

Article

# Optimized Weighted Ensemble Approach for Enhancing Gold Mineralization Prediction

M. M. Zaki <sup>1,2,3</sup> , Shaojie Chen <sup>1,2,\*</sup>, Jicheng Zhang <sup>1,2</sup>, Fan Feng <sup>1,2</sup>, Liu Qi <sup>1,2</sup>, Mohamed A. Mahdy <sup>4</sup> and Linlin Jin <sup>1,2</sup>

- <sup>1</sup> State Key Laboratory Breeding Base for Mining Disaster Prevention and Control, Shandong University of Science and Technology, Qingdao 266590, China; mohamed.zaki@azhar.edu.eg (M.M.Z.); zhangjicheng@sdust.edu.cn (J.Z.); fengfan0213@126.com (F.F.); xihuancicou@163.com (L.Q.); jinlinlin@sdust.edu.cn (L.J.)
- <sup>2</sup> College of Energy and Mining Engineering, Shandong University of Science and Technology, Qingdao 266590, China
- <sup>3</sup> Mining and Petroleum Engineering Department, Faculty of Engineering, Al-Azhar University, Cairo 11884, Egypt
- <sup>4</sup> Faculty of Computers and Artificial Intelligence, Beni-Suef University, Beni-Suef 62521, Egypt; marafa.mahdy@fcis.bsu.edu.eg
- \* Correspondence: chensj@sdust.edu.cn; Tel.: +86-13730951965

**Abstract:** The economic value of a mineral resource is highly dependent on the accuracy of grade estimations. Accurate predictions of mineral grades can help businesses decide whether to invest in a mining project and optimize mining operations to maximize the resource. Conventional methods of predicting gold resources are both costly and time-consuming. However, advances in machine learning and processing power are making it possible for mineral estimation to become more efficient and effective. This work introduces a novel approach for predicting the distribution of mineral grades within a deposit. The approach integrates machine learning and optimization techniques. Specifically, the authors propose an approach that integrates the random forest (RF) and k-nearest neighbor (kNN) algorithms with the marine predators optimization algorithm (MPA). The RFKNN\_MPA approach uses log normalization to reduce the impact of extreme values and improve the accuracy of the machine learning models. Data segmentation and the MPA algorithm are used to create statistically equivalent subsets of the dataset for use in training and testing. Drill hole locations and rock types are used to create each model. The suggested technique's performance indices are superior to the others, with a higher R-squared coefficient of 59.7%, a higher R-value of 77%, and lower MSE and RMSE values of 0.17 and 0.44, respectively. The RFKNN\_MPA algorithm outperforms geostatistical and conventional machine-learning techniques for estimating mineral orebody grades. The introduced approach offers a novel solution to a problem with practical applications in the mining sector.

**Keywords:** machine learning; kriging; MPA optimizer; log normalization; hybrid algorithm



**Citation:** Zaki, M.M.; Chen, S.; Zhang, J.; Feng, F.; Qi, L.; Mahdy, M.A.; Jin, L. Optimized Weighted Ensemble Approach for Enhancing Gold Mineralization Prediction. *Appl. Sci.* **2023**, *13*, 7622. <https://doi.org/10.3390/app13137622>

Academic Editor: Yosoon Choi

Received: 17 May 2023

Revised: 19 June 2023

Accepted: 23 June 2023

Published: 28 June 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Reserve estimation is critical in mine planning and development, impacting profitability and sustainability. Accurately estimating ore grade is crucial for reserve estimation, but challenges arise due to ore deposits' internal variability and complex formation processes [1].

Several mathematical and statistical techniques have been developed to overcome the challenges of estimating metal grades. These techniques range from more conventional geostatistical approaches, such as ordinary and indicator kriging, to innovative machine learning algorithms, such as artificial neural networks and random forests. The geometry, geology, and grade distribution of the ore deposit are all essential factors when choosing a technique [2].

Traditional geostatistical methods, such as ordinary and indicator kriging, are commonly used in grade estimation. These techniques are based on spatial interpolation of data and the assumption that the spatial pattern of grade distribution is stationary. While these methods have been used successfully for a long time, they have some significant drawbacks. For example, they require that the data meet certain assumptions, they do not use all of the available data, and they cannot find complex, nonlinear correlations [3]. The difficulties and limitations of conventional geostatistical methods for estimating ore grade have prompted researchers to investigate alternative approaches. In recent years, various computational learning approaches have been developed that can forecast ore grades more correctly without relying on underlying assumptions. These techniques have proved successful in the mining industry [4].

Machine learning (ML) models are a category of computational algorithms with significant potential for use in ore grade estimation in the mining industry. ML models learn from data and then use that knowledge to predict previously unknown data. They can be more adaptable than conventional statistical models when they include and combine various variables, such as geological, geophysical, and geochemical data [5]. This improves mining ore grade estimate. Nevertheless, one of the limitations of ML models is that they need to be fed an adequate amount of data to extract useful features. In addition, machine learning models may be computationally costly, and training and optimizing them might require a lot of computing resources.

Machine learning (ML) is a promising approach for spatial estimates in various domains by a large number of research. The most often used machine learning models include artificial neural networks, support vector machines, k-nearest neighbors, cubist, Gaussian process regression, and random forests [6–15]. These models have been utilized in the process of spatial estimate in a variety of different domains, including mineral prediction. However, typical ML approaches may have performance difficulties when working with non-linear, inconsistent, or higher-dimensional data. Researchers have developed hybrid and ensemble models as a solution to these problems. These models combine traditional ML methods with additional soft computing approaches or use bagging and boosting methodology. The benefits of many soft computing approaches may be combined into a single method through the process of hybridization, which can result in increased productivity. These hybrid approaches have shown to be particularly effective when dealing with huge, multi-dimensional, and non-stationary datasets [16–24]. The need for a more reliable, consistent, and objective method of estimating gold content in exploration projects has led to the development of a hybrid system based on feature optimization and prediction. Both the feature optimization technique and the prediction strategy employ a hybrid model that integrates different ML models to achieve improved accuracy. The feature optimization approach is responsible for selecting the most pertinent characteristics for prediction.

The MPA is one of the recent optimization algorithms; it is inspired by the optimal foraging strategy between predator and prey in marine ecosystems to solve the optimization. The foraging strategy enables the MPA to converge to the optimal global solution within a convenient time even in multidimensional search spaces. Accordingly, based on the experimental results illustrated in [25], the MPA showed outstanding performance against the old algorithms GA, PSO, and the well-known GSA, CS, SSA, and CMA-ES; furthermore, the MPA was initially evaluated on twenty-nine test functions, test suite of CEC-BC-2017, three engineering benchmarks, and two real-world engineering design problems. These features have motivated researchers to employ the MPA in many various applications, i.e., in [26], the MPA was used to optimize the control parameters of four machine learning (ML) models for predicting the compressive strength (CS) of concrete. In another study [27], authors employed the MPA algorithm to perform a feature selection process; they developed a robust human action recognition framework by integrating deep learning and swarm intelligence. In [28], the MPA was used to optimize three ML models for predicting and analyzing parks and attached green spaces based on a dataset of urban green space. These points motivated us to use the MPA algorithm in the proposed methodology.

The primary purpose of this piece of study is to present an innovative method for estimating the mineral resources (prediction of gold content in the Quartz Ridge area). The suggested approach, which will be referred to as RFKNN-MPA, would increase the accuracy of grade estimations by combining the Random Forest (RF) and k-nearest neighbor (kNN) models with optimized using the Marine Predators Optimization Algorithm (MPA). The method incorporates geological data, such as lithology, and geographical data, such as easting, northing, and elevation, into the model creation process. The model-making process consists of three steps. In the first step, the random forest and kNN models generate initial predictions. In the second step, the MPA method is used to optimize the parameters of both the kNN and the random forest models. In the third step, the predictions generated by the optimized versions of the Random Forest and kNN models are merged to provide a final forecast. Once the model has been trained with enough data, it can learn the relationship between the input factors (rock type and geographic position) and the gold grade, the outcome variable of interest. The RFKNN-MPA technique improves grade forecasts and mineral exploration estimates by integrating different models and refining their parameters using the MPA algorithm. To the best of our knowledge, this study is the first attempt to estimate the grade using the optimized weighted integration between KNN and RF regressions optimized by the MPA; further, the experimental results investigated how well the proposed RFKNN-MPA predicted outcomes compared to the results of other machine learning models and geostatistical approaches. The outcomes demonstrated that the RFKNN-MPA method performed better in accuracy and reliability than the other models.

## 2. Method

The following section provides an overview of the research methods used in the investigation and the efficiency evaluation that was conducted.

### 2.1. Predication Approaches

Estimating a gold grade can be carried out in various ways, the most common of which fall into one of three broad categories: geostatistical, traditional, or hybrid. This proposed study aims to develop a hybrid system for gold-grade prediction. The suggested hybrid method, dubbed RFKNN-MPA, combines the Random Forest algorithm with the K-nearest neighbor strategy and MPA optimization. The experiments were carried out using the commercial program MATLAB, while the geostatistical methods were performed on the Surpac (6.6.2) software.

In light of this, a novel hybrid method known as RFKNN-MPA was evaluated and compared to the following techniques for geostatistics strategies (OK and IK) and machine learning approaches (RF, KNN, GPR, DT, and FCN). The underlying principles are briefly discussed in this part.

#### 2.1.1. Geostatistical Technique

##### Ordinary Kriging

Geostatistics is a better prediction method than traditional methods. Kriging algorithms are widely regarded as the best linear unbiased estimator capable of producing reliable estimates even for locations that have not been sampled [29]. A linear combination of a number of nearby values is used to determine the overall result, with weights given to minimize the variation in the predicted value. Lagrangian multipliers are a valuable tool for determining the weights that should be used when normalization is a consideration.

Kriging uses the variogram model to quantify the autocorrelation pattern in the research region [30]. The variogram is produced by taking the average of the semi-variances, which are computed as follows for each combination of locations and distance bin, and then plotting the results:

$$\gamma(h) = \frac{1}{2} E[\{Z(x) - Z(x+h)\}^2] \quad (1)$$

According to this formula, the semi-variance between any two locations separated by the vector  $h$  is denoted by  $\gamma(h)$  (the lag distance). The semi-variance is half of the anticipated square difference between two values of the variable  $Z$ , specifically  $Z(x)$  and  $Z(x+h)$ , which are differentiated by the vector  $h$ . This is how the semi-variance is defined. An essential part of kriging is the sampling strategy used. This method needs samples to be divided by various lag lengths, from minor to very large, to ensure that the variogram can accurately portray the spatial complexity of the area. This is crucial for obtaining a dependable variogram model and enabling kriging to achieve its optimum precision.

On the other hand, kriging assumes a linear relationship between samples. Additionally, the structural data required for the Kriging method (i.e., the covariance or variogram) is based on two-point statistics. As a result, kriging may not always be sufficient to represent the comprehensive nature of spatial nonlinearity and complex features. Second-order stationarity is when the mean and standard deviation are both the same across the study area, and the covariance depends only on the lag distance  $h$  (where  $h$  is a vector that includes the size of the distance between the data pairs and the direction of the line that connects them) [31,32].

The formula for the OK estimator is as follows:

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (2)$$

Since extremely positive values significantly affect semi-variance calculations, kriging is extremely sensitive to skewed distributions. This makes it possible for Kriging estimations to be unreliable [30].

