ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to the following individuals and organizations who have been instrumental in the completion of this thesis. I am immensely thankful to my thesis supervisors, Professor Tsuyoshi Adachi and Professor Youhei Kawamura for their unwavering guidance, expertise, and patience throughout this research journey. Their mentorship and constructive feedback have been invaluable. My heartfelt thanks go to my family for their unconditional love and unwavering support. Their encouragement and belief in me have been my driving force.

I want to express a special and heartfelt thank you to my mother, Matlhodi Tsae. Her unwavering support, encouragement, and belief in me have been the cornerstone of my academic journey. Her endless patience and understanding during the challenging phases of this thesis provided me with the strength to persevere. Her love and support have been an invaluable source of inspiration. Thank you, Mom, for being my rock throughout this endeavor. I am immensely grateful to my entire family for their unwavering support, love, and understanding throughout this academic endeavor. Their constant encouragement, sacrifices, and belief in my abilities have been the foundation of my success. To my family, thank you for standing by me, cheering me on, and providing the strength and motivation I needed. Your support has been instrumental in reaching this milestone, and I am forever grateful for your presence in my life."

I am grateful to my friends for their encouragement, understanding, and occasional distractions that provided much-needed breaks during this intense period of research. I want

to acknowledge my lab members, whose discussions and insights enriched my understanding of the subject matter.

I extend my appreciation to MEXT scholarship for providing financial support for this research through their support made this project possible. I am thankful to Mr. Bradley Ackroyd for helping with data collection for this research. Your willingness to contribute to this study was crucial. The dedicated staff at Akita University and the Leading Program deserve recognition for their assistance in accessing resources and materials. I would like to acknowledge the academic advisors, professors, and mentors who have shaped my academic journey and provided valuable insights. This thesis would not have been possible without the collective support and encouragement of these individuals and organizations. Thank you for being a part of this journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Mineral ore grade estimation is a critical process in the mining industry, involving the assessment of the concentration or quality of valuable minerals within an ore deposit. Accurate grade estimation is essential for several reasons, including resource evaluation, mine planning, optimization of extraction methods, and economic feasibility studies. Here, we discuss mineral ore grade estimation, its limitations, and the potential of artificial neural networks (ANNs) and machine learning algorithms in improving grade estimation. Accurate grade estimation influences investment decisions and determines the economic viability of mining operations. It helps in optimizing resource extraction, reducing waste, and increasing resource recovery. Precise grade estimation contributes to minimizing the environmental impact of mining activities.

Conventional methods such as geometric and geostatistical are the most popular techniques in mineral resource estimation, but they fail to capture the complexity of the orebody. Ore deposits are often characterized by complex spatial variability, making it challenging to interpolate grades accurately. Data from drilling and sampling may be sparse and irregularly distributed, leading to uncertainty in grade estimation. Additionally, human interpretation and judgment in conventional methods can introduce subjectivity and errors. Due to these limitations, the grades are incorrectly estimated leading to inaccurate mine plans and costly financial decisions.

The advancements in technology have shown an immense potential of machine learning (ML) algorithms over other interpolation techniques for ore grade estimation because

artificial neural networks (ANNs) and machine learning algorithms can handle large and complex datasets, capturing nonlinear relationships and patterns that may be missed by conventional methods. These algorithms can automatically select relevant features, improving the accuracy of grade estimation by focusing on the most influential factors. ANNs and machine learning algorithms can achieve high levels of accuracy, reducing the risk of costly errors in mining operations. Furthermore, machine learning models can be continuously updated with new data, allowing for real-time grade estimation and adjustment of mining plans. ANNs and machine learning algorithms offer a promising alternative by leveraging their data-driven, spatial modeling, and feature selection capabilities, ultimately improving the precision and reliability of grade estimation in the mining industry.

In this study, we propose ore grade prediction using artificial neural network for copper and gold grades. Fourteen thousand, two hundred and ninety-four (14294) datasets were collected from Jaguar mine in Western Australia. First, the proposed model was developed by incorporating lithology, alteration, eastings, northings, altitude, dip, and azimuth to predict the copper ore grade. The performance evaluation metrics were measured based on mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE) and coefficient of determination (R^2) and ANN model outperformed the classic machine learning methods s 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 conducted to evaluate the feature importance of the input parameters on copper ore grade prediction. Lithology demonstrated the highest influence on ore prediction while the eastings showed the less impact on the output. This proposed approach is promising in ore model predictions.

The study then shifts the attention to the prediction of gold ore grade. We present a multilayer feed-forward artificial neural network that incorporates correlation coefficient feature selection methods for estimating gold ore grade. The study aims to identify the critical features or variables that can enhance the accuracy of predicting gold ore grade using artificial neural networks (ANNs). By pinpointing these important features, the research seeks to reduce prediction errors and improve the overall performance of the model. The feature selection method yielded six significant input features, namely lithology, alteration, dip, azimuth, and the coordinates X and Z, and these essential features were used to train the ANN model for the prediction of gold ore grade.

A comparative analysis was conducted to evaluate the performance of predictive models with and without feature selection. The performance of the ANN model was assessed using various metrics, including the correlation coefficient, mean absolute error, mean square error, and root mean square error. The findings indicate that the ANN model with feature selection exhibited superior performance compared to the model without feature selection with R, MAE, MSE, and RMSE, of 0.720 ,0.264,0.374 and 0.653 respectively. In addition to improving model accuracy, the correlated features model had lower computing efficiency. This comparison provides valuable insights into the actual benefits of using feature selection techniques in tasks related to predicting mineral grades.

The study's final comparison of the use of ANN for gold and copper grade prediction revealed that the suggested model accurately predicts copper grades. The predicted outcomes revealed higher accuracy and fewer errors for copper ore grade predictions. The proposed model is

considered innovative because it integrates various types of data that were not commonly combined in existing literature. This innovative approach has the potential to advance techniques for mining exploration and ore grade estimation, suggesting that it may contribute to improving the efficiency and accuracy of mining operations. Although the ANN model moderately predicted the mineral ore grade, it did not consider the geological structure of the orebody, faults, and discontinuities. Future research can explore alternative feature selection techniques to enhance the accuracy and precision of the model and to ensure comprehensive model analysis and these results can be compared to existing methods in literature.

CHAPTER 1: INTRODUCTION

1.1 Motivation and Problem definition for Grade Estimation

Mineral grade estimation is a fundamental process in the mining industry that involves determining the quality or concentration of valuable minerals in an ore deposit. It plays a critical role in decision-making processes related to mining operations, resource evaluation, and economic feasibility studies. Accurate estimation of mineral grades is essential for several reasons. First and foremost, it helps in assessing the economic viability of a mining project by determining the potential value of the mineral resources. It also guides mining engineers and geologists in planning and optimizing the extraction process, as higher-grade zones are typically prioritized for mining to maximize profitability. Additionally, grade estimation aids in resource classification, mine design, and mineral reserve estimation, which are vital for regulatory compliance and reporting purposes.

The process of mineral grade estimation involves the integration of geological data, sampling techniques, laboratory analysis, statistical methods, and computer modeling. Grade estimation is one of the most complicated aspects of mining. It is described as a prediction problem in statistical terms and the complexity of grade estimation stems from geological uncertainty and human judgement error. The most popular techniques used in mineral resource estimation are the traditional methods, namely geometric and geostatistical methods. Kriging is a well-known estimation technique in the mining industry and has gained enormous recognition as the accurate estimator for mineral resources over the years. Kriging is an ideal spatial regression technique mainly designed for the regional or local estimation of block grades as a linear combination of available data and the estimation error is kept at a

minimum (Akbar., 2012). There are various kriging techniques applied in mineral resource estimation such as simple kriging (SK), indicator kriging (IK), and Ordinary kriging (OK). Ordinary kriging, also known as BLUE (Best Linear Unbiased Estimator), is the most widely used estimation technique in mineral resource estimation (Isaaks and Srivastava, 1989). This technique can be used to estimate a value at an unsampled location in a region of interest using data in the region and a variogram model interpreted from all the data within the region, additionally kriging can also be used to estimate block grades (Badel et al.,2011). Doing so minimizes the expected error between the estimated grade and the actual grade. Even though the supremacy and efficiency of these methods have been exemplified in several studies (Wackernagel.,1998; Cressie., 1989; Paithankar et al.,2018), the major limitation of these conventional techniques is that they require assumptions to be made based on the spatial correlation between samples to be estimated at unsampled location (Yamamoto.,2005; Abuntori et al.,2021; Sadeghi et al.,2021).

The spatial distribution of kriging estimates tends to be smooth and overestimates low-grade values and underestimates high-grade values. Deutsch and Journel, 1998 introduced sequential gaussian simulation (SGS) as a solution to the smoothing problem of kriging. Pan et al., (1992) concluded that the conventional approaches may not give the best estimates of the grade because of the complex relationship between the spatial pattern variability and grade distribution. However, the difficulty in estimating grade for ore deposits with few data points through geometrical and geostatistical methods have paved way for the application of artificial intelligence in grade estimation.

Over the past decades, researchers (Yama and Lineberry., 1999; Wu and Zhou.,1993; Jafrasteh and Fathianpour.,2017; Jafrasteh et al.,2018; Li et al.,2013] have applied neural networks in ore grade prediction. The advancements in technology have shown an immense potential of machine learning (ML) algorithms over other interpolation techniques for ore grade estimation because of their ability to learn any linear or non-linear relationship between inputs and outputs. The use of neural network method is very appealing and has become one of the versatile techniques in grade prediction. Also, machine learning-based resource estimation techniques are efficient and cheaper than traditional resource estimation approaches (Zhang et al.,2021). Moreover, ML contributes to the understanding of the diverse types of ore deposits by modernizing the hypothesis testing and geological modeling (Zhang et al.,2021). Machine learning techniques address different operational challenges in the mining industry, mineral exploration, drilling and blasting as well as mineral processing.

Olmos-De-Aguilera et.al., (2023) 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 term planning and short-term planning. They used feedforward neural network (FNN), one-dimensional (1D) convolutional, neural network (CNN), and long short-term memory (LSTM) models were analyzed. Matias et al., (2004) examined the precision of kriging, regularization networks (RN), multilayer perceptron (MLP), and radial basis function (RBF) networks when determining the slate quality. 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 (Koike et al., 2002).

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., (2009) 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., (2016) 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., (1998) 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., (2020) 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.

1.2 Purpose of the Study

The purpose of this study is to present an innovative ore grade prediction approach based on ANN model. This approach aims to improve the accuracy of predicting metal grades (copper and gold) in ore deposits by incorporating a range of input variables, including spatial information (eastings, northings, and altitude), drilling parameters (dip and azimuth), and geological information (lithology and alteration). By including these diverse sets of factors, the model can capture a broader range of variables that may affect copper and gold grades in ore deposits. One of the strengths of ANN models is their ability to learn complex, nonlinear relationships between input variables and output variables. In this case, the ANN is expected to identify intricate connections between the geological, drilling, and sample location data and the metal grades of copper and gold. This is valuable because ore deposit formation is a multifaceted process with non-obvious relationships. Additionally, incorporating a wide array of input variables and leveraging the power of ANN, the proposed approach has the potential to provide more accurate predictions of copper and gold grades. This can be immensely valuable in the mining industry, where precise ore grade estimations are essential for resource planning and extraction efficiency.

This study is not only focused on prediction but also contributes to a deeper understanding of ore deposits. By exploring the intricate relationships between geological features, drilling data, and metal grades, this approach may shed light on the complexities and variations in ore deposit types. This insight can be useful for geologists and mining professionals in characterizing ore bodies more comprehensively.

This work also focusses on feature selection techniques which are essential in machine learning and data analysis to identify the most relevant features or variables that contribute to the predictive power of the model on gold ore grade. These techniques aim to improve the model performance, reduce overfitting, enhance interpretability, and reduce computational complexity, also it is important to note that the choice of feature selection technique depends on the data characteristics and the chosen machine learning algorithm. Additionally, the study compares the performance of models with and without feature selection and this provides valuable insights into the actual benefits of these techniques in mineral grade prediction tasks. The proposed model is innovative because it combines various types of data which were not common in the existing literature, therefore this study can potentially lead to advancements in mining exploration and estimation techniques. However, it is essential to ensure rigorous data collection and model validation processes to achieve reliable results and make a meaningful impact in the field of mining and geology.

1.3 Relevance of the study

The main contribution of this study is to launch an innovative approach for predicting ore grade, which is crucial in mining and resource exploration. This approach incorporates various input variables including geological attributes, spatial location and drilling data to enhance the accuracy of ore grade predictions. This study employs ANN as the primary tool for ore grade prediction. ANNs are a type of machine learning model inspired by the human brain's neural structure and are known for their ability to handle complex patterns and relationships in data. This study also compared the efficacy of ANNs with five classic machine learning techniques and ANN outperformed these methods.

Researchers have combined optimization algorithms, the generic algorithm, k-means clustering, and Levenberg-Marquardt, and the combination of kNN and ANNs adopted for grade prediction over the years. Among various optimization algorithms available for training ANNs, we chose the Bayesian regularization algorithm due to its precision and performance. The proposed technique is not limited to a specific mineral resource. It can be applied to assess grades for a wide range of mineral resources, making it a versatile tool in the field of mining and resource exploration. The study acknowledges some limitations of the technique. It does not account for geological discontinuities, faults, and joints in mineral estimation. This means that it may not capture certain geological complexities that can affect ore grade distribution. Additionally, since ANN performance relies on data, having a sufficient amount of data is essential for accurate predictions.

Accurate grade estimation is of paramount importance in the mining industry for several critical reasons and these are illustrated in Figure 1.

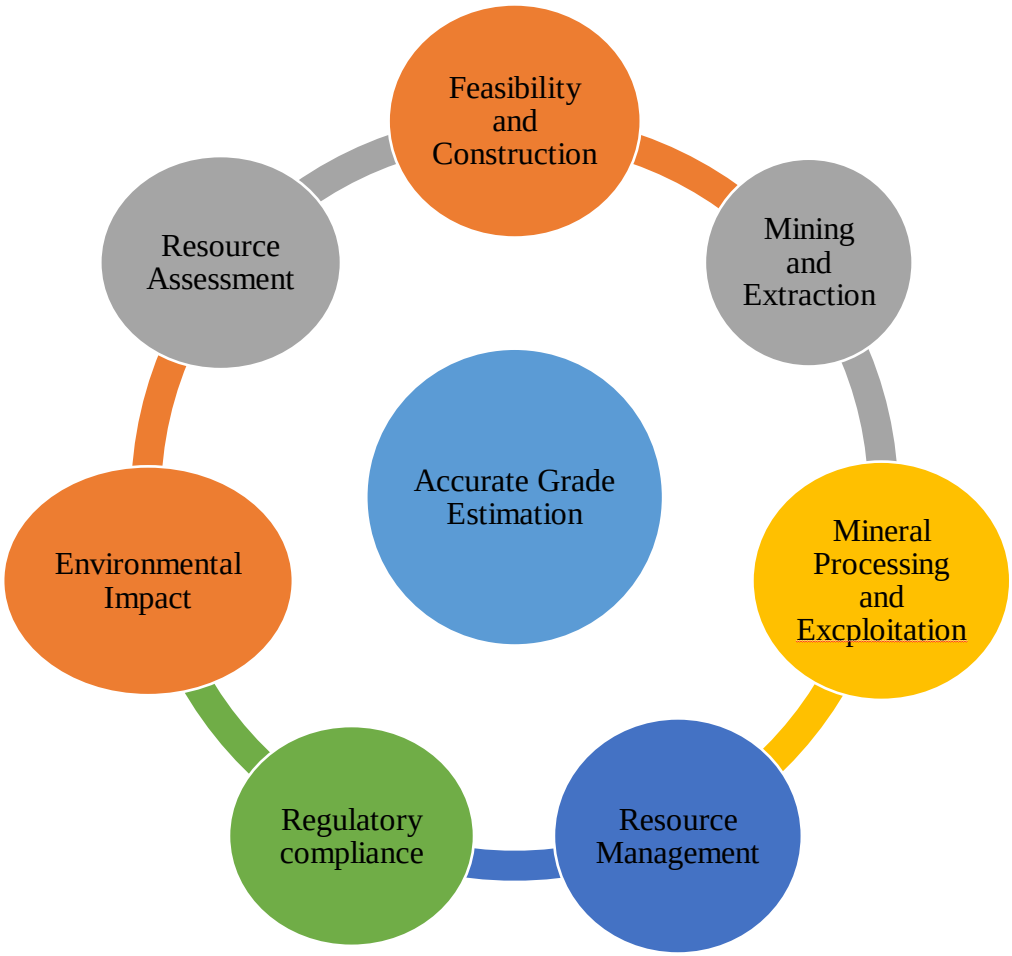


Figure 1: Significance of accurate grade estimation in the mining industry

This study is not only critical for the economic success of mining operations but also for environmental sustainability, safety, and regulatory compliance. Accurate grade estimation serves as the foundation upon which mining companies make decisions about resource

development, processing methods, and market positioning, ultimately influencing the industry's overall sustainability and profitability.

This study may shed some light in different areas of the mine. It may be essential in determining the quantity and quality of mineral resources in a deposit. This information is fundamental for making informed decisions about whether to develop a mining project, as it directly impacts the economic viability and potential profitability of the venture. Mining operations are costly and resource intensive thus knowing the precise grade of ore allows mining companies to assess whether a particular deposit can be mined profitably. High-grade ore can be extracted more economically than low-grade ore, and accurate estimation helps optimize the allocation of resources.

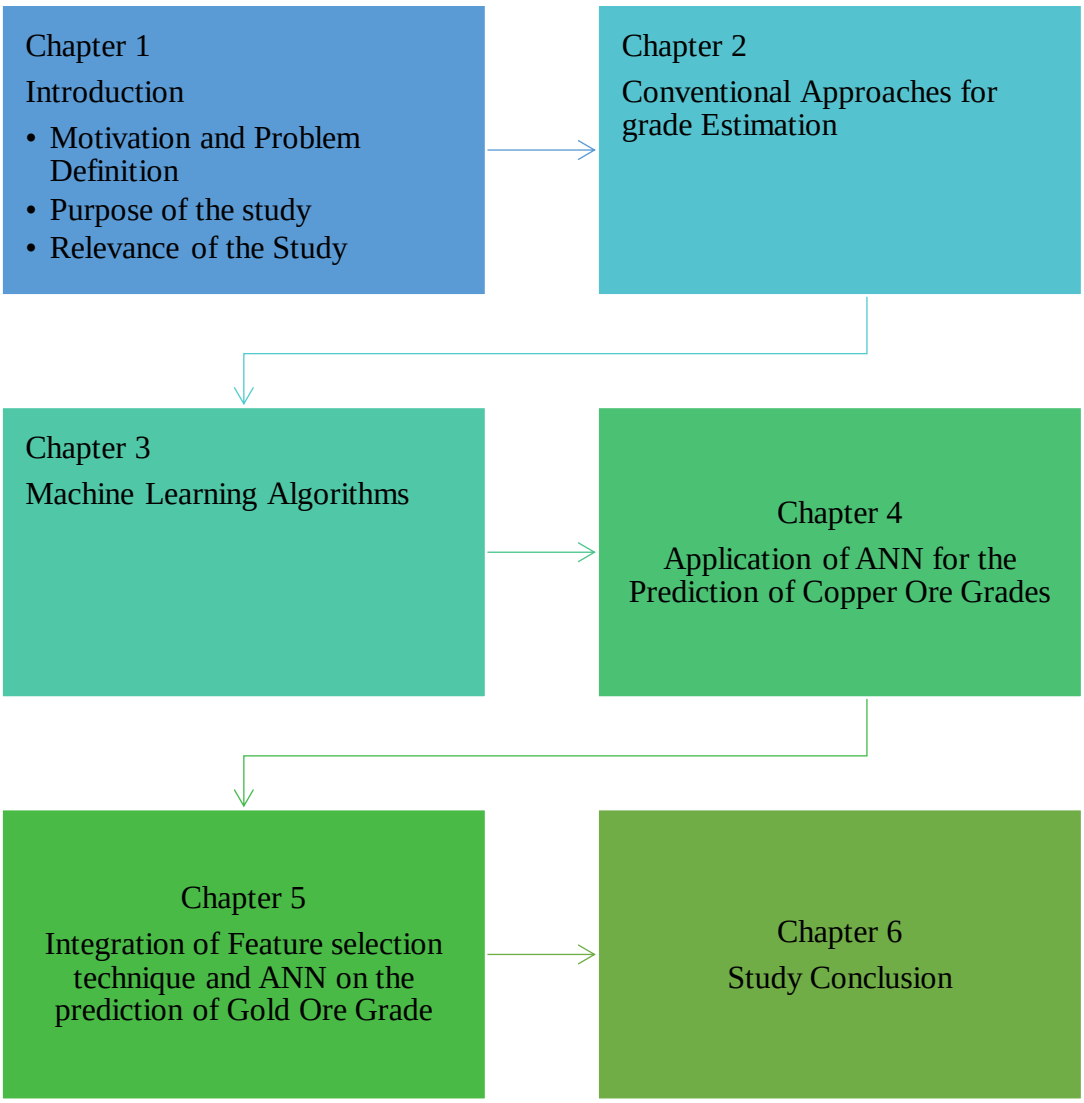
Ore grades directly affect the efficiency of mineral processing operations. Processing low-grade ore requires more energy and resources, which can increase operational costs, therefore grade estimation enables companies to optimize processing techniques, reduce waste, and improve overall efficiency. The precise prediction of grade aids in efficient resource management. By knowing the grade distribution within a deposit, mining companies can plan their extraction strategies, minimize waste, and maximize ore recovery. This reduces environmental impact and enhances sustainability. Since mining operations can be hazardous, and safety is a top priority, accurate grade estimation helps in better mine planning and design, reducing the risk of accidents and ensuring the safety of workers. Furthermore, the quality of ore determines its marketability and price. High-grade ores are often in higher demand and command better prices in the market, therefore adopting the proposed technique may allow

mining companies to market their products effectively and negotiate favorable contracts with buyers.

