Article

Application of Artificial Neural Network for the Prediction of Copper Ore Grade

Ntshiri Batlile Tsae [1,*]**, Tsuyoshi Adachi** [1] **and Youhei Kawamura** [2]

[1] Graduate School of International Resource Science, Akita University, Akita 010-8502, Japan;
    adachi.t@gipc.akita-u.ac.jp
[2] Division of Sustainable Resources Engineering, Faculty of Engineering, Hokkaido University,
    Sapporo 060-8628, Japan; kawamura@eng.hokudai.ac.jp
*   Correspondence: ntshiritsae@gmail.com

Abstract: Precise prediction of ore grade is essential in feasibility studies, mine planning, open-pit and underground optimization, and ore grade control. Conventional methods, such as geometric and geostatistical methods, are the most popular techniques for mineral resource estimation but fail to capture the complexity of orebodies. Due to this limitation, grades are incorrectly estimated, leading to inaccurate mine plans and costly financial decisions. Here, we propose an ore grade prediction method using an artificial neural network (ANN). We collected 14,294 datasets from the Jaguar mine in Western Australia. The proposed model was developed by incorporating lithology, alteration, eastings, northwards, altitude, dip, and azimuth to predict the grade, and the performance evaluation metrics were measured based on the mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), correlation coefficient, R, and coefficient of determination (R^2). The proposed ANN model outperformed classic machine learning methods with R^2, R, MAE, MSE, and RMSE of 0.584, 0.765, 0.0018, 0.0016, and 0.041, respectively. The Shapley technique was used to evaluate the feature importance of the input variables for the grade prediction. Lithology demonstrated the highest influence on ore prediction, whereas eastings had the least impact on output. The proposed approach is promising for ore model prediction.

Keywords: ore prediction; artificial neural network; feature importance

1. Introduction

Precise prediction of ore grade is important for mineral resource estimation and many mine operations, such as ore grade control, underground operations, open-pit optimization, and mine planning and design. Ore grade estimation plays a vital role in the economic evaluation of mining projects, capital allocation, sustainability, depletion rates, and mining feasibility. Estimating ore grade is complicated and problematic because of the multifaceted processes involved in ore deposition. Traditional methods, namely geometric and geostatistical methods, are the most popular in mineral resource estimation. Kriging is a well-known estimation technique in the mining industry and has gained enormous recognition as an accurate estimator of mineral resources. Kriging is an ideal spatial regression technique designed for the regional or local estimation of block grades as a linear combination of available data, which minimizes the estimation error [1]. Various kriging techniques have been applied for mineral resource estimation, such as simple kriging (SK), indicator kriging (IK), and ordinary kriging (OK). Ordinary kriging, also known as Best Linear Unbiased Estimator, is the most widely used technique for estimating mineral resources [2]. This technique can be used to estimate a value at an unsampled location in a region of interest using data from the region and a variogram model interpreted from all data within the region, which minimizes the expected error between the estimated and actual grades [3]. In addition, kriging can be used to estimate the mining block grades. This minimizes the expected error between the estimated and actual grades. Although the supremacy and

efficiency of these methods have been demonstrated in several studies [4–6], the major limitation of these conventional techniques is that they require assumptions based on the spatial correlation between samples to be estimated at unsampled location [7–9].

The spatial distribution of the kriging estimates tended to be smooth; they overestimated the low-grade values and underestimated the high-grade values. Deutsch and Journel [10] introduced sequential Gaussian simulation (SGS) as a solution to the smoothing problem of kriging. Pan et al. [11] concluded that conventional approaches may not provide the best grade estimates because of the complex relationship between the spatial pattern variability and grade distribution. However, the difficulty in estimating the grade of ore deposits with few data points using geometrical and geostatistical methods has paved the way for the application of artificial intelligence in grade estimation.

Over the past few decades, researchers [12–16] have applied neural networks to ore grade prediction. Advancements in technology have shown the immense potential of machine learning (ML) algorithms over other interpolation techniques for ore grade estimation because of their ability to learn any linear or nonlinear relationship between inputs and outputs. The neural network method is appealing and has become a versatile technique for grade prediction. Additionally, machine learning-based resource estimation techniques are more efficient and cheaper than traditional resource estimation approaches [17]. Moreover, ML contributes to the understanding of diverse types of ore deposits by modernizing hypothesis testing and geological modeling [17]. Machine learning techniques address various operational challenges in the mining industry, including mineral exploration, drilling and blasting, and mineral processing.

Aguilera et al. [18] studied the performance of deep learning- (DL) based models in ore grade estimation for a copper mine in Chile to reduce these differences in long- and short-term planning. They analyzed feed-forward neural network (FNN), one-dimensional (1D) convolutional neural network (CNN), and long short-term memory (LSTM) models. Matias et al. [19] examined the precision of kriging, regularization networks (RN), multi-layer perceptron (MLP), and radial basis function (RBF) networks when determining the slate quality. Schnitzler et al. [20] assessed the Random Forest performance with varying numbers of instances and input variables. The MLP network performed well in terms of test error and training speed. Samantha et al. [21] estimated ore grade values using an RBF network and compared the results to feed-forward neural networks and conventional ordinary kriging. They concluded that feed-forward neural networks provided better results than ordinary kriging. Chatterjee et al. [22] suggested the use of a genetic algorithm (GA) and k-means clustering techniques for ensemble neural network modeling of a lead–zinc deposit. Two types of ensemble neural network models were investigated: a resampling-based neural ensemble and a parameter-based neural ensemble. K-means clustering was used to select diversified ensemble members. The GA was used to improve the accuracy by calculating the ensemble weights. The results were compared with the average ensemble, weighted ensemble, best individual networks, and ordinary kriging models. The GA-based model outperformed all other methods that were considered. An artificial neural network (ANN) was trained to recognize the relationship between a sample point's location, lithology, and major metal content because the spatial correlation structures could not be extracted from the semi-variograms or cross-variograms between two major and minor elements [23]. Based on sample data, the network model can generate a model with many high-content zones.

The development of multi-layered ANN with multiple input variables has resulted in considerable advances in ANN accuracy, and numerous studies have been conducted on this topic. Mahmoudabadi et al. [24] suggested a hybrid method that combines the Levenberg–Marquardt (LM) method and a GA to identify the optimal initial weights of the ANN. Jalloh et al. [25] integrated an ANN and geostatistics for an optimum mineral reserve estimation. The drilling spatial locations (X, Y, and Z) and sample length were used to predict the grade of the mineral sand. They concluded that the model showed precise predictions of the ore grade; however, the major drawback of this approach was

that the model underestimated high-grade values that had relatively few training sets. Alawi et al. [26] predicted the grades of bauxite deposits from 163 drillholes by developing a multilayer feed-forward ANN model using a backpropagation algorithm. X and Y were used as input variables, whereas the thickness of the mineralized lengths of the deposit and the corresponding silica and alumina contents were used as target variables. The results show that the input variables could only explain 79% of the output variables. To make grade assessments of mineral deposits, Kaplan and Topal [27] suggested a modeling strategy that included k-nearest neighbor (kNN) and ANN. The kNN model predicted rock types and alteration levels before estimating the grades and estimates of geological information at non-sampled locations. In the second step, the ANN model uses the geological information predictions provided by the kNN model and the geographic information as input variables. Although existing literature highlights the efficiency and potential benefits of machine learning algorithms for the accurate prediction of grades, there are some shortcomings associated with these techniques. The most significant problem is that there are no set rules for determining the network hyperparameters to achieve the correct model structure; additionally, the method requires a computer-intensive procedure that involves trial and error to obtain the results.

The purpose of this study is to present an ore grade prediction approach based on an ANN model that incorporates spatial information (eastings, northings, and altitude), drilling parameters (dip and azimuth), and geological information (lithology and alteration) as model input variables and copper ore grade as an output variable. Previous researchers used sample locations and geological attributes (lithology and alterations); however, the proposed technique goes beyond the use of sample location and geological attributes by incorporating drilling parameters into ore grade prediction. The proposed technique is unique because of its ability to learn nonlinear relationships between input variables based on a combination of geological, drilling, and sample location information and the output variable, that is, copper grade. Seven input variables were selected as essential features for ore grade estimation, because they provided the relevant information required for the model to accurately predict the ore grade. The alteration and lithology are related to the mineralization of the orebody, whereas the sample location shows the exact coordinates of where the sample was collected. The dip and azimuth angles indicate the angles at which the drillhole was drilled. The proposed approach contributes to a better understanding of the complexities and types of ore deposits.