### Indicator Kriging

According to Journel (1983) [33], indicator kriging is a nonparametric approach for evaluating the likelihood of multiple threshold values,  $Z_k$ , given geographical data. The first thing that must be performed in order to use the kriging indicator (IK) is to convert it into indicators [34]. This is accomplished by giving the values 0 and 1 in the binary system to grades that are, respectively, above and below a particular cut-off. The explanation is the binomial coding of the data as either 1 or 0, depending on how it is connected to the specified  $Z_k$  cut-off number [35]. For any given value,  $Z(x)$ :

$$i(x, Z_k) = \begin{cases} 1, & \text{if } Z(x) \leq Z_k \\ 0, & \text{if } Z(x) > Z_k \end{cases} \quad (3)$$

The indicator variogram is combined with a weighted average of the indicator values at the sample sites to estimate the IK at the unsampled point. A variogram of indicators is the distribution of the variance of the indicator function at varying distances and orientations from a sample point (1 if the variable is present and 0 otherwise).

Indicator kriging has several advantages, including the ability to use complex data and soft data and the ability to estimate the probabilities of occurrence for a binary or categorical variable. However, it requires many samples to provide reliable estimates and can be affected by outliers. It also requires software proficiency and a strong understanding of spatial statistics and geostatistics [36]. In particular, it is a suitable choice when the data distribution is significantly skewed, which is commonly the situation in ore grade estimation. This is because it allows for more accurate predictions of the actual values of the variables.

### 2.1.2. Machine Learning Algorithms in Grade Prediction

During the scope of this research, a unique hybrid approach known as RFKNN-MPA was evaluated alongside six different machine learning algorithms. These algorithms included RF, K-NN, GPR, DT, and FCN. These algorithms were selected because of their ability to represent non-linear correlations with little to no user input, lack of assumption requirements, long track records of success, and proven robustness.

### Random Forest (RF)

The foundation of RF is an ensemble of regression trees, a family of algorithms that divide the training data into classes of probability based on a set of if-then rules. The CART algorithm (Breiman, 1984) serves as the foundation for RF. This algorithm fits a single tree, utilizing the whole training data. Unlike other tree approaches that only create a small number of trees, RF generates hundreds or thousands of decision trees [37]. The fundamental concept of RF is to repeatedly choose bootstrap samples from the given data, construct a decision tree (DT) for each bootstrap sample, and combine the estimates from several DTs into the prediction model. A single DT can overfit the data, resulting in a lower bias but greater prediction variance. Averaging the predictions from multiple DTs can achieve a trade-off between bias and variance [8].

The strategy relies on identifying the most significant data error as the primary error and correlating other data inaccuracies with it. Selecting a general error based on the most data gives us a more accurate and reliable forecast. This prediction becomes even more accurate when the data fed into the model does not change [38]. It is widely acknowledged that this algorithm is among the most effective learning algorithms due to its ability to accurately classify various inputs [39]. One of the strengths of this classifier is that it performs quite well even when used in massive datasets [40]. The most crucial quality of random forests is their superior performance in assessing the significance of variables, which establishes the contribution of each variable to the prediction of outcomes [41].

### K-Nearest Neighbors (K-NN)

Classification and regression problems can be tackled with KNN, a non-parametric instance-based learning technique [42], by finding the k-number of points closest to a given data point and using the mean of their values as the projected value for that point.

The KNN algorithm starts by calculating the distance between the new data point and the training points. The distance can be determined using a variety of distance metrics, such as the Euclidean distance, the Manhattan distance, or the Minkowski distance. After calculating the distances, the KNN algorithm selects the k-nearest neighbors from the training set that are most similar to the new data point. In regression, the algorithm calculates the average of the k-nearest training points to estimate the new data point's value. The mathematical expression that may be used to describe KNN is as follows: First, determine how far away the new data point is from each position in the training group. Second, Select the k-nearest neighbors that best describe the new data point based on the computed distances. Finally, predict the value of the new data point by taking the average of the values of the k closest training points [22].

The KNN algorithm is a flexible, simple, and easy-to-implement algorithm that can be applied to various data types, including those with incomplete values and categorical factors. However, it can be computationally expensive due to the need to calculate the distance between each new data point and each training point. It can also be susceptible to unimportant or noisy features. To prevent overfitting and underfitting, selecting the k value carefully is vital. In this research, the KNN algorithm was used with a k value of 13, which was found to be the optimal number through experimental results for predicting gold concentration.

### Gaussian Process Regression (GPR)

Gaussian processes are non-parametric techniques utilized in the Bayesian method of machine learning. They can solve both supervised and unsupervised learning problems, such as regression and classification. The GPR-based system offers various practical benefits over other supervised learning algorithms, like flexibility, abilities to generate uncertainty predictions, and learning noisy and smooth parameters through training data. Other supervised learning methods are limited in their ability to learn these parameters. In addition, GPR performs well on small datasets and can provide [43]. Therefore, it can be employed in continuous variable forecasting, modeling purposes, mapping, and interpolating [44],

and it can be guided from training data to represent the intrinsic nonlinear relationship in any dataset. Due to the presence of this characteristic, it is suitable for evaluating the grade of the ore.

$$F(x) \sim GP(m(x), k(x, x')) \tag{4}$$

where  $m(x) = E[f(x)]$  is the mean function at input  $x$ , and  $k(x; x') = E[(f(x) - m(x))(f(x') - m(x'))^T]$  is the covariance function that represents the dependency between the function values for various input points  $x$  and  $x'$ .

The squared exponential covariance function is a popular option for the kernel in the statistical technique known as Gaussian process regression. This kernel assumes that the function being modeled is continuous and smooth and that data points that are close together strongly correlate with one another. Gaussian process regression uses zero mean and squared exponential covariance functions. This is seen in Equation (5).

$$K(x, x') = \sigma_f^2 \exp\left(-\frac{r}{l}\right) \tag{5}$$

where  $r$  is equal to  $(|x - x'|^2/l^2)$  and  $\sigma_f$  and  $l$  are hyperparameters that have a major impact on the performance of the GP algorithm. Model noise is  $\sigma_f$ , and length scale is  $l$ . Input parameters have a high covariance regardless of their proximity, but this value decreases exponentially with distance. Several covariance and kernel functions can be used to optimize the GPR rational quadratic. The model of the distribution of predictions is conditioned on the training data. This is made possible via

$$p(f^* | x^*, D) = GP(f^* | \mu^*, \sigma^{*2}) \tag{6}$$

The mean prediction  $\mu^*$  may be calculated using the formula below.

$$\mu^* = k(x^*)^T (K(X, X) + \sigma_N^2 I_N)^{-1} * t \tag{7}$$

Furthermore, the variance prediction  $\sigma^*$  may be computed as

$$\sigma^{*2} = \sigma_N^2 - k(x^*)^T (K(X, X) + \sigma_N^2 I_N)^{-1} k(x^*) + k(x^*, x^*) \tag{8}$$

The training dataset's covariance matrix is represented by the matrices  $K(X, X)$ , and the testing dataset's covariance matrix is represented by the matrices  $K(X^*, X^*)$ . The previous equations make it possible to demonstrate that the linear combination of the observed target may be proven to be the mean forecast. The observed target did not influence the variance; instead, it depended entirely on the observed inputs. Gaussian distributions have this attribute.

GPR is a robust technique that can be used to make predictions and estimate uncertainty. It assumes the function to be a random sampling from a *Gaussian process (GP)* and returns a probability density for the function's value at each given input position. It is flexible enough to include information from multiple sources and models and can deal with non-linear and non-stationary functions. The choice of kernel function can be challenging and requires prior information, and it can be computationally expensive for massive datasets or high-dimensional input spaces. In mineral exploration, GPR can predict the *geographical distribution of minerals* using information gathered from *geological, geophysical, and geochemical studies*.

#### Decision Tree (DT)

Decision tree regression predicts dataset output after developing a decision tree structure. As a supervised learning method, this model analyses the data to create a model and forecast the output based on the input values [45].

A DT is a data structure that can visualize and analyze data. It splits the data according to specified criteria and assigns each choice a probability. The data are first partitioned into nodes or points, and then branches are drawn between them to represent the decisions that need to be made. This process, known as “splitting,” is determined by the available data results. Each decision tree branch represents a possible outcome and is assigned a probability based on the values of the other branches. The model is evaluated using new data, and its accuracy and reliability are assessed based on the results [46].

Decision tree regression equations are often complex algorithms, such as randomized decision trees, gradient-boosted trees, and random forests. These equations use a variety of mathematical processes to build a model and make predictions.

Decision tree regression is a valuable tool for non-technical people because it is easy to use, does not require much computing power, and can be very accurate. However, it may be less precise and reliable than other models, and it can take a very long time to train. The model may also struggle with overly complex datasets or situations where the data must be balanced. Finally, the trees may grow to be too big and unreadable [47].

#### Fully Connected Neural Network (FCN)

A fully connected neural network (FCN) is a type of artificial neural network (ANN) that employs direct connections between nodes in various layers. It is a popular deep learning framework. In this architecture, all nodes from one layer are linked to those from the next. FCN layers are commonly used in supervised (classification, regression) and unsupervised (density estimation) learning tasks.

By analyzing its structure and composition, FCNs can accurately predict a material’s chemical, mineral, and physical characteristics. They work by developing mathematical models that are meant to describe the relationship between the inputs (features) and the outputs (results) (mineral content). The network of neurons compresses and processes information from the input to generate an output. Each neuron is connected to every other neuron with weights, and a learning process called backpropagation is used to update the weights [48].

FCNs can be used to predict mineral composition using a simple equation and the backpropagation algorithm. Backpropagation is a method for updating the weights of each neuron based on the degree of discrepancy between the input and the desired output. This helps the neural network minimize error through data-driven learning. Backpropagation is used to change the weights between randomly allocated neurons.

FCNs are well-suited for estimating mineral content because of their accuracy, adaptability to different input/output formats, and ability to capture non-linear correlations. However, they also have drawbacks, including overfitting and poor interpretability. Overfitting occurs when a neural network learns the training data too well, leading to poor performance on new data. Additionally, FCNs can be difficult to interpret, making it challenging to understand why a particular prediction was made.

#### 2.2. Marine Predators Optimization Algorithm (MPA)

Biologically inspired metaheuristic algorithms are a type of algorithm that was developed to address complicated optimization issues by utilizing natural events and biological processes. Algorithms like the whale optimization algorithm (WOA), genetic algorithm (GA), ant colony optimization (ACO), differential evolution (DE), lion optimization algorithm (LOA), grasshopper optimization algorithm (GOA), sine cosine algorithm (SCA), bat algorithm (BA), particle swarm optimization (PSO), grey wolf optimizer (GWO), simulated annealing (SA), and marine predator algorithm (MPA) are all examples of this kind. They have solved several engineering, financial, and computer science optimization challenges. One of the biological metaheuristic algorithms developed by [25] will be shown in this work. This method is known as the Marine Predator Algorithm (MPA). Optimization strategies have seen a dramatic uptick in popularity in recent years. When dealing with the

many different optimization sectors, such as machine learning and feature selection, they are more effective than other strategies currently in use [49].

The foraging strategies of marine predators in the wild serve as the basis for the MPA mathematical model. MPA can accommodate both the Lévy and Brownian statistical distributions. The Lévy search technique involves traversing space with a series of prominent hops, while the Brownian method makes a systematic and consistent progression across the search space. While the Brownian technique guarantees visits to far-flung places, the Lévy strategy's strength lies in its thorough and precise search. This collaboration has greatly enhanced the search capabilities of MPA.

The movement equation is an essential equation employed in the MPA method. It governs how the predators move about the solution space and is thus very important. The equation for movement is defined as follows:

$$x_i(t+1) = x_i(t) + v_i(t) \quad (9)$$

where  $x_i(t)$  represents the location of the  $i$ th predator at time  $t$ ,  $v_i(t)$  represents the velocity of the  $i$ th predator at time  $t$ , and  $t$  represents the current iteration of the algorithm.

The MPA algorithm's strengths lie in its rapid convergence to optimal solutions and adaptability to multimodal and massively parallel optimization problems. The method might become trapped in the local optimum and requires careful parameter tuning.

