Estimators	100	Number of trees in the model.
	Max depth	30	Limits tree complexity.
	Min samples per leaf	1	Minimum samples required to form a leaf.
	Min samples to split	2	Seed for reproducibility.
	Random state	42	Semilla para la reproducibilidad.
	Verbosity	0	Model's output message level.
Extreme Gradient Boosting (XGBoost)	Booster	Gbtree	Uses decision trees as base.
	Eta (learning rate)	0.05	Learning rate for updates.
	Max Depth	10	Depth of each tree.
	Min child weight	5	Controls overfitting.
	Evaluation metric	Rmse	Prediction accuracy metric.
	Estimators	1000	Number of trees constructed.
	Objective	Reg: squared error	Minimizes squared error.
	Random state	42	For consistent results.
	Positive weight	1	Balances imbalanced classes.
Support Vector Regression (SVR)	C	1	Balance between margin and error.
	Kernel	Rbf	For projection into higher-dimensional space.
	Gamma	Scale	Defines influence of a single sample.
	Epsilon	0.1	Margin of error without penalty.
	Degree	3	For polynomial kernels.
	Coef	0	Independent term in the kernel.
	Shrinkage	True	For optimization efficiency.
	Tolerance	1.0 e-03	Convergence precision.
	Cache size	200	Memory for kernel calculations.

3.1. Training of machine learning models

Figure 4 displays the loss function minimization trajectory during the training of the Multi-layer Perceptron Artificial Neural Network (ANN-MLP), illustrating a reduction from high initial values to a stabilized low level as the epochs increase. This behavior indicates a well-fitted model, where the loss function significantly decreases in the first 20 epochs and converges around an approximate value of 0.05. This demonstrates the model's ability to learn from the data without overfitting.

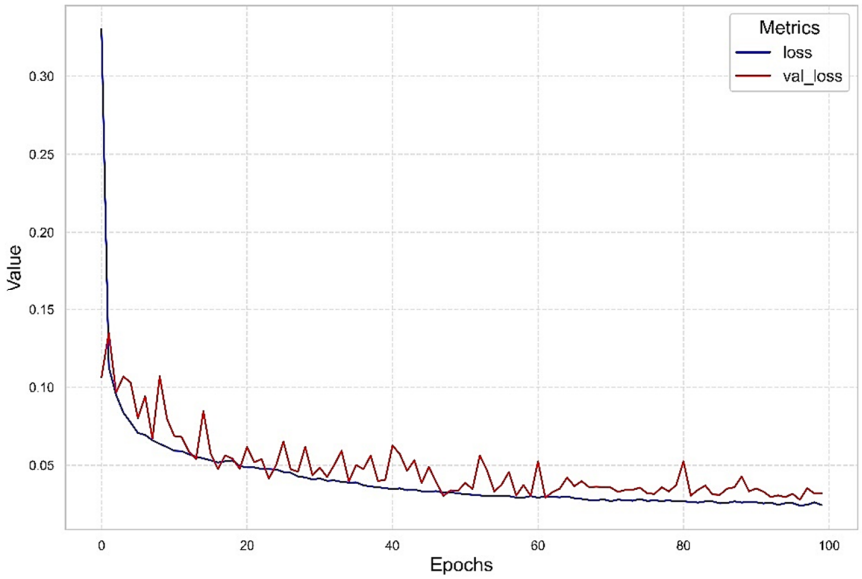

Figure 4. Loss evolution during ANN-MLP training.

Figure 5 explores the impact of the number of trees on the performance of the Random Forests (RFs) model, illustrating a significant decrease in Mean Squared Error (MSE) as the number of trees increases up to a threshold of 100, at which point the improvement in accuracy stabilizes. The initially high MSE, close to 0.0305, rapidly decreases with the first 50 trees before reaching a plateau, suggesting resource optimization by avoiding unnecessary increases in the number of trees.