Mining operations are subject to regulations and environmental standards so precise grade prediction helps in complying with these regulations by ensuring that only the permitted quantities of ore are extracted and processed. Moreover, it is crucial for attracting investors. Potential investors and stakeholders require confidence in the resource estimates to assess the financial viability and risk associated with a mining project. Inaccurate estimations can erode investor confidence and deter funding.

Mining is a long-term endeavor that requires careful planning and investment thus correct grade estimation provides the foundation for long-term mine planning, helping mining companies make strategic decisions about the lifespan of a mine, expansion, and future exploration. Low-grade ore extraction can have a higher environmental impact due to increased waste generation and energy consumption. Accurate prediction of ore grade supports environmentally responsible mining practices by minimizing waste and resource usage. The findings of this study are promising and have the potential to assist mining corporations to achieve their set goals.

1.4 Outline of the Study



Chapter 2 summarizes the existing/conventional methods used in mineral estimation.

Chapter 3 entails the explanation of how machine learning algorithms work. It starts with a general overview of machine learning and then continues with a more detailed take on the

core parts of machine learning. It discusses evaluation metrics, feature selection techniques and Shapley values. It covers how artificial neural networks operate the structure of artificial neural networks and explains why Artificial neural networks are expected to perform better.

Chapter 4 contains the application of ANN on the prediction of the copper ore grades. It examines the data and methods used. It contains the geology of the study area and data collection and pre-processing. The first section shows the results for application of artificial neural network on copper ore grades, hyperparameter tuning, performance analysis of the proposed ANN models in comparison with classical machine learning algorithms and the feature importance analysis by Shapley values.

Chapter 5 This chapter demonstrates the application of artificial neural networks by considering feature selection techniques for estimating gold ore grades. The final section discusses the comparative analysis of gold and copper ore grade estimations using Artificial Neural Network.

Chapter 6 concludes the key points of the feasibility of Artificial Neural Network in the prediction of metal grades.

REFERENCES

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.

Isaaks, H.; Srivastava, R.M. Ordinary Kriging. In Applied Geostatistics; Oxford University Press Inc.: New York, NY, USA, 1989; pp. 288–295.

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.

Wackernagel, H. Multivariate Geostatistics: An Introduction with Applications; Springer: Berlin, Germany, 1998; p. 256.

Cressie, N. Spatial prediction and ordinary kriging. J. Int. Assoc. Math. Geol. 1989, 21, 493–494.

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.

Yamamoto, J.K. Correcting the Smoothing Effect of Ordinary Kriging Estimates. J. Int. Assoc. Math. Geol. 2005, 37, 69–94.

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.

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.

Deutsch, C.V.; Journel, A.G. GSLIB: Geostatistical Software Library and User's Guide; Oxford University Press: New York, NY, USA, 1998.

Pan, G.; Harris, D.P.; Heiner, T. Fundamental issues in quantitative estimation of mineral resources. Nat. Resour. Res. 1992, 1, 281–292.

Yama, B.; Lineberry, G. Artificial neural network application for a predictive task in mining. Min. Eng. 1999, 51, 59–64.

Wu, X.; Zhou, Y. Reserve estimation using neural network techniques. Comput. Geosci. 1993, 19, 567–575.

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.

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

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.

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.

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.

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.

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.

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.

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.

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.

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

CHAPTER 2: CONVENTIONAL APPROACHES FOR ORE GRADE ESTIMATION

The conventional methods such as geostatistical methods play a crucial role in ore grade estimation in the mining industry. These methods utilize the principles of spatial statistics to analyze and model the spatial variability of ore grades within a deposit. By understanding this variability, geologists and mining engineers can make more informed decisions regarding resource estimation, mine planning, and production optimization. Numerous researchers (Journel and Huijebregts.,1978; Rendu.,1979; Goovaets.,1997, Chiles) have proposed diverse resource estimation strategies based on models such as stochastic simulations, kriging, and inverse distance weighting (IDW). However, in order for these methods to be properly implemented, these geostatistical techniques necessitate certain knowledge, expertise, and mathematical skills (Badel.,2011, Tamasebi.,2011). Each approach has unique qualities that make it more suitable for specific types of deposits.

Here are some key concepts and methods used in geostatistical ore grade estimation:

Stationarity: Geostatistical methods assume that the spatial properties of the ore deposit remain relatively consistent within certain domains or zones. This assumption is known as stationarity, and it implies that the statistical properties of the data, such as mean and variance, do not change significantly across the deposit.

2.1 Kriging

Kriging: Kriging is a spatial interpolation technique that predicts values at unsampled locations based on the values at neighboring sampled locations. It assumes that the spatial correlation between samples decreases with increasing distance, allowing for the estimation

of grades at untested locations. Kriging considers both the spatial structure of the variable of interest and the spatial arrangement of the samples, providing a smooth estimate that honors the observed data while capturing the underlying trends. The kriging process involves several steps:

1. **Variogram modeling:** The first step is to create a variogram which is a fundamental tool in geostatistics that measures the spatial variability between pairs of sample points. It quantifies the spatial correlation and provides information about the range, sill, and nugget effect. The variogram is computed by calculating the variance of the differences in grade values at various lag distances between pairs of points. Figure 2 shows the variogram.

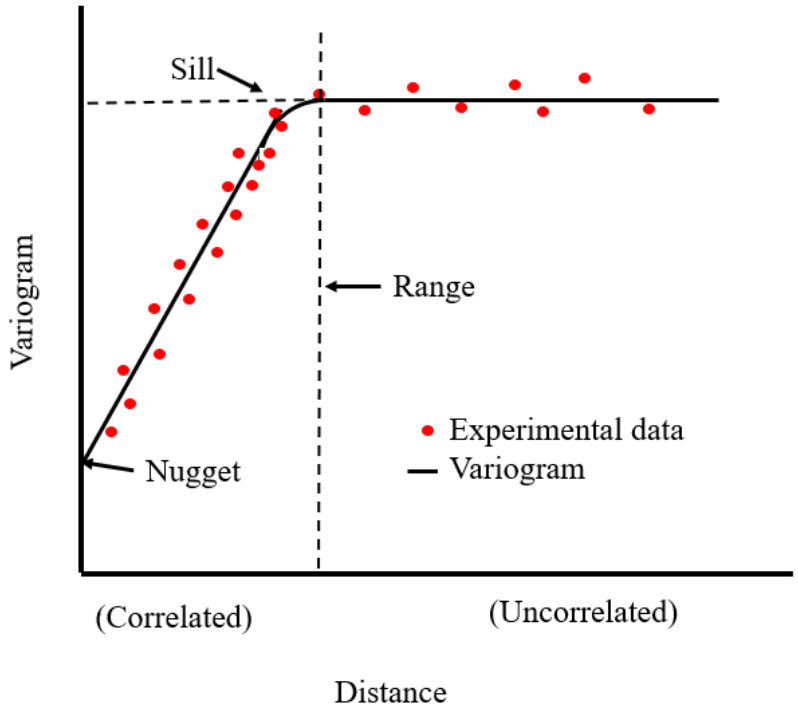


Figure 2: Variogram model

2. **Model fitting:** A mathematical model is then fitted to the variogram. Common models include spherical, exponential, and Gaussian variogram models.

3. **Prediction:** Using the variogram model, kriging estimates the value at an unsampled location by considering the values at surrounding sampled locations. The weights assigned to neighboring points are determined by the spatial structure revealed by the variogram. The unsampled datapoint is treated as a random. The kriging model then assumes that the mean of the random variable being estimated is constant throughout the area of interest leading to the smoothing effect as shown in Figure.

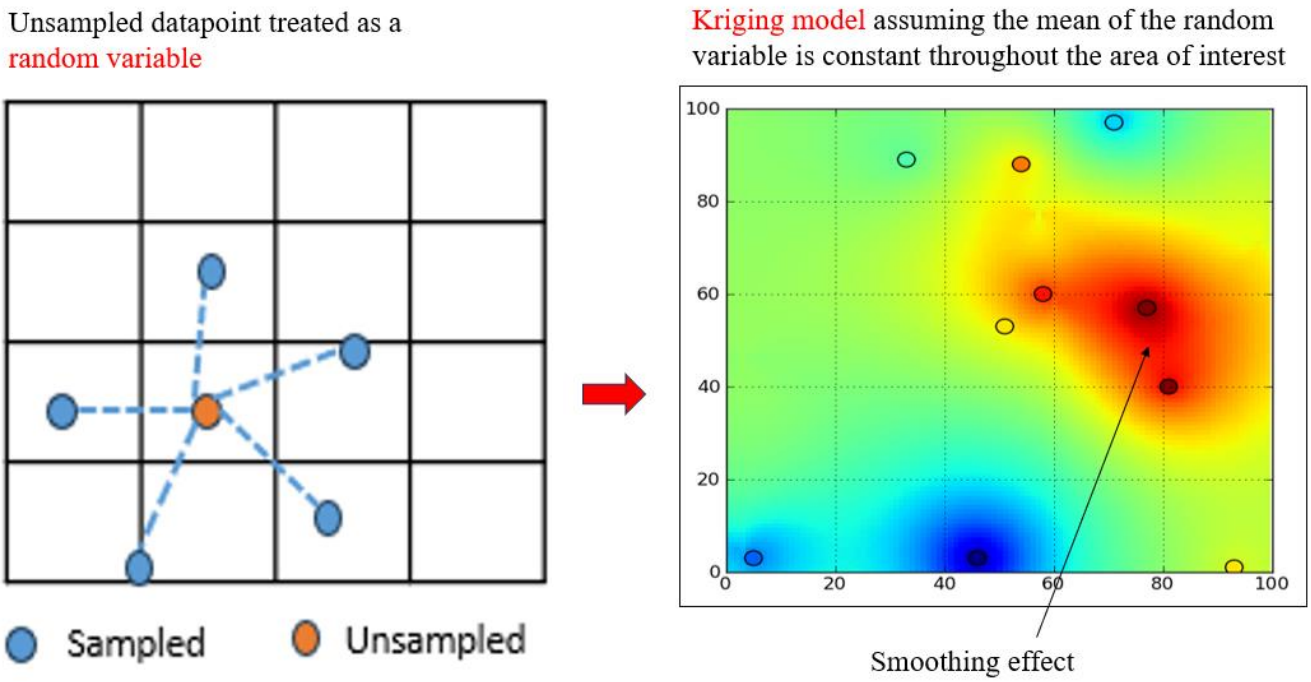


Figure 3: Kriging model

Variants of kriging include ordinary kriging, which assumes a known mean but estimates the local spatial variation, and universal kriging, which also estimates the mean in addition to the

local variation. These methods can be particularly useful when dealing with data sets that exhibit trends or non-stationary behavior.

Ordinary Kriging is the most commonly used kriging method in ore grade estimation. It assumes that the mean of the variable being estimated is constant across the study area. OK provides unbiased estimates that are linearly unbiased, meaning that on average, the estimation errors are zero (Isaaks.,1989).

Co-kriging: Co-kriging is used when additional auxiliary variables (e.g., lithology, geophysical data) are available to improve the estimation accuracy. By incorporating the spatial relationship between the primary variable (ore grade) and the auxiliary variable(s), co-kriging produces more reliable estimates.

Indicator Kriging: In cases where the ore deposit is discontinuous or exhibits distinct geological domains, indicator Kriging is employed. This method estimates the probability of a certain threshold being exceeded at unsampled locations, allowing for the modeling of discrete ore/non-ore boundaries.

In addition to kriging, other geostatistical methods such as inverse distance weighting (IDW) and conditional simulation techniques like sequential Gaussian simulation (SGS) are also employed for ore grade estimation.

2.2. Inverse Distance Weighting (IDW)

Inverse Distance Weighting (IDW) is a spatial interpolation method used in geographic information systems (GIS) and geostatistics to estimate values at unobserved locations based

on values observed at known locations. The basic idea behind IDW is to give more weight to nearby sample points and less weight to more distant ones when making predictions. IDW estimates values based on the inverse of the distance to nearby samples, assigning more weight to closer samples. While IDW is computationally simple, it does not consider the spatial correlation structure as effectively as kriging.

$$Z(\mathbf{u}) = \frac{\sum_{i=1}^{n} w_i Z_i}{\sum_{i}^{n} w_i}$$

where:

$Z(\mathbf{u})$ is the estimated value at the unobserved location $\mathbf{u}$.

n is the number of observed sample points.

Z_i is the measured value at the i-th sample point.

w_i is the weight assigned to the i-th sample point, calculated based on the inverse of the distance between the unobserved location and the i-th sample point.

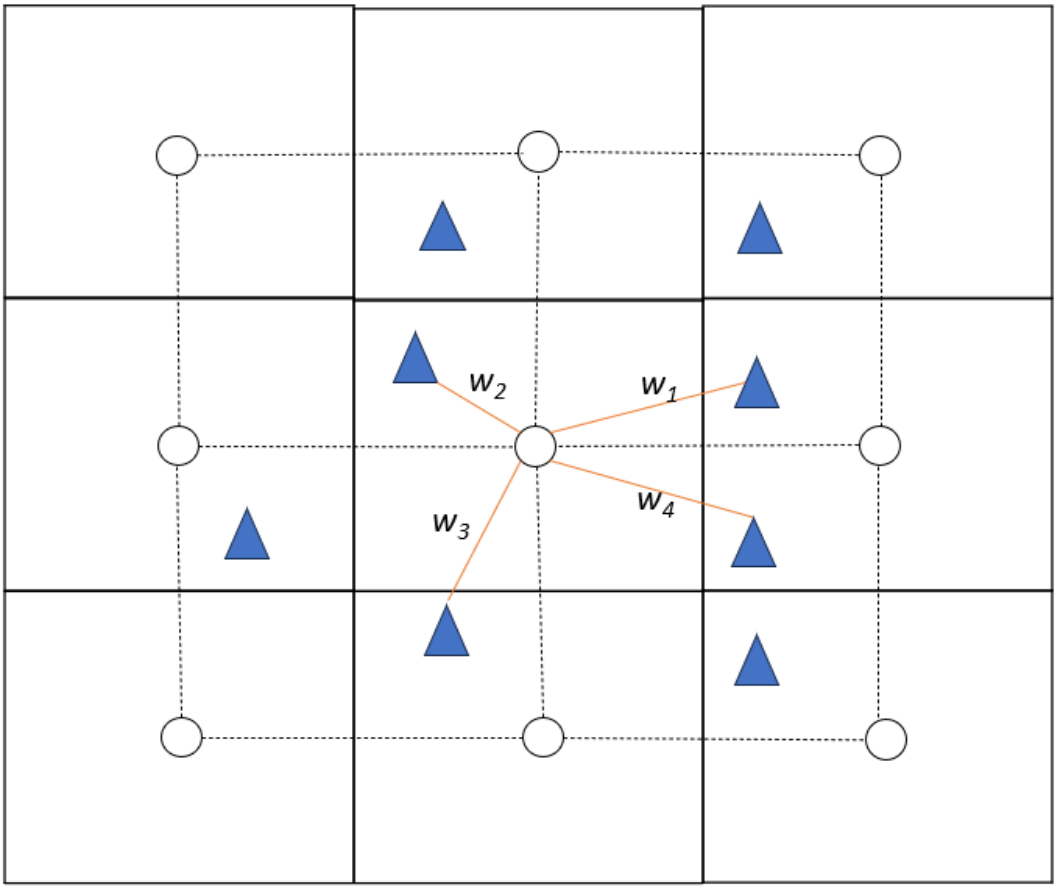


Figure 4: Schematic diagram of the inverse distance weighted interpolation with the blue triangles showing the data points.

The weight w_i is often defined as:

$$w_i = \frac{1}{d_i^p}$$

where:

d_i is the distance between the unobserved location and the i-th sample point.

p is a user-defined positive power parameter. This parameter controls how quickly the influence of a sample point diminishes with distance. The IDW method assumes that values at unobserved locations are influenced more by nearby observations than by those farther away. The choice of the power parameter *p* and the method of defining distance can affect the results, and these parameters are often determined based on the specific characteristics of the data and the problem at hand.

Sequential Gaussian Simulation (SGS)

SGS is a stochastic geostatistical method that generates multiple equiprobable realizations of the ore grades at unsampled locations. It honors the data constraints and reproduces the spatial variability characteristics of the deposit, providing a range of possible grade distributions.

Pros of Sequential Gaussian Simulation:

- SGS maintains the spatial continuity and correlation patterns observed in the sampled data, producing simulated models that reflect the actual spatial behavior of the phenomenon.
- It provides a measure of uncertainty by generating multiple realizations that represent plausible scenarios based on the available data and spatial variability.
- The method allows for the incorporation of different variogram models and can handle non-stationary and anisotropic spatial patterns, offering adaptability to various spatial datasets.

- SGS can honor hard data constraints (like known sample values) while simulating values at unsampled locations, ensuring that the simulated data respects the observed data values.

Cons of Sequential Gaussian Simulation:

- For large datasets or high-resolution simulations, SGS can be computationally intensive and time-consuming, especially if multiple realizations are required.
- SGS tends to produce smooth simulations, which might oversimplify the heterogeneity or abrupt changes in the real phenomenon, potentially smoothing out important details.
- Generating realistic simulations near the boundary of the observed area can be challenging, often resulting in less accurate predictions or biased estimations.
- SGS relies on assumptions such as stationarity, isotropy, and the validity of the variogram model, which might not hold true for all datasets, leading to potential inaccuracies in the simulated results.

Overall, while Sequential Gaussian Simulation is a valuable tool for spatial modeling and uncertainty assessment, its effectiveness depends on various factors, and it's essential to carefully consider its limitations and assumptions when applying it to real-world spatial data analysis.

David., (2012) argued that when applying geostatistical methods for ore grade estimation, it is essential to consider several factors. These include the quality and representativeness of the sample data, the selection of an appropriate variogram model that describes the spatial

correlation, the choice of estimation method, and the validation of the estimated grades against independent data. Additionally, careful consideration should be given to addressing potential biases and uncertainties associated with geostatistical modeling. The difficulty in determining grade for ore deposits with few data points using geometrical and geostatistical methods, on the other hand, has paved the way for the use of artificial intelligence in grade estimation.

REFERENCES

Journel, A.G., Huijbregts., C.J, Mining Geostatistics, Vol. 600, Academic press, London, 1978.

Rendu, J, Kriging, logarithmic kriging, and conditional expectation: Comparison of theory with actual results, in Proceedings 16th APCOM Symposium, Tucson, Arizona, 1979. p. 199–212.

Goovaerts, P, Geostatistics for Natural Resources Evaluation, NY: Oxford University Press, 1997.

Chiles, J.P, Definer., Modeling spatial uncertainty, Geostatistics, Wiley series in p 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. 37 (4) (2011), pp. 530–540. doi:10.1016/j. cageo.2010.07.009.

Tahmasebi, P., Hezarkhani, A., Application of a modular feedforward neural network for grade estimation, Nat. Resour. Res. 20 (1) (2011), pp. 25–32.

Isaaks, H.; Srivastava, R.M. Ordinary Kriging. In Applied Geostatistics; Oxford University Press Inc.: New York, NY, USA, 1989; pp. 288–295.

David, M., Geostatistical Ore Reserve Estimation, Tucson, Arizona: Elsevier, 2012.

CHAPTER 3: THEORETICAL FRAMEWORKS FOR MACHINE LEARNING

3.1 Introduction

Machine learning plays a crucial role in the prediction of ore grade estimation by leveraging statistical algorithms and computational power to analyze large datasets. It enables the development of predictive models that can learn patterns and relationships within the data, allowing for accurate estimation of ore grades based on various input features. Traditional methods of ore grade estimation relied on geological and statistical techniques, which often involved manual data analysis and assumptions. With the advancements in computing power and availability of large datasets, machine learning algorithms have emerged as powerful tools for ore grade estimation (Ewusi.,2021).

Machine learning is data-driven, meaning it can adapt to the underlying patterns in the data. If the data is heterogeneous, noisy, or contains outliers, machine learning algorithms can often handle it better than kriging, which assumes that the data follows a specific spatial correlation structure. They can capture non-linear patterns and handle a wide range of data types, making them suitable for cases where the relationship between variables is not well-understood or is non-linear. Moreover, machine learning algorithms can easily handle multi-dimensional datasets, allowing the incorporation of a wider range of input features into the grade estimation model, which can improve prediction accuracy. Furthermore, machine learning provides a wide array of algorithms to choose from, allowing the selection of the most appropriate model for grade estimation. Feature selection and hyperparameter tuning can be employed to improve the accuracy and performance of the various models.