The remainder of this paper is organized as follows: Section 2 outlines the geology of the study area, dataset information, methodology, and data pre-processing. Section 3 describes the proposed ANN, network training, and its implementation. Section 4 presents the results and discussion, and Section 5 presents the conclusions.

2. Dataset and Methods

2.1. Geology of the Study Area

The dataset used in this case study was collected from the Jaguar mine, located 60 km north of Leonora in the Eastern Goldfields region of Western Australia. The Jaguar deposit lies four kilometers to the south of the historic Teutonic Bore mine. A map of the location is presented in Figure 1. The deposit consists of a steep west-dipping massive sulfide lens of pyrite/pyrrhotite, chalcopyrite, and sphalerite mineralization hosted in a succession of basaltic and andesitic flow sills. Mineralization occurs in basalts that lie above a thick basal rhyolitic sequence with an overlying andesite. The rhyolitic sequence comprises rhyolitic mass flow units and lavas that vary in nature from massive and locally flow-banded to highly auto-brecciated. The Jaguar stratigraphy strikes from north northwest (NNW) to south southeast (SSE) and dips steeply from 75° to 80° to the west. Drilling extends to a maximum downhole depth of 870 m. However, in this study, drilling extends to 190 m for simplicity of analysis. The data contained copper grade values measured as percentages of 185 drillholes with a drill spacing of 20 m.
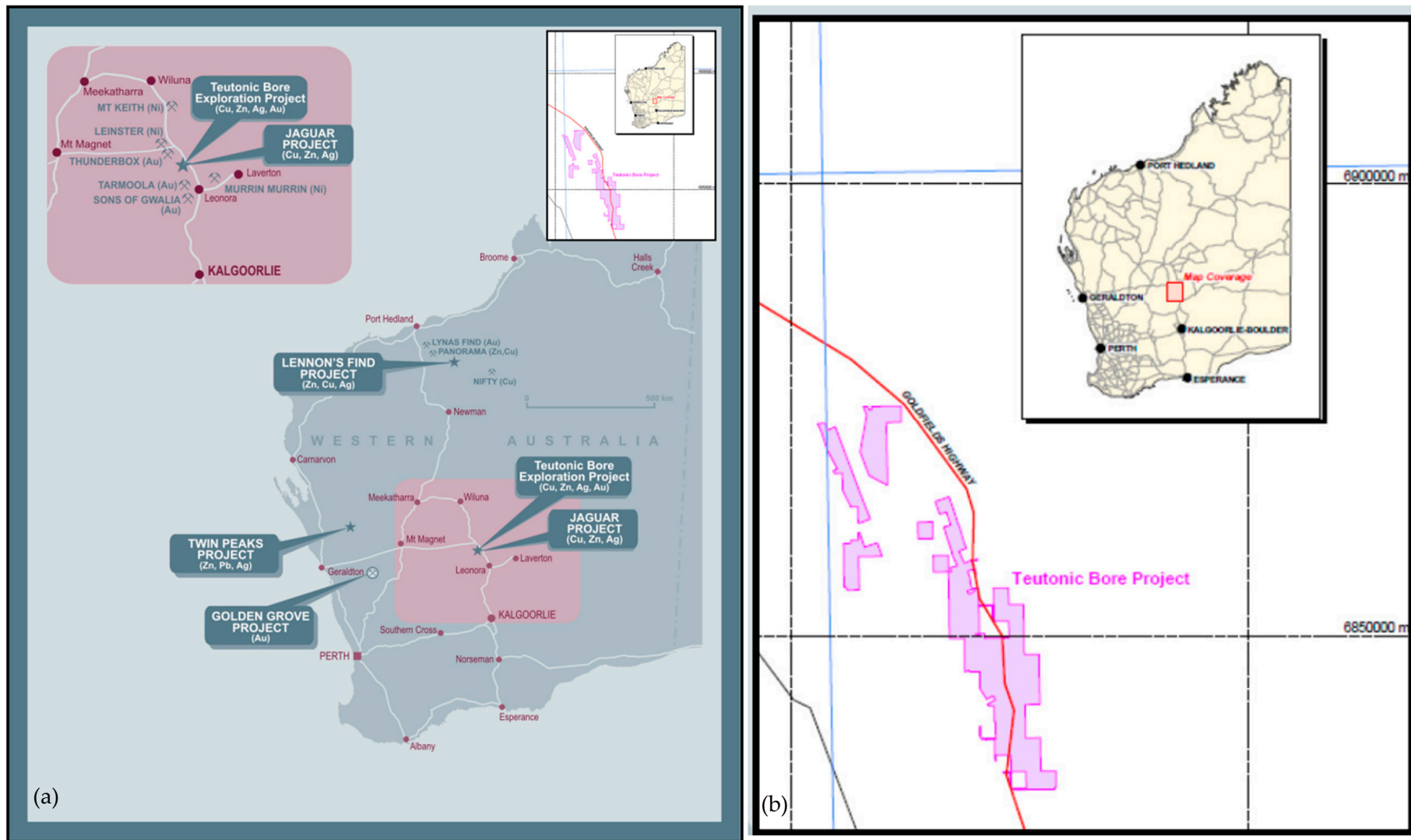
Figure 1. Jaguar Mine Location Map (**a**,**b**).

2.2. Dataset Collection and Pre-Processing

Grade estimation is based primarily on geological attributes, spatial information, and drilling parameters. The aim of this study was to develop a model that demonstrates the effects of sample location and geological and drilling parameters in accurately predicting ore grade. Samples from the drillholes were collected at 1-m intervals. The raw drillhole data were composited based on lithology and 14,294 samples were produced. Seven input variables, i.e., dip, azimuth, eastings, northings, altitude, lithology, and alteration, were investigated with only one output: the copper grade. In this study, lithologies that displayed similar characteristics were grouped into five categories, namely dolerite, basalt, andesite, massive sulfides, and sediments, to minimize estimation errors. The ore was extracted from four major alteration types: sericitization, chloritization, silification, and carbonatization. The alterations and lithologies are related to the chemical composition of the mineral deposits. East, north, and altitude indicate the location at which the sample was collected. The dip indicates the direction in which the drillholes are inclined from the horizontal plane, whereas the azimuth is the inclination angle measured from north during drilling. A list of unweighted ore grade prediction variables and the corresponding ore grades is displayed in Table 1.

Table 1. List of unweighted input variables and corresponding copper ore grade.

X	Y	Z	Dip	Azimuth	Lithology	Alteration	Cu Grade (%)
9879.20	55,954.33	4056.84	312	356	Basalt	Chloritization	0.0230
9878.16	55,954.46	4057.86	304	357	Dolerite	Sericitization	0.5896
↓	↓	↓	↓	↓	↓	↓	↓
9892.56	56,101.79	3965.46	321	330	Andesite	Chloritization	0.0090
9806.73	56,245.590	3879.88	329	333	Massive sulfides	Chloritization	0.0330

Dip and Azimuth were measured in degrees. ↓: There are other drillhole information in between the drillholes.

Figure 2 shows a flowchart of the overall analysis. The following steps were followed to prepare the ANN model: firstly, since this dataset used raw drillhole data, it was normalized to avoid spatial grade variability and noise caused by outliers, which differed greatly from other observations. Normalization was also performed to improve the learning performance of the model and avoid overfitting. Table 2 shows the descriptive statistics of the dataset, and Equation (1) shows the formula for data normalization. The numerical variables were normalized by using the mean and standard deviation as shown in Equation (1) where the normalized variable z is obtained by subtracting the mean μ from each value in x and all divided by the standard deviation σ. The data were transformed into numerical values, because neural networks only work with numbers. A hold-out method was used to split the data into two sets: training and testing. 14,179 datasets were used for training. To show how the well the model performed across the drillholes, a set of data from the entire single drillhole was excluded from the dataset and used as a test case.

Table 2. Descriptive statistics of the dataset.