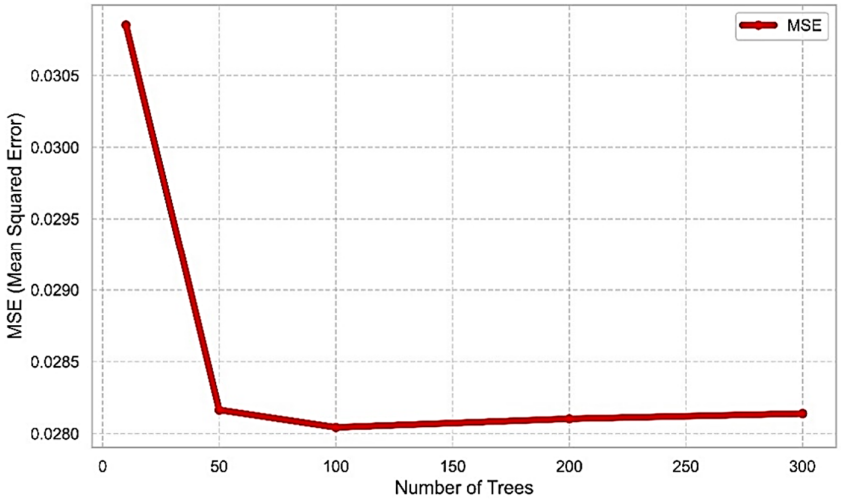

Figure 5. Optimization of the number of trees in an RFs ensemble model.

Figure 6 reveals the dynamics of training and validation error in the Extreme Gradient Boosting (XGBoost) model over a sequence of boosting iterations. A critical inflection point is observed around iteration 0.1, where the validation RMSE reaches its minimum value before increasing thereafter. This indicates an optimal point to halt training and prevent overfitting. The graph displays fluctuations in the validation RMSE, rising from approximately 0.1505 to 0.1525, highlighting the model's sensitivity to hyper-parameter settings.

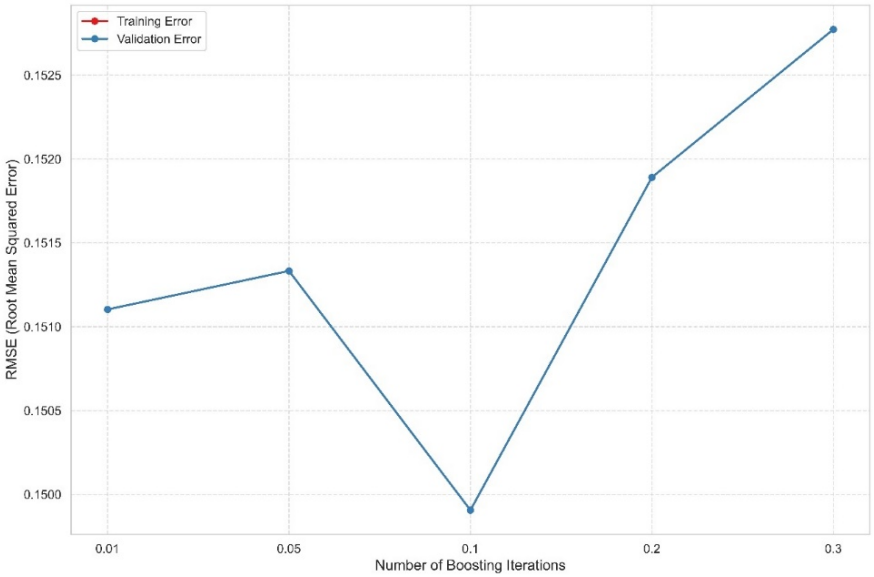

Figure 6. Dynamics of training and validation error with boosting iterations in XGBoost.

Figure 7 provides a detailed view of how the hyper-parameter C affects the performance of the Support Vector Regression (SVR) model. It demonstrates a significant decrease in both training and validation Mean Squared Error (MSE) as the value of C is increased from 0.1 to 10. This variation in C showcases an optimization in the balance between model complexity and generalization ability, with a stabilization point in the training MSE around 0.0275. Meanwhile, the validation MSE decreases and then slightly expands, suggesting an optimal value of C close to 1 to avoid overfitting.