Theoretical Background:

Supervised Learning: Supervised learning is a fundamental concept in machine learning. In the context of ore grade estimation, it involves training a model using labeled data, where each sample is associated with a known ore grade. The model learns from these labeled examples to make predictions on unseen data. Supervised learning algorithms commonly used for ore grade estimation include regression and classification algorithms. Some of the commonly used machine learning algorithms in ore grade estimation are regression algorithms, decisions trees, K-Nearest Neighbors (KNN) and neural networks. As illustrated in Figure 5. Linear regression and more advanced techniques like support vector regression or random forest regression can be employed to model the relationship between input features. Neural networks can learn complex patterns and relationships in the data through a process known as training. Decision trees are non-parametric models that partition the input space into regions based on feature values and make predictions based on the majority class in each region. Decision tree-based algorithms, such as random forests and gradient boosting machines (GBM), are widely used in ore grade estimation. KNN is a simple and intuitive algorithm that makes predictions based on the nearest neighbors in the feature space. In ore grade estimation, KNN can be used to estimate the grade of a sample based on the grades of its k nearest neighbors.

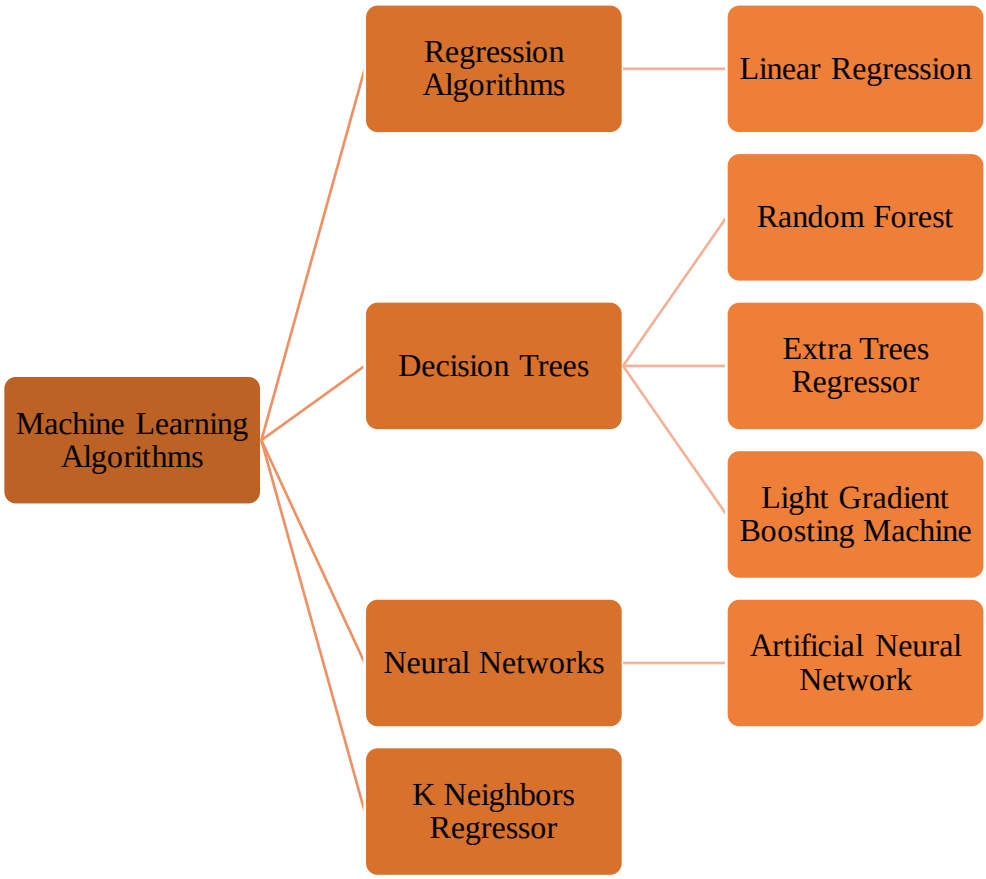


Figure 5: Machine learning Algorithms

It is important to note that the choice of machine learning algorithm depends on several factors, including the nature of the dataset, available computational resources, interpretability requirements, and the specific goals of the ore grade estimation task. Additionally, feature engineering, data preprocessing, and model evaluation techniques are crucial for obtaining reliable and accurate predictions in ore grade estimation using machine learning algorithms. The machine learning techniques used in this study are discussed in detail in this section.

3.2 Linear Regression

Linear regression is a fundamental and widely used technique in machine learning for predicting numerical values based on a set of input variables or features. It assumes a linear relationship between the input variables and the target variable and aims to find the best-fit line that minimizes the difference between the predicted values and the actual values.

The basic idea behind linear regression is to model the relationship between the input variables (denoted as X) and the target variable (denoted as y). Figure 6 shows a linear regression graph.

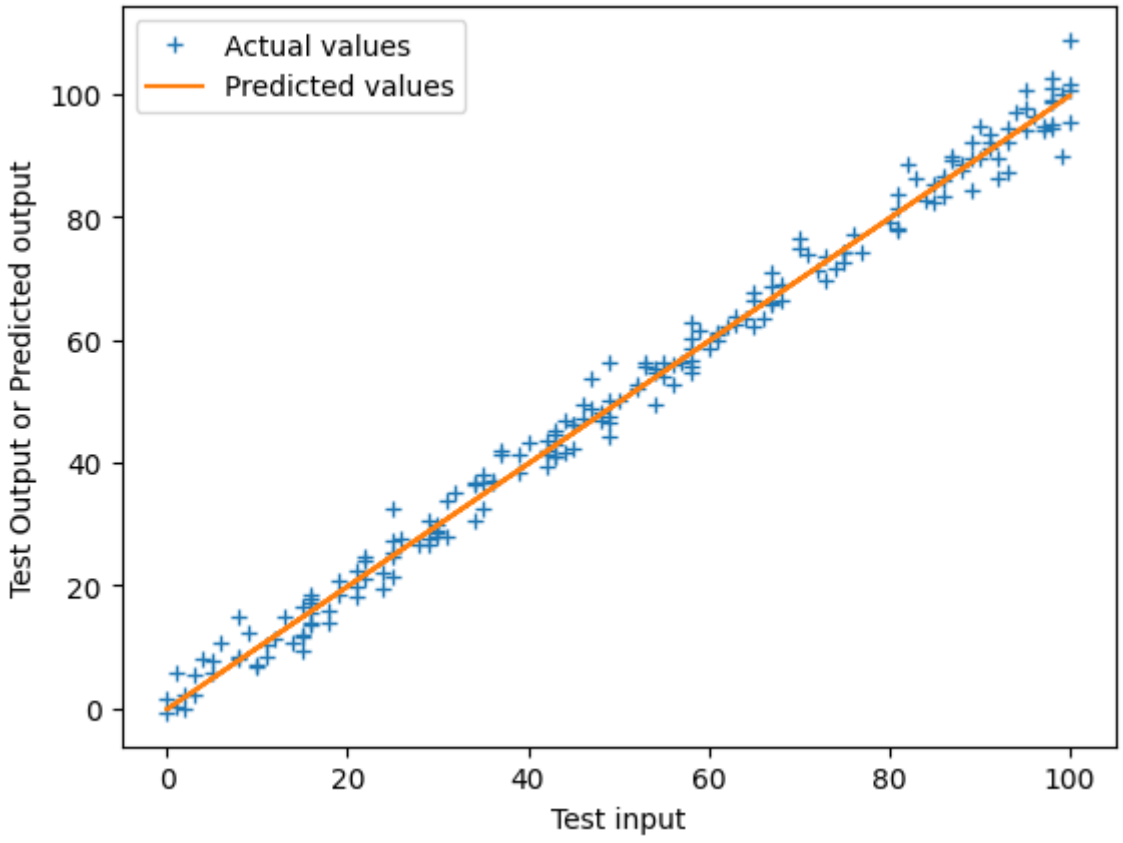


Figure 6: Best linear regression line with actual values

The goal is to find the optimal values for the coefficients that minimize the error between the predicted values and the actual values. The process of fitting a linear regression model involves estimating the coefficients using a method called "ordinary least squares" (OLS). OLS calculates the coefficients that minimize the sum of squared differences between the predicted values and the actual values. Once the coefficients are determined, the model can be used to make predictions on new data by plugging in the values of the input variables. The key advantage of linear regression is that it is a straightforward and easy-to-understand algorithm, The coefficients in linear regression can be interpreted to understand the impact of each input variable on the target variable. It can be computed efficiently, even with large datasets.

The major limitation of linear regression is that it assumes a linear relationship between the input variables and the target variable. If the relationship is nonlinear like in ore grade distribution, it may not provide accurate predictions. It is very sensitive to outliers, as it tries to minimize the squared differences. Outliers can significantly impact the estimated coefficients and the overall performance of the model. Linear regression assumes that the input variables are independent of each other. If there is high multicollinearity (correlation) among the input variables, it can lead to unreliable coefficient estimates. Linear regression is a powerful and widely used tool in machine learning for solving regression problems, making it a fundamental technique for data analysis and prediction.

3.3 Random Forest Regressor

Random forest was proposed by Breiman., (2001) as an ensemble classifier or regression tree based on decision trees. Random Forest Regressor is a popular machine learning algorithm used for regression tasks. It is an ensemble method that combines multiple decision trees to make predictions as displayed in Figure 7. Each tree in the random forest is trained on a random subset of the training data, and the final prediction is obtained by averaging the predictions of all the individual trees. The algorithm selects a random subset of the training data with replacement. This random sampling is called bootstrapping. It ensures that each tree is trained on a slightly different subset of the data, introducing diversity in the ensemble. A decision tree is constructed using the bootstrapped subset of data. At each node of the tree, the algorithm selects the best feature and split point based on certain criteria (such as minimizing the variance of the target variable).

The tree continues to grow until a stopping criterion is met, typically when a maximum depth is reached or when further splits do not improve the model significantly. Multiple decision trees are trained using different bootstrapped subsets of the training data. The number of trees in the random forest is a parameter that can be tuned. Having a larger number of trees generally improves the model's performance, but it also increases the computational cost. To make a prediction, the random forest regressor takes the average prediction of all the individual decision trees. For regression tasks, this averaging step produces a continuous output. Figure shows the random forest diagram.

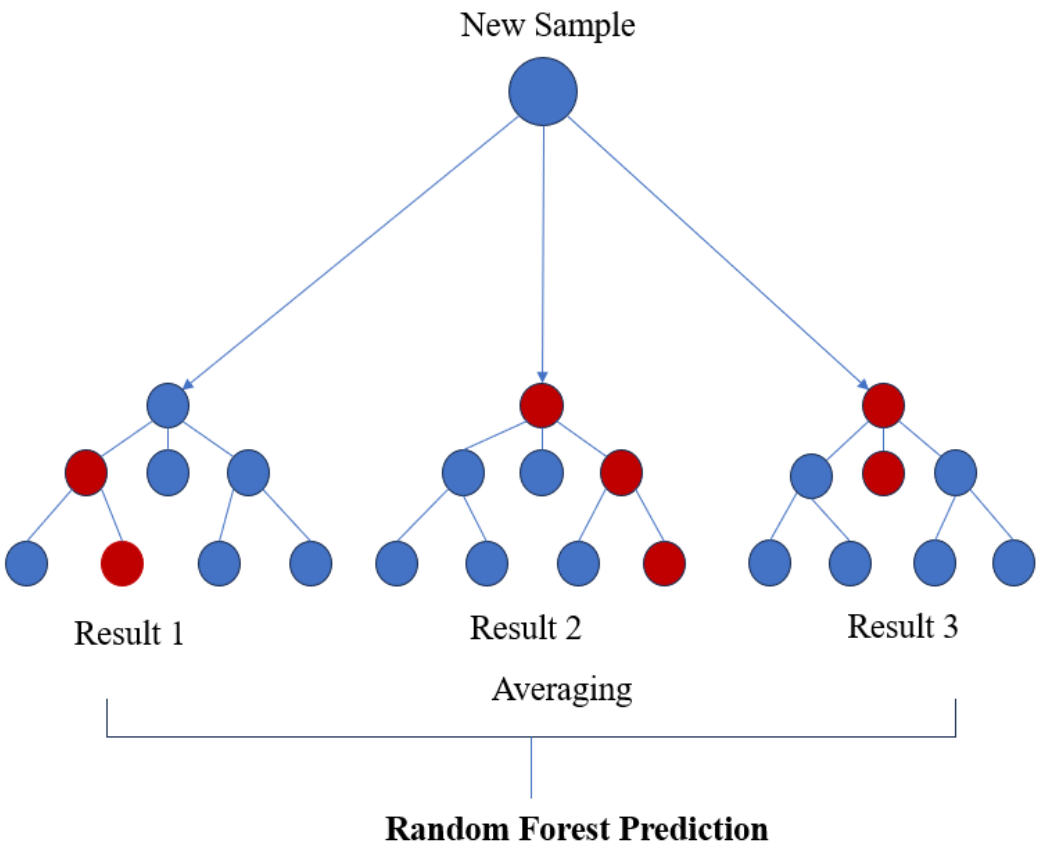


Figure 7: Random Forest Regressor

Random Forest Regressor offers several advantages, random forests are less prone to overfitting compared to individual decision trees because of the random sampling and averaging process. The ensemble of diverse trees helps to reduce variance and improve generalization. Also, they can provide a measure of feature importance. By evaluating the contribution of each feature in the ensemble of trees, it can help identify the most relevant features for the regression task. Additionally, random forests can capture nonlinear relationships between the features and the target variable. They are capable of modeling

complex patterns and interactions in the data. They can handle outliers and missing data effectively. They make use of the majority vote or averaging mechanism, which mitigates the impact of individual extreme values or missing values.

However, random forests also have some limitations. Interpretability of individual trees within the random forest is challenging, and the overall model can be considered a black box. Understanding the exact relationship between features and predictions may be difficult. Random forests with a large number of trees can be computationally expensive, especially when dealing with large datasets. Training and making predictions can take longer compared to simpler models. Random Forest Regressor is a powerful algorithm for regression tasks, known for its robustness, ability to handle complex relationships, and feature importance analysis. It is commonly used in various domains, including mineral resource estimations and environmental science.

Schnitzler et al., (2019) assessed the Random Forest performance with varying numbers of instances and input variables. The MLP network performed admirably in terms of test error and training speed. Samantha et al. (2009) estimated ore grade values using an RBF network and compared the results to feed-forward neural networks and conventional ordinary kriging and they concluded that feed forward neural networks provided better results than ordinary kriging. In the porphyry copper deposit in south-eastern Iran, Jafrasteh et al. studied the grade prediction performance of machine learning approaches such as ANN, random forest (RF), and Gaussian process (GP). Their findings revealed that GP performed best overall, followed by the widely used traditional estimation technique, indicator kriging; however, developing

a multilayered ANN model for grade estimation necessitates the selection of network complexity, which has a direct impact on prediction performance. The complexity of a network is determined by (a) neuron complexity, (b) the number of neurons in each layer, (c) the number of layers, and (d) the number and type of linking weights.

3.4 Extra Trees Regressor

Extra Trees Regressor is an ensemble learning method used for regression tasks. It is an extension of the Random Forest algorithm and shares many similarities with it. In Extra Trees Regressor, multiple decision trees are trained on random subsets of the training data and random subsets of the features as shown in Figure 8. The final prediction is obtained by averaging the predictions of all the individual trees. One key characteristic of the Extra Trees Regressor is that it introduces additional randomness compared to Random Forests. In Random Forests, the optimal split point for each feature is determined by searching for the best split among a subset of randomly selected features. In contrast, Extra Trees Regressor randomly selects the split points without searching for the best split. This randomness leads to a higher level of diversity among the trees, which can reduce overfitting and improve generalization performance.

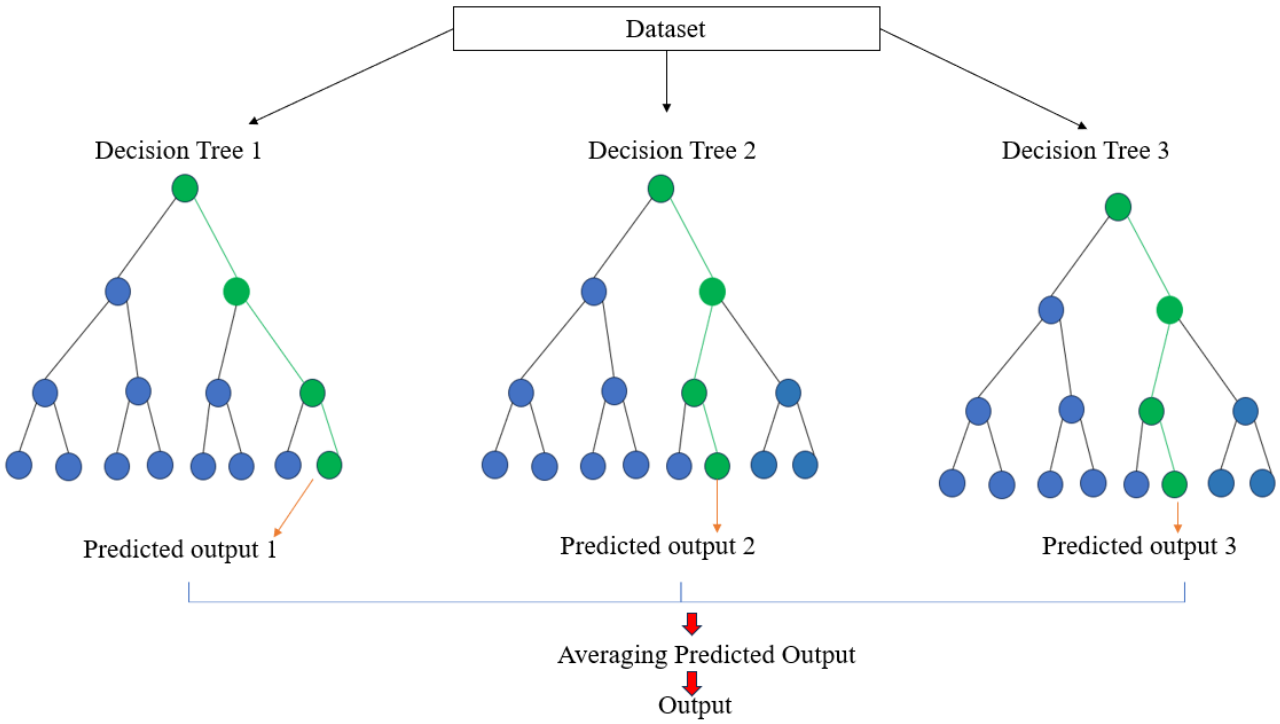


Figure 8: Extra trees regressor

The main advantage of Extra Trees Regressor is its ability to handle high-dimensional datasets with noisy or irrelevant features. By randomly selecting features for each split, it becomes less sensitive to noisy or irrelevant attributes, allowing it to focus on the more informative features. Moreover, the additional randomness introduced in Extra Trees Regressor reduces the bias of the individual trees, making it more robust to outliers.

However, the increased randomness in Extra Trees Regressor comes at the cost of increased computational complexity compared to Random Forests. Since each tree is constructed without searching for the best split, the training process is faster but may result in slightly higher variance. Additionally, the interpretability of Extra Trees Regressor may be lower

compared to single decision trees, as it becomes harder to trace the contributions of individual features. In practice, the performance of Extra Trees Regressor can vary depending on the dataset and the specific problem at hand. It is generally recommended to compare its performance with other regression algorithms, such as Random Forests and Gradient Boosting, to determine the most suitable model for a given task.

3.5 Light Gradient Boosting Machine (LGBM)

Gradient Boosting Machines (GBMs) are a powerful class of machine learning algorithms that have gained significant popularity in recent years. They are particularly effective for solving regression and classification problems and have proven to be highly successful in various domains, including finance, healthcare, and computer vision. One variant of GBMs that has gained prominence is LGBM. LGBM is an efficient implementation of the gradient boosting framework, developed by Microsoft. It is designed to be fast, scalable, and memory-efficient, making it suitable for working with large datasets.

It is worth noting that LGBM, like other GBM algorithms, works by sequentially training a series of weak learners, typically decision trees, in an ensemble manner as shown in Figure 9.