### 2.3. The Proposed RFKNN-MPA Methodology

A combined model is useful when the predictions of different ensemble members do not match up or when errors of the different members are not closely linked. The primary concerns when using a hybrid model are the construction or selection of component models and the manner of combining their outputs. A hybrid model's ensemble members should be chosen so their predictions do not match up [50].

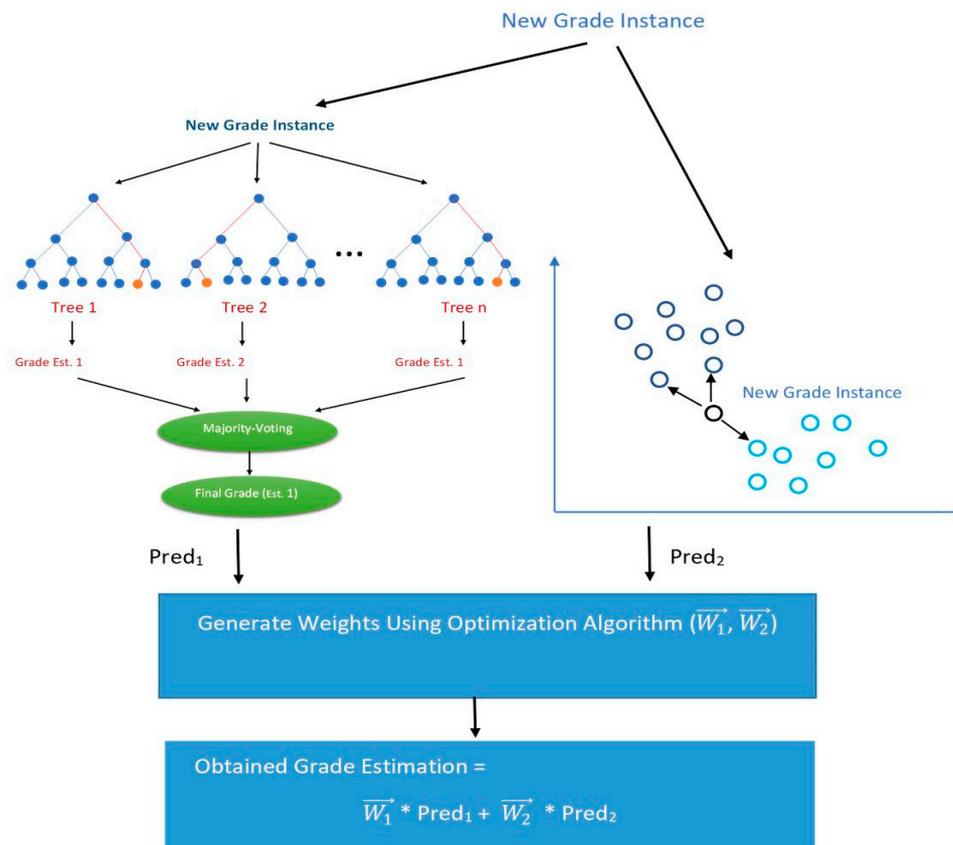
A hybrid ensemble network combines the results of several different models to make a single prediction. The subcomponents of the hybrid model work in parallel, but each component is self-sufficient in making predictions. This can improve system performance. The random forest and  $k$ -nearest neighbors hybrid algorithm with the marine predators optimization algorithm (MPA) is a strategy that improves the accuracy of predictions by combining the strengths of both RF and KNN as the MPA algorithm.

The method that combines random forest (RF) and  $k$ -nearest neighbors (KNN) with MPA is a three-stage procedure that optimizes data. Figure 1 depicts the algorithm's stages.

In the first stage, predictions are made using RF and KNN. The original RF and KNN algorithms are trained on the training data and then evaluated on the validation data. In the second stage, the parameters of the RF and KNN models are optimized using the MPA method. In this case, MPA aims to find the set of parameters that minimizes the mean squared error (MSE) between the predicted and actual ore grades. In the third stage, the final prediction is made by combining the predictions from the improved RF and KNN models. The predictions from the RF and KNN models are weighted according to their performance on the validation data. The final prediction is then computed as the sum of the weighted predictions.

Specifically, the RF regressor produces Pred 1, and the KNN regressor produces Pred 2. These two predictions are then optimized so that the strengths of both regressions can be exploited. An optimization algorithm is used to find the optimal weights for Pred 1 and Pred 2. The optimal weights are then used to compute the final prediction,  $P$ . The MPA optimizer employs a population of solutions with two parameters ( $W_1$  and  $W_2$ ) that are initialized randomly. Throughout the optimization operation's iterations, these generated solutions evolve towards the promising solution to reach a proper objective value; in our case, the objective function is the final prediction obtained with a minimized MSE error from the corresponding actual regression value. Compared to other algorithms, the hybrid RFKNN-MPA algorithm is superior in several respects, making it an effective tool for both machine learning and data analysis. It uses the Random Forest, KNN, and

Marine Predators Optimization Algorithm (MPA) to create more accurate predictions than individual models. In addition to this, it can deal with high-dimensional data, which is a typical obstacle in machine learning. It is also adaptable and can be applied to classification and regression issues.



**Figure 1.** Shows the various steps that make up the proposed approach.

#### 2.4. Efficiency Evaluation

Cross-validation is a popular ML model evaluation approach that divides data into multiple groups for training and validation. The k-fold cross-validation approach often splits the data into k equal parts and trains the model on k-1 parts while validating on the rest. Each part becomes the validation set once during this process.

Cross-validation helps to prevent overfitting and underfitting by accurately assessing the model's performance on unseen data. This research evaluates the regression model's performance using 5-fold cross-validation, which divides the data into 5 equal parts and repeats the training and validation process 5 times [51].

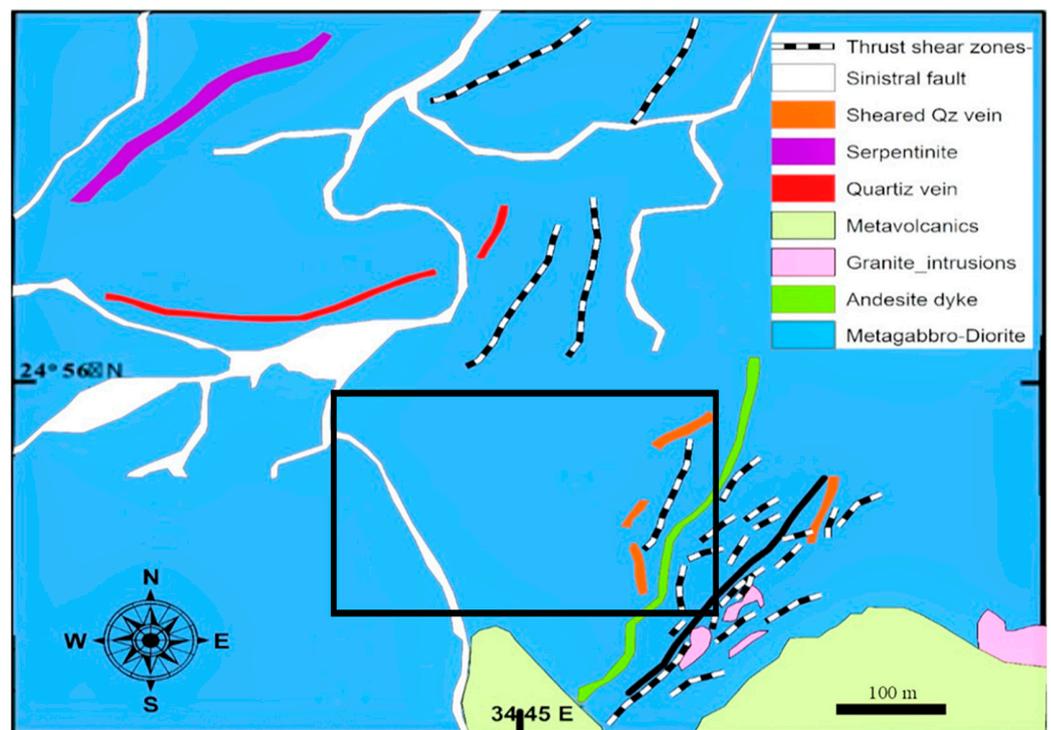
The approaches' performance was assessed using prediction errors, and four mathematical metrics were employed to control and compare them. The first metric, the correlation coefficient  $R$ , assesses the strength of two variables' association as they vary. The coefficient of determination ( $R^2$ ) measures model fitting accuracy by dividing real data variance by predicted value variation.  $R^2$  values between 55% and 75% are considered satisfactory, but  $R^2$  values below 30% are cause for concern. The mean squared error (MSE) takes bias and error variation into account but is more susceptible to outliers than the mean absolute error. Moreover, root mean square error (RMSE) is a measure of the average distance between the predicted values and the actual values. It is calculated by taking the square root of the mean of the squared residuals. The lower the RMSE, the better the model predicts the actual values. RMSE is often used to compare the performance of different models [52].

### 3. Results and Discussion

#### 3.1. Case Study Area for Grade Estimation

Quartz Ridge is a newly discovered field at 24°56′ North and 34°45′ East in the Central Eastern Desert (CED), just a short distance to the east of the Sukari gold mine (roughly 5 km). It is important to note that the Sukari gold mine has resources totaling more than 14.3 million ounces of gold, with an emphasis on the unexplored reserves in the neighborhood [53,54]. The Sukari concession contains several interesting and potentially lucrative options, but Quartz Ridge stands out among them. As a result of this, it garners more attention, which, in the end, leads to more substantial revenue and returns on investment.

The top domain is the southeastern domain, which contains the most investigated resources. This domain consists of a well-defined granodiorite intrusion intersected by dominant ENE steep fault zones and flat dipping thrusts. Extension veins are evident from the drilling as stacked extension veins with a flat angle and a thickness of up to 1 m. Ankerite makes up most of the surface alteration in the Southeastern domain, whereas a silica-chlorite assemblage dominates the area under the surface. The visual assessment and the mineralization patterns within this zone are shown in Figure 2 [55].



**Figure 2.** Depicts a geological map of the area of the Quartz Ridge Prospect. The territory that was investigated is denoted by the dark rectangle in the image at the middle.

#### 3.2. Exploratory Analysis of the Data

The exploration data used in this study come from unpublished exploration reports of gold prospecting by Centamin plc in the Quartz Ridge region of the eastern desert. It includes details on the lithology and geographic location of boreholes and Au assay values. This region is a valuable and virgin resource for exploration because it has never been mined. The information is based on 523 exploratory boreholes, with an average drilling interval of 20 m. In total, 27,505 samples from the boreholes were obtained as a result, which were not arranged in a regular grid. The samples, which come from various rock types, were taken at various intervals. Reverse circulation (RC) and diamond drilling (DD) drill holes, which are used in mineral prospecting for gathering samples from deep underground locations, were used to collect data over a number of years.

Overall, this exploration data offers essential knowledge about the regional geology and mineral potential of Quartz Ridge. Machine learning and statistical methods can be used to analyze these data to reveal information about the distribution of mineral grades and the likelihood of economic mineralization in the region.