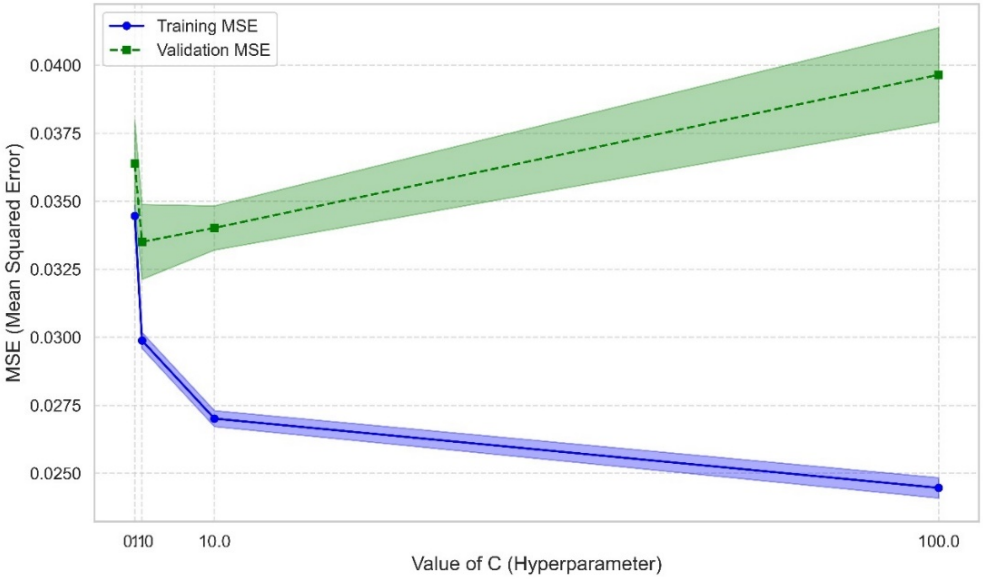

Figure 7. Impact of the hyper-parameter C on training and validation MSE in SVR.

3.2. Validation and test results of machine learning models

Figure 8 provides a comparative analysis between predicted and actual values obtained through the Multi-layer Perceptron Artificial Neural Network (ANN-MLP) model, broken down by the test, training, and validation datasets. In the test dataset, a correlation of 0.76 was achieved, indicating a considerable adequacy between the model's predictions and the observed copper grade.

Accuracy improves in the training dataset, with a correlation of 0.85, which is consistent with the trend of models learning specifically from the data they are trained on. The correlation in the validation dataset is 0.79, suggesting that the model maintains good generalization beyond the training sample.

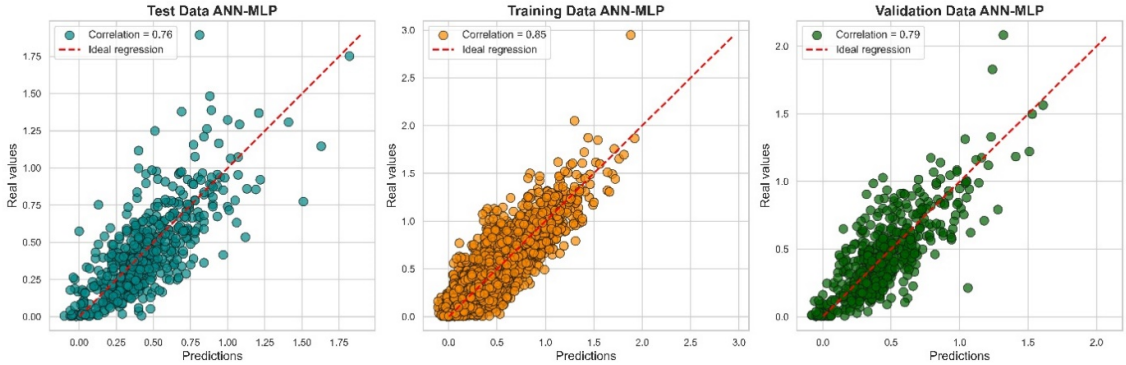

Figure 8. Comparison of predictions and actual values with ANN-MLP.

Figure 9 provides a comparative analysis of predicted results versus actual values for the Random Forests (RFs) model. An exceptionally high correlation of 0.98 is observed in the training dataset, indicating a high level of accuracy in this set. However, this accuracy is slightly lower in the test and validation datasets, with correlations of 0.80 and 0.84, respectively. This could suggest a potential overfitting to the training data while still maintaining robust predictive capability.

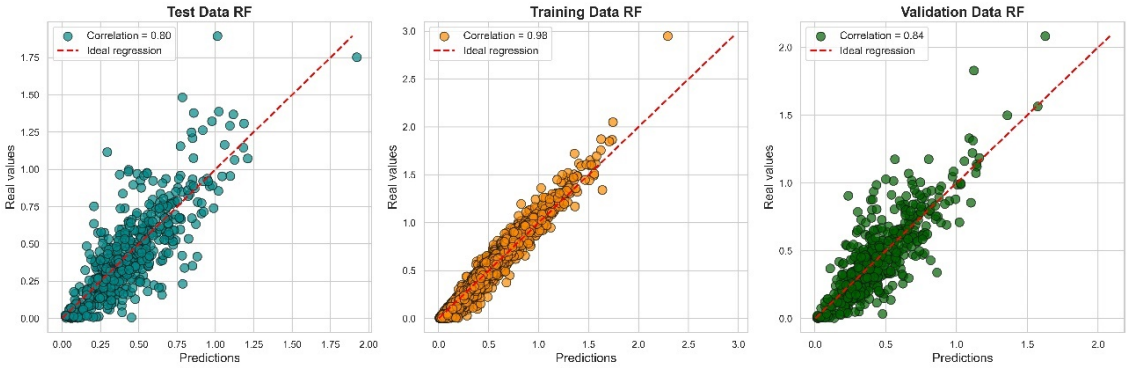

Figure 9. Comparison of predictions and actual values with RFs.