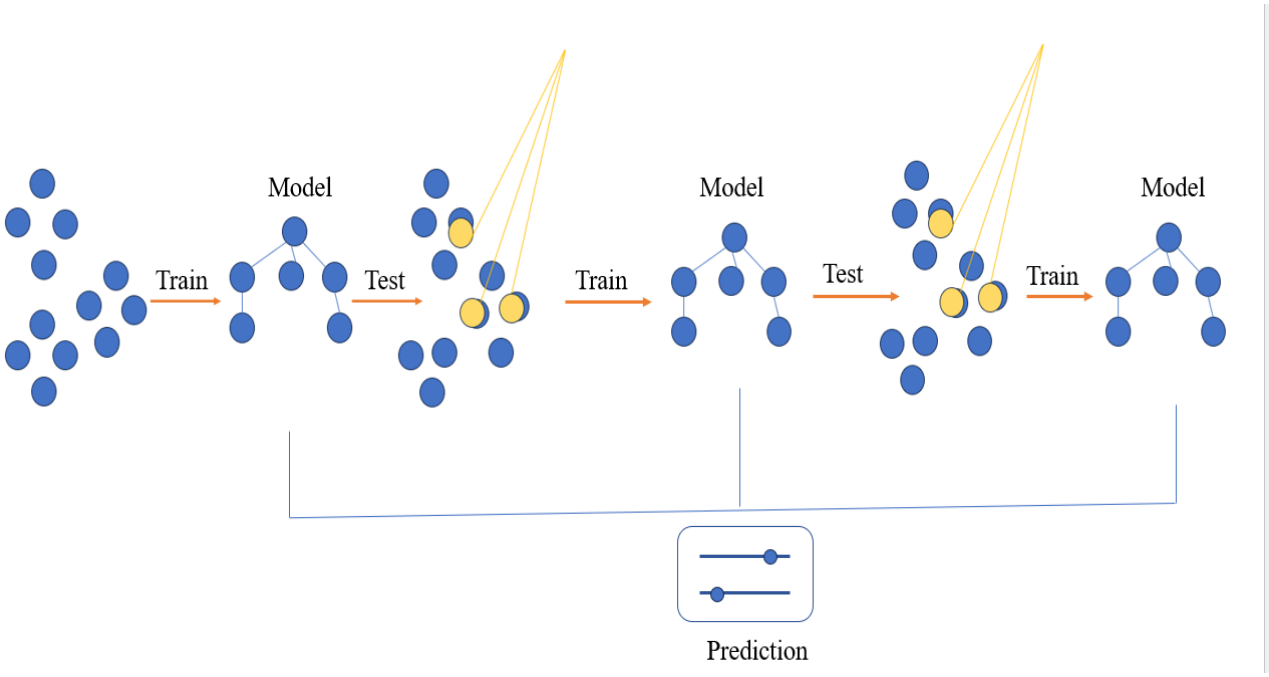


Figure 9: Light Gradient Boosting Machine

It builds the model in an iterative fashion, where each subsequent weak learner is trained to correct the mistakes made by the previous learners. This process allows the model to focus on the samples that are difficult to predict. uses a novel technique called Gradient-based One-Side Sampling (GOSS) to select a subset of data instances that are more informative for the model's training. This approach significantly reduces memory usage and training time without compromising accuracy.

Additionally, LGBM supports efficient handling of categorical features and provides built-in support for parallel and distributed computing. LGBM employs a leaf-wise tree growth strategy, as opposed to the level-wise growth strategy used by most other implementations. In a leaf-wise strategy, the algorithm grows the tree by expanding the leaf with the highest loss reduction, resulting in a more balanced and deeper tree structure. This approach

generally leads to better accuracy but can potentially overfit if not carefully controlled through regularization techniques.

LGBM offers a wide range of hyperparameters that can be tuned to optimize model performance. These include parameters related to tree structure (e.g., maximum depth, minimum number of samples per leaf), learning rate, number of boosting iterations, and regularization parameters. Hyperparameter tuning plays a crucial role in achieving the best possible performance and preventing overfitting. It is important to note that while LGBM is a powerful tool, it is not a one-size-fits-all solution. The choice of algorithm depends on the specific problem, dataset characteristics, and computational constraints. Other gradient boosting libraries, such as XGBoost and CatBoost, also have their unique strengths and may be more suitable in certain scenarios. In summary, LGBM is a valuable addition to the gradient boosting landscape, offering speed, scalability, and memory efficiency. Its innovative techniques and extensive feature set make it a popular choice for machine learning practitioners and researchers alike.

Kaplan et al., (2021) presented a grade estimation workflow using gradient boosting-based machine learning methods, namely, XGBoost, LightGBM and CatBoost. The case study demonstrated that the three-gradient descent-based models performed better than the OK method. XGBoost model demonstrated the best estimation performance with an R^2 of 0.728 accuracies, whereas traditional Ordinary Kriging (OK) model yielded 0.651 for R^2.

3.6 K Neighbors Regressors

K Neighbors Regressor is a machine learning algorithm that belongs to the family of supervised learning algorithms. It is primarily used for regression tasks, where the goal is to predict a continuous numerical value rather than a discrete class label. The algorithm works by finding the K nearest neighbors to a given data point in the feature space and then making predictions based on the average or weighted average of the target values of those neighbors.

When using K Neighbors Regressor algorithm works a labeled dataset consisting of input features and corresponding target values is required. Figure 10 shows how a K neighbors regressor works. The pink variables to be predicted circles and the green circles indicates the neighboring variables used to estimate the target. The features are the independent variables that will be used to predict the target variable, which is a continuous numerical value. The parameter K represents the number of neighbors to consider when making predictions. It is an important parameter that is needed to specify before training the model. The value of K can be chosen based on domain knowledge or through cross-validation techniques. The algorithm calculates the distance between the input data point and each of the training data points using a distance metric such as Euclidean distance. The distance metric determines the similarity between data points and plays a crucial role in finding the nearest neighbors.

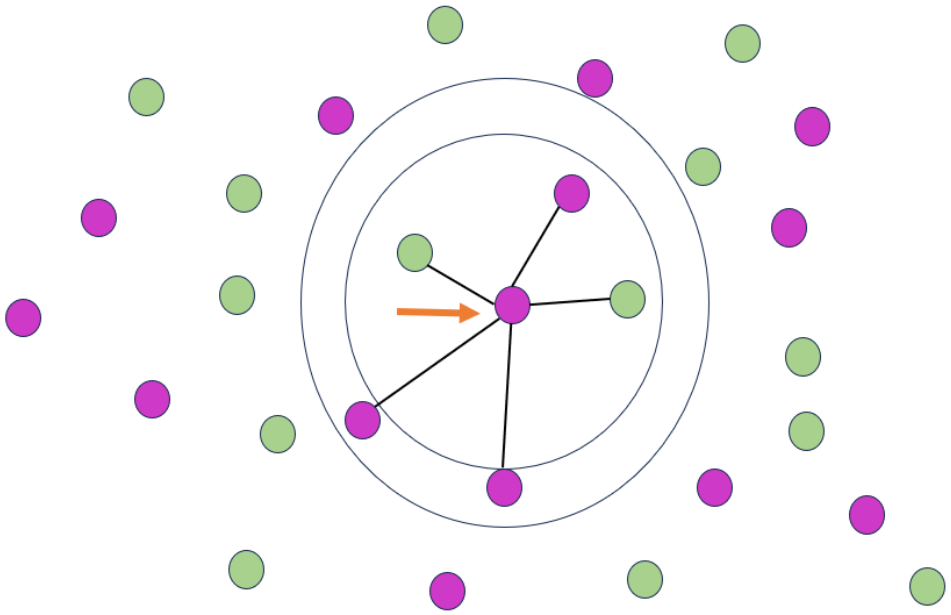


Figure 10: K neighbors regressor

The algorithm identifies the K nearest neighbors to the input data point based on the calculated distances. These neighbors are the K training instances with the closest feature values to the input data points. Once the nearest neighbors are identified, the algorithm makes a prediction for the target value of the input data point. In the case of regression, the prediction is usually the average or weighted average of the target values of the nearest neighbors. The weight of each neighbor can be determined by the inverse of their distance to the input data point. After making predictions for all the data points in the test set, the model's performance is evaluated using appropriate regression evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), or R^2.

Certain considerations are needed when using KNN. The choice of K is critical and can significantly impact the performance of the model. A small value of K can lead to overfitting, where the model becomes too sensitive to the noise in the data. On the other hand, a large value of K can lead to underfitting, where the model oversimplifies the relationship between the features and the target variable. It is essential to experiment with different values of K and use techniques like cross-validation to select the optimal value. Since K Neighbors Regressor uses distance-based calculations, it is generally recommended to normalize the input features.

Normalization ensures that each feature contributes equally to the distance calculations and prevents features with larger scales from dominating the distance calculation. The K Neighbors Regressor algorithm has a high computational cost during prediction since it requires calculating the distances between the input data point and all the training instances. This can be particularly slow for large datasets or high-dimensional feature spaces. The curse of dimensionality refers to the fact that as the number of dimensions (features) increases, the density of the data becomes sparse in the feature space. This can lead to a degradation in the performance of the K Neighbors Regressor algorithm, as it becomes difficult to find meaningful neighbors. Feature selection or dimensionality reduction techniques can be applied to mitigate the curse of dimensionality. To summarize, K Neighbors Regressor is a simple yet powerful algorithm for regression tasks. Chatterjee et al. (2010) 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 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.

3.7 Evaluation Performance Metrics

Evaluation metrics are quantitative measures used to assess the performance of a model, in various fields, including machine learning and data analysis. Evaluating the performance of a machine learning model is crucial to ensure its reliability and generalizability. These metrics provide a way to objectively gauge how well a model is performing and are crucial for making informed decisions and improvements. The choice of evaluation metrics depends on the specific task and goals. Selection of the appropriate evaluation metrics takes place after careful consideration of potential tradeoffs between the different metrics in relation to ore grade estimation, the evaluation metrics were chosen for this study are mean absolute error (MAE), root mean square error (RMSE), mean absolute error, correlation coefficient (R) and coefficient of determination (R^2).

Coefficient of Determination, R^2

The coefficient of determination, denoted as R^2, is a statistical measure used to assess the goodness of fit of a regression model. It represents the proportion of the variance in the dependent variable that can be explained by the independent variables in the model. Equation 1 shows how R^2 is computed.

$$\boldsymbol{R^2} = 1 - \frac{RSS}{TSS} \qquad (1)$$

Where:

RSS= sum of squared residuals.

TSS = sum of squares.

R^2 is a value between 0 and 1. A value of 0 indicates that the model does not explain any of the variability in the dependent variable, while a value of 1 indicates that the model explains all of the variability. In general, the closer the R^2 value is to 1, the better the model fits the data. However, it is important to note that R^2 has some limitations and should not be solely relied upon to evaluate the effectiveness of a regression model. R^2 measures the proportion of the total variation in the dependent variable that is explained by the independent variables. It gives an overall indication of how well the model fits the data. A higher R^2 suggests that the model captures a larger proportion of the variability.

R^2 tends to increase as more independent variables are added to the model, even if those variables are not truly meaningful. Therefore, it is important to consider other metrics to account for the number of variables and prevent overfitting. R^2 does not provide information about the causal relationships between variables or the statistical significance of the coefficients. It only measures the overall fit of the model. Therefore, it is important to also examine the significance of individual coefficients and conduct hypothesis tests.

In conclusion, the coefficient of determination (R^2) is a useful measure to assess the goodness of fit of a regression model. However, it should be used in conjunction with other evaluation metrics and should be interpreted carefully, taking into account the limitations and context of the specific problem.

Correlation Coefficient, R

The correlation coefficient is a statistical measure that quantifies the relationship between two variables (in this case the input variables and the ore grade). It is used to determine the strength and direction of the linear association between the variables. The expression for calculation R is shown in the equation.

$$\mathbf{R} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum x_i-\bar{x})^2\,(\sum y_i-\bar{y})^2}} \quad (2)$$

Where:

x_i and y_i are the individual data points.

$\bar{X}$and $\bar{Y}$ are the means of the X and Y data, respectively.

Figure 11 shows a correlation coefficient structure.

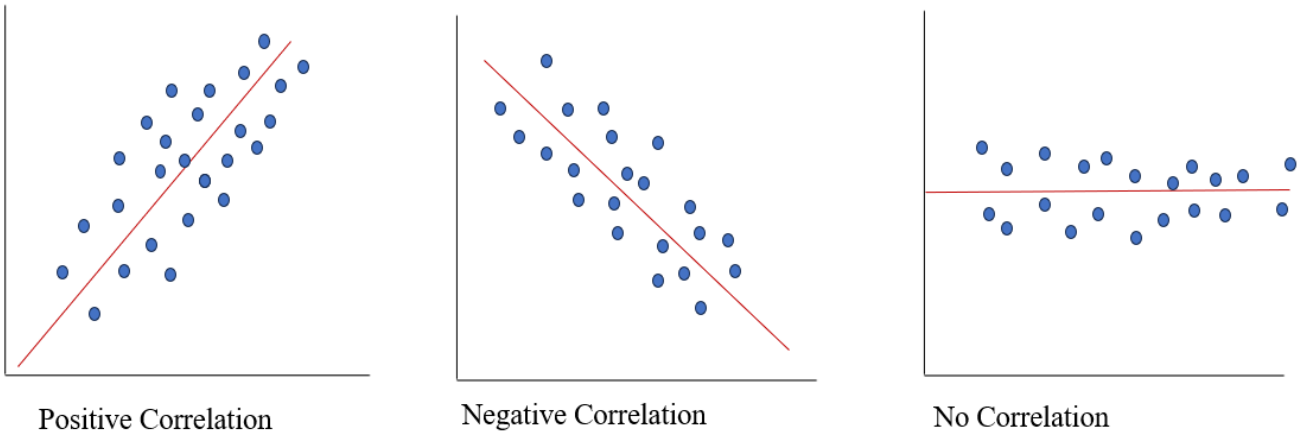


Figure 11: Correlation coefficient structure

Correlation coefficient can range from -1 to 1. A correlation coefficient of 1 indicates a perfect positive correlation, meaning that as one variable increases, the other variable increases proportionally. On the other hand, a correlation coefficient of -1 indicates a perfect negative correlation, whereas one variable increases, the other variable decreases proportionally. A correlation coefficient of 0 indicates no linear relationship between the variables. The correlation coefficient is widely used in various fields, including statistics and economics, to examine relationships between variables, identify patterns, make predictions, and assess the strength of associations. However, it is vital to mention that correlation does not imply causation. Even if two variables are strongly correlated, it does not necessarily mean that one variable causes the other to change. Other factors and relationships need to be considered to establish causation.

Mean Square Error, MSE

Mean squared error (MSE) is a commonly used loss function in machine learning, particularly for regression problems. It measures the average squared difference between the predicted and actual values. The MSE expression is shown by equation 3:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2 \quad (3)$$

where MSE is the mean square error, n is the number of observations, y_i the observed value, and $\widehat{y_i}$ = predicted value.

MSE is a differentiable function, which makes it suitable for optimization algorithms that rely on gradient-based techniques, such as gradient descent. Squaring the differences between

predicted and actual values amplifies larger errors, making the model more sensitive to such cases during training. MSE treats overestimation and underestimation equally since the squared difference is always positive. This balanced approach can be desirable in certain scenarios.

However, MSE also has some limitations, the squared differences in MSE can be heavily influenced by outliers, as they contribute significantly to the overall loss. This can lead to models that are overly sensitive to extreme values. The squared values in MSE make it less interpretable compared to other loss functions. Since MSE squares the errors, it tends to prioritize reducing larger errors over smaller ones. In situations where small errors are more critical, alternative loss functions may be more appropriate. MSE is a widely used loss function in regression problems due to its simplicity and differentiability. However, it's important to consider the specific characteristics of the problem at hand and evaluate whether MSE aligns well with the requirements and constraints of the task. 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.

Root Mean Square Error, RMSE

Root Mean Square Error (RMSE) is a commonly used metric for evaluating the accuracy or performance of a prediction model. It measures the average deviation between the predicted values and the actual values, providing a single number that represents the overall error of the model. The RMSE is calculated by taking the square root of the mean of the squared

differences between the predicted values and the actual values. It follows a specific formula 4:

$$\text{RMSE} = \text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (4)$$

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.

By squaring the differences between the predicted and actual values, we eliminate any negative signs and emphasize larger errors. Taking the mean and then the square root ensures that the RMSE is on the same scale as the original data, making it easier to interpret and compare across different models or datasets. A lower RMSE value indicates better predictive accuracy, as it reflects a smaller average deviation between predicted and actual values. Conversely, a higher RMSE suggests a larger average deviation and poorer model performance. RMSE is a versatile metric that can be applied to various prediction tasks, such as regression, time series analysis, and forecasting. It provides a quantitative measure of the model's predictive power and can be used to compare different models or assess the improvement of a model over time.

Mean Absolute Error, MAE

Mean Absolute Error (MAE) is a commonly used metric in statistics and machine learning to measure the average magnitude of errors between predicted and actual values. It provides a simple and straightforward way to assess the accuracy of a predictive model.

MAE is particularly useful when dealing with continuous numerical data, where the difference between predicted and actual values is meaningful. It is expressed in the same units as the data being measured, making it easy to interpret.

To calculate the MAE, take the absolute difference between each predicted value and its corresponding actual value, sum up these differences, and then divide by the total number of observations. The formula for MAE can be represented as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \widehat{y_i}| \quad (5)$$

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

The MAE value indicates the average magnitude of errors made by the model, regardless of the direction (overestimation or underestimation). A lower MAE value indicates better model accuracy, as it signifies smaller average errors. MAE is a popular choice in various domains such as regression analysis as it provides a simple and intuitive understanding of how well a model performs in terms of absolute errors, making it a valuable tool for assessing and comparing different models or techniques.

3.8 Feature Importance

Feature importance refers to the process of determining the relative importance or contribution of different features in a given dataset or model. It helps in understanding which features have the most influence on the outcome or target variable. Feature importance can be useful in various machine learning tasks, such as feature selection, model interpretation,

and identifying the key factors driving predictions. Additionally, feature importance should be used as a tool for gaining insights rather than as the sole basis for decision-making. There are different methods to calculate feature importance, and the choice of method depends on the type of model and data. For this study Shapley values were used to determine feature importance.

Shapley values are a concept from cooperative game theory that has been adapted and applied to machine learning to understand the contribution of individual features or variables in a predictive model (Lundberg and Lee., 2017). They provide a way to fairly distribute the credit for the model's prediction among its input features. In the context of machine learning, Shapley values quantify the importance or influence of each feature in the model's predictions by considering all possible combinations of features and computing the marginal contribution of each feature to the prediction. The idea is to measure the change in the model's prediction when a particular feature is added or removed, taking into account all possible orderings or coalitions of features. To calculate Shapley values, one typically considers a reference model, such as an average prediction or a baseline prediction, and evaluates the average contribution of each feature across all possible feature combinations. This involves calculating the model's prediction for every possible subset of features, comparing it to the reference prediction, and attributing the difference to the corresponding features. The Shapley value for each feature is the average contribution it provides across all possible orderings of features.

Shapley values offer several advantages in interpreting and understanding machine learning models. They provide a quantitative measure of the importance or influence of each feature,

allowing for feature selection and prioritization. Shapley values ensure that the credit for the model's prediction is distributed fairly among the features, accounting for their individual contributions. They provide insights into the relationship between features and predictions, helping to explain the model's decision-making process. Additionally, they can help identify and diagnose issues in the model by highlighting features that have inconsistent or unexpected contributions. Shapley values can guide feature engineering efforts by indicating which features have the most impact on the model's predictions, enabling more informed feature transformations or interactions. It is worth noting that computing Shapley values exactly can be computationally expensive due to the exponential number of possible feature combinations. Various approximations and sampling techniques have been developed to address this computational challenge and provide reasonable estimates of the Shapley values. Shapley values offer a principled and interpretable approach to feature importance and can enhance the transparency and understanding of machine learning models.

3.9 Why ANN

An Artificial Neural Network (ANN) is a computational model inspired by the structure and functioning of biological neural networks, such as the human brain. 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 (Dumakor-Dupey and Arya.,2021). It consists of three major interconnected layers: the input, hidden, and output layers, which determine the

network architecture. The input layer receives input data. Hidden layers are intermediate layers between the input and output layers. They perform computations and extract features from the input data while the output layer produces the network output. Each neuron has an activation function that determines its output based on the weighted sum of its inputs.

Some of the key concepts and components of ANN are discussed. The process of passing data through the network from the input layer to the output layer is called feedforward. During this process, computations are performed layer by layer. Backpropagation is the training algorithm used to update the weights of the connections in the network. It involves calculating the error between the predicted output and the actual target and then adjusting the weights to minimize this error. The loss function quantifies how well the network's predictions match the actual target values. During training, the goal is to minimize this loss function. The learning rate is a hyperparameter that controls the step size during weight updates in the training process. It affects how quickly or slowly the network converges to a solution,

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. Yama and Lineberry., (1999) analyzed sulfur levels in 1152 samples collected from a 7315 x 4572-m coal site in northern West Virginia. It should be noted that the utilization of real data in similar studies is quite unusual. Due to computer memory constraints, the property was

divided into 25 regions (914 x 914 m). A network was trained for each region, utilizing the easting and northing as inputs and the sulfur readings as outputs. All data values were standardized before being used for network training and testing. The data were regularly distributed, which causes networks to produce results that are close to the mean value. Presenting the tails of the distribution more often to the network and with a higher learning rate has reduced this effect. The results obtained from the ANNs were compared with results from kriging. Clarici et al. (1993) also presented a similar strategy using a single hidden layer network. However, in that study, just one neural network was used for the full sampling area. 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 (Koike et al.,2002) Based on sample data, the network model can generate a model with many high-content zones.