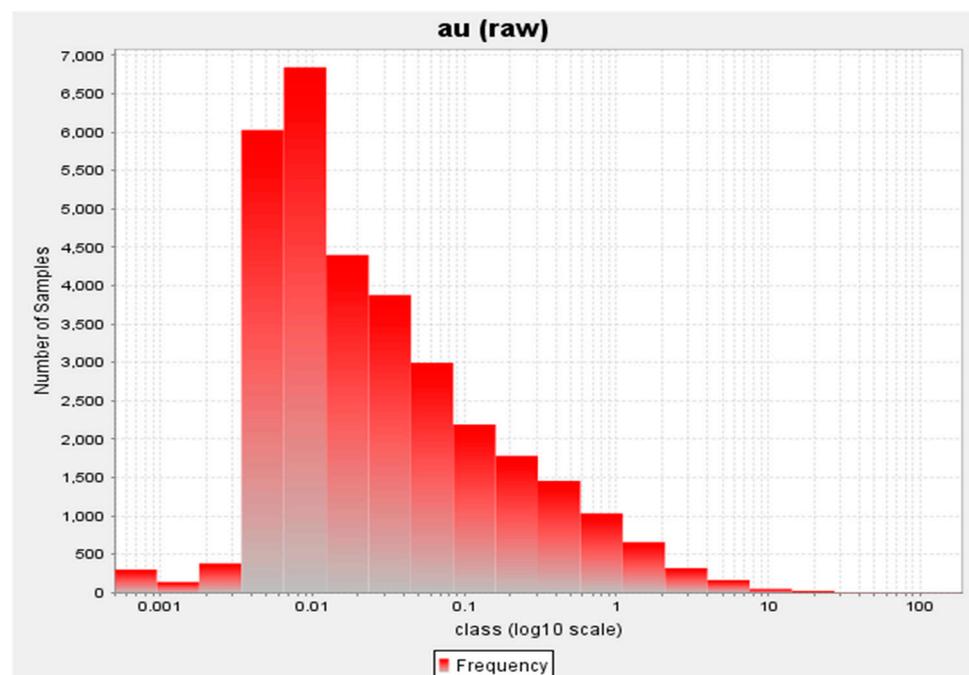
### 3.2.1. Data Analysis and Descriptive Statistics

The statistical analysis performed on the raw sample before regularization is summarized in Table 1 for convenience. The deposit mean and standard deviation were 0.22 and 1.72 ppm, respectively. The coefficient of was 7.74, indicating a relatively high degree of variability in the gold grades. These results indicate that the deposit has a complicated grade distribution with a wide range of values.

**Table 1.** Summarizes data on the gold concentration in the area under study.

N	Min	Max	Mean	Variance	S.D.	CoV	Skewness	Kurtosis
27505	0.005	187	0.22	2.95	1.72	7.74	64.94	6010.3

Figure 3 shows the raw samples' histogram plot. The histogram indicates that the drilling dataset contains a large number of low-quality values and a small number of high-quality values. Most of the information is clustered at the lower end of the scale, with only a few outliers at the other extreme, indicating that the data distribution is positively skewed.



**Figure 3.** Shows a logarithmic plot of the gold concentration in the deposit.

However, skewed distributions can make analysis and estimation challenging because the frequencies of the sample values are unbalanced. A regularization technique may be required to balance the data and improve prediction accuracy. The statistical analysis offers valuable information about the deposit's characteristics and the challenges of analyzing skewed data.

### 3.2.2. Data Regularization

To enhance the accuracy of grade estimates, the dataset was regularized by using equal-length samples in both geostatistical and machine-learning models. Regularization helps to reduce sampling bias, leading to more reliable grade estimates. The regularization process

was performed using SURPAC, and the data were regularized to a 1 m spacing. During the regularization process, only drill holes within the ore body were used to estimate the grade. This approach ensures that the data accurately represents the grade distribution of the deposit, thereby avoiding the risk of overestimating the amount of ore and underestimating the grade.

By excluding the waste zone, the composites used for grade estimation may underestimate the ore grade and volume. Therefore, this regularization technique has been thoroughly examined and applied to ensure the accuracy and reliability of the grade estimates.

### 3.2.3. Outlier Detection and Data Enhancement

The current study involves data processing, which includes identifying outliers based on geological domains and statistical populations. These anomalous values, known as outliers, can be identified by the fact that they deviate significantly from the average. The histogram plot technique enhances the data representation and modeling of the dataset. A top cut of ten ppm was implemented because there were outliers in the Au values; as a result, 29 outliers were found. Geostatistical analysis and using the dataset for machine learning training are made more accessible. The composite data were statistically analyzed, and the results showed significant in-depth variations (Table 2 presents the summary statistic). The regularized statistical data on gold grade reveals a mean of 0.19 ppm, with a variance of 0.48 ppm and a standard deviation of 0.69 ppm. The CoV is 3.72, which shows a moderate degree of variability in the gold grades [56,57].

**Table 2.** Summarizes compositing data on the gold concentration in the area under study.

N	Min	Max	Mean	Variance	S.D.	CoV	Skewness	Kurtosis
32980	0.005	10	0.18	0.43	0.69	3.72	8.9	99.87

### 3.2.4. The Lithological Analysis

The non-stationarity of gold grade in the deposit refers to the fact that the distribution of gold grades within the deposit is not constant or uniform. Instead, it varies geographically and is affected by many geological features, including lithology, structure, alteration, and mineralization type.

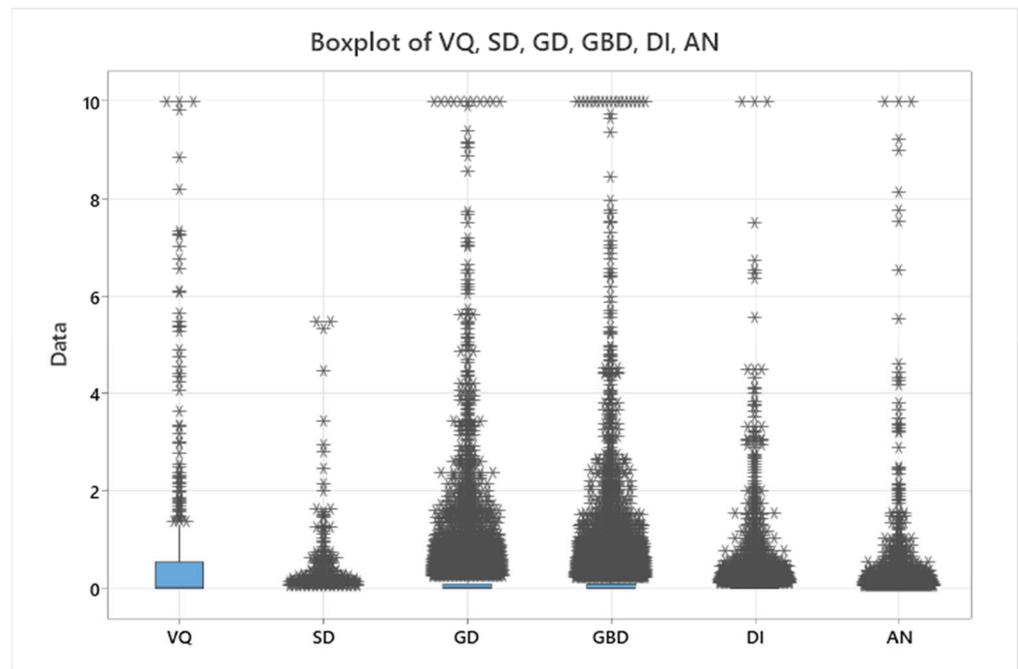
In geology, it is common practice to group similar lithologies to simplify the categorization and interpretation of geological data. Reducing the number of rock types to six (VQ: Quartz vein, SD: Sedimentary rocks, GD: Granodiorite, GBD: Gabbro Diorite, DI: Diorite, and AN: Andesite) created a more manageable set of categories that can be used to understand the geological distribution in a more general way and to help predict gold content.

The categorization of rocks based on their genetic category, structure, composition, and grain size is known as rock type [58].

The data in Table 3 and Figure 4 suggest that the area under investigation has a wide range of gold contents across the various rock types. This study identifies six distinct rock types, namely VQ, SD, GD, GBD, DI, and AN, corresponding sample sizes of 392, 1742, 10,733, 10,004, 4967, and 5142, respectively.

**Table 3.** Statistical summary of the data on gold content in various types of rocks.

Variable	N	Mean	SD	Variance	Min	Q1	Median	Q3	Max	Skewness	Kurtosis
VQ	392	0.76	1.75	3.06	0.005	0.01	0.03	0.55	10	3.26	11.2
SD	1742	0.08	0.34	0.12	0.005	0.005	0.01	0.03	5.5	10.80	143.4
GD	10,733	0.2	0.67	0.45	0.005	0.01	0.03	0.11	10	8.38	92.7
GBD	10,004	0.21	0.74	0.54	0.005	0.01	0.02	0.10	10	8.21	86.3
DI	4967	0.13	0.51	0.25	0.005	0.005	0.02	0.05	10	9.91	137.3
AN	5142	0.08	0.46	0.21	0.005	0.005	0.01	0.03	10	14.38	253.6



**Figure 4.** Boxplot of gold distribution through rock types.

The VQ rock type has the highest average gold content, with a mean of 0.76 and a maximum of 10. This highlights that VQ is a significant lithology for gold mineralization in the region. SD and AN have the lowest average gold concentrations, with mean values of 0.08 and 0.13, respectively, and maximum values of 5.5 and 10. GD, GBD, and DI rock types exhibit moderate mean gold concentrations, varying between 0.13 and 0.21.

The presented data indicate significant variability in the distribution of gold concentration across different rock types, with standard deviations from 0.34 to 1.75. This indicates that gold mineralization may be affected by various geological factors, including lithology, structure, and alteration. According to the skewness and kurtosis values for each kind of rock, the distribution of gold content is not normally distributed and might be affected by outliers or non-linear correlations.

The data presented in this study can be used to analyze patterns and trends in the distribution of gold within different rock types. This can provide a valuable overview of the gold concentration in the region. However, further investigation is required to fully understand the geological constraints on gold mineralization and give a complete picture of the distribution of gold in the area.

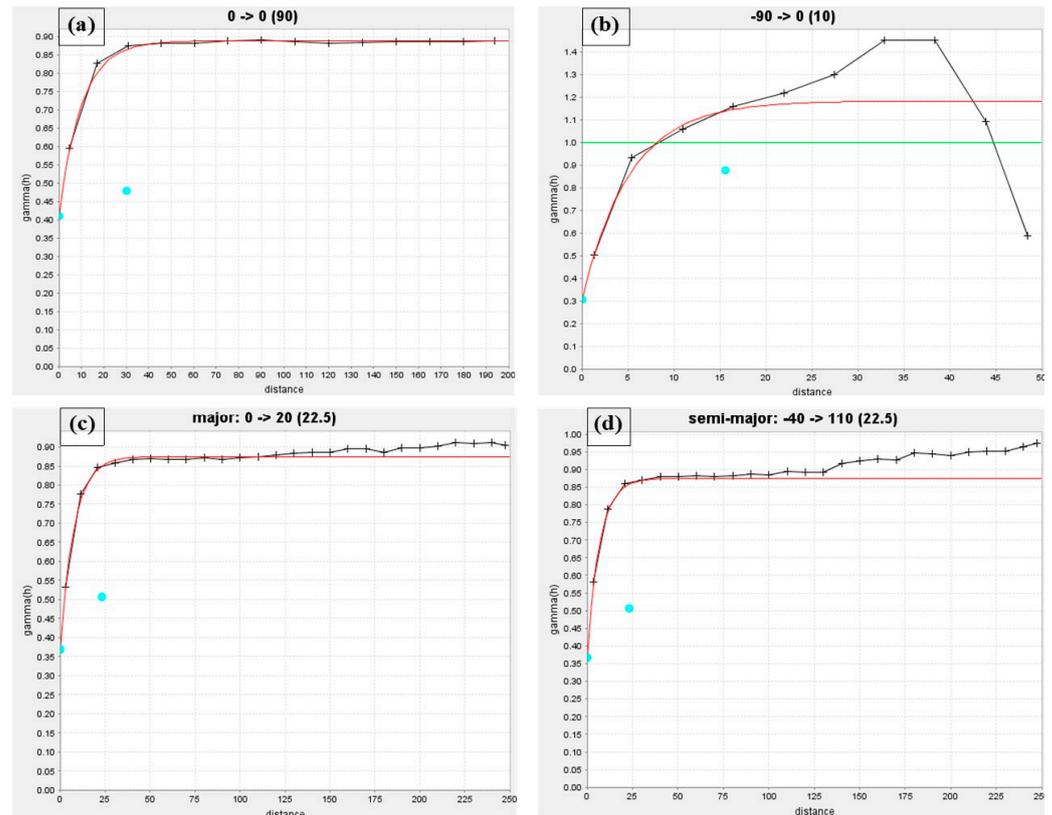
### 3.3. Variography

A variogram represents the degree of spatial dependence or correlation between two data points in a spatial domain. It measures how the variance of two points' values varies with distance. Geostatisticians use variograms to model spatial autocorrelation in datasets by plotting pairs of points versus their separation distances. The variogram curve's shape can reveal the dataset's spatial properties. A variogram that rises quickly and then levels off indicates a short-range spatial relationship, while one that grows slowly and steadily suggests a long-range one. The variogram range shows where data points have no spatial association.