Figure 10 details the relationship between the values predicted by the Extreme Gradient Boosting (XGBoost) model and the actual values. With a perfect correlation of 1.00 in the training dataset, the model exhibits unparalleled ability to capture the variability of the training data. This level of accuracy remains high in the test and validation datasets, with correlations of 0.82 and 0.85, respectively, demonstrating XGBoost's outstanding ability to generalize and provide reliable predictions on unseen data.

Figure 10. Comparison of predictions and actual values with XGBoost.

Figure 11 illustrates the predictive capability of the Support Vector Regression (SVR) model by comparing predicted values to actual values across the test, training, and validation datasets. The model achieves a correlation of 0.75 in the test dataset, 0.80 in the training dataset, and 0.81 in validation. This demonstrates consistency in predicting copper mineral grades and adequate generalization, with a slight decrease in the test dataset correlation suggesting room for further optimization.

Figure 11. Comparison of predictions and actual values with SVR.

For comparative purposes, Figure 12 and Figure 13 display the prediction results for the machine learning models on an independent test dataset. Furthermore, Table 3 presents the prediction results, where the proposed Extreme Gradient Boosting (XGBoost) model yielded an RMSE value of 0.17 and an R^2 value of 0.66, while the Random Forests (RFs) model yielded an RMSE of 0.17 and an R^2 of 0.65. The Support Vector Regression (SVR) model resulted in an RMSE of 0.19 and an R^2 of 0.57, and finally, the Multi-layer Perceptron Artificial Neural Network (ANN-MLP) model had an RMSE of 0.19 and an R^2 of 0.55.