ANNs have proven to be a prominent technique for estimating mineral resources, and studies (Chatterjee., 2010; Kaplan and Topal.,2020) 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 2.

The tanh activation function is used for the hidden layers, whereas a linear function is used for the output layer.

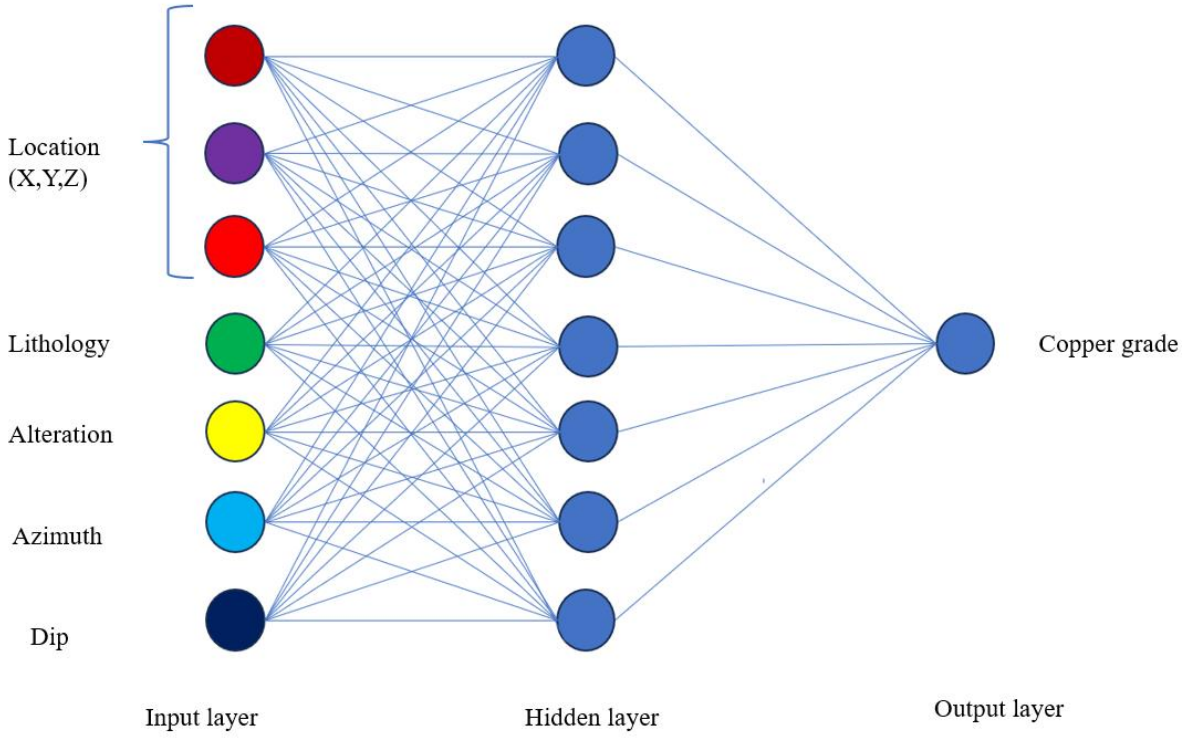


Figure 12:The proposed ANN model architecture for copper ore grade

Network training and Implementation

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. It is designed to minimize the error between the actual and the predicted model (Parker., 1982; Rumelhart.,1986). In this study, an ANN was trained using a Bayesian regularization backpropagation algorithm. Bayesian regularization is a powerful technique in machine learning and statistical modeling. It incorporates Bayesian principles into the process of regularization, which helps to overcome overfitting and uncertainty estimation in model parameters. Bayesian regularization (BR) exploits a mathematical process that converts nonlinear regression into a well-posed statistical problem in the manner of a ridge regression (Burden and Winkler, 2008). 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 (Awan et al.,2020),

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. It is a versatile technique that can enhance model robustness and interpretability.

\

REFERENCES

Ewusi, A., Ahenkorah, I., Aikins, D., 2021. Modelling of total dissolved solids in water supply systems using regression and supervised machine learning approaches. Appl. Water Sci. 11, 1–6. doi:10.1007/s13201-020-01352-7.

Breiman, L. Random Forests. Mach. Learn. 2001, 45, 5–32.

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.

Samantha, 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.

Kaplan,U.E, Y. Dagasan, and E. Topal, "Mineral grade estimation using gradient boosting regression trees," *International Journal of Mining, Reclamation and Environment*, vol. 35, no. 10, pp. 728–742, Nov. 2021, doi: 10.1080/17480930.2021.1949863.

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.

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.

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

Yama, B.R., and Lineberry, G.T., Artificial Neural Network Application for a Predictive Task in Mining. SME, Mining Engineering, February 1999, pp 59-64.

Clarici, E., Owen, D., Durucan, S., and Ravenscroft, P., Recoverable Reserve Estimation Using a Neural Network. 24th International Symposium on the Application of Computers and Operations Research in the Minerals Industries (APCOM), Montreal, Quebec, Canada, 1993.

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.

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

Parker, D.B. Learning Logic. Invention Report S81-64, File 1; Oce of Technology Licensing; Stanford University: Paolo Alto, CA, USA, 1982.

Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. Nature 1986, 323, 533–536.

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

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.

APPLICATION OF ANN FOR THE PREDICTION OF MINERAL ORE GRADES

This chapter is divided into two distinct parts. The opening chapter focuses on the use of ANN in predicting copper ore grade, while the subsequent chapter focuses on gold ore grade prediction. The effectiveness of the feature selection technique is demonstrated in the second section that concludes the second chapter.

Copper without Feature Selection

To predict copper ore grade from seven input variables.

Gold without Feature selection

To predict gold ore grade

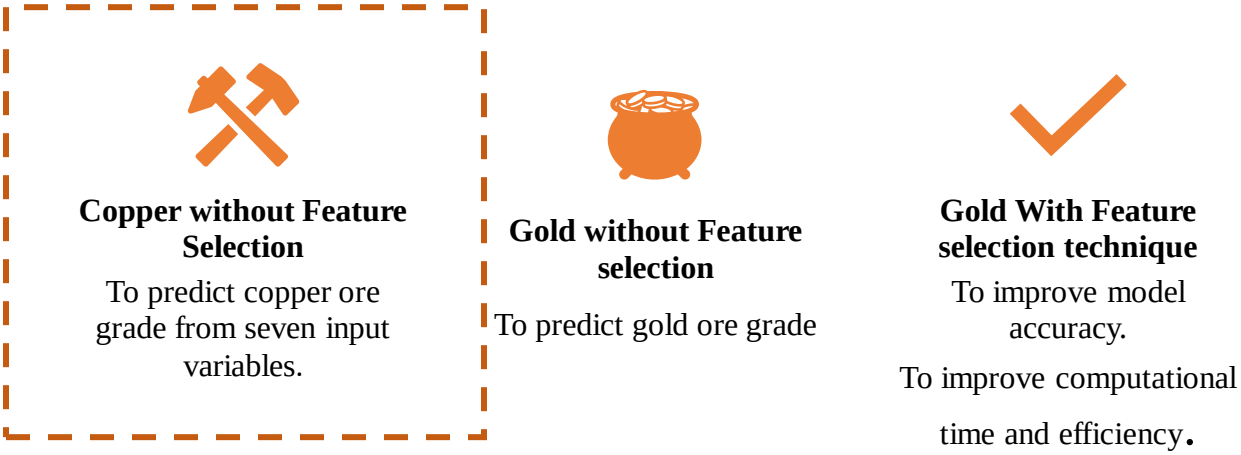
Gold With Feature selection technique

To improve model accuracy.

To improve computational time and efficiency.

CHAPTER 4: APPLICATION OF ANN FOR THE PREDICTION OF COPPER ORE GRADE

This chapter focuses on the application of ANN on copper grade prediction. It explains how the methodology was executed. The first section of the results and discussion highlights the efficacy of ANN and compares it to other machine learning techniques based on copper grade prediction and accuracy.



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 13. 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 and gold grade values of 185 drillholes with a drill spacing of 20 m. The raw drillhole data are shown on Figure 14.

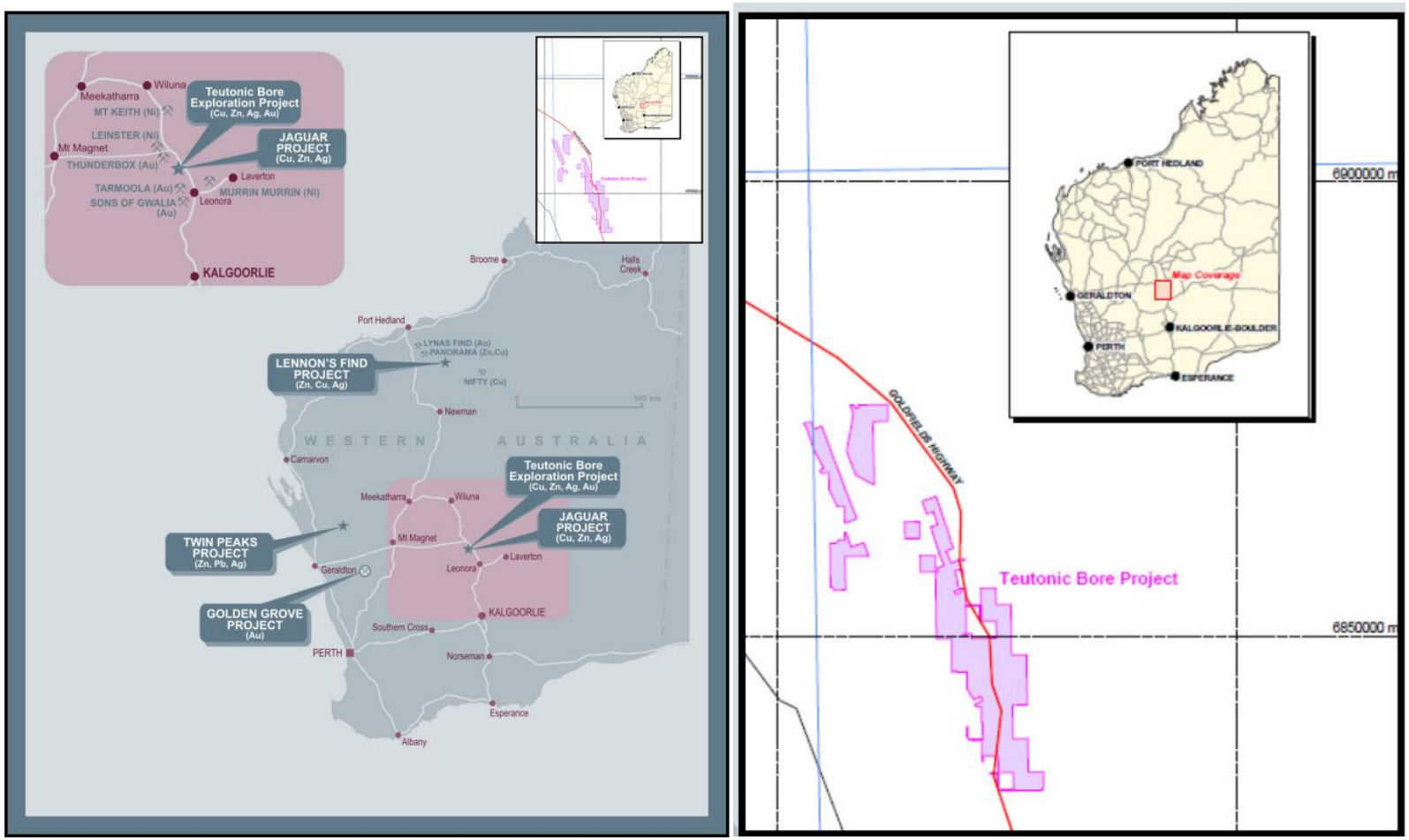
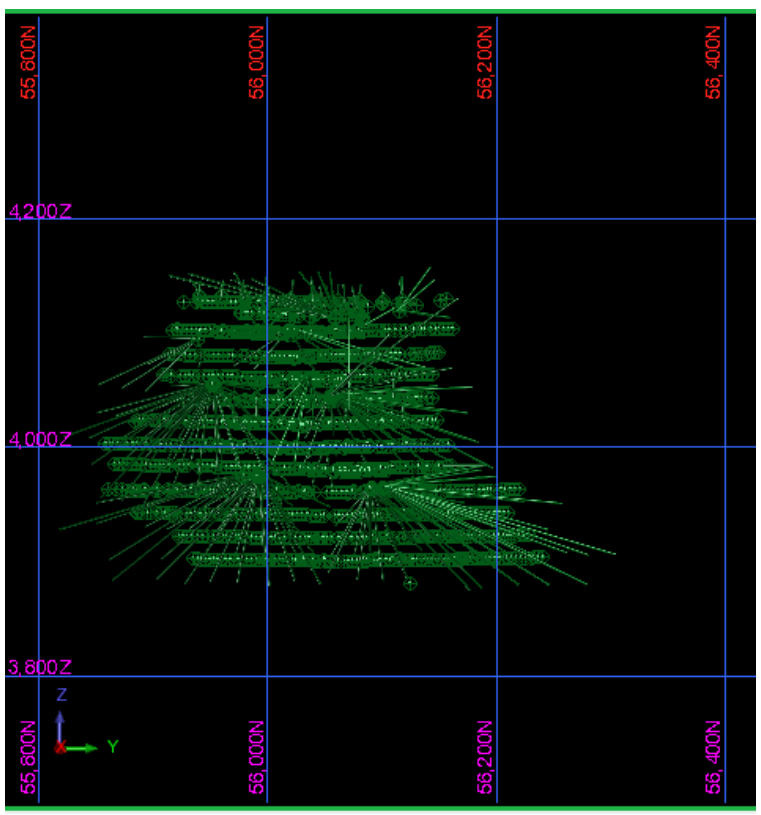
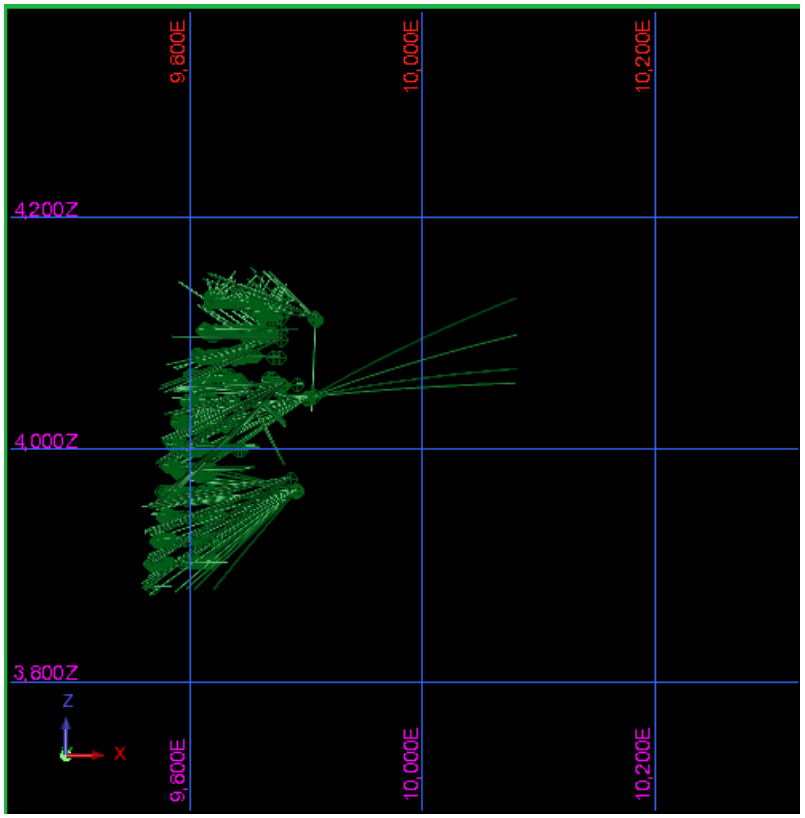


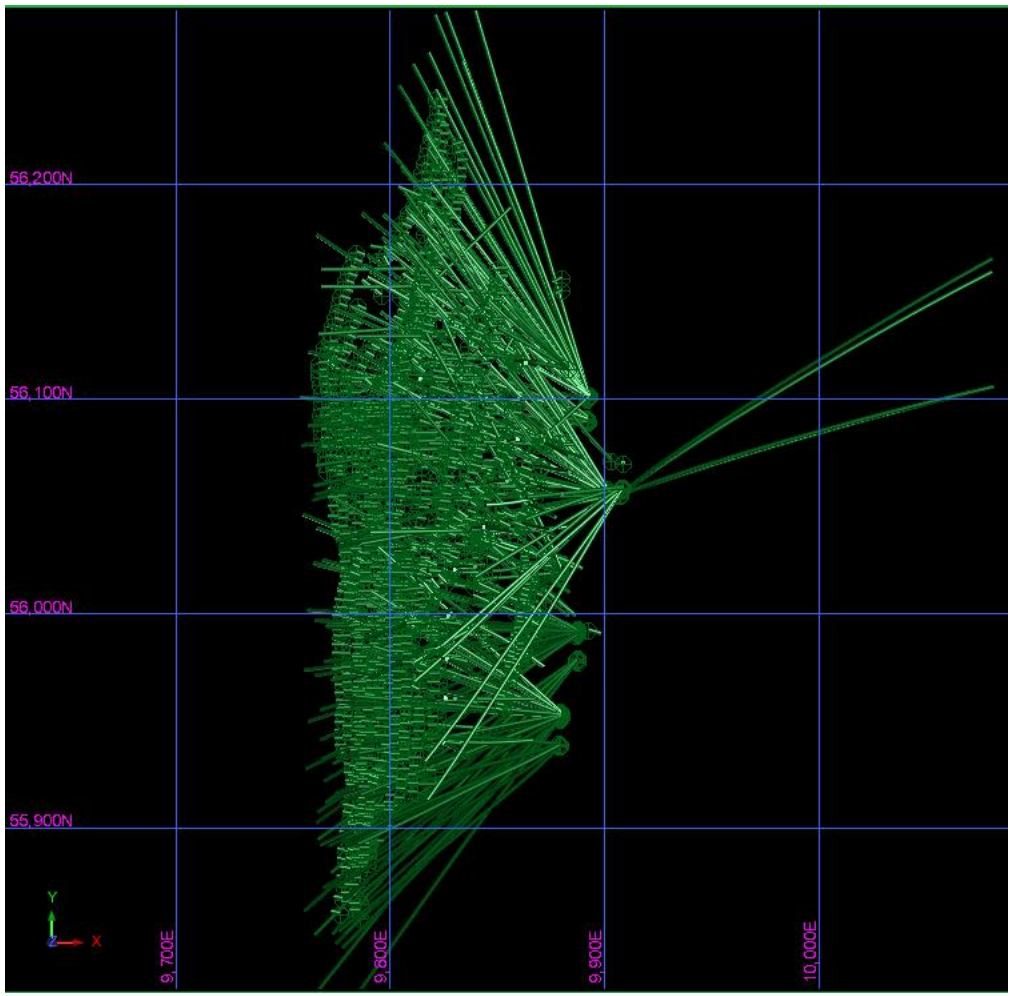
Figure 13:Jaguar Mine Location Map


55,800N
56,000N
56,200N
56,400N
4,200Z
4,000Z
3,800Z
Z
Y

(a)

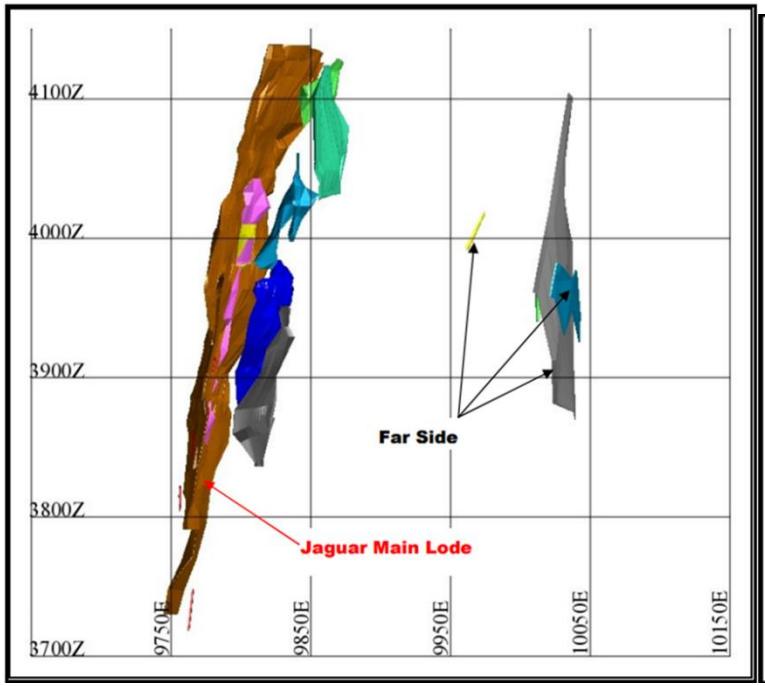

9,800E
10,000E
10,200E
4,200Z
4,000Z
3,600Z
Z
X

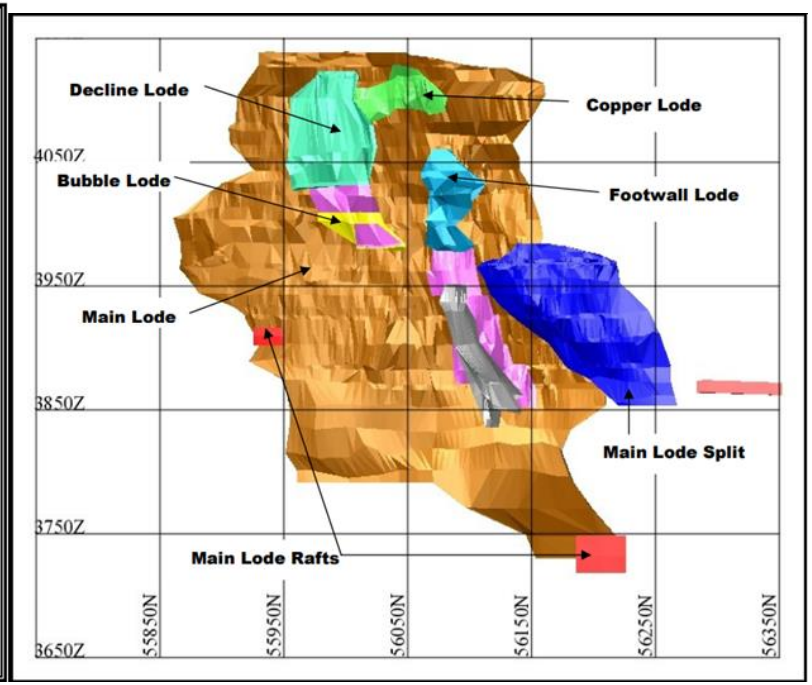
(b)



(c)

Figure 14: Raw Drill hole views showing (a) Z-Y (b) Z-X and (c) Y-X directions


4100Z
4000Z
3900Z
3800Z
3700Z
9750E
9850E
9950E
10050E
10150E
Far Side
Jaguar Main Lode


Decline Lode
Copper Lode
4050Z
Bubble Lode
Footwall Lode
3950Z
Main Lode
3850Z
Main Lode Split
3750Z
Main Lode Rafts
3650Z
55850N
55950N
56050N
56150N
56250N
56350N

4.1 Data description and Methods

The following steps were taken to construct the proposed ANN model for the prediction of metal ore grade as indicated by the flowchart in Figure 15.