Anisotropic spatial correlation in datasets can be caused by geometric and zonal anisotropy. Anisotropy can be accounted for by modeling a variogram in multiple directions or using a directional model with spatial structural orientation. Anisotropic variograms can estimate and replicate subsurface conditions, including faults and fractures, which affect mineral deposit distribution in the geostatistical modeling of subsurface geological formations. [59–61].

### 3.3.1. Analysis of Grade Variography

Considering geographical variation and randomness, the variogram function may depict the spatial variable structure of regional variables. Variography was conducted in multiple directions and dips to examine gold anisotropy. Most variograms showed various ranges and sills, indicating anisotropy in the gold deposit. An omnidirectional model is shown in Figure 5. The analytical model featured a nugget effect of 0.41, an exponential model with a sill value of 0.48, and a range of 29.2 m. According to the spherical model with a range of 29.2 m, the spatial correlation appears to decrease with increasing distance, but it may still be observed up to 29.2 m.



**Figure 5.** Variography for gold content three types of variograms: (a) omnidirectional, (b) downhole, and (c,d) directional variograms. The black line shows an experimental variogram and the red line represents the fitting model.

Table 4 displays an exponential variogram-based directional model. The relatively small nugget effect and large sill value in the directional model indicate a significant spatial dependency in the data. With a range of 23.4 m, this model’s spatial correlation appears to cover a relatively short distance, and the exponential variogram model indicates that the spatial correlation drops quickly with distance. Because most of the drilling composite assay values are of a relatively low grade, a normalized nugget of roughly 0.31 is not surprising.

**Table 4.** Characteristics of Variogram Models for Estimation.

Direction Model	Type	Nugget (ppm <sup>2</sup> )	Sill (ppm <sup>2</sup> )	Range (m)
Omnidirectional		0.41	0.48	29.2
Downhole	Exponential	0.31	0.88	15.6
Directional		0.37	0.51	23.4

The inspection of the variogram revealed the presence of anisotropy. The Directional model accounts for the anisotropy of the spatial correlation, which can potentially increase the accuracy of interpolation and prediction in regions with directional variability [62].

### 3.3.2. Indicator Variography for Spatial Variability Analysis and Modeling

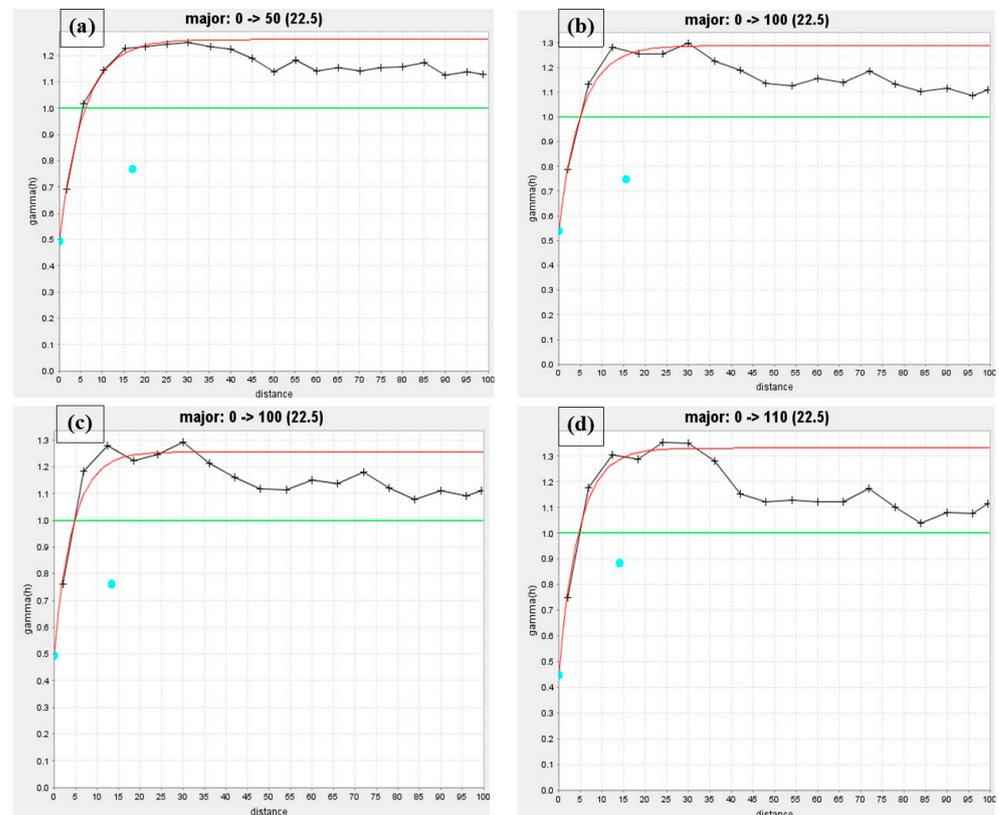
The study of directional indicator variograms considers the spatial correlation between indicator variables, which are binary variables that show the existence or absence of a particular characteristic in a specific area. In the presented scenario, directional indicator variograms were utilized to evaluate the geographical continuity of four possible cut-offs, namely 0.3, 0.6, 0.9, and 1.5 parts per million (ppm).

The lag spacing and the angular tolerance had to be adjusted so that there would be sufficient numbers of samples. Lag spacing is the distance between the data points used to make the variogram, and angular tolerance is the biggest angle between them. In the variogram analysis, an accurate representation of the spatial correlation of the data is essential, and these factors play a significant role in that.

Table 5 lists the parameters of the variogram models fitted to the experimental variograms, while Figure 6 displays experimental variograms and their corresponding fitted models. The spatial correlation and variability of the data can be better understood by examining the variograms and their associated parameters, allowing for more accurate predictions of gold values in areas that were not sampled.

**Table 5.** Cut-off indicator variogram parameters.

Cut-off	Variogram Model	Nugget Effect	Sill	Range	Azimuth	Dip
0.3	Exponential	0.49	0.77	17.1	50	−60
0.6	Exponential	0.54	0.75	15.67	100	−50
0.9	Exponential	0.49	0.76	13.4	100	−50
1.5	Exponential	0.45	0.88	14.14	110	−60



**Figure 6.** Indicator variograms at different cut-offs (a) 0.3 ppm, (b) 0.6 ppm, (c) 0.9 ppm, and (d) 1.5 ppm.

Using an exponential variogram model, Table 5 shows the directed variogram parameters for four different cut-off values, from 0.3 to 1.5. The nugget effect is a representation of the data's variation when there is no distance involved. In this scenario, the nugget effect drops marginally with increasing cut-off value, which may indicate that measurement or sampling error reduces data variability.

The value of sill represents the entire variation in the data. The relative consistency of the sill values across the various thresholds suggests that the spatial correlation of the data is robust.

The range is the distance where spatial correlation becomes minimal. Range values fluctuate across cut-off values, demonstrating that spatial data correlation varies with cut-off value.

Last but not least, the azimuth and dip values indicate, respectively, the direction and angle of the connection that is the strongest. Variation in the directional dependency of the data is also seen in the azimuth and dip values as a function of the cut-off value.

### 3.4. Block Model Creation and Validation for Resource Estimation

This section describes generating and validating block models for resource estimation in gold mineralization deposits. The block model was created by dividing the deposit into small, similar-sized blocks using empty block models within closed wireframe models of the mining bodies. To prevent overestimation, sub-blocking was employed. The SURPAC program incorporated constraints into the block model, such as specific gravity and deposit geology, to ensure an accurate estimate of the target tonnage and minimize potential inaccuracies.

The validity of the block model provided quantitative evidence for the accuracy of the estimating process. Because it results in the slightest standard error, a minor confidence interval, and the highest level of confidence or the least amount of risk, kriging was selected as the method for performing the estimation [63]. Ordinary kriging was utilized to estimate the deposit grade, as it has been found to perform effectively in previous models. Moreover, the spatial correlation of the binary data can be leveraged to estimate the likelihood of exceeding a particular grade cut-off at unsampled locations through indicator kriging. This information can help identify areas with a high potential for mineralization, which can help guide future exploration and mining efforts.

The block model's  $20 \times 20 \times 5$  m parent blocks encompassed the mineralization domain. Compositing, drilling spacing, and geological structure all factored into the decision-making process for selecting the properties of the parent blocks. Sub-blocking ensured volume representation was accurate for resource estimation [64].

All blocks within the deposit were assigned an average specific gravity of 2.67 to facilitate tonnage calculations based on the deposit's geology. The resource domain was used to generate ordinary and indicator kriging models, which were then utilized to estimate the grade of each block.

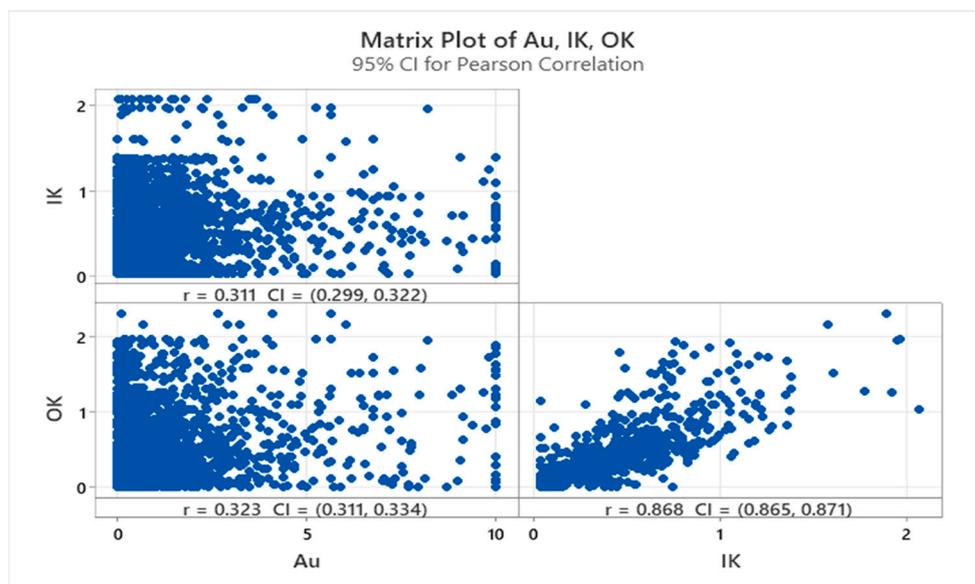
Building and validating a block model are crucial aspects of the resource estimation process since they provide a solid foundation for generating reliable mineralization estimates for subsequent exploration and mining efforts. Restrictions and sub-blocking variables were employed to prevent overestimation and ensure that the block model accurately represents the deposit's mineralization.

In this study, mathematical metrics and prediction errors were used to assess the performance of various approaches. These methods can provide valuable insights into the effectiveness and suitability of different models or techniques for estimating gold content.