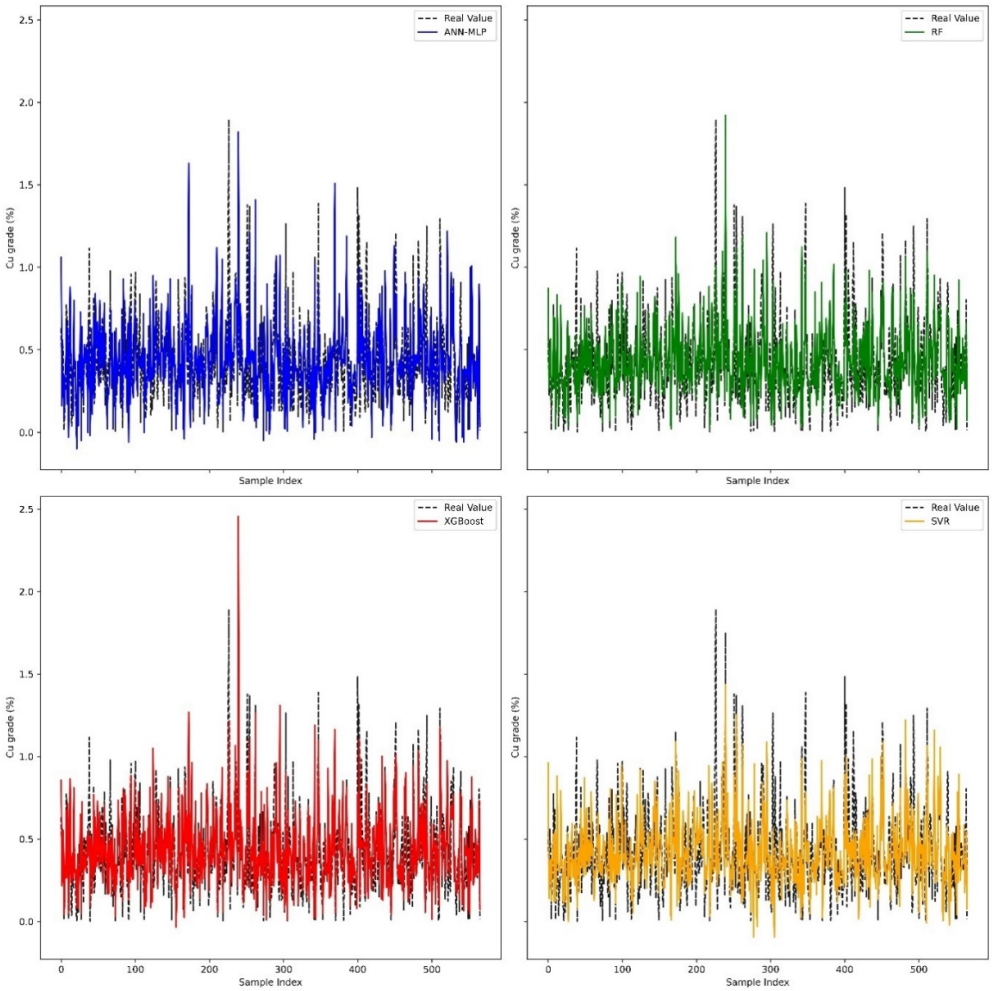

Figure 12. Prediction results of the models on an independent test dataset.

Table 3. Model performance of machine learning methods.

Method	R^2	MAE	MSE	RMSE
Multi-layer Perceptron Artificial Neural Network (ANN-MLP)	0.80	0.19	0.09	0.30
Random Forests (RFs)	0.76	0.17	0.11	0.33
Extreme Gradient Boosting (XGBoost)	0.55	0.18	0.20	0.44
Support Vector Regression (SVR)	0.81	0.14	0.09	0.30

This demonstrates that machine learning algorithms can estimate mineral grades, with Extreme Gradient Boosting (XGBoost) being the best model with the highest coefficient of determination, followed by Random Forests (RFs), Support Vector Regression (SVR), and Multi-layer Perceptron Artificial Neural Network (ANN-MLP). These machine learning models are more adaptable and can be updated for improved results by feeding the model with more data when new observations are available.

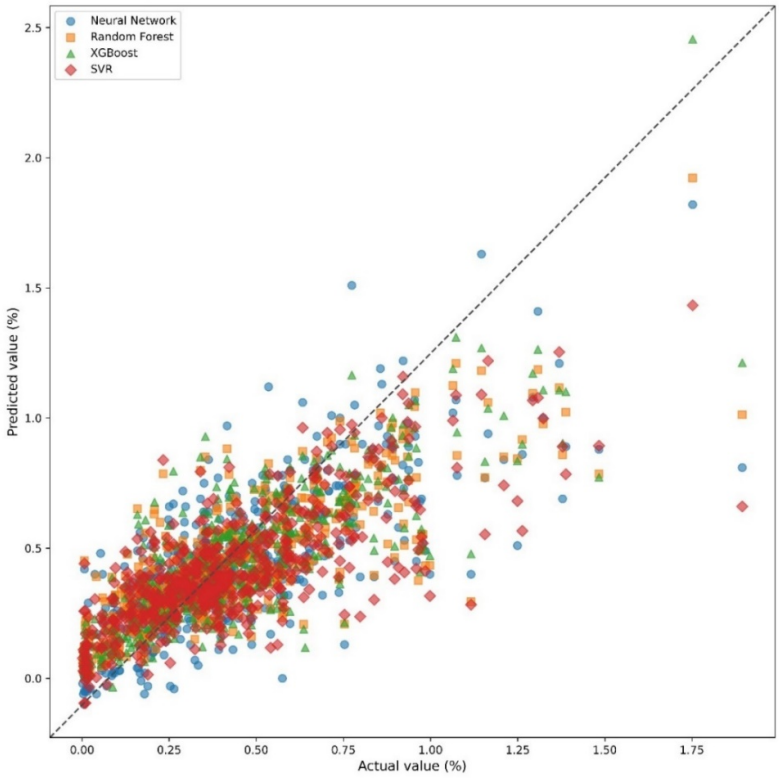

Figure 13. Actual grade *vs.* predicted grade in test data for each machine learning model.

4. Conclusions

In this research work, machine learning models including Multi-layer Perceptron Artificial Neural Network (ANN-MLP), Random Forests (RFs), Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR) were employed to predict the copper grade in a copper deposit. It has been demonstrated that the XGBoost and RFs models can be successfully used for grade estimation. While the SVR and ANN-MLP models performed well, there is still room for improvement in terms of hyper-parameters. The results of this study indicate that:

All four gradient-based models demonstrated a satisfactory level of accuracy (R^2) exceeding 0.55. XGBoost outperformed the other three methods with an accuracy (R^2) of 0.66, SSE of 15.67, RMSE of 0.17, and NMSE of 0.34. The RFs model also exhibited good performance with an accuracy (R^2) of 0.65, SSE of 16.40, RMSE of 0.17, and NMSE of 0.36. The SVR model achieved an R^2 of 0.57, SSE of 19.94, RMSE of 0.19, and NMSE of 0.43. Finally, the ANN-MLP model had the lowest prediction performance with an R^2 of 0.55, SSE of 21.00, RMSE of 0.19, and NMSE of 0.46.