Data preparation: Grade estimation is based primarily on geological attributes, spatial information, and drilling variables. 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 in Surpac software 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 or gold 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. Eastings, northings, and altitude indicate the location at which the sample wa14s 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 for each sample is displayed in Table 1.

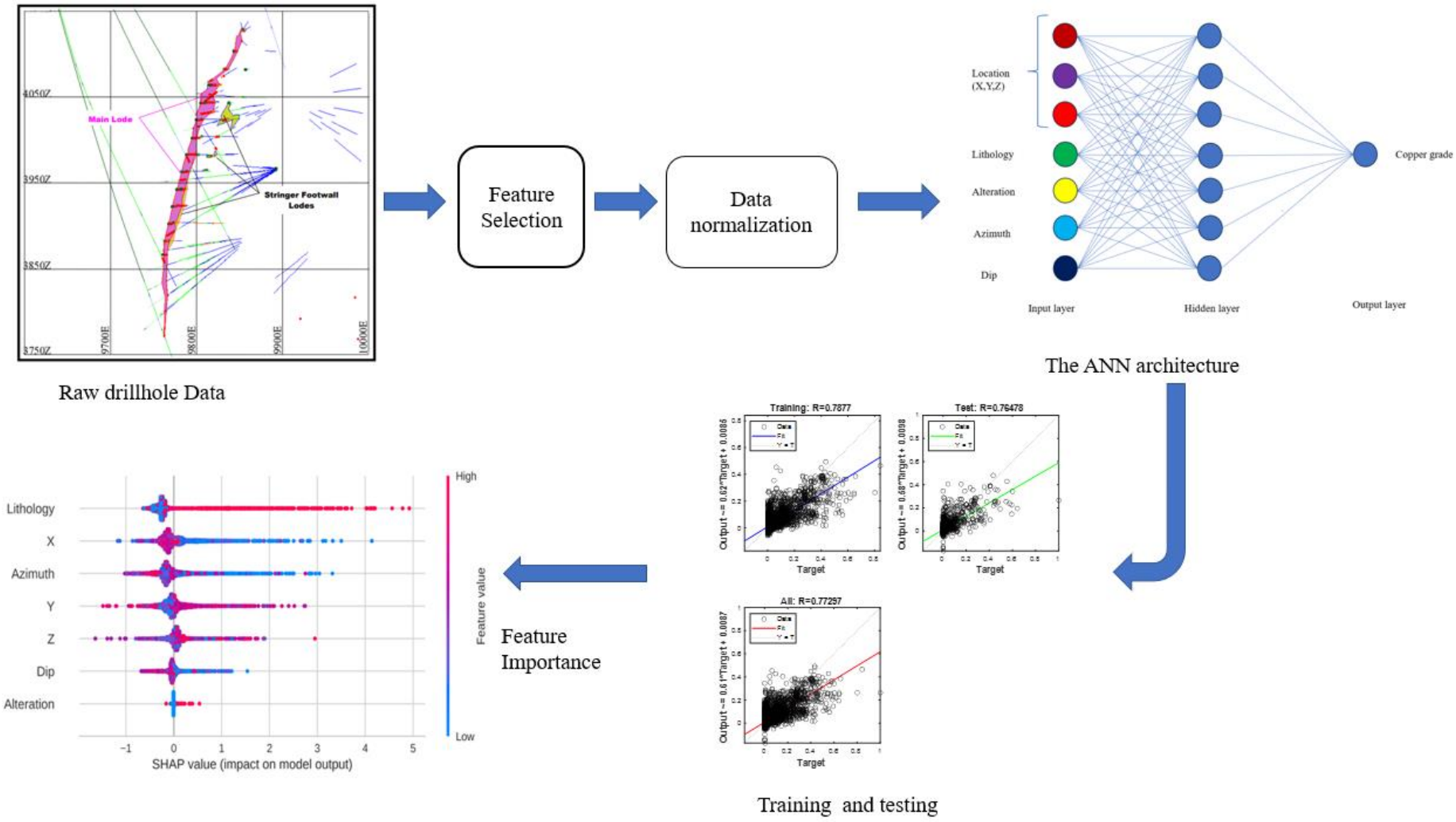


Figure 15: Flowchart of the overall proposed approach for copper ore grade estimation

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

Hole ID	X	Y	Z	Dip	Azimuth	Lithology	Alteration	Cu Grade (%)
08JUDD001	9879.20	55954.33	4056.84	312	356	Basalt	Chloritization	0.0230
08JUDD002	9878.16	55954.46	4057.86	304	357	Dolerite	Sericitization	0.5896
↓	↓	↓	↓	↓	↓	↓	↓	↓
08JUDD184	9892.56	56101.79	3965.46	321	330	Andesite	Chloritization	0.0090
08JUDD185	9806.73	56245.590	3879.88	329	333	Massive sulfides	Chloritization	0.0330

Table 2:. Descriptive statistics of the dataset.

	X	Y	Z	Dip	Azimuth	Cu Grade (%)
Count	14 294	14 294	14294	14 294	14294	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

Table 3: Raw drillhole data

hole_id	depth_from	depth_to	lithology	alteration_type	au_ppm	cu_pct
08JUDD001	0	1	basalt	chloritization	0.000	0.023
08JUDD001	1	2	basalt	chloritization	0.000	0.013
08JUDD001	2	3	basalt	chloritization	0.000	0.006
08JUDD001	3	4	basalt	chloritization	0.000	0.007
08JUDD001	4	5	basalt	chloritization	0.000	0.007
08JUDD001	5	6	basalt	chloritization	0.000	0.015
08JUDD001	6	7	basalt	chloritization	0.000	0.014
08JUDD001	7	8	basalt	chloritization	0.000	0.009
08JUDD001	8	9	basalt	chloritization	0.000	0.002
08JUDD001	9	10	basalt	chloritization	0.000	0.010
08JUDD001	10	11	basalt	chloritization	0.000	0.012
08JUDD001	11	12	basalt	chloritization	0.000	0.013
08JUDD001	12	13	basalt	chloritization	0.000	0.022
08JUDD001	13	14	basalt	chloritization	0.000	0.015
08JUDD001	14	15	dolerite	chloritization	0.000	0.011
08JUDD001	15	16	dolerite	chloritization	0.000	0.010
08JUDD001	16	17	dolerite	chloritization	0.000	0.011
08JUDD001	17	18	dolerite	chloritization	0.000	0.014
08JUDD001	18	19	dolerite	chloritization	0.000	0.010
08JUDD001	19	20	dolerite	chloritization	0.000	0.010
08JUDD001	20	21	dolerite	chloritization	0.000	0.010
08JUDD001	21	22	dolerite	chloritization	0.000	0.009
08JUDD001	22	23	basalt	chloritization	0.000	0.075
08JUDD001	23	24	basalt	chloritization	0.000	0.275
08JUDD001	24	25	basalt	chloritization	0.000	0.052
08JUDD001	25	26	basalt	chloritization	0.000	0.185
08JUDD001	26	27	basalt	chloritization	0.000	0.302
08JUDD001	27	28	basalt	chloritization	0.000	0.056
08JUDD001	28	29	basalt	chloritization	0.000	2.310
08JUDD001	29	30	basalt	chloritization	0.000	1.410
08JUDD001	30	31	basalt	chloritization	0.000	0.081
08JUDD001	31	32	basalt	chloritization	0.000	0.518
08JUDD001	32	33	basalt	chloritization	0.000	1.540
08JUDD001	33	34	basalt	chloritization	0.000	0.082
08JUDD001	34	35	basalt	chloritization	0.000	0.235
08JUDD001	35	36	basalt	chloritization	0.000	0.127
08JUDD001	36	37	basalt	chloritization	0.000	0.189
08JUDD001	37	38	basalt	chloritization	0.000	0.000
08JUDD001	38	39	basalt	chloritization	0.000	0.031
08JUDD001	39	40	basalt	chloritization	0.000	0.419
08JUDD001	40	41	basalt	chloritization	0.000	0.205
08JUDD001	41	42	basalt	chloritization	0.000	0.004
08JUDD001	42	43	basalt	chloritization	0.000	0.004
08JUDD001	43	44	basalt	chloritization	0.000	0.004
08JUDD001	44	45	basalt	chloritization	0.000	0.002
08JUDD001	45	46	basalt	chloritization	0.000	0.004
08JUDD001	46	47	basalt	chloritization	0.000	0.008
08JUDD001	47	48	basalt	chloritization	0.000	0.006
08JUDD001	48	49	basalt	chloritization	0.000	0.018
08JUDD001	49	50	basalt	chloritization	0.000	0.011
08JUDD001	50	51	basalt	chloritization	0.000	0.009
08JUDD001	51	52	basalt	chloritization	0.000	0.010
08JUDD001	52	53	dolerite	chloritization	0.000	0.010
08JUDD001	53	54	dolerite	chloritization	0.000	0.015
08JUDD001	54	55	dolerite	chloritization	0.000	0.011
08JUDD001	55	56	dolerite	chloritization	0.000	0.009
08JUDD001	56	57	dolerite	chloritization	0.000	0.012
08JUDD001	57	58	dolerite	chloritization	0.000	0.013
08JUDD001	58	59	dolerite	chloritization	0.000	0.013

Data Preprocessing: The necessary data preprocessing steps such as data cleaning, handling missing values, and feature scaling (normalization or standardization) were performed to ensure the data is in a suitable format for training the ANN. ANN method performs numerical calculations, it can only work with numbers, categorical data was transformed into numerical values. 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 4 shows the formula for data normalization. The numerical variables were normalized by using the mean and standard deviation as shown in Equation 4 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.

$$z = \frac{x-\mu}{\sigma} \tag{6}$$

Table 3 shows the raw drillhole data for Hole ID JUDD001 for the depth of 60 m. The data was transformed to values between 0 and 1. For example, the maximum value (360 degrees) of the dip/azimuth value of drillholes was assigned 1 and the minimum value was assigned zero. All the other dip/azimuth of values were divided by 360 degrees in order for them to fall within the range of 0 to 1.

$$Dip/azimuth = \frac{a_i}{360^o} \tag{7}$$

Where a_i is the dip or azimuth in the clockwise direction.

Splitting the dataset: 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.

ANN Architecture Design: The architecture ANN was determined, including the number of layers, number of neurons per layer, and activation functions. The proposed ANN architecture comprises one input layer consisting of seven neurons, one hidden layer, and one output layer.

Model Training: The ANN was trained using the training set. This involved feeding the input features through the network, noting the output, comparing it with the actual ore grade, and updating the network weights using backpropagation. This process is repeated iteratively for a specified number of epochs or until the model converges.

Model Evaluation: After training, the performance of the trained ANN was evaluated using the testing set of a single drillhole. The relevant evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), coefficient of determination (R^2), correlation coefficient (R) and Mean Absolute Error (MAE) were computed to assess the accuracy of the predictions.

Model Tuning: To improve the results, hyper parameter tuning was carried out. This process involved adjusting hyperparameters and retraining the model.

It is worth noting that the success of an ANN-based prediction model for ore grade relies heavily on the quality and representativeness of the dataset. Gathering accurate and diverse data is crucial for training a robust and reliable model. Once the ore grade was predicted, Shapley values were used to determine the feature importance of the input parameter on the model output.

Copper Grade Distribution

Figure 16 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.

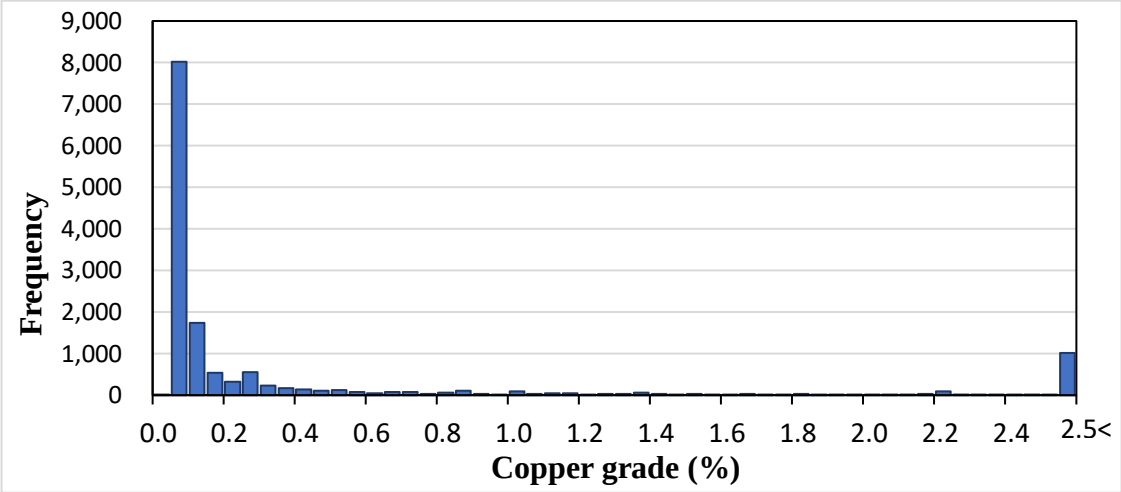


Figure 16:The histogram of copper ore grade distribution

4.2 Results and Discussion

4.2.1 The proposed Artificial Neural Network analysis for Copper Grade Prediction

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.

The dataset used in this study consists of 185 drillholes. This indicates that the study has a substantial amount of data to work with, which can be both advantageous and challenging. The primary challenge associated with working with a large dataset of drillhole samples, is the significant variations in the samples and erratic distribution of geochemical anomalies. This variability could stem from drilling methods and geological factors. Such variations can make it challenging to create a predictive model that accurately represents the underlying patterns in the data. Additionally, certain elements or compounds may not be evenly distributed throughout the sample, and this could problematic and may result in biases and inaccuracies in the predictive model. Thus, careful selection of data partitioning procedure is very crucial as it can significantly impact the accuracy of the predictive model. It is vital to

ensure that the data partitions are representative of the overall dataset. Individual samples were modeled along the z-axis based on core sample interval of 1m. This indicates a specific approach to handling the three-dimensional nature of drillhole data, which can be crucial in geological and geochemical studies.

A hold-out method was used to split data into two sets: training and testing. The 14,179 dataset was used for training and a single set of drillhole with 115 dataset was used for testing. The performance of the model across drillholes was validated by using an independent and unused testing dataset. This is crucial to ensure that the performance of the model is not over-optimized for the training data and can make accurate predictions on new, unseen data.

Figure 17 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 coefficient, R for the training data is 0.788, indicating a relatively strong positive correlation between the predicted and actual values in the training dataset. The R for the testing data is 0.765, which suggests a reasonably strong positive correlation between predicted and actual values in the testing dataset whereas the R for the overall model, considering both training and testing data, is 0.773, indicating that the model's performance is consistent across both datasets.

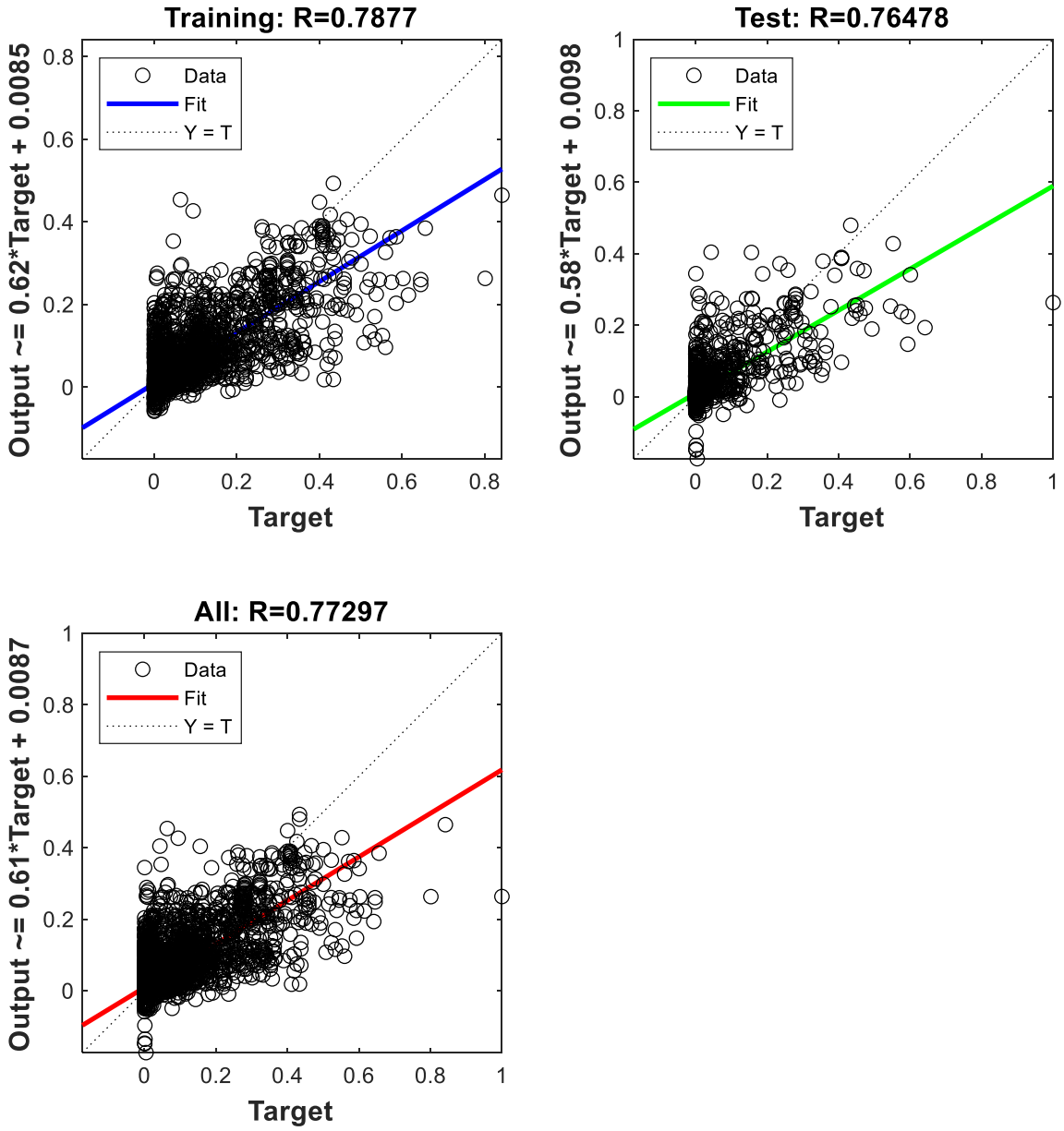


Figure 17: ANN regression models for copper ore grade showing actual and predicted data distribution with the blue, green, and red showing best line fit for training, test, and the overall model data, respectively. The white circles represent the dataset.