	X	Y	Z	Dip	Azimuth	Cu Grade (%)
Count	14,294	14,294	14,294	14,294	14,294	14,294
Mean	0.9960	0.9983	0.9840	0.8011	0.7012	0.5893
Std	0.0015	0.0094	0.0040	0.3394	0.2219	1.7594
Min	0.9925	0.9548	0.9765	0.0000	0.1321	0.0000
25%	0.9947	0.9818	0.9806	0.8969	0.6509	0.0140
50%	0.9958	0.9914	0.9833	0.9411	0.7460	0.0330
75%	0.9972	0.9961	0.9880	0.9691	0.8711	0.2110
Max	1.0000	1.0000	1.0000	0.9556	1.0000	26.996

Raw drillhole Data

Feature Selection

Data normalization

The ANN architecture

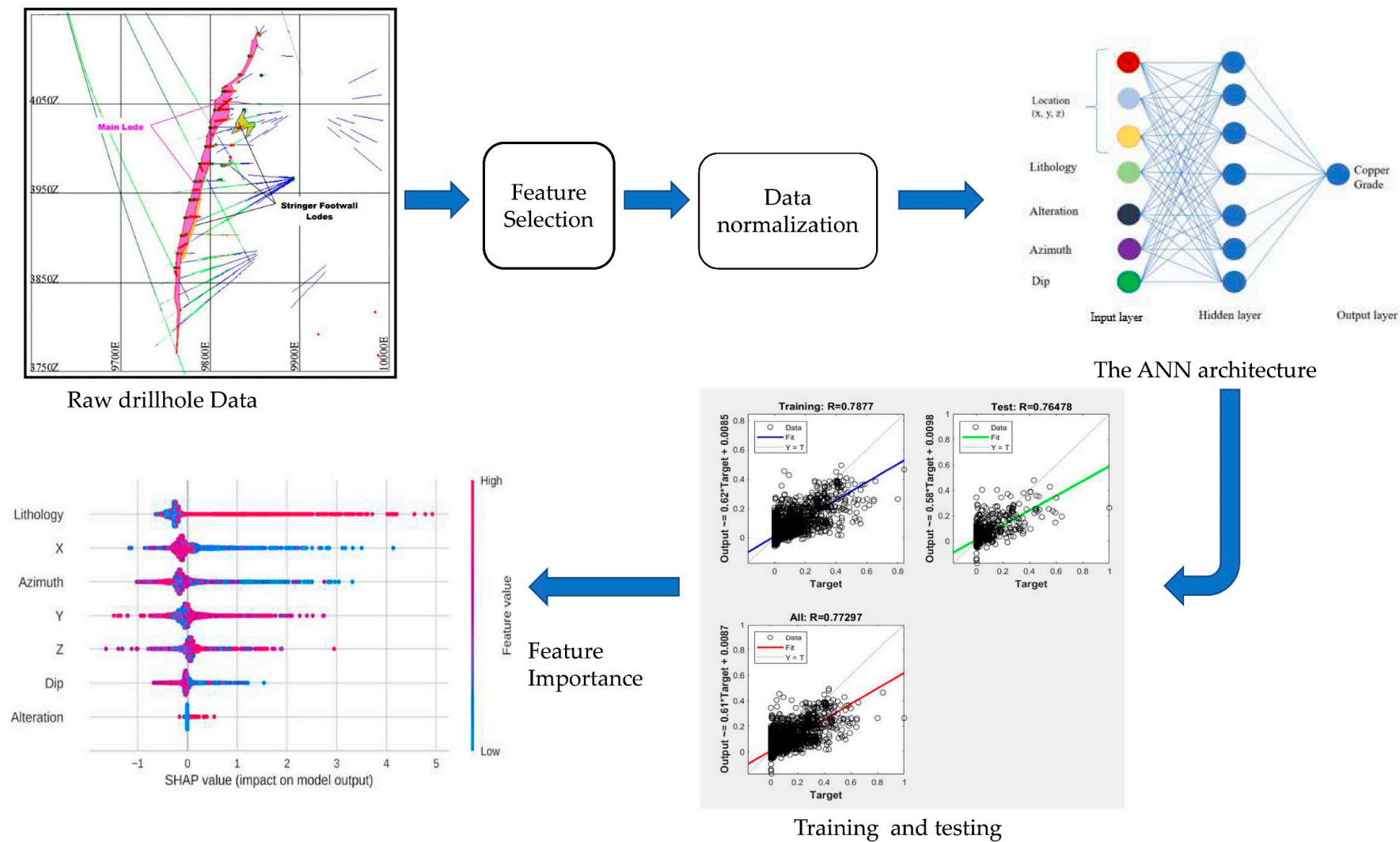Feature Importance

Training and testing

Figure 2. Flowchart of the proposed technique for ore grade estimation.

Finally, Shapley values were used to determine the feature importance of the input parameter on the model output.

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

Figure 3 illustrates the copper ore grade histogram, with a minimum value (y_{min}) of 0, maximum value (y_{max}) of 26.996, and mean (y_{mean}) of 0.5893. The copper grade distribution is positively skewed, with a high coefficient variation of (σ/y_{mean}) = 2.986, indicating the presence of extreme values in the dataset. Some areas of the ore deposit were rich in Cu, whereas others had low Cu grades. A high ratio between the mean and y_{max} is one of the features that renders accurate grade prediction difficult, because it requires a model to identify high-grade areas among low-grade areas. The proposed multilayer feed-forward ANN can help solve this problem by learning the nonlinear relationships between the inputs and outputs.
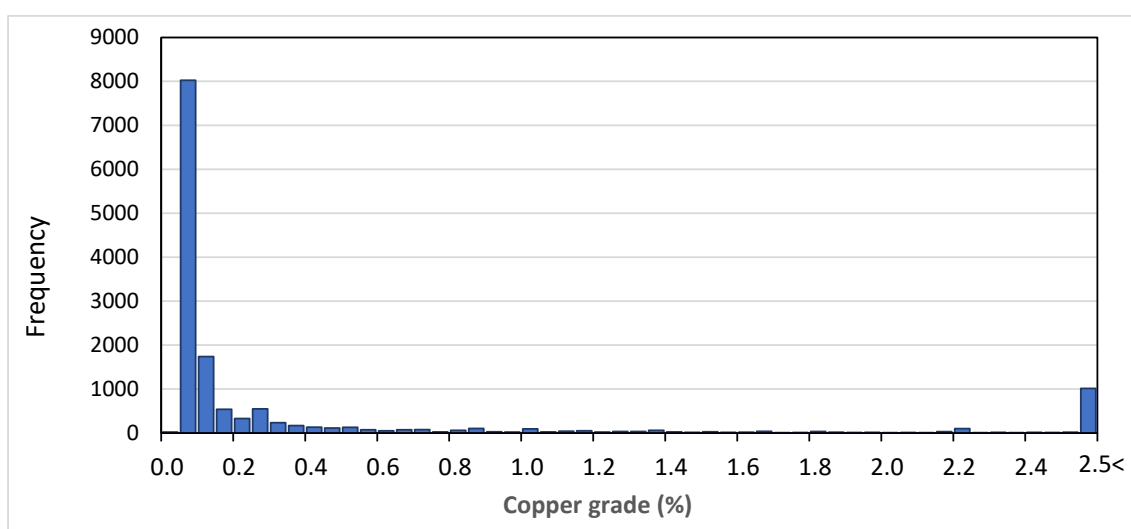
Figure 3. The histogram of the copper ore grade distribution.

3. The Proposed ANN for Grade Estimation

ANNs are composed of 'neurons', which are programming constructs that simulate the properties of biological neurons. A network of weighted connections allows information to propagate through the network to solve artificial intelligence problems without the network designer having a model of a real system. An ANN is a robust machine learning technique that can be applied to model complicated patterns, solve prediction issues by recognizing existing relationships in a dataset, and predict the output values for a given input dataset [28]. It consists of three major interconnected layers: the input, hidden, and output layers, which determine the network architecture. ANNs have been widely used in different fields, and the recognition of this approach has been attributed to their ability to learn and model nonlinear complex relationships. Over the years, ANNs have gained significant attention in mineral resource estimation because of the outstanding learning and generalization performance of the model from given parameters. ANNs have proven to be a prominent technique for estimating mineral resources, and studies [22,27] have gone beyond the use of raw drillhole spatial positions to include critical geological parameters such as lithology and alteration.