Table 6 shows that the model's predictions are inaccurate for the kriging methods (OK and IK). A low R (Figure 7) means that the predicted and actual values do not have a strong linear relationship, and a low  $R^2$  means that the predicted values only explain a small amount of the variation in the actual values. A large MSE and RMSE suggest that the estimates are, on average, far from the observed values.

**Table 6.** Geostatistical (OK and IK) performance for predicting gold content.

Metric	R	R <sup>2</sup>	MSE	RMSE
OK	0.32	0.104	0.40	0.63
IK	0.31	0.096	0.43	0.65

**Figure 7.** Shows the correlation coefficients for OK and IK kriging methods.

Based on the study's findings, the gold content in the research region exhibits significant skewness and large coefficients of variation. This may explain why the models perform poorly in predicting the gold content, particularly compared to the results of variogram modeling. The variogram modeling revealed rising nugget/sill values for each model. However, the  $R^2$  values for all models used in the research were less than 0.5, indicating poor performance. This can be attributed to two main factors. Firstly, the poor spatial correlation between the gold dataset and the study region may have affected the accuracy of the models. Secondly, the skewed data distribution may have impacted the accuracy of the model's predictions.

The study anticipated poor results from using ordinary kriging to forecast gold content due to its tendency to produce locally linear estimates and smoothing effects. Kriging calculates each grid point by taking a weighted average of surrounding samples, which can hinder its ability to detect local differences in the data. To overcome these limitations, the study employed machine learning algorithms and hybrid methods to construct a gold deposit model. These algorithms can identify non-linear associations and patterns within the data, improving the accuracy and reliability of models that predict gold content.

### 3.5. Data Preparation and Normalization

The features of the data that are accessible, as well as the output that is intended, should serve as the basis for choosing an acceptable normalizing procedure. Since the choice of a normalizing approach may substantially influence the accuracy of the model's predictions, it is vital to examine the efficacy of various normalization strategies on the performance of the model. This evaluation should be performed as thoroughly as possible [65].

For this analysis, a composite dataset was used that had been processed to control outliers. Prior to training the ML model, the dataset was normalized, which helped to reduce noise and improve prediction accuracy. Normalization is particularly advantageous since it ensures that the statistical distribution of the data is uniform, allowing the model to predict values more accurately for each input and output [18,66].

Log normalization was chosen because it can make data less skewed and more regularly distributed by reducing the effect of extreme values. This could enhance the overall performance of some machine-learning models. This is consistent with Zaki's conclusion [8].

Log normalization's ability to stabilize the variance of the data is helpful for models that depend on the variance being constant. In addition, log normalization can make interpreting the model's results simpler, as the transformed values may be more intuitive or relevant to the specific application. Log normalization may not be appropriate for all data types or machine learning models. The impact of log normalization on model performance should be carefully assessed, and other normalizing methods should be considered.

#### Experiments with Random Training and Test Partitions

In many previous studies on predicting ore grades, the available data are typically randomized before being split into training and testing subsets. For this analysis, the dataset was divided into training (70%) and testing (30%) subsets for model evaluation. It is critical to emphasize that these two subsets should exhibit comparable statistical characteristics. Data segmentation and the MPA are optimization techniques for splitting a dataset into training and testing groups [10]. After partitioning the dataset into primary segments labeled low, medium, or high based on their gold values, they were further divided into sub-segments through a visual inspection of the histogram plot.

Data segmentation was employed to improve the accuracy of resource estimates by ensuring that both the training and testing subsets of the dataset were representative of the range of gold values in the deposit (see Table 7 for more details). The study aimed to identify the combination of input variables for the resource estimating model that produced the most accurate estimates using the MPA as an optimization method. The analysis, depicted in Table 7, describes the statistical analysis of the original dataset following splitting. The evaluation of how well the data partitioning approach utilized in the study worked requires that this comparison of statistics be performed. The resource estimate model may be inaccurate if the training and testing datasets have very varying gold contents. The study shows that the data partitioning approach worked and that the resource estimate model is likely accurate and dependable by showing that the training and testing datasets had similar gold contents.

**Table 7.** The statistical analysis of the original dataset after partitioning.

Variable	N	Mean	SD	Min	Q1	Median	Q3	Max
Test Data	9894	0.175	0.655	0.0005	0.01	0.02	0.07	10
Train Data	23,086	0.176	0.657	0.0005	0.01	0.02	0.07	10

Hyperparameters were tuned to obtain the optimal parameters for each ML algorithm utilized. The optimization algorithm searched for the best values for each method's parameters in the ML table (Table 8). It began by generating a set of starting parameters and then iteratively refined them until it reached the optimal parameter values for the prediction issue. Various hyperparameters, such as the number of hidden units, learning rate, training iterations, and network depth, must be considered. Although this step is crucial for optimal algorithm performance, it can be arduous and time-consuming [67].

Rerunning machine learning studies may lead to different findings. Therefore, it is crucial to record parameters, assess sensitivity, and employ statistical tools to evaluate significance and uncertainty. This is because the computer randomly selects the training and testing datasets. Bayesian optimization, a valuable technique for improving models, was employed in this study [68]. Additionally, a heuristic method was used to select the best network to avoid overfitting in terms of network performance.

**Table 8.** Outcomes from optimizing hyperparameters to improve gold predictions on the Quartz Ridge dataset.

Model	Hyperparameters
Typical configurations	Five k-folds CV Bayesian optimization Iterations: 30
RF	Number of trees = 200
K-NN	K = 13 Metric: Euclidean distance
GPR	Basis function: constant Isotropic Rational Quadratic kernel
DT	leaf size: 6 Minimum leaf size: 1-9041
FCN	Three layers used Iterations: 1000 Activation: ReLU
RFKNN-MPA	Number of trees = 100 K = 13

3.6. Comparative Analysis

To accurately predict the grade value of the gold resource in the Quartz Ridge region, several other model methods, including RF, K-NN, GPR, DT, and FCN, were tested against the RFKNN-MPA model developed using the same dataset and optimized features as input. To compare the results, the model’s performance was analyzed by applying the same indices (R, RMSE, and R<sup>2</sup>) to the training and testing datasets.

Table 9 summarizes the findings obtained from the comparison research for the training data. According to the results presented in Figure 8, the random forest (RF), RFKNN-MPA, and K-NN models exhibit significantly higher R values compared to the other models. This finding suggests a stronger linear relationship between the predicted and actual values. Among these models, the RFKNN-MPA model exhibits the highest R-value of 0.74, indicating the best linear correlation performance.

**Table 9.** Performance predicting gold content using the proposed method and machine learning models on the training dataset.

Metric	RF	K-NN	GPR	DT	FCN	RFKNN-MPA
R	0.69	0.66	0.63	0.50	0.39	0.74
R <sup>2</sup>	0.47	0.43	0.40	0.25	0.15	0.54
MSE	0.36	0.47	0.48	0.51	0.58	0.31
RMSE	0.67	0.69	0.69	0.74	0.78	0.59

The R<sup>2</sup> values for all models are relatively low, suggesting that only a small percentage of the variation in the actual values can be accounted for by the predicted values. However, the RFKNN-MPA model exhibits the highest R<sup>2</sup> value, indicating that it performs better than the other models in explaining the variation in the data. This model has an R<sup>2</sup> value of 0.54.

Based on the mean square error and root mean square error values, the RFKNN-MPA model exhibits the lowest prediction error compared to the other models. The mean square error (MSE) and root mean square error (RMSE) for this model are both 0.31 and 0.59, respectively, lower than the corresponding values for the other models. This finding provides evidence that the RFKNN-MPA model outperforms the other models in accurately forecasting the amount of gold present.

Table 10 presents the performance of the proposed approach and machine learning models for estimating gold content on the testing dataset. The results show that the RFKNN-MPA model exhibits the highest R-value (0.77), indicating a robust linear relationship between the predicted and actual values. Another model that performs well is the K-NN model, with an R-value of 0.71. Although their R-values are lower, the other models demonstrate a linear relationship between the predicted and observed values.

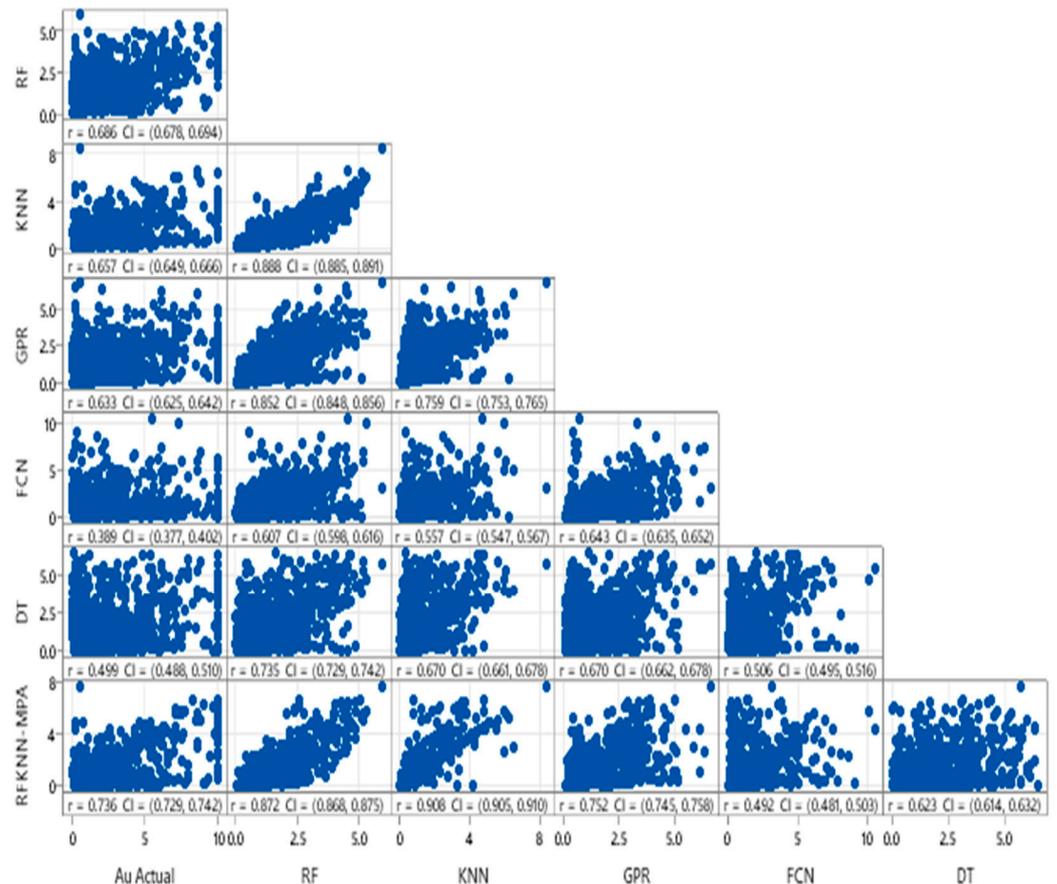


Figure 8. Scatterplots comparing actual gold content vs. predicted values for RF, KNN, GPR, FCN, DT, and the proposed method RFKNN-MPA for training data.

Table 10. Performance predicting gold content using the proposed method and machine learning models on the testing dataset.