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 17, 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 input layers, one hidden layer, and one output layer. A correlation coefficient of around 0.765 suggests that the model has some ability to capture the relationship between the variables, but it is important to keep in mind that correlation does not imply causation, and the quality of predictions may depend on other factors as well. Further analysis, such as assessing the model's mean squared error or other relevant metrics, can provide a more comprehensive evaluation of the model's performance. Additionally, external validation on completely unseen data is essential to ensure the model's generalizability to new drillhole data.

Figure 18 displays the learning performance of the model. It typically includes a plot of the MSE over different iterations or epochs. MSE is a common metric used to measure the quality of a regression model's predictions, with lower values indicating better performance. It can be noted that the best performance for both the training and test data was achieved at epoch 1000 steps of iteration. This suggests that the model achieved its lowest MSE on both the training and test datasets after 1000 iterations. At epoch 1000, the corresponding MSE and gradient values were 0.0016 and 0. 00066.

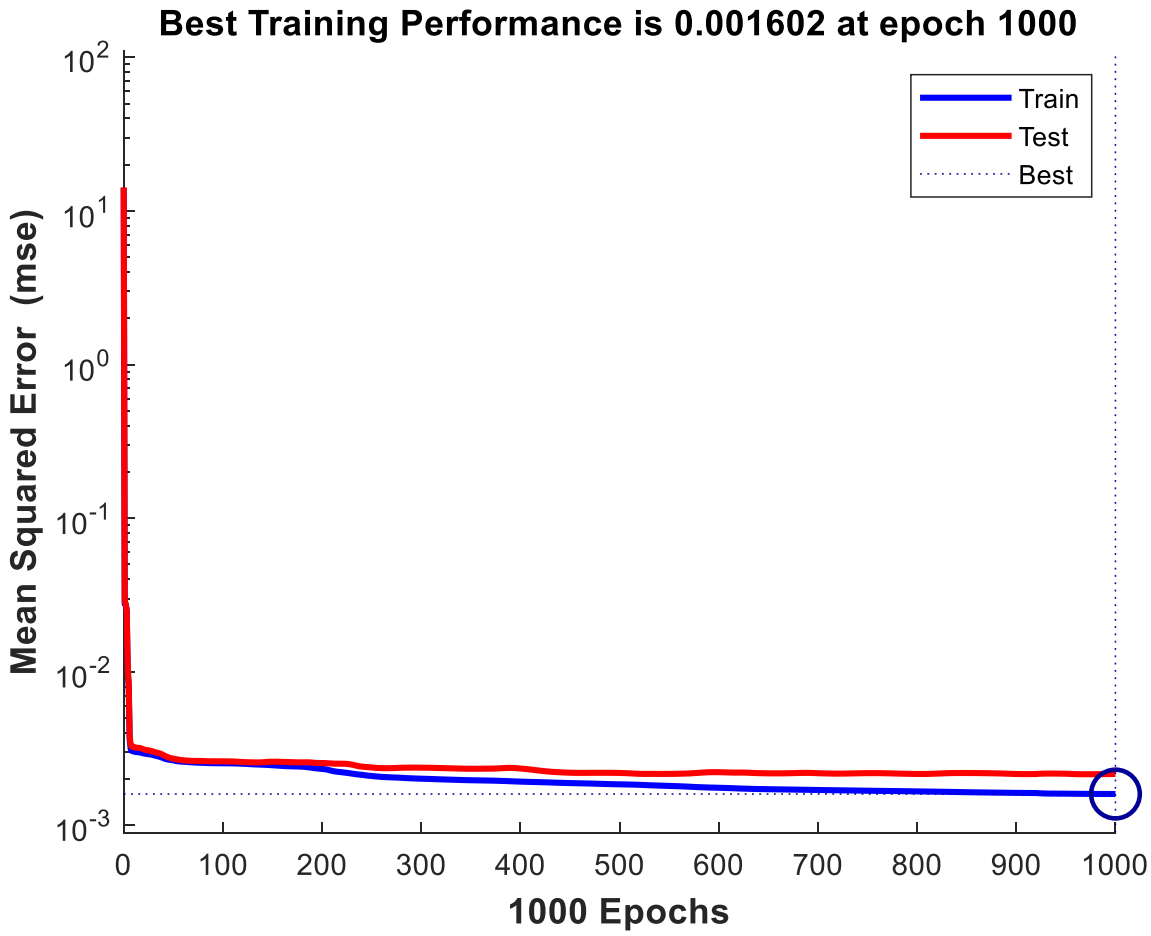


Figure 18:Actual versus predicted values of the ANN model with the red and blue lines showing the predicted and actual values, correspondingly.

The findings suggest that the ANN model achieved its best performance after 1000 epochs, with an MSE of 0.0016, MAE of 0.018, and RMSE of 0.041. A lower MAE indicates better accuracy while RMSE penalizes larger errors more heavily than MAE, and a lower RMSE also indicates better accuracy. These metrics indicate that the model's predictions are relatively close to the actual values, and it has learned to represent the underlying patterns in the data reasonably well.

An error histogram was generated to visualize the distribution of errors for both the training and testing datasets. This histogram helps understand the spread and concentration of errors in the model's predictions. Figure 19 shows the prediction error distribution. The orange line, representing zero error, suggests that the model's errors are primarily concentrated in the range of ±0.08. This means that the majority of the ANN model's predictions are within this range of deviation from the actual values.

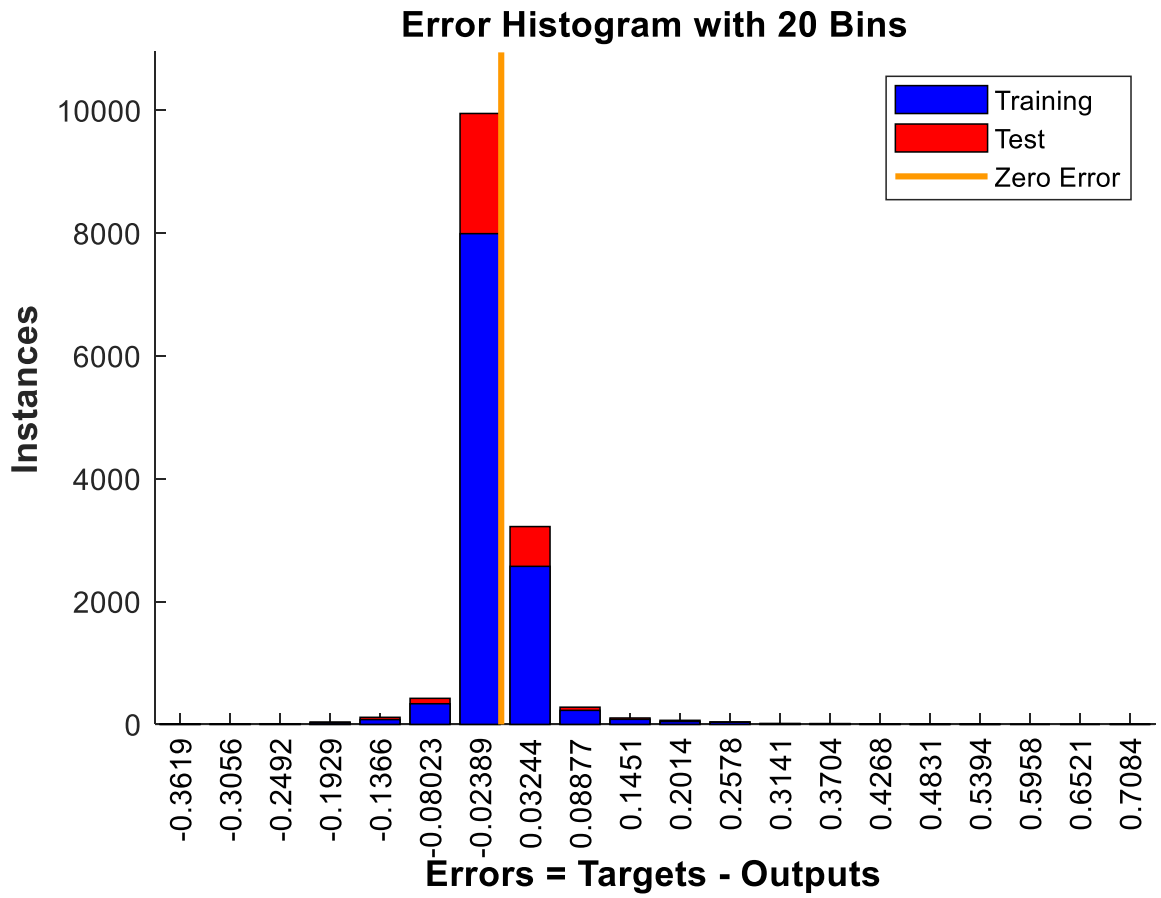


Figure 19: Error Histogram for copper ore grade with orange line showing the concentration of the errors.

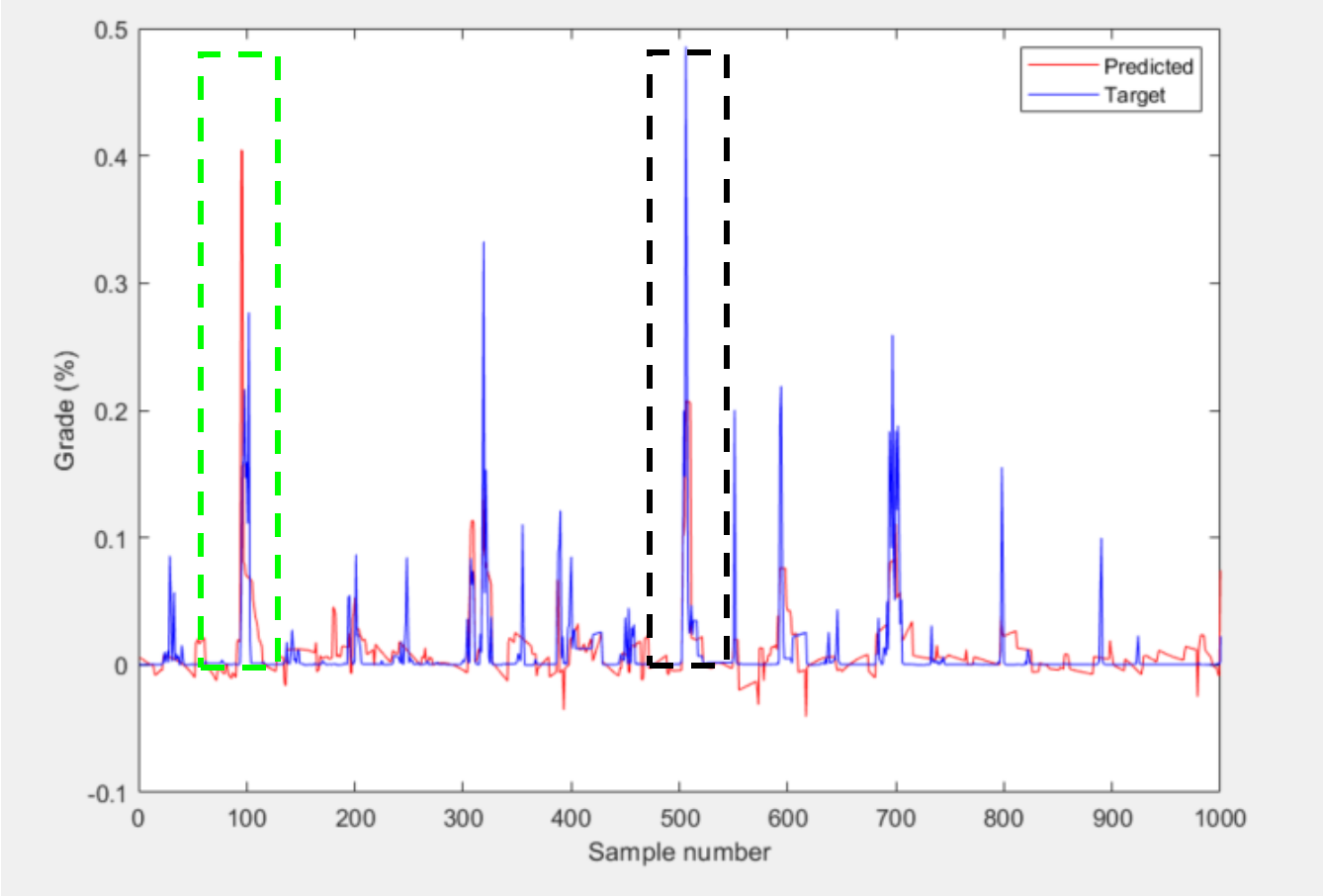


Figure 20:Actual versus predicted values of the ANN model with the red and blue lines showing the predicted and actual values, correspondingly.

Figure 20 displays the data distribution of the actual versus predicted grades of the model. This figure provides a visual representation of how well the ANN model's predictions align with the actual data. The model overestimated the grade on sample 100 and underestimated the copper ore grade on sample 500. It can be noted that the copper grade can be moderately estimated by the proposed ANN model. This suggests that the model captures the underlying

patterns in the data to provide reasonably accurate predictions. The overall model results showed minimum errors, indicating a high degree of correlation between the input and output variables. The error analysis, visualization of error distribution, and data distribution between predicted and actual grades all indicate that the ANN model is accurate, powerful and reliable. These findings support the model's applicability in the mining industry for predicting copper grades, which can have important implications for optimizing mining operations and resource management.

4.2.1 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.

Hyperparameter Choice

Hyperparameter optimization is a crucial step in the development of machine learning models, including artificial neural networks (ANNs). It plays a significant role in improving a model's

predictive performance and ensuring its robustness and credibility. The choice of hyperparameters can significantly impact a model's performance. By tuning these parameters, one can often achieve better accuracy and precision. Hyperparameter tuning helps prevent overfitting, a common problem in machine learning. Overfitting occurs when a model learns to perform exceptionally well on the training data but fails to generalize to unseen data. Proper hyperparameter settings can help strike the right balance between model complexity and generalization. For this study, Bayesian optimization was chosen to yield better results.

Additionally, hyperparameter optimization techniques, such as grid search or randomized search, can be employed to fine-tune the model's parameters and improve its performance.

Overfitting and Cross-Validation

Overfitting is a common challenge in machine learning, where a model performs well on the training data but fails to generalize to new, unseen data. To mitigate overfitting, techniques like cross-validation are used to assess the model's performance on multiple subsets of the data. Cross-validation helps estimate how well the model will perform on unseen data, enabling more robust predictions. Ensemble methods combine multiple models to make more accurate predictions. Techniques like bagging (e.g., random forests) or boosting (e.g., gradient boosting machines) can be employed to aggregate the predictions of multiple models, reducing bias and variance and improving overall performance.

By leveraging these theoretical concepts, machine learning algorithms can analyze large datasets, identify patterns and relationships, and make accurate predictions for ore grade

estimation. The utilization of these techniques helps optimize mining operations, minimize costs, and maximize resource utilization in the mining industry.

To evaluate the prediction performance of the classic methods, hyperparameter tuning was performed to optimize the classic machine learning models, making them more robust and credible in predictive modeling. Hyperparameter optimization is a crucial step in machine learning model development as it helps find the best settings for the model. Table 4 summarizes how the selected classic machine learning methods were optimized. It includes information about the specific hyperparameters that were tuned for each model.

Table 4.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 5 shows the results of the statistical methodologies used to predict the copper ore grade. The results show that the ANN surpasses the classic machine learning methods in terms of, R^2, R, MAE, MSE, and RMSE, with values of, 0.584, 0.765, 0.018, 0.0016, and 0041, respectively. The coefficients of determination 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. R^2 values indicate the quality of the predictions made by these models. Higher R^2 values are generally desirable as they indicate a better fit to the data. In this case, the ANN model outperformed these classic methods because it had a higher R^2 value of 0.584.

Table 5. 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

Linear regression showed poor performance with the lowest R^2 value of 0.123. This poor performance is attributed to the fact that linear regression assumes a linear relationship

between the independent and dependent variables, which may not be appropriate for modeling the variation in ore grade, as it can have nonlinear patterns. The ore grade is a varying component, implying that it may exhibit nonlinear relationships with the input variables. This is why linear regression, which assumes linearity, may not perform well in predicting grades.

It is of paramount importance to report the standard deviation (SD) for a fair statistical comparison of the models. A lower SD indicates that the data points are closer to the mean, which is considered ideal. The proposed artificial neural network model is presented in Table 4 as the best-performing model with the lowest SD of 0.041 compared to the basic machine learning methods.

Figure 21 displays prediction error plots for classic machine learning approaches using R^2 evaluation metrics. These plots compare the actual values of copper grades with the predicted values generated by different models. R^2 measures how well the model explains the variance in the data. It is worth noting that despite having high correlation coefficients, some machine learning models (random forest regressor, extra tree regressor, light gradient boosting machines, and K neighbor regressor) have significant errors in their predictions. This suggests that these models may not be accurate in capturing the underlying patterns in the data. 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.

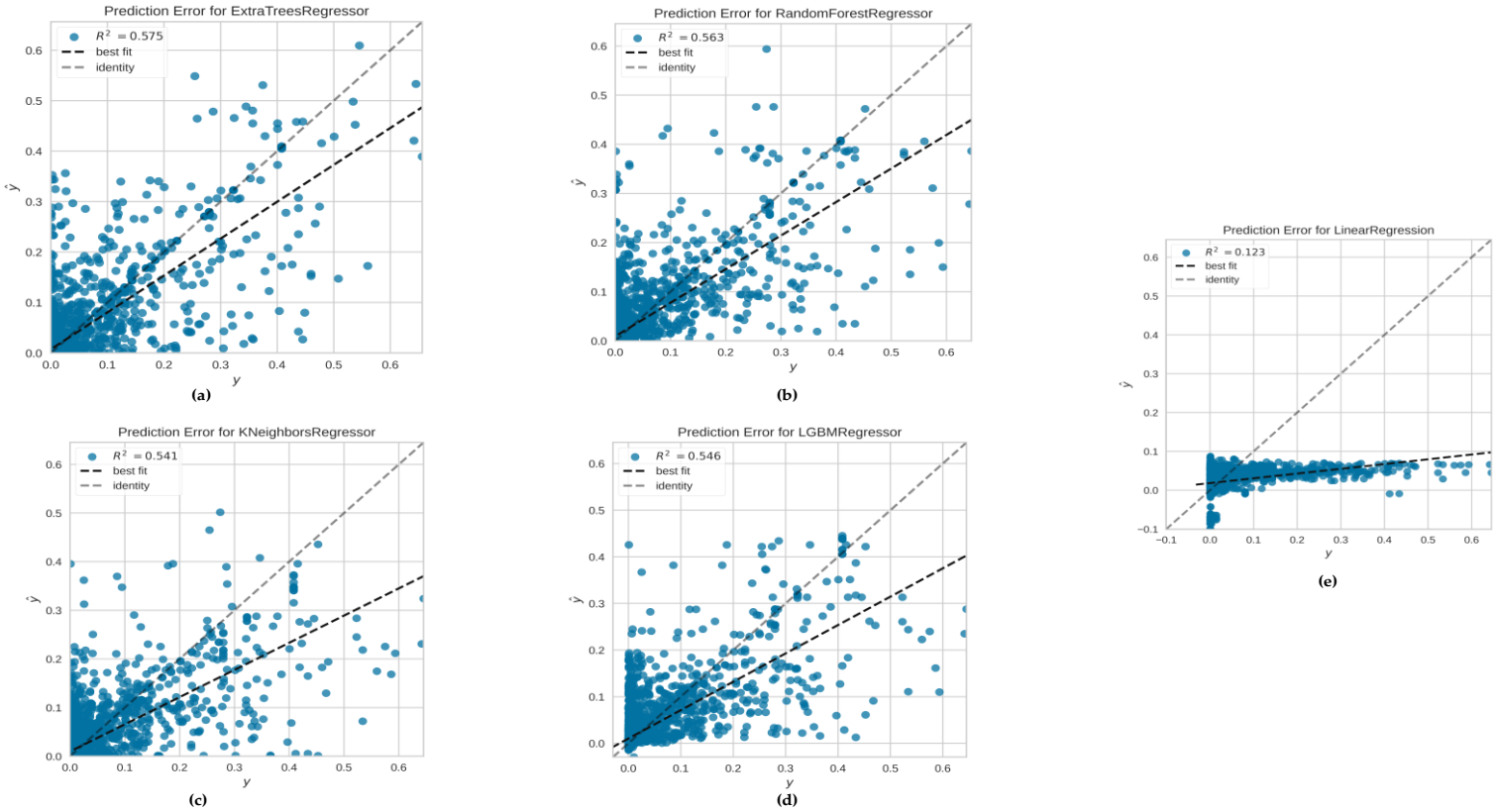


Figure 21:The prediction error plots for the classic machine learning techniques, with y and y^i showing the data distribution of actual and predicted values. (a) Prediction error plot for extra trees regressor; (b) Prediction error plot for random forest regressor; (c) Prediction error plot for K neighbors regressor; (d) Prediction error plot for light gradient boosting machine; (e) Prediction error plot for linear regression

The ANN model has 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) perform less accurately, with higher MAE values of 0.332, 0.369, 0.415, and 0.821, respectively. 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 indicating its potential as a more reliable and powerful method for predicting copper grade than the classic machine learning methods.