In this study, the sample location (X, Y, and Z), geological attributes (lithology and alteration), and drilling parameters (dip and azimuth) were combined to predict the copper grade. The ANN architecture was determined by trying several neural network configurations and selecting the one with the lowest error rate. The proposed ANN architecture comprises one input layer consisting of seven neurons, one hidden layer, and one output layer, as shown in Figure 4. The tanh activation function is used for the hidden layers,

whereas a linear function is used for the output layer. The mean square error (MSE) is a popular regression model evaluation technique that utilizes the squared difference between the predicted and actual values and averages them. The MSE sums the actual and predicted values and divides them by the total number of observations. The MSE was used as a cost function because it ensures that the model does not have outlier predictions with large errors because the MSE assigns higher weights to these errors in the squaring part of the function. The MSE expression is shown in Equation (2); however, a major limitation of this method is that the squared part magnifies the errors if the model contains extreme values. The lower the MSE, the better the results are. Gradient descent and momentum terms were used to train the ANN. In this case, gradient descent was used as an optimization algorithm to determine the local minimum of a differentiable function and minimize the cost function, i.e., MSE.

$$\text{MSE} \; = \; \frac{1}{n}\sum_{i\,=\,1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

where MSE is the mean square error, n is the number of observations, y_i the observed value, and $\hat{y}_i$ = predicted value.
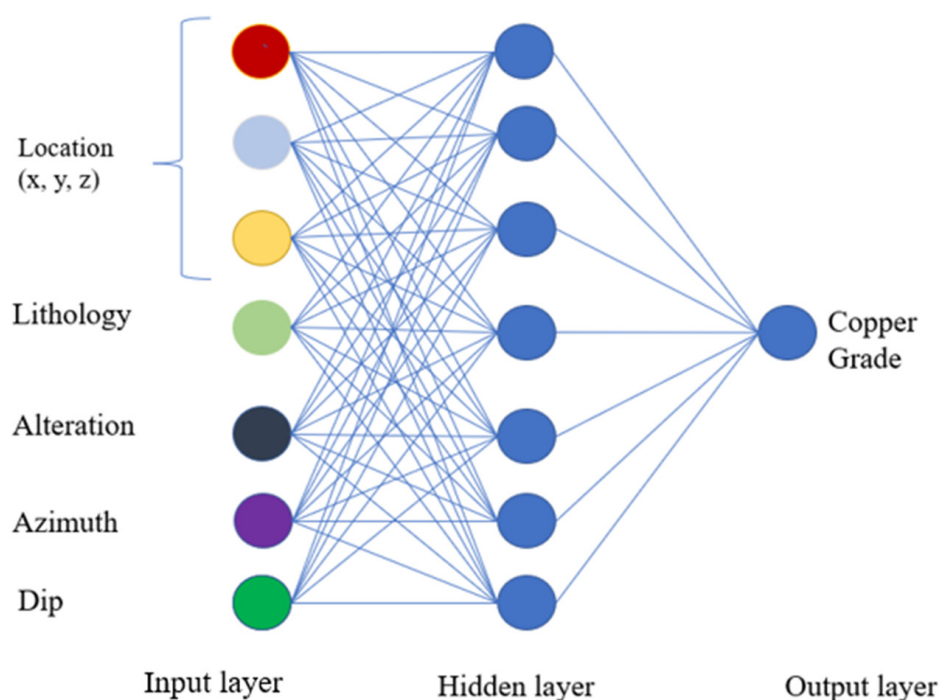
Figure 4. The proposed ANN model architecture.

Network Training and Implementations

The input and output data were normalized from zero to one to supplement the learning performance of the ANN. The data were split into training and testing, respectively, using the hold-out cross validation method. The training process was performed using MATLAB (R2020b) with a deep-learning toolbox on a workstation with a Windows 10 64-bit operating system, Intel Core i7-8750H CPU @ 2.2 central processing unit, 16 GB memory, and NVIDIA GeForce GTX graphics processing (Mouse Computer Co., Ltd, Akita city, Akita, Japan). Although there are numerous methods to train neural networks, the backpropagation method is the most adaptable and powerful. For multilayer neural networks, learning in this manner is most effective. Backpropagation algorithms are widely used because they are excellent at overcoming prediction issues. In this study, an ANN was trained using a Bayesian regularization backpropagation algorithm. Bayesian regularization (BR) exploits a mathematical process that converts nonlinear regression into a well-posed statistical problem in the manner of a ridge regression [29]. Essentially, BR

generates a network that minimizes the combination of errors and squared weights to determine the correct combination and achieve a generalized model. Since evidence procedures provide an objective Bayesian criterion for determining when to stop training, they are difficult to overtrain. They are also difficult to overfit because BRANN only calculates and trains on a small number of effective network parameters or weights, effectively turning off those that are no longer relevant [30]. In most cases, this effective number is less than the number of weights in a typical fully connected backpropagation neural network, which was adopted in this study because of the performance and accuracy of the predicted models. It can also handle uncertainties in the model parameters, which contribute significantly to accurate prediction.

4. Results and Discussion

All models were trained based on similar parameters of data splitting, including learning rate, training ratio, and epochs. The performance metrics used for the evaluation of prediction performance are mean squared error (MSE), mean absolute error (MAE), Root mean square error (RMSE), correlation coefficient (R), and coefficient of determination (R^2). MAE is the average of all absolute errors. RMSE is a parameter that can be used to evaluate a model's performance by determining the amount of deviation between the predicted and observed values. The key advantage of MSE and RMSE is that they account for uncertainty in predictions; however, their primary downside is that the methods are problematic when there are a lot of extreme values. Even though MAE is an absolute measure like MSE, its outstanding feature over MSE is that it is less influenced by outliers. RMSE and MAE are defined by Equations (3) and (4). The correlation coefficient is an evaluation approach used to measure the relationship between variables, while the coefficient of determination measures how well the model predicts the outcome. It measures the goodness of fit and is the proportion of variance in the dependent variable that the model explains.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

where RMSE is the root mean square error, n is the number of observations, y_i the observed value, and $\hat{y}_i$ = predicted value.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4}$$

where MAE = mean absolute error, n = number of observations, y_i = observed value, and $\hat{y}_i$ = predicted value.

4.1. The Proposed ANN Model Analysis

The dataset used in this study comprises 185 drillholes. The primary issue with such a large dataset is the significant variation in drillhole samples and the erratic distribution of geochemical anomalies; thus, careful selection of the data partitioning procedure is crucial in order to improve the accuracy of the prediction model. Individual samples were modelled along the *z*-axis based on the cores sample intervals of 1 m. A hold-out method was used to split data into two sets: training and testing. The 14,179 dataset was for training and 115 dataset for testing. The performance of the model across drillholes was validated by using an independent and unused testing dataset. Figure 5 shows the regression analysis diagram for the training data, testing data and the overall data for the drillhole. It can be noted that the correlation coefficients, R of training, testing, and overall model, are 0.788, 0.765, and 0.773, respectively. As BR was used to train the ANN model, there was no overfitting or underfitting because it has an objective function that stops training whenever necessary. The numbers of layers, neurons, and activation functions were optimized. As indicated by the green line in Figure 5, a set of a single drillhole was used for testing to provide an unbiased evaluation of the final model fit to the training