Metric	RF	K-NN	GPR	DT	FCN	RFKNN-MPA
R	0.70	0.71	0.72	0.55	0.36	0.77
R <sup>2</sup>	0.49	0.497	0.52	0.29	0.13	0.597
MSE	0.24	0.26	0.25	0.30	0.37	0.17
RMSE	0.49	0.51	0.50	0.57	0.64	0.44

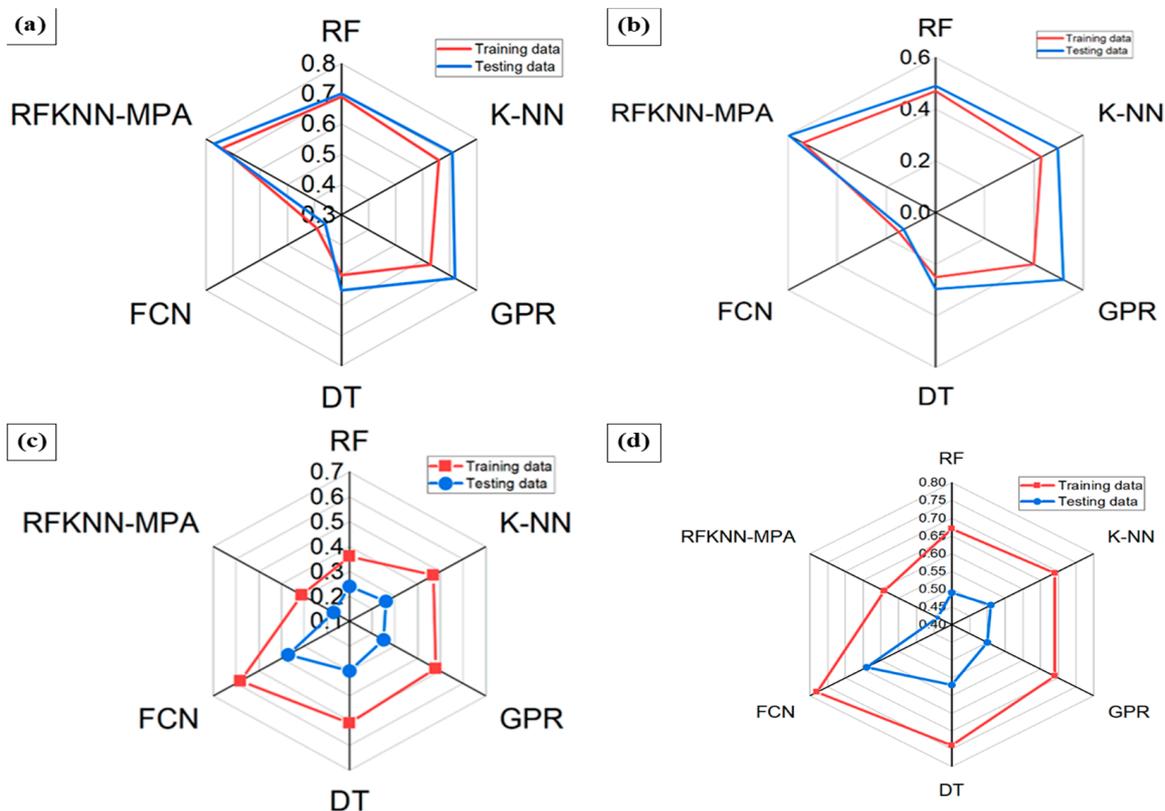
The models explain testing dataset variance better than the training dataset because their R<sup>2</sup> values are greater. Compared to the other models, the RFKNN-MPA model has the greatest R<sup>2</sup> value of 0.597, which indicates that this model is superior in explaining the variation seen in the data.

The RFKNN-MPA model exhibits the lowest prediction error (as measured by MSE and RMSE) compared to the other models. The mean squared error (MSE) and root mean squared error (RMSE) for this model are 0.17 and 0.44, respectively, lower than the corresponding values for the other models. This finding suggests that the RFKNN-

MPA model is more accurate than the other models in predicting the gold content of the testing dataset.

The RFKNN-MPA model performed best on the testing dataset for gold content prediction using the proposed technique. These results are encouraging and indicate that the model has excellent generalizability. However, it is required to perform further validation and testing on other datasets to conduct a comprehensive analysis of the model's performance.

Figure 9 depicts a radar map that compares the performance of six models, including RF, KNN, GPR, FCN, and DT, and the proposed RFKNN-MPA hybrid model. The study evaluated the accuracies of several models for predicting gold content using four metrics:  $R$ ,  $R^2$ , MSE, and RMSE. The radar plot displays different lines for training and testing data for each model, with testing data exhibiting better performance than training data. The results show that the RFKNN-MPA hybrid model outperformed the other models in all four metrics for training and testing data. As a result, the hybrid model is the most accurate and reliable model for estimating gold content in the study region.



**Figure 9.** Radar plots for (a)  $R$ , (b)  $R^2$ , (c) MSE, and (d) RMSE.

In this comparative study of models, the RFKNN-MPA model outperformed the other models (including both machine learning methods and geostatistical approaches) in terms of  $R$ ,  $R^2$ , MSE, and RMSE values, exhibiting higher  $R$  and  $R^2$  values than the other models. These results suggest that, compared to the other models, the RFKNN-MPA model was able to explain a more significant proportion of the variance in the dependent variable and produce more accurate predictions.

It is essential to bear in mind that the success of machine learning models can be influenced by a wide range of factors, including the quality and quantity of data, the features and hyperparameters used, and the specific application. Therefore, it is essential to conduct a thorough analysis of the performance of various models and select the one best suited for the particular task.

#### 4. Conclusions

The proposed hybrid model is an innovative approach for predicting mineral grade distribution that can capture nonlinearity and spatial heterogeneity. It has produced superior estimates compared to standalone machine learning or kriging models. The comparative analysis with various machine learning models and geostatistical methods, such as OK and IK, highlights the hybrid model's advantages in reducing errors resulting from incomplete or inconsistent drill hole assay data. The study focuses on the hybrid model's ability to enhance accuracy.

MPA optimization algorithms play a crucial role in the proposed hybrid model, reducing computational time and enhancing model efficiency and scalability. The MPA optimizer can improve the hybrid method's predictions by adjusting the integration weights, which is particularly beneficial when data are limited or uncertain.

Log normalization can improve the overall performance of ML models by reducing the impact of extreme values and reducing data skewness. It stabilizes data variance, which is beneficial for models that require a constant variance, and simplifies model interpretation with more intuitive converted values.

Incorporating factors such as rock types into the model, in addition to the coordinates, can enhance its accuracy by providing more information about the mineral deposit's properties and accounting for spatial variations that may impact the distribution of mineral grades. This is particularly beneficial for complex or heterogeneous mineral deposits.

The RFKNN-MPA algorithm is a robust and accessible method for estimating mineral orebody grades. It exhibits higher R-squared coefficients and lower RMSE and R values, resulting in superior estimations. These statistics indicate that it can lead to more confident decision making in mining operations. The proposed method preserves patterns within the data by considering geological characteristics (rock type) and chemical composition (gold grade).

Overall, the suggested hybrid model is a promising approach for predicting the distribution of mineral grades. This method can potentially improve decision making in mining operations and has broader implications in various industries. However, it is essential to continue evaluating and improving the method to ensure its accuracy and usefulness in various settings.

**Author Contributions:** Conceptualization, M.M.Z. and J.Z.; methodology, M.M.Z.; software, M.M.Z. and M.A.M.; validation, M.M.Z. and S.C.; analysis, M.M.Z. and S.C.; data curation, M.M.Z., F.F., and L.Q.; writing—original draft preparation, M.M.Z.; writing—review and editing, M.M.Z., S.C., J.Z., F.F., M.A.M., L.Q. and L.J.; visualization, M.M.Z.; supervision S.C.; funding acquisition S.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the National Natural Science Foundation of China (No. 52074169, 52004143, and 52174159).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** This data is confidential and owned by the company, and we do not have the right to publish it.

**Acknowledgments:** The authors wish to extend their appreciation to Centamin PLC for supplying the case study dataset that was instrumental in the success of this research project. Additionally, the authors are thankful to Mohamed Bedair and Mohamed Elgharib for their enthusiastic support and insightful feedback during the manuscript review process.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Ehteram, M.; Khozani, Z.S.; Soltani-Mohammadi, S.; Abbaszadeh, M. *The Necessity of Grade Estimation BT—Estimating Ore Grade Using Evolutionary Machine Learning Models*; Ehteram, M., Khozani, Z.S., Soltani-Mohammadi, S., Abbaszadeh, M., Eds.; Springer Nature: Singapore, 2023; pp. 1–6. ISBN 978-981-19-8106-7.
2. Isaaks, E.H.; Srivastava, R.M. *Applied Geostatistics*; Oxford University Press: New York, NY, USA, 1989; p. 561.
3. Rodriguez-Galiano, V.; Sanchez-Castillo, M.; Chica-Olmo, M.; Chica-Rivas, M. Machine Learning Predictive Models for Mineral Prospectivity: An Evaluation of Neural Networks, Random Forest, Regression Trees and Support Vector Machines. *Ore Geol. Rev.* **2015**, *71*, 804–818. [[CrossRef](#)]
4. Jang, H.; Topal, E. A Review of Soft Computing Technology Applications in Several Mining Problems. *Appl. Soft Comput.* **2014**, *22*, 638–651. [[CrossRef](#)]
5. Kanevski, M.; Pozdnukhov, A.; Timonin, V. *Machine Learning for Spatial Environmental Data. Theory, Applications and Software*; With CD-ROM; EPFL Press: New York, NY, USA, 2009. [[CrossRef](#)]
6. Asgari Nezhad, Y.; Moradzadeh, A.; Kamali, M.R. A New Approach to Evaluate Organic Geochemistry Parameters by Geostatistical Methods: A Case Study from Western Australia. *J. Pet. Sci. Eng.* **2018**, *169*, 813–824. [[CrossRef](#)]
7. Gilardi, N.; Bengio, S. Comparison of Four Machine Learning Algorithms for Spatial Data Analysis. In *Mapping Radioactivity in the Environment—Spatial Interpolation Comparison*; Office for Official Publications of the European Communities: Luxembourg, 2003; Volume 97, pp. 222–237.
8. Zaki, M.M.; Chen, S.; Zhang, J.; Feng, F.; Khoreshok, A.A.; Mahdy, M.A.; Salim, K.M. A Novel Approach for Resource Estimation of Highly Skewed Gold Using Machine Learning Algorithms. *Minerals* **2022**, *12*, 900. [[CrossRef](#)]
9. Samanta, B.; Bandopadhyay, S.; Ganguli, R.; Dutta, S. A Comparative Study of the Performance of Single Neural Network vs. Adaboost Algorithm Based Combination of Multiple Neural Networks for Mineral Resource Estimation. *J. S. Afr. Inst. Min. Metall.* **2005**, *105*, 237–246.
10. Dutta, S.; Bandopadhyay, S.; Ganguli, R.; Misra, D. Machine Learning Algorithms and Their Application to Ore Reserve Estimation of Sparse and Imprecise Data. *J. Intell. Learn. Syst. Appl.* **2010**, *2*, 86–96. [[CrossRef](#)]
11. Misra, D.; Samanta, B.; Dutta, S.; Bandopadhyay, S. Evaluation of Artificial Neural Networks and Kriging for the Prediction of Arsenic in Alaskan Bedrock-Derived Stream Sediments Using Gold Concentration Data. *Int. J. Min. Reclam. Environ.* **2007**, *21*, 282–294. [[CrossRef](#)]
12. Jafrasteh, B.; Fathianpour, N.; Suárez, A. Comparison of Machine Learning Methods for Copper Ore Grade Estimation. *Comput. Geosci.* **2018**, *22*, 1371–1388. [[CrossRef](#)]
13. Kapageridis, I.; Denby, B. *Ore Grade Estimation with Modular Neural Network Systems—A Case Study*; Panagiotou, G.N., Michalakopoulos, T.N., Eds.; AA Balkema Publishers: Rotterdam, The Netherlands, 1998; p. 52.
14. Afeni, T.B.; Lawal, A.I.; Adeyemi, R.A. Re-Examination of Itakpe Iron Ore Deposit for Reserve Estimation Using Geostatistics and Artificial Neural Network Techniques. *Arab. J. Geosci.* **2020**, *13*, 657. [[CrossRef](#)]
15. Zhang, X.; Song, S.; Li, J.; Wu, C. Robust LS-SVM Regression for Ore Grade Estimation in a Seafloor Hydrothermal Sulphide Deposit. *Acta Oceanol. Sin.* **2013**, *32*, 16–25. [[CrossRef](#)]
16. 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](#)]
17. Dutta, S.; Misra, D.; Ganguli, R.; Samanta, B.; Bandopadhyay, S. A Hybrid Ensemble Model of Kriging and Neural Network for Ore Grade Estimation. *Int. J. Min. Reclam. Environ.* **2006**, *20*, 33–45. [[CrossRef](#)]
18. Tahmasebi, P.; Hezarkhani, A. A Hybrid Neural Networks-Fuzzy Logic-Genetic Algorithm for Grade Estimation. *Comput. Geosci.* **2012**, *42*, 18–27. [[CrossRef](#)]
19. Jahangiri, M.; Ghavami Riabi, S.R.; Tokhmechi, B. Estimation of Geochemical Elements Using a Hybrid Neural Network-Gustafson-Kessel Algorithm. *J. Min. Environ.* **2018**, *9*, 499–511.
20. Hariharan, S.; Tirodkar, S.; Porwal, A.; Bhattacharya, A.; Joly, A. Random Forest-Based Prospectivity Modelling of Greenfield Terrains Using Sparse Deposit Data: An Example from the Tanami Region, Western Australia. *Nat. Resour. Res.* **2017**, *26*, 489–507. [[CrossRef](#)]
21. 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](#)]
22. Kaplan, U.E.; Topal, E. A New Ore Grade Estimation Using Combine Machine Learning Algorithms. *Minerals* **2020**, *10*, 847. [[CrossRef](#)]
23. Samson, M.J. Mineral Resource Estimates with Machine Learning and Geostatistics. Master’s Thesis, University of Alberta, Edmonton, AB, Canada, 2019; pp. 1–112. [[CrossRef](#)]
24. Li, X.; Li, L.H.; Zhang, B.L.; Guo, Q. jin Hybrid Self-Adaptive Learning Based Particle Swarm Optimization and Support Vector Regression Model for Grade Estimation. *Neurocomputing* **2013**, *118*, 179–190. [[CrossRef](#)]
25. Faramarzi, A.; Heidarinejad, M.; Mirjalili, S.; Gandomi, A.H. Marine Predators Algorithm: A Nature-Inspired Metaheuristic. *Expert Syst. Appl.* **2020**, *152*, 113377. [[CrossRef](#)]
26. Ben Seghier, M.E.A.; Golafshani, E.M.; Jafari-Asl, J.; Arashpour, M. Metaheuristic-based Machine Learning Modeling of the Compressive Strength of Concrete Containing Waste Glass. *Struct. Concr.* **2023**. [[CrossRef](#)]