The practical applications of these models in mining are significant. By improving the accuracy of mineral resource estimation, these techniques facilitate more effective and efficient mine planning. This leads to optimal resource allocation, time reduction, and more accurate and less invasive exploration. In addition, accurate grade prediction can improve risk management and decision making at all stages of mining. For future work, it is recommended to consider incorporating additional geological variables into the model training, as the current model only used coordinates, rock type, and secondary mineral grade (Mo). Geological features such as alterations and mineralization zones should be included. Additionally, a comparison of these models with traditional methods like kriging could be explored in further research.

References

[1]. Reza, M. and Aghajani, A. (2019). Cutoff grades optimization in open pit mines using meta-heuristic algorithms. *Resources Policy, 60*, 72-82.

[2]. Batlile, N., Adachi, T., and Kawamura, Y. (2023). Application of Artificial Neural Network for the Prediction of Copper Ore Grade. *Minerals, 13* (5).

[3]. Battalgazy, N., Valenta, R., Gow, P., Spier, C., and Forbes, G. (2023). Addressing Geological

Challenges in Mineral Resource Estimation: A Comparative Study of Deep Learning and Traditional Techniques. Minerals, 13(7).

[4]. Alimoradi, A., Maleki, A., Karimi, M., Sahafzadeh, S., and Abbasi. (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 (JME)*, *11* (3), 865-879.

[5]. Zhang, S., Nwaila, G., Bourdeau, J., Ghorbani, Y., and Carranza, E. (2023). Machine Learning-based Delineation of Geodomain Boundaries: A Proof-of-Concept Study using Data from the Witwatersrand Goldfields. *Natural Resources Research, 32,* 879-900.

[6]. Deutsch, J., Szymanski, J., and Deutsch, C. (2014). Checks and measures of performance for kriging estimates. *Journal of the Southern African Institute of Mining and Metallurgy, 114* (3).

[7]. Ro, Y. and Yoo, C. (2022). Numerical Experiments Applying Simple Kriging to Intermittent and Log-Normal Data. *Water, 14* (9).

[8]. Park, N., Kyriakidis, P., and Young, S. (2016). Spatial Estimation of Classification Accuracy Using Indicator Kriging with an Image-Derived Ambiguity Index. *Remote Sens, 8* (4).

[9]. Giraldo, R., Delicado, P., and Mateu, J. (2011). Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics, 18*, 411-426.

[10]. Dumakor-Dupey, N. and Arya, S. (2021). Machine Learning—A Review of Applications in Mineral Resource Estimation. *Energies, 14*, 4079.

[11]. Ghasemitabar, H., Alimoradi, A., Hemati, H., and Fathi, M. (2024). Intelligent Borehole Simulation with python Programming. *Journal of Mining and Environment, 15* (2), 707-730.

[12]. Lloyd, C. and Atkinson, P. (2001). Assessing uncertainty in estimates with ordinary and indicator kriging. *Computers & Geosciences, 27* (8), 929-937.

[13]. Gia, T., Kappas, M., Van, C., and Khanh, L. (2019). Application of Ordinary Kriging and Regression Kriging Method for Soil Properties Mapping in Hilly Region of Central Vietnam. *ISPRS Int. J. Geo-Inf, 8* (3).

[14]. Emrah, U. and Topal, E. (2020). A New Ore Grade Estimation using Combine Machine Learning Algorithms. *Minerals, 10* (10).

[15]. Li, X., Xie, Y., Guo, Q., and Li, L. (2010). Adaptive ore grade estimation method for the mineral deposit evaluation. *Mathematical and Computer Modelling, 52* (11-12), 1947-1956.

[16]. Chatterjee, S., Bandopadhyay, S., and Machuca, D. (2010). Ore Grade Prediction using a Genetic Algorithm and Clustering-based Ensemble Neural Network Model. *Mathematical Geosciences, 42* (3), 309-326.

[17]. Guerra, C., Souza, C., and Muico, E. (2023). Ore-Grade Estimation from Hyperspectral Data using Convolutional Neural Networks: A Case Study at the Olympic Dam Iron Oxide Copper-Gold Deposit, *Australia. Economic Geology, 118* (8), 1899-1921.

[18]. Nagpal, G., Shrikant, S., Krishna, N., Nagpal, A., and Mohan, G. (2022). Ore Grade Estimation in Mining Industry from petro-physical data using neural networks. *ICIMMI '22: Proceedings of the 4th International Conference on Information Management & Machine Intelligence, 75,* 1-5.