dataset. The best model architecture consists of seven inputs, one hidden layer, and one output. Although the results clearly show that the input variables are highly correlated with the output, with high accuracy, a conclusion should not be drawn solely on the basis of the high correlation coefficient; further investigation of statistical analysis must also be considered. Figure 6 shows the learning performance of the models based on the MSE. It can be highlighted that the best performance of the training and test data was attained at epoch 1000 steps of iteration with a corresponding MSE value and gradient value of 0.0016 and 0.00066, correspondingly. The MAE and RMSE of the ANN model prediction were 0.018, and 0.041, respectively. The MAE has a lower value because it does not place too much emphasis on outliers, and this loss function provides a generic and even measure of how well the model performs. This finding suggests that the proposed model performed well based on the MAE when considering the variability of the copper.
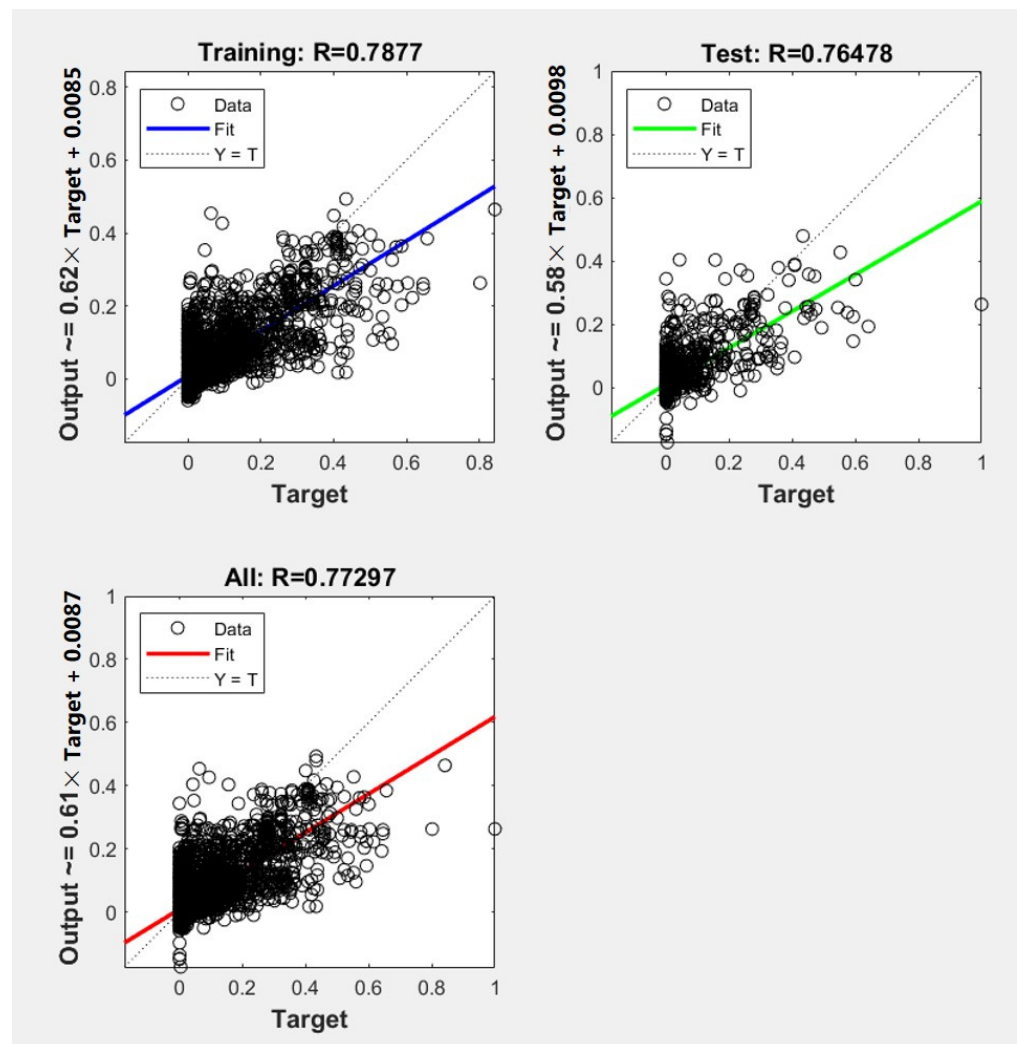
Figure 5. ANN regression models showing actual and predicted data distribution with the blue, green, and red showing best line fit for training, test, and the overall model data, respectively. The white circles represent the data set.
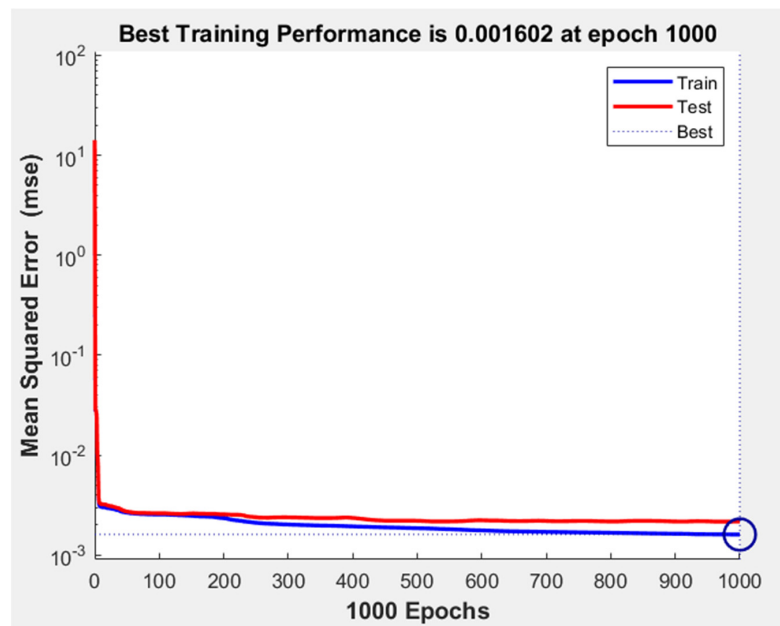
Figure 6. ANN learning curve based on mean square error with the blue and red line showing the training and testing error.

Additionally, an error histogram was generated to show the distribution of errors of the training and testing dataset. Figure 7 shows the prediction error distribution. The zero error, represented by the orange line in this histogram, indicates that the error is largely concentrated in the region of ±0.08. Figure 8 shows the data distribution of the actual versus predicted grade of the model. This figure shows that the copper grade can be moderately estimated by the proposed ANN model. The overall model results showed minimum errors, indicating that the input and output variables were highly correlated. The results show that the proposed ANN is a reliable and powerful tool for ore grade prediction and can be applied to mining operations.
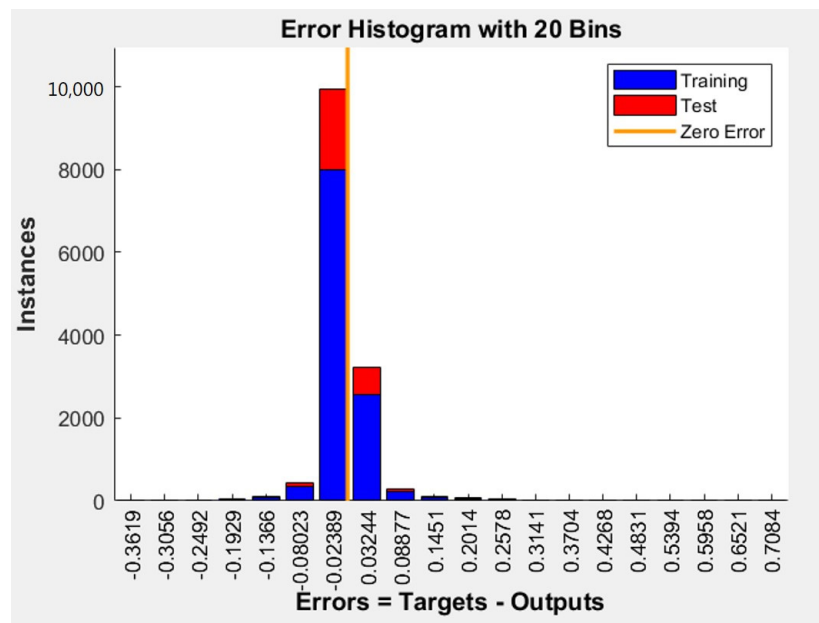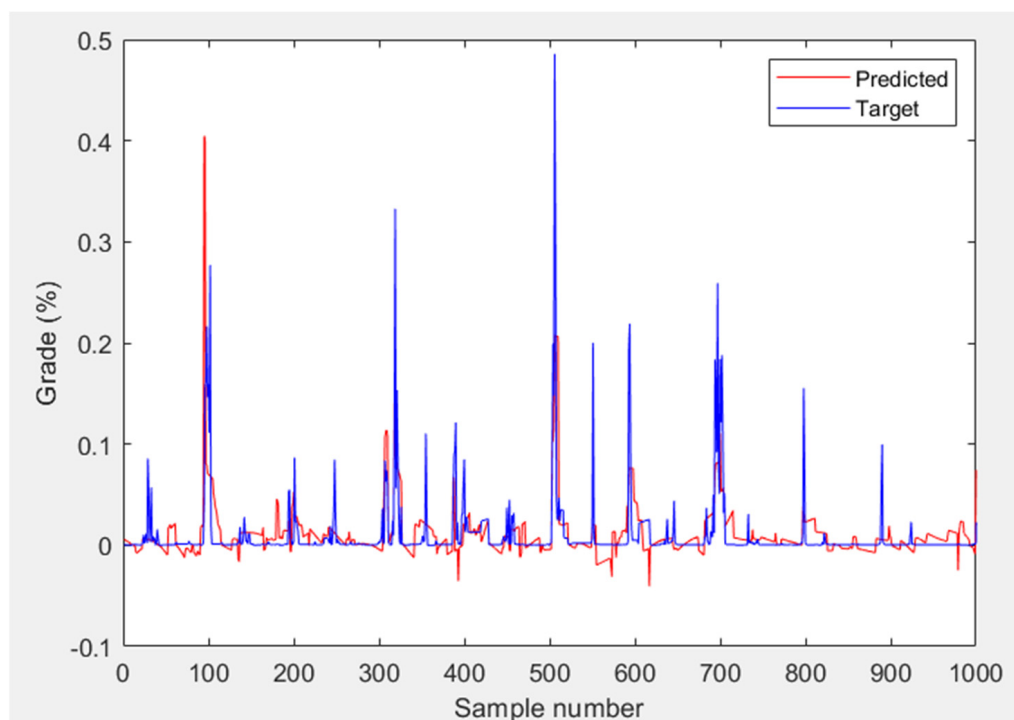

Figure 7. Error histogram.

Figure 8. Actual versus predicted values of the ANN model with the red and blue lines showing the predicted and actual values, correspondingly.