27. Helmi, A.M.; Al-qaness, M.A.A.; Dahou, A.; Abd Elaziz, M. Human Activity Recognition Using Marine Predators Algorithm with Deep Learning. *Future Gener. Comput. Syst.* **2023**, *142*, 340–350. [CrossRef]
28. Yan, J.; Liu, H.; Yu, S.; Zong, X.; Shan, Y. Classification of Urban Green Space Types Using Machine Learning Optimized by Marine Predators Algorithm. *Sustainability* **2023**, *15*, 5634. [CrossRef]
29. McKinley, J.M.; Atkinson, P.M. A Special Issue on the Importance of Geostatistics in the Era of Data Science. *Math. Geosci.* **2020**, *52*, 311–315. [CrossRef]
30. Oliver, M.A.; Webster, R. A Tutorial Guide to Geostatistics: Computing and Modelling Variograms and Kriging. *Catena* **2014**, *113*, 56–69. [CrossRef]
31. Olea, R.A. *Geostatistics for Engineers and Earth Scientists*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; ISBN 1461550017.
32. Dai, L.; Wang, L.; Liang, T.; Zhang, Y.; Li, J.; Xiao, J.; Dong, L.; Zhang, H. Geostatistical Analyses and Co-Occurrence Correlations of Heavy Metals Distribution with Various Types of Land Use within a Watershed in Eastern Qinghai-Tibet Plateau, China. *Sci. Total Environ.* **2019**, *653*, 849–859. [CrossRef]
33. Wackernagel, H. *Multivariate Geostatistics: An Introduction with Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2003; ISBN 3540441425.
34. Sinclair, A.J.; Blackwell, G.H. *Applied Mineral Inventory Estimation*; Cambridge University Press: Cambridge, UK, 2009.
35. Hawkins, D.M.; Rivoirard, J. Introduction to Disjunctive Kriging and Nonlinear Geostatistics. *J. Am. Stat. Assoc.* **1996**, *91*(433), 337–340. [CrossRef]
36. Hohn, M. *Geostatistics and Petroleum Geology*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1998; ISBN 041275780X.
37. Englund, C.; Verikas, A. A Novel Approach to Estimate Proximity in a Random Forest: An Exploratory Study. *Expert Syst. Appl.* **2012**, *39*, 13046–13050. [CrossRef]
38. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
39. Calderoni, L.; Ferrara, M.; Franco, A.; Maio, D. Indoor Localization in a Hospital Environment Using Random Forest Classifiers. *Expert Syst. Appl.* **2015**, *42*, 125–134. [CrossRef]
40. Zabihi, M.; Pourghasemi, H.R.; Behzadfar, M. Groundwater Potential Mapping Using Shannon’s Entropy and Random Forest Models in the Bojnourd Township. *Iran. J. Ecolhydrol.* **2015**, *2*, 221–232.
41. Catani, F.; Lagomarsino, D.; Segoni, S.; Tofani, V. Landslide Susceptibility Estimation by Random Forests Technique: Sensitivity and Scaling Issues. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 2815–2831. [CrossRef]
42. Rokach, L. *Pattern Classification Using Ensemble Methods*; World Scientific: Singapore, 2009; Volume 75, ISBN 978-981-4271-06-6.
43. MacKay, D.J.C. Gaussian Processes—a Replacement for Supervised Neural Networks? *Semant. Sch.* **1997**. Lecture notes for a tutorial at NIPS.
44. Firat, M.; Gungor, M. Generalized Regression Neural Networks and Feed Forward Neural Networks for Prediction of Scour Depth around Bridge Piers. *Adv. Eng. Softw.* **2009**, *40*, 731–737. [CrossRef]
45. Fürnkranz, J. *Decision Tree BT—Encyclopedia of Machine Learning and Data Mining*; Sammut, C., Webb, G.I., Eds.; Springer: Boston, MA, USA, 2017; pp. 330–335. ISBN 978-1-4899-7687-1.
46. Rokach, L.; Maimon, O. *Decision Trees BT—Data Mining and Knowledge Discovery Handbook*; Maimon, O., Rokach, L., Eds.; Springer: Boston, MA, USA, 2005; pp. 165–192. ISBN 978-0-387-25465-4.
47. Sutton, C.D. Classification and Regression Trees, Bagging, and Boosting. *Handb. Stat.* **2005**, *24*, 303–329.
48. Wang, Y.-T.; Zhang, X.; Liu, X.-S. Machine Learning Approaches to Rock Fracture Mechanics Problems: Mode-I Fracture Toughness Determination. *Eng. Fract. Mech.* **2021**, *253*, 107890. [CrossRef]
49. Shrivastava, P.; Shukla, A.; Vepakomma, P.; Bhansali, N.; Verma, K. A Survey of Nature-Inspired Algorithms for Feature Selection to Identify Parkinson’s Disease. *Comput. Methods Programs Biomed.* **2017**, *139*, 171–179. [CrossRef]
50. Dutta, S. Predictive Performance of Machine Learning Algorithms for Ore Reserve Estimation in Sparse and Imprecise Data. Ph.D. Thesis, University of Alaska Fairbanks, Fairbanks, AK, USA, 2006; p. 189.
51. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 26.
52. Zorlu, K.; Gokceoglu, C.; Ocakoglu, F.; Nefeslioglu, H.A.; Acikalin, S. Prediction of Uniaxial Compressive Strength of Sandstones Using Petrography-Based Models. *Eng. Geol.* **2008**, *96*, 141–158. [CrossRef]
53. Kriewaldt, M.; Okasha, H.; Farghally, M. *Sukari Gold Mine: Opportunities and Challenges BT—The Geology of the Egyptian Nubian Shield*; Hamimi, Z., Arai, S., Fowler, A.-R., El-Bialy, M.Z., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 577–591. ISBN 978-3-030-49771-2.
54. Available online: <https://www.Centamin-Plc-Annual-Report-2022-Web-Ready-Secured.Pdf> (accessed on 5 April 2022).
55. Bedair, M.; Aref, J.; Bedair, M. *Automating Estimation Parameters: A Case Study Evaluating Preferred Paths for Optimisation*; International mining geology Conference: Perth, Australia, 2019.
56. Babakhani, M. Geostatistical Modeling in Presence of Extreme Values. Master’s Thesis, University of Alberta, Edmonton, AB, Canada, 2014.
57. Kim, S.M.; Choi, Y.; Park, H.D. New Outlier Top-Cut Method for Mineral Resource Estimation via 3D Hot Spot Analysis of Borehole Data. *Minerals* **2018**, *8*, 348. [CrossRef]

58. Agriculture United States Department of Agriculture. *Part 631 Geology National Engineering Handbook—Chapter 4 Engineering Classification of Rock Materials*; Agriculture United States Department of Agriculture: Washington, DC, USA, 2012.
59. Emery, X.; Maleki, M. Geostatistics in the Presence of Geological Boundaries: Application to Mineral Resources Modeling. *Ore Geol. Rev.* **2019**, *114*, 103124. [[CrossRef](#)]
60. Heuvelink, G.B.M.; Webster, R. Spatial Statistics and Soil Mapping: A Blossoming Partnership under Pressure. *Spat. Stat.* **2022**, *50*, 100639. [[CrossRef](#)]
61. Lark, R.M.; Minasny, B. Classical Soil Geostatistics. In *Pedometrics*; Springer: Cham, Switzerland, 2018; Chapter; pp. 291–340.
62. Van Groenigen, J.W. The Influence of Variogram Parameters on Optimal Sampling Schemes for Mapping by Kriging. *Geoderma* **2000**, *97*, 223–236. [[CrossRef](#)]
63. Vann, J.; Jackson, S.; Bertoli, O. Quantitative Kriging Neighbourhood Analysis for the Mining Geologist—a Description of the Method with Worked Case Examples. In *Proceedings of the 5th International Mining Geology Conference, Bendigo, Australia, 17–19 November 2003*; Australian Inst Mining & Metallurgy: Melbourne, Australia, 2003; Volume 8, pp. 215–223.
64. Tercan, A.E.; Sohrabian, B. Multivariate Geostatistical Simulation of Coal Quality Data by Independent Components. *Int. J. Coal Geol.* **2013**, *112*, 53–66. [[CrossRef](#)]
65. Singh, D.; Singh, B. Investigating the Impact of Data Normalization on Classification Performance. *Appl. Soft Comput.* **2020**, *97*, 105524. [[CrossRef](#)]
66. Beale, M.H.; Hagan, M.T.; Demuth, H.B. Neural Network Toolbox. *User's Guide MathWorks* **2010**, *2*, 77–81.
67. Yang, L.; Shami, A. On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice. *Neurocomputing* **2020**, *415*, 295–316. [[CrossRef](#)]
68. Greenhill, S.; Rana, S.; Gupta, S.; Vellanki, P.; Venkatesh, S. Bayesian Optimization for Adaptive Experimental Design: A Review. *IEEE Access* **2020**, *8*, 13937–13948. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.