[19]. Mostafaei, K. and Ramazi, H. (2018). 3D model construction of induced polarization and resistivity data with quantifying uncertainties using geostatistical methods and drilling (Case study: Madan Bozorg, Iran). *Journal of Mining & Environment*, *9* (4), 857-872.

[20]. Mostafaei, K. and Ramazi, H. (2019). Mineral resource estimation using a combination of drilling and IP-Rs data using statistical and cokriging methods. *Bulletin of the Mineral Research and Exploration, 160*, 177-195.

[21]. Jafrasteh, B., Fathianpour, N., and Suárez, A. (2018). Comparison of machine learning methods for copper ore grade estimation. *Comput Geosci, 22,* 1371-1388.

[22]. Hekmantnejad, A., Emery, X., and Alipour, M. (2019). Comparing linear and non-linear kriging for grade prediction and ore/waste classification in mineral deposits. *International Journal of Mining, Reclamation and Environment, 33* (4).

[23]. Goswami, A., Mishra, M., and Patra, D. (2022). Evaluation of machine learning algorithms for grade estimation using GRNN & SVR. *Engineering Research Express, 4* (3).

[24]. Fathi, M., Alimoradi, A., and Hemati, H. (2021). Optimizing Extreme Learning Machine Algorithm using Particle Swarm Optimization to Estimate Iron Ore Grade. *Journal of Mining and Environment (JME), 12* (2), 397-411.

[25]. Alimoradi, A., Hajkarimian, H., Hemati, H., and Salsabili, M. (2022). Comparison between the performances of four metaheuristic algorithms in training a multilayer perceptron machine for gold grade estimation. *International Journal of Mining and Geo-Engineering, 56* (2), 97-105.

[26]. Sarantsatsral, N., Ganguli, R., Pothina, R., and Tumen, B. (2021). A Case Study of Rock Type Prediction using Random Forests: Erdenet Copper Mine, Mongolia. *Minerals*, *11* (10).

[27]. Afzal, P., Farhadi, S., Boveiri, M., Shamseddin, M., and Daneshvar, L. (2022). Geochemical Anomaly Detection in the Irankuh District using Hybrid Machine Learning Technique and Fractal Modeling. *Geopersia, 12* (1), 191-199.

[28]. Patel, A., Chatterjee, S., and Gorai, A. (2019). Development of a machine vision system using the support vector machine regression (SVR) algorithm for the online prediction of iron ore grades. *Earth Sci Inform, 12,* 197-210.

[29]. Farhadi, S., Afzal, P., Boveiri, M., Daneshvar, L., and Sadeghi, B. (2022). Combination of Machine Learning Algorithms with Concentration-Area Fractal Method for Soil Geochemical Anomaly Detection in Sediment-Hosted Irankuh Pb-Zn Deposit, Central Iran. *Minerals, 12* (6), 689.

[30]. Xie, J., Wang, Q., Liu, P., and Li, Z. (2021). A hyperspectral method of inverting copper signals in mineral deposits based on an improved gradient-boosting regression tree. *International Journal of Remote Sensing, 42* (14), 5474-5492.

[31]. Sola, J. and Sevilla, J. (1997). Importance of input data normalization for the application of neural networks to complex. *IEEE Trans. Nucl. Sci, 44,* 1464-1468.

[32]. Larsen, R. and Marx, M. (2005). An Introduction to Mathematical Statistics and its Applications, UK: Prentice Hall: London.

[33]. Singh. D. and Singh, B. (2019). Investigating the impact of data normalization on classification performance. *Appl. Soft Comput.,* 105524.

[34]. Irie, B. and Miyake, S. (1988). Capabilities of three-layered perceptrons. *In: IEEE International Conference on Neural Networks,* 1988, 641-648.

[35]. Gomes, G., Ludermir, T., and Lima, L. (2011). Comparison of new activation functions in neural network for forecasting financial time series. *Neural Comput. Applic, 20* (3), 417-439.

[36]. Criminisi, A., Shotton, J., and Konukoglu, E. (2012). Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Computer Graphics and Vision, 7* (223), 81-227.

[37]. Trehan, S., Carlberg, K., and Durlofsky, L. (2017). Error modeling for surrogates of dynamical systems using machine learning. *International Journal for Numerical Methods in Engineering.*

[38]. Fernández-Delgado, M., Cernadas, E., Barro, S., and Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res,* 15, 3133-3181.