4.2. Model Comparison with Other Machine Learning Methods

A comparative analysis of various ore grade estimation techniques was performed to determine the best copper grade prediction. MSE, MAE, RMSE, R, and R^2 were used as evaluation performance measures to compare the ANN model with other machine learning techniques. Although the R^2 provides some useful insights into the regression model, one should not rely solely on the measure in assessing a statistical mode because it does not reveal information about the causal relationship between the independent and dependent variables, nor does it indicate the correctness of the regression model, which is why the other evaluation performance metrics were considered in this study. The MSE, MAE, and RMSE indicate the accuracy and precision of the model. The best model was chosen based on the highest correlation R^2 and the lowest MAE and MSE errors.

To evaluate the prediction performance of basic machine learning approaches to the ANN model, hyperparameter optimization was performed in order to produce a robust and credible predictive model. Hyperparameter tuning is very important in model development. Table 3 shows the summary of how the chosen classic machine learning methods were optimized. Table 4 shows the results of the statistical methodologies used to predict the copper grade. The coefficients of determination, R^2 for the classic methods—extra trees regressor, random forest regressor, light gradient boosting machines (LGBM), K neighbor regressor, and linear regression—were 0.575, 0.563, 0.546, 0.541, and 0.123, respectively. The results indicated that these statistical methods exhibited moderate correlation coefficients, whereas linear regression performed poorly. Linear regression showed the worst performance, with the lowest correlation, R^2 of 0.123, which is not surprising given that linear regression does not account for nonlinear relationships. Since the ore grade is a varying component, this linear regression method cannot produce a strong model.

**Table 3.** Hyperparameter tuning results for classic machine learning methods.

Method	Parameter	
Extra trees regressor	bootstrap	false
	criterion	Square_error
	n_estimator	100
	random_state	211
Random forest regressor	bootstrap	true
	criterion	Square_error
	n_estimators	100
	random_state	4822
Light gradient boosting machine	boosting_type	Gbdt
	min_child_samples	20
	min_child_weight	0.001
	n_estimators	100
	num_leaves	31
	random_state	179
	subsample_for_bin	200,000
K neighbors regressor	Leaf_size	30
	Metric	Euclidean
	n_neighbors	28
	p	2
	weights	distance
Linear regression	fit_intercept	True
	n_jobs	−1
	positive	false

Table 4. Model performance of the machine learning statistical methods.

Methodology	R^2	R	MAE	MSE	RMSE	SD
Artificial neural network	0.584	0.765	0.018	0.0016	0.041	0.0414
Extra trees regressor	0.575	0.756	0.319	0.0020	0.0448	0.0761
Random forest regressor	0.563	0.746	0.332	0.0021	0.0458	0.0758
Light gradient boosting machine	0.546	0.723	0.369	0.0022	0.0463	0.0663
K neighbors regressor	0.541	0.665	0.415	0.0024	0.0485	0.0779
Linear regression	0.123	0.315	0.821	2.725	1.643	1.512

Figure 9 presents the prediction error plots for the classic machine learning approaches using R^2 evaluation metrics. The prediction error graphs show the actual values versus the predicted values generated by the models. These models show us how much variance there is in the model. Figure 9 clearly shows that, despite having high correlation coefficients, the actual and predicted values for the random forest regressor, extra tree regressor, light gradient boosting machines, and K neighbor regressor have significant errors around them. The data distribution of linear regression model appears rather poor, and it should be emphasized that the model is not a good fit for the existing dataset. To perform a fair statistical comparison of the models, it is interesting to report the standard deviation (SD) of each model. The standard deviation measures the spread of data around the mean, with an SD around zero being ideal. As can be seen in Table 4, the proposed ANN model has the lowest SD of 0.041 when compared to the other machine learning approaches. The proposed ANN outperformed the other machine learning methods with R^2, R, MAE, MSE, and RMSE of, 0.584, 0.765, 0.018, 0.0016, and 0041, respectively. Hence, better prediction accuracy was achieved by the ANN. The proposed ANN model had the lowest MAE, MSE, and RMSE followed by the extra tree regressor, with MAE, MSE and RMSE values of 0.319, 0.0020, and 0.0448, respectively. The subsequent models–random forest regressor, light gradient boosting machine, K neighbors regressor, and linear regression–showed MAE of 0.332, 0.369, 0.415, and 0.821, respectively. This clearly indicates the superior performance

of the ANN model compared with other machine learning methods. It can be concluded that the results from our proposed approach can moderately predict the copper grade because of the high coefficient of determination, and R^2 and the standard deviation of this model was optimal as it was closer to zero. Moreover, the data for ANN is well-distributed which makes it a more reliable and powerful method than the classic machine learning methods.
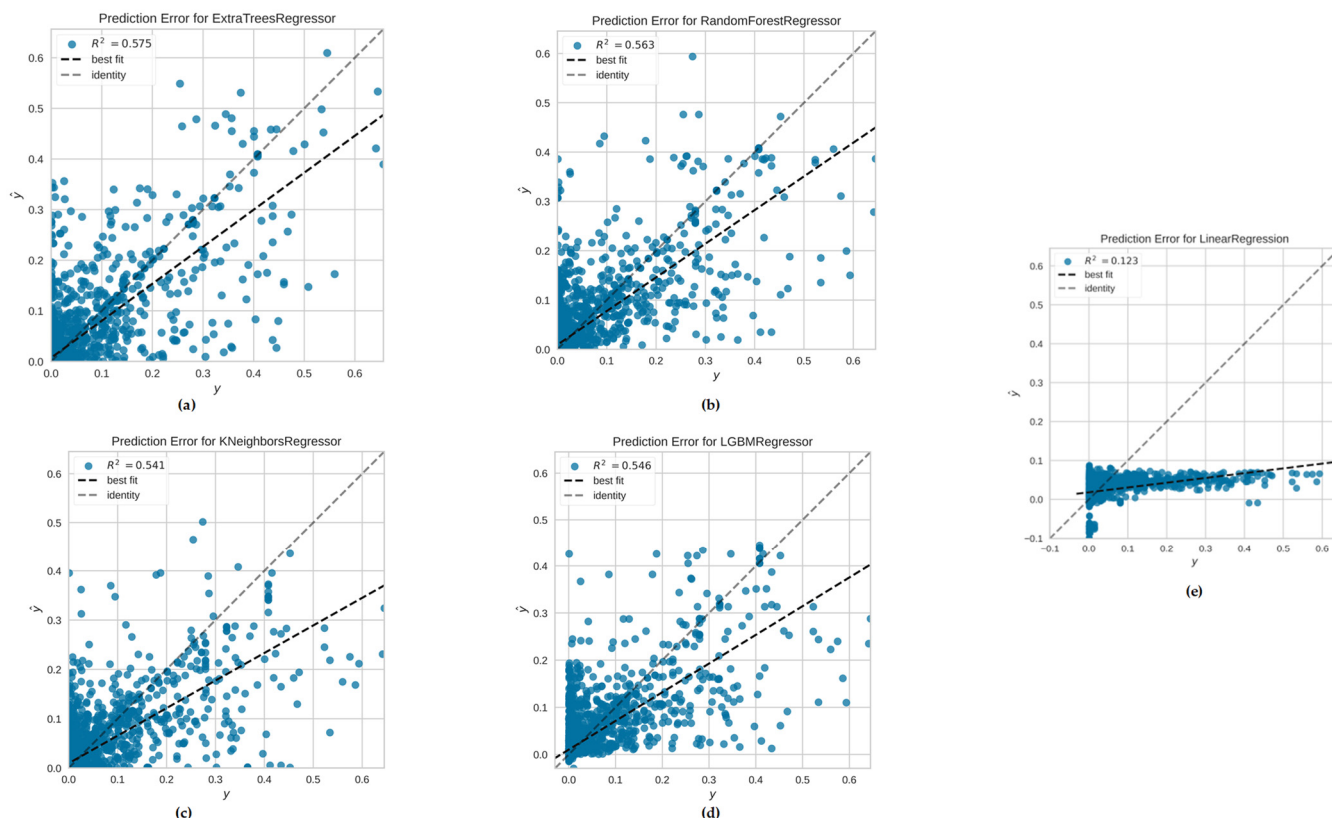
Figure 9. The prediction error plots for the classic machine learning techniques, with y and $\hat{y}_i$ showing the data distribution of actual and predicted values. (**a**) Prediction error plot for extra trees regressor; (**b**) Prediction error plot for random forest regressor; (**c**) Prediction error plot for K neighbors regressor; (**d**) Prediction error plot for light gradient boosting machine; (**e**) Prediction error plot for linear regression.