[39]. Ho, T. (1998). The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell, 20* (8), 832-844.

[40]. Breiman, L. (1996). Bagging predictors. *Mach. Learn, 24* (2), 123-140.

[41]. Friedman, J. (2001). Greedy boosting approximation: A gradient boosting machine. *Annal. Stat*, *29* (5), 1189-1232.

[42]. Breiman, L., Friedman, J., and Olshen, R. (1984). Classification and regression trees. *Wadsworth int, Group, 37* (15), 237-251.

[43]. Song, K., Yan, F., and Ding, T. (2020). A steel property optimization model based on the xgboost algorithm and improved pso. *Comput. Mater. Sci, 174*, 109472.

[44]. Krizhevsky, A., Sutskever. I., and Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems,* 1097-1105.

[45]. Emrah, U., Dagasan, Y., and Topal, E. (2021). Mineral grade estimation using gradient boosting regression trees. *International Journal of Mining, Reclamation and Environment, 35* (10), 728-742.

[46]. Rivas-Perea, P., Cota-Ruiz, J., Chaparro, D., Venzor, J., Carreón, A., and Rosiles, J. (2013). Support vector machines for regression: a succinct review of large-scale and linear programming formulations. *Int J Intell Sci, 3*, 5-14.

[47]. Prasad, K., Gorai, A., and Goyal, P. (2016). Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time. *Atmos Environ, 128*, 246-262.

پیش‌بینی عیار سنگ معدن مس با استفاده از تکنیک‌های یادگیری ماشین در یک انبار مس

جایرو مارکینا-آراوجو[1*]، مارکو کوترینا-تئاتینو[1]، خوزه مامانی-کوئیسپه[2]، ادواردو نوریگا-ویدال[1]، خوان وگا-گونزالس[3]، و خوان کروز-گالوز[3]

۱. گروه مهندسی معدن، دانشکده مهندسی، دانشگاه ملی تروخیلو، تروخیلو، پرو
۲ .گروه مهندسی شیمی، دانشکده مهندسی، دانشگاه ملی آلتیپلانو پونو، پونو، پرو
۳. گروه مهندسی متالورژی، دانشکده مهندسی، دانشگاه ملی تروخیلو، تروخیلو، پرو

ارسال ۲۰۲۴/۰۱/۰۶، پذیرش ۲۰۲۴/۰۱/۲۷

* نویسنده مسئول مکاتبات: jairomarquina@unitru.edu.pe

چکیده:

هدف از این کار تحقیقاتی استفاده از تکنیک‌های یادگیری ماشینی از جمله شبکه‌های عصبی مصنوعی پرسپترون چندلایه (ANN-MLP)، جنگل‌های تصادفی (RFs)، افزایش گرادیان شدید (XGBoost) و رگرسیون برداری پشتیبانی (SVR) برای پیش‌بینی عیار سنگ مس در یک ذخایر مس واقع در پرو مدل ها با استفاده از ۵۶۵۴ کامپوزیت حاوی اطلاعات زمین شناسی موجود (نوع سنگ)، و همچنین مکان نمونه ها (شرق، شمال و ارتفاع) و عیار سنگ ثانویه (Mo) به دست آمده از چاه های حفاری توسعه داده شدند. داده ها به ۱۰ درصد (۵۶۵ کامپوزیت) برای آزمایش، ۱۰ درصد (۵۶۵ کامپوزیت) برای اعتبار سنجی و ۸۰ درصد (۴۵۲۳ کامپوزیت) برای آموزش تقسیم شدند. معیارهای ارزیابی شامل SSE (مجموع خطاهای مربعی)، RMSE (ریشه میانگین مربعات خطا)، NMSE (میانگین مربعات خطای عادی) و R^2 (ضریب تعیین) بود. مدل XGBoost می تواند عیار سنگ معدن را با SSE 15.67، RMSE = 0.17، NMSE = 0.34، و R^2 = 0.66، مدل RFs با SSE 16.40، RMSE = 0.17، NMSE = 0.36، و R^2 = 0.30 پیش بینی کند. مدل SVR با SSE 19.94، RMSE = 0.19، NMSE = 0.43، و R^2 = 0.57، و مدل ANN-MLP با SSE = 21.00، RMSE = 0.19، NMSE = 0.46، و R^2 = 0.55. در نتیجه، مدل XGBoost موثرترین مدل در پیش‌بینی عیار سنگ مس بود.

کلمات کلیدی: شبکه عصبی مصنوعی پرسپترون چند لایه، جنگل های تصادفی، تقویت گرادیان شدید، رگرسیون بردار پشتیبانی.