4.3. Feature Importance Analysis

The correlation matrix provides the relevant information for feature importance analysis. Figure 10 depicts the correction matrix based on the correlation coefficient, which measures the linear relationship between two variables. The color variation represents the correlation relationship between two variables, with dark blue indicating a significant negative correlation and contribution and dark red indicating a strong positive correlation. The correlation matrix normally has values ranging from -1 to 1, with 1 indicating a perfectly positive linear correlation between two variables, 0 indicating that there is no linear correlation between the two variables, and -1 indicating a completely negative linear correlation between two variables. Lithology correlates with copper grade more strongly than the other variables. Lithology correlates positively with altitude and alteration but negatively with eastings, northings, altitude, azimuth, and dip. Eastings correlate positively with azimuth, northings, dip, and alteration but negatively with lithology. Figure 3 illustrates the other variable relationships.
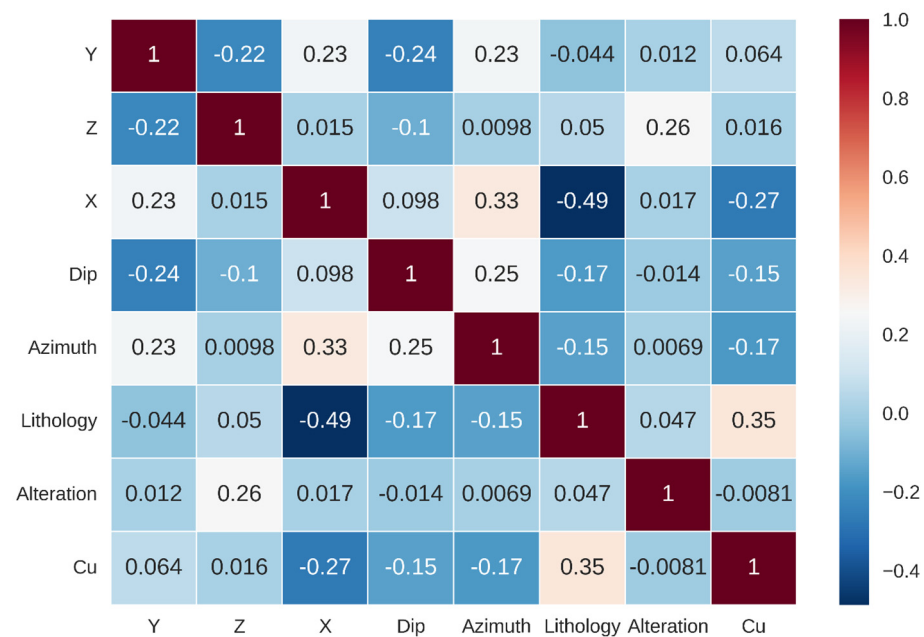
Figure 10. Correlation matrix based on predictors and output variable with dark red showing strong positive correlation and blue showing strong negative correlation between variables.

Researchers are often reluctant to adopt machine learning algorithms because of the complexities associated with evaluating the mechanism inside the model. Therefore, an ANN is often treated as a black box, where the connection weights of the neurons are highly volatile over the amount of data. To verify the soundness of this study, the Shapley Additive Explanation (SHAP) was used for feature importance. SHAP is the most prominent technique adapted from cooperative game theory, it is a useful tool for feature importance, and it supports explainable machine learning [31]. The Shapley value approach was used to reveal and understand the feature importance or contribution of the input parameters to the grade prediction of copper. This was also performed to avoid the black box issue. The kernel explainer way of the SHAP was used to determine important features of the model. Kernel SHAP is a technique that generates the relevance of each feature employing a particular weighted linear regression. The significant outcomes produced are Shapley values from game theory as well as coefficients from a local linear regression. It is of utmost importance to note that Kernel SHAP can interpret any machine learning model regardless of its nature, which is why it was used for this study. The SHAP library's KernelExplainer computes SHAP values using 10,000 background samples. However, the n_samples parameter can be used to change this. Fewer samples may be sufficient for smaller datasets or less complex models, whereas more samples may be required for larger datasets or more sophisticated models to obtain accurate results.. The background dataset is used for feature integration. To determine the impact of a feature, it should be set it to "missing" and the change should be monitored in model output. Due to the fact that most models are not built to handle random missing data during testing, we mimic "missing" by replacing the feature with the values it takes from the background dataset. So, if the background dataset is a simple sample of all zeros, we can approximate a missing feature by setting it to zero. For simple problems, the entire training set can be used as the background dataset, but for larger problems, we considered using a single reference value or the k-means function to summarize the dataset. It is worth noting that for sparse situations, we accept any sparse matrix but converted to LIL format for efficiency reasons.

Figure 11 depicts the feature importance of the input features. The color variation indicates the impact of the features on the model output, with blue showing the least contribution and red the most. Lithology had the greatest influence on copper grade prediction, with a SHAP value of 5. This research showed that lithology significantly affects

grade prediction because it is linked to the geochemical formation and mineralization of the deposit. Altitude was the second most influential input parameter. This is because the samples were collected at 1 m intervals, allowing the model to simulate the spatial distribution along the drillholes and improve the performance and accuracy. The eastings had the least impact on the prediction because the drillhole samples extended along the *x*-axis. Consequently, the model performance may have been skewed because closer holes tended to exhibit characteristics similar to those of other holes. The dip and azimuth did not show much significance in the grade prediction.
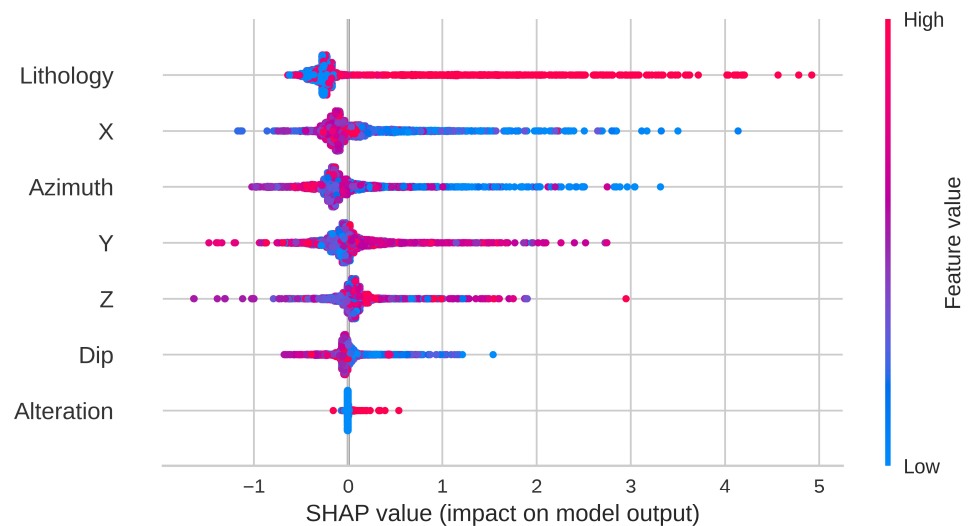
Figure 11. Feature importance based on the SHAP values. Red represents higher feature values and blue represents lower feature values.

The main contribution of this study is the launch of an innovative and novel ore prediction approach that uses seven input variables that incorporate geological attributes, spatial locations, and drilling parameters to predict ore grade using ANNs. This study also compared the efficacy of ANNs with five classic machine learning techniques. All these classic methods were outperformed by the ANN. Researchers have combined optimization algorithms, the generic algorithm, k-means clustering, the generic algorithm and Levenberg–Marquardt, and the combination of kNN and ANNs adopted for grade prediction over the years. However, for this research, we adopted the Bayesian regularization algorithm over other algorithms because of its precision and performance. It should be noted that the proposed technique can be used to assess grades for a wide range of mineral resources. Despite its promising potential, the main drawback of the technique is that it does not account for geological discontinuities, faults, and joints in mineral estimation. Furthermore, because ANN performance is data-driven, an adequate amount of data is required for an accurate grade prediction model.

5. Conclusions

Accurate ore grade prediction is challenging because of the multifaceted processes associated with geological formation and ore deposition. Precise grade prediction plays a significant role in mine planning, ore grade control, and feasibility studies. In this study, we propose a multilayer feed-forward ANN that combines seven input variables, sample locations (X, Y, and Z), geological attributes (alteration and lithology), and dip and azimuth, for ore grade estimation of the Jaguar mine in Australia. The proposed technique is data-driven and learns the relationship between the input and output values to predict the grade. The performance metrics, R^2, R, MAE, MSE, and RMSE, were used to evaluate the prediction performance of the ANN model and the other machine learning techniques: linear regression, K neighbors regressor, random forest regressor, light gradient boosting

machine, and extra tree regressor. The ANN model outperformed these classical approaches with R^2, R, MAE, MSE, and RMSE of 0.584, 0.765, 0.018, 0.0016, and 0.041, respectively. Moreover, the standard deviation of the proposed ANN model was the lowest with an SD of 0.0414. Shapley values were used to assess the input variables to measure feature importance. Lithology has the greatest influence on copper grade prediction because it is associated with the mineral composition of the orebody. It is important to note that this study presents the implementation of a robust and powerful methodology for ore grade estimation by learning the relationship between the input and output variables. The developed ANN model demonstrates that this technique can be used to supplement exploration activities, thereby reducing drilling requirements. It can also be used for mine-planning analysis as an efficient mineral resource evaluation approach that generates the best block model for mine design, resulting in extensive savings. Although the ANN model accurately predicted the ore grade, it did not consider the geological structure of the orebody, faults, and discontinuities. The presented results are promising and pave the way for further research in the future. In future research, it would be worthwhile to compare the proposed model to the established geostatistical methods such as kriging. Furthermore, future approaches should integrate feature selection in the data preprocessing step in machine learning as an effective way to remove unnecessary variables and reduce the dimensionality of input features. The best input variables can then be used to accurately predict the grade.

Author Contributions: Conceptualization, N.B.T., T.A. and Y.K.; methodology, N.T and Y.K; software, N.B.T.; validation, T.A. and Y.K.; Formal Analysis, N.B.T.; Investigation, N.B.T.; Resources, T.A.; Data Curation, N.B.T. and T.A.; Writing—Original Draft Preparation, N.B.T.; Writing—Review and Editing, T.A. and Y.K.; Visualization, T.A.; Supervision, T.A. and Y.K.; Project Administration, T.A.; Funding Acquisition, T.A. All authors have read and agreed to the published version of the manuscript.

References

1. Akbar, D.A. Reserve estimation of central part of Choghart north anomaly iron ore deposit through ordinary kriging method. *Int. J. Min. Sci. Technol.* **2012**, *22*, 573–577. [CrossRef]
2. Isaaks, H.; Srivastava, R.M. Ordinary Kriging. In *Applied Geostatistics*; Oxford University Press Inc.: New York, NY, USA, 1989; pp. 288–295.
3. Badel, M.; Angorani, S.; Panahi, M.S. The application of median indicator kriging and neural network in modeling mixed population in an iron ore deposit. *Comput. Geosci.* **2011**, *37*, 530–540. [CrossRef]
4. Wackernagel, H. *Multivariate Geostatistics: An Introduction with Applications*; Springer: Berlin, Germany, 1998; p. 256.
5. Cressie, N. Spatial prediction and ordinary kriging. *J. Int. Assoc. Math. Geol.* **1989**, *21*, 493–494. [CrossRef]
6. Paithankar, A.; Chatterjee, S. Grade and Tonnage Uncertainty Analysis of an African Copper Deposit Using Multiple-Point Geostatistics and Sequential Gaussian Simulation. *Nat. Resour. Res.* **2018**, *27*, 419–436. [CrossRef]
7. Yamamoto, J.K. Correcting the Smoothing Effect of Ordinary Kriging Estimates. *J. Int. Assoc. Math. Geol.* **2005**, *37*, 69–94. [CrossRef]
8. Abuntori, C.A.; Al-Hassan, S.; Mireku-Gyimah, D. Assessment of Ore Grade Estimation Methods for Structurally Controlled Vein Deposits—A Review. *Ghana Min. J.* **2021**, *21*, 31–44. [CrossRef]
9. Sadeghi, B.; Madani, N.; Carranza, E.J.M. Combination of geostatistical simulation and fractal modeling for mineral resource classification. *J. Geochem. Explor.* **2015**, *149*, 59–73. [CrossRef]
10. Deutsch, C.V.; Journel, A.G. *GSLIB: Geostatistical Software Library and User's Guide*; Oxford University Press: New York, NY, USA, 1998.
11. Pan, G.; Harris, D.P.; Heiner, T. Fundamental issues in quantitative estimation of mineral resources. *Nat. Resour. Res.* **1992**, *1*, 281–292. [CrossRef]
12. Yama, B.; Lineberry, G. Artificial neural network application for a predictive task in mining. *Min. Eng.* **1999**, *51*, 59–64.
13. Wu, X.; Zhou, Y. Reserve estimation using neural network techniques. *Comput. Geosci.* **1993**, *19*, 567–575. [CrossRef]

14. Jafrasteh, B.; Fathianpour, N. A hybrid simultaneous perturbation artificial bee colony and back-propagation algorithm for training a local linear radial basis neural network on ore grade estimation. *Neurocomputing* **2017**, *235*, 217–227. [CrossRef]

15. Jafrasteh, B.; Fathianpour, N.; Suárez, A. Comparison of machine learning methods for copper ore grade estimation. *Comput. Geosci.* **2018**, *22*, 1371–1388. [CrossRef]

16. Li, X.-L.; Li, L.-H.; Zhang, B.-L.; Guo, Q.-J. Hybrid self-adaptive learning based particle swarm optimization and support vector regression model for grade estimation. *Neurocomputing* **2013**, *118*, 179–190. [CrossRef]

17. Zhang, S.E.; Nwaila, G.T.; Tolmay, L.; Frimmel, H.E.; Bourdeau, J.E. Integration of Machine Learning Algorithms with Gompertz Curves and Kriging to Estimate Resources in Gold Deposits. *Nat. Resour. Res.* **2021**, *30*, 39–56. [CrossRef]

18. Olmos-De-Aguilera, C.; Campos, P.G.; Risso, N. Error reduction in long-term mine planning estimates using deep learning models. *Expert Syst. Appl.* **2023**, *217*, 119487. [CrossRef]

19. Matías, J.M.; Vaamonde, A.; Taboada, J.; González-Manteiga, W. Comparison of Kriging and Neural Networks With Application to the Exploitation of a Slate Mine. *J. Int. Assoc. Math. Geol.* **2004**, *36*, 463–486. [CrossRef]

20. Schnitzler, N.; Ross, P.-S.; Gloaguen, E. Using machine learning to estimate a key missing geochemical variable in mining exploration: Application of the Random Forest algorithm to multi-sensor core logging data. *J. Geochem. Explor.* **2019**, *205*, 106344. [CrossRef]

21. Samanta, B.; Bandopadhyay, S. Construction of a radial basis function network using an evolutionary algorithm for grade estimation in a placer gold deposit. *Comput. Geosci.* **2009**, *35*, 1592–1602. [CrossRef]

22. Chatterjee, S.; Bandopadhyay, S.; Machuca, D. Ore Grade Prediction Using a Genetic Algorithm and Clustering Based Ensemble Neural Network Model. *Math. Geosci.* **2010**, *42*, 309–326. [CrossRef]

23. Koike, K.; Matsuda, S.; Suzuki, T.; Ohmi, M. Neural Network-Based Estimation of Principal Metal Contents in the Ho-kuroku District, Northern Japan, for Exploring Kuroko-Type Deposits. *Nat. Resour. Res.* **2002**, *11*, 135–156. [CrossRef]

24. Mahmoudabadi, H.; Izadi, M.; Menhaj, M.B. A hybrid method for grade estimation using genetic algorithm and neural networks. *Comput. Geosci.* **2009**, *13*, 91–101. [CrossRef]

25. Jalloh, A.B.; Kyuro, S.; Jalloh, Y.; Barrie, A.K. Integrating artificial neural networks and geostatistics for optimum 3D geological block modeling in mineral reserve estimation: A case study. *Int. J. Min. Sci. Technol.* **2016**, *26*, 581–585. [CrossRef]

26. Al-Alawi, S.M.; Tawo, E. Application of Artificial Neural Networks in Mineral Resource Evaluation. *J. King Saud Univ. Eng. Sci.* **1998**, *10*, 127–138. [CrossRef]

27. Kaplan, U.E.; Topal, E. A New Ore Grade Estimation Using Combine Machine Learning Algorithms. *Minerals* **2020**, *10*, 847. [CrossRef]

28. Dumakor-Dupey, N.K.; Arya, S. Machine Learning—A Review of Applications in Mineral Resource Estimation. *Energies* **2021**, *14*, 4079. [CrossRef]

29. Burden, F.; Winkler, D. Bayesian regularization of neural networks. *Artif. Neural Netw.* **2008**, *458*, 23–42.

30. Awan, S.E.; Raja, M.A.Z.; Gul, F.; Khan, Z.A.; Mehmood, A.; Shoaib, M. Numerical Computing Paradigm for Investigation of Micropolar Nanofluid Flow Between Parallel Plates System with Impact of Electrical MHD and Hall Current. *Arab. J. Sci. Eng.* **2020**, *46*, 645–662. [CrossRef]

31. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 25 November 2017; pp. 1–10.