4.2.2 Feature Importance Analysis

The correlation matrix provides valuable insights into how variables in the dataset are related. A correlation matrix is a valuable tool in statistics and data analysis that helps understand the relationships between different variables. Figure 22 depicts the correlation matrix based on correlation coefficient. The correlation coefficient quantifies the strength and direction of the linear relationship between two variables. It can range from -1 to 1, where -1 indicates a perfect negative linear correlation, 0 indicates no linear correlation, and 1 indicates a perfect positive linear correlation. In Figure 22, color variation is used to visually represent the strength and direction of correlations. The dark blue represents a strong negative correlation, while dark red indicates a strong positive correlation. This color coding makes it easier to identify relationships quickly.

Lithology appears to have a significant correlation with other variables. It correlates positively with altitude and alteration but negatively with eastings, northings, altitude, azimuth, and dip. This information suggests that as lithology increases, altitude and alteration tend to increase as well, while eastings, northings, azimuth, and dip tend to decrease. Eastings show positive correlations with azimuth, northings, dip, and alteration but a negative correlation with lithology. The dip correlates positively with azimuth and eastings but shows negative correlation with alteration, lithology, altitude and northings, on the other hand the azimuth shows positive correlations with northings, altitude, eastings, dip and alteration but a negative correlation with lithology.

The correlation matrix gives valuable details about how variables in the dataset are related. Positive correlations indicate that as one variable increases, the other tends to increase as well, while negative correlations suggest that as one variable increases, the other tends to decrease Understanding these relationships can be crucial for feature selection, data exploration, and modeling in various data analysis tasks.

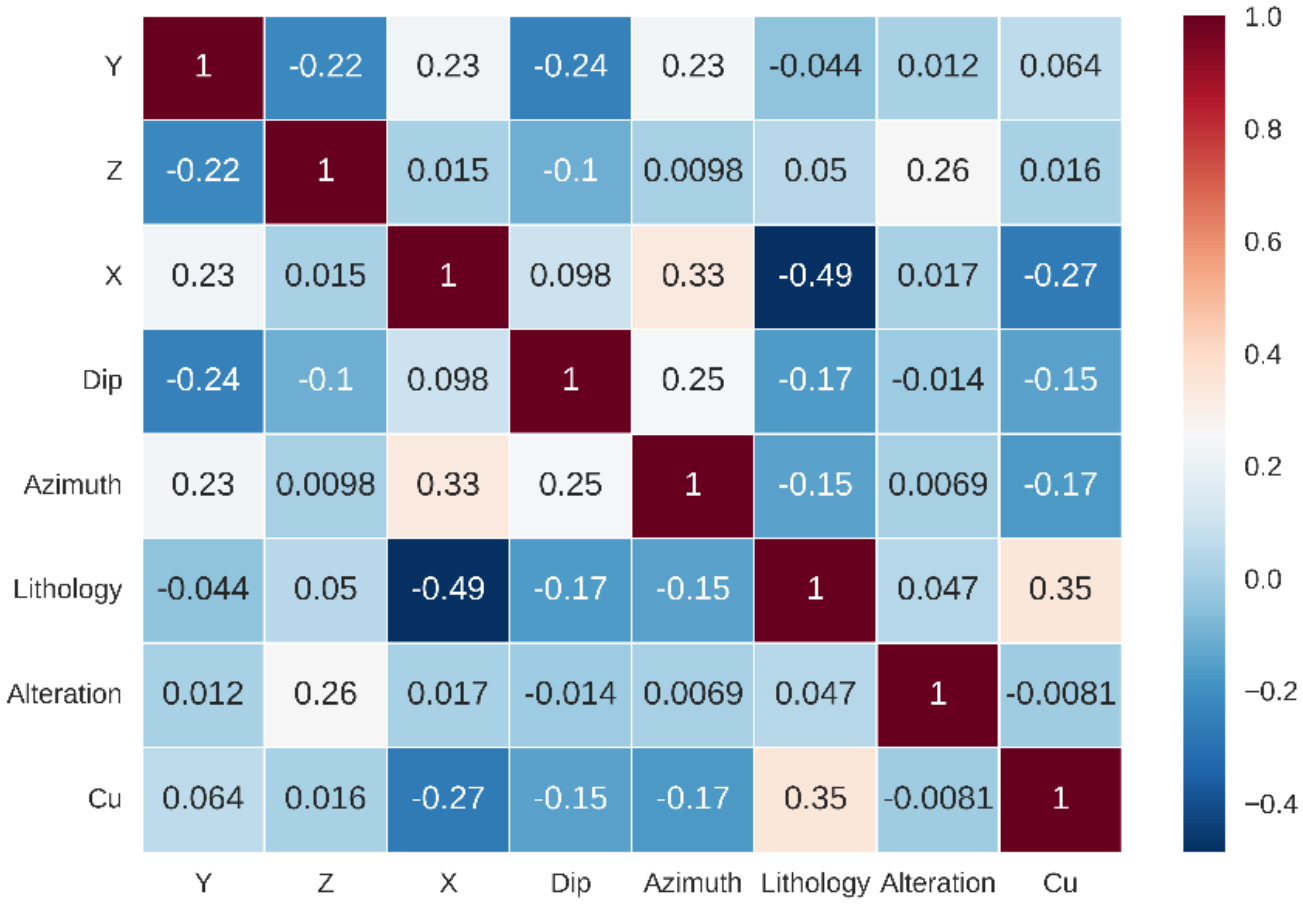


Figure 22: Correlation matrix based on predictors and output variable with dark red showing strong positive correlation and blue showing strong negative correlation between variables.

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 (Lundberg and Lee.,2017). 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. Equation shows the formula used to calculate the SHAP values.

$$\phi_i(\mathrm{f}) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!|N| - |S| - 1)!}{|N|!} \quad [f(S \cup \{i\}) - f(S)$$

Where:

$\phi i(f)$ represents the Shapley value for feature *i*.

N is the set of all features.

S represents a subset of features that does not include feature *i*.

$f(S)$ is the model's output when using the features in subset *S*.

$|S|$ represents the number of features in subset *S*.

$|N|$ represents the total number of features.

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 23 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 because it is linked to the geochemical formation and mineralization of the deposit. Altitude was the second most influential input variable. 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.

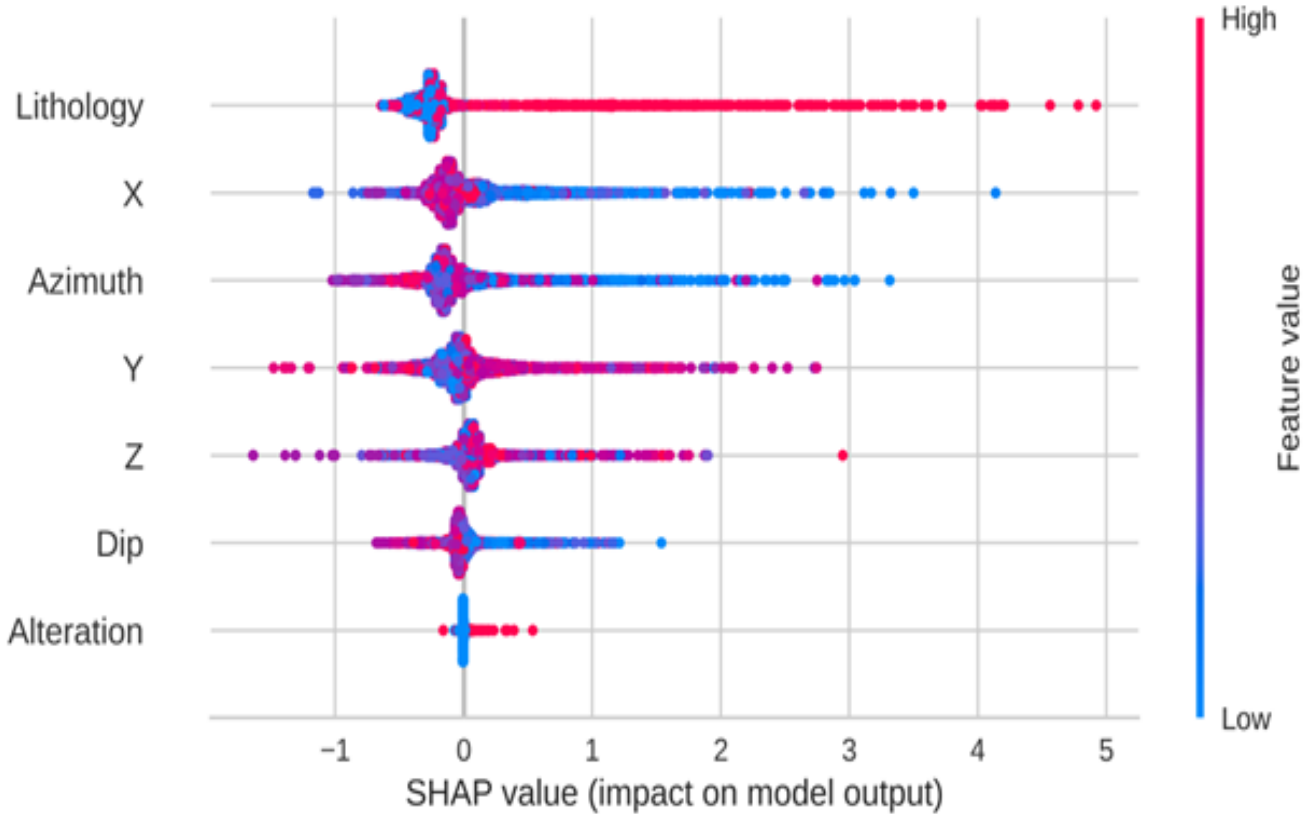


Figure 23:Feature importance based on the SHAP values. Red represents higher feature values and blue represents lower feature values.

4.2.3 Conclusion

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. The proposed technique is data-driven and learns the relationship between the

input and output values to predict the grade. The performance metrics, R2, 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 to predict the ore grade has potential benefits in the mining industry as can be used to supplement exploration activities, thereby reducing drilling requirements. It can also aid in mine planning and design and mineral resource evaluation potentially resulting in extensive cost savings.

Although the ANN model moderately 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. Future research and improvements can address some of the identified limitations and further enhance the performance and applicability of the model. In the future directions, 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.

CHAPTER 5: INTERGRATION OF FEATURE SELECTION TECHNIQUE INTO THE APPLICATION OF ARTIFICIAL NEURAL NETWORK ON THE PREDICTION OF GOLD ORE GRADE

This chapter shifts the attention to gold prediction using ANN. It begins with the introduction of the existing literature to predict mineral ore grades. The second section of this chapter focuses on prediction of gold ore grade without feature selection while the third section takes feature selection into consideration when predicting gold ore grade. This section emphasizes the critical relevance of feature selection, and this study conducted a comparative analysis of the performance of ANN models, both with and without feature selection on gold grades .

Introduction

Gold without Feature selection

Gold With Feature selection technique

5.1 Introduction

Machine learning algorithms (MLAs) have gained prominence in spatial data prediction in recent years. Contrary to previous approaches, MLAs may simulate complicated non-linear correlations between mineral occurrences and evidentiary aspects (Brown et al., 2000; Ewusi et al.,2021). Nonetheless, increasing the number of input features may result in greater complexity and over-fitting (the curse of dimensionality) (Rodriguez-Galiano et al., 2012; Bellman, 1966). Numerous studies in the literature demonstrated that MLAs such as Random Forest (RF), Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) are efficient and robust for accurate mineral grade estimation. Carranza et al.,2016 carried out research that demonstrated Random Forest as a viable technique for data-driven predictive modeling of mineral prospectivity. They concluded that RF modeling outperforms evidentiary belief (EB) modeling of prospectivity for hydrothermal Au-Cu deposits in Catanduanes Island, which has 17 recognized hydrothermal Au-Cu prospects. Furthermore, RF modeling, like EB modeling, enables the investigation of spatial correlations between known prospects and individual layers of predictor data. Some studies have successfully employed k nearest neighbors and combined kNN and ANN (Kaplan and Topal, 2021). Several studies (Samantha et al.2005; Chatterjee et al.,2006; Li et al.,2010; Mlynarczuk et al.,2017; Singh et al.;2018) have applied ANN in mineral resource estimation. Tsae et al.,2023 investigated the application of ANN in the prediction of copper ore grade by incorporation seven input variables and they concluded that ANN outperformed all the other machine learning algorithms based on five evaluation metrics, correlation coefficient of

determination, correlation coefficient, mean absolute error, mean squared error and root mean squared error).

Although the existing literature demonstrate the potential of machine learning algorithms in accurately estimating various mineral ore grades, these studies do not highlight the significance of feature selection techniques in accurate mineral grade prediction. Feature selection techniques are essential in machine learning and data analysis to identify the most relevant features or variables that contribute to the predictive power of a model. These techniques aim to improve model performance, reduce overfitting, enhance interpretability, and reduce computational complexity. By removing irrelevant or noisy features, models become less prone to overfitting and can generalize better to new, unseen data (Aman,2023). An additional benefit of feature selection is the reduced computational time; fewer features mean faster model training and inference, which can be crucial in real-time or resource-constrained applications. Collecting and maintaining data on all possible features can be expensive and time-consuming, therefore feature selection can help in reducing data collection costs by focusing only on the most important variables.

Feature selection is a critical step in the prediction of mineral ore grade, as it can significantly impact the accuracy and efficiency of the predictive models. Mineral ore grade prediction is essential in the mining industry as it helps optimize the extraction process, reduce waste, and improve overall profitability. Before selecting features, it is crucial to have a deep understanding of the ore extraction process and the factors that influence ore grade to be able

to identify the most relevant features. To have a comprehensive dataset, a wide range of potential features such geological, mineralogy, and chemical should be considered.

To address this gap in the existing literature, this study aims to incorporate feature selection techniques into the proposed prediction approach of the gold ore grade. This study explores correlation feature based feature selection methods and feature importance to determine the most valuable features for accurate mineral grade estimation. Additionally, the study compares the performance of models with and without feature selection and this provides valuable insights into the actual benefits of these techniques in mineral grade prediction tasks. models using all features to ensure feature selection benefits the model.

Filter Methods

Filter methods are a category of feature selection techniques used in machine learning and data analysis to identify and select the most relevant features (variables or attributes) from a dataset. These methods operate independently of any specific machine learning algorithm and are typically applied before the actual model training process. Filter methods assess the intrinsic characteristics of each feature and rank or select them based on predefined criteria (Sánchez-Maroño.,2007). They are valuable tools for feature selection in machine learning, offering simplicity, efficiency, and interpretability. The most popular filter methods are correlation-based feature selection. This method calculates the correlation between each feature and the target variable (or among features) and selects the features with the highest correlation scores. Features that are highly correlated with the target variable are considered more important.

In this study, we focus on the prediction of gold ore grade by considering the feature selection techniques. The aims of this study are to:

1. Understand the relationship between the diverse set of input variables such as geological, spatial information and drilling parameters that are relevant for the accurate prediction of gold ore grade.
2. Demonstrate the performance of the feature selection approaches in analyzing the high dimensional gold grade.
3. To identify the optimal feature selection strategies for capturing relevant information and improving the accuracy of gold ore grade.

After completing these essential processes, the optimal input features will be employed to train the artificial neural network (ANN) model. The objective is to predict the grade of gold ore and evaluate the model's reliability by using various sets of selected features. The primary contribution of this study is the identification of key features that have the potential to enhance prediction accuracy and minimize prediction errors in artificial neural network used for evaluating the grade of gold ore.

5.2. Gold Without Feature Selection

Introduction

Gold without Feature selection

Gold With Feature selection technique

5.2.1 Data description and Method

A total dataset of 14 294 from 185 drillholes was collected at Jaguar mine. Samples were collected from drillholes at 1-metre intervals. This study investigated seven input variables namely, dip, azimuth, eastings, northings, altitude, and lithology while the primary output of interest in this study is the gold ore grade. The lithologies that had comparable characteristics were categorized into five groups using the same criteria as the copper grade (sediments, massive sulfides, dolerite, basalt, andesite). In this study, the four main forms of alteration,

sericitization, chloritization, silification, and carbonatization were considered. Figure 13 shows the flowchart of the ANN model for the estimation of gold ore grade. The necessary data preprocessing steps such as data cleaning, handling missing values, and feature scaling were performed to ensure the data is in a suitable format for training the ANN model. 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. *Z*-score normalization was computed. Table 6 shows the descriptive statistics of the dataset. 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 a single drillhole was excluded from the dataset and used as a test case. The architecture ANN was determined, including the number of layers, number of neurons per layer, and activation functions.

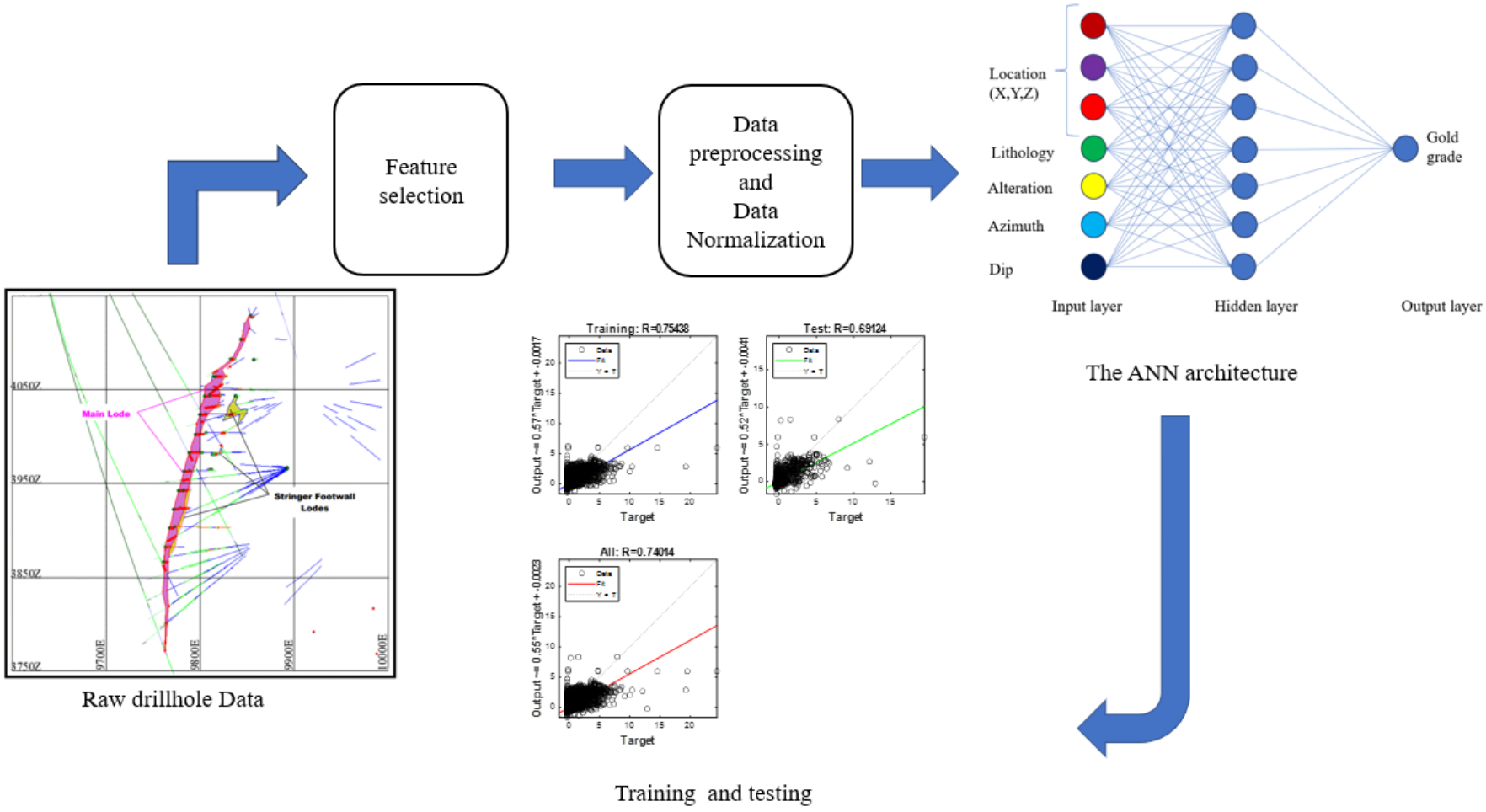


Figure 24:Flowchart of the proposed ANN model for gold grade without feature selection.

Table 6: Descriptive statistics of the dataset

	X	Y	Z	Dip	Azimuth	Au Grade (ppm)
Count	14 294	14 294	14,294	14 294	14,294	14 294
Mean	0.9960	0.9983	0.9840	0.8011	0.7012	20.57
Std	0.0015	0.0094	0.0040	0.3394	0.2219	51.89
Min	0.9925	0.9548	0.9765	0.0000	0.1321	0.0000
25%	0.9947	0.9818	0.9806	0.8969	0.6509	1.250
50%	0.9958	0.9914	0.9833	0.9411	0.7460	5.000
75%	0.9972	0.9961	0.9880	0.9691	0.8711	10.00
Max	1.0000	1.0000	1.0000	0.9556	1.0000	1287

5.2.2 Model built with all features

The dataset used in this study consists of 185 drillholes. This indicates that the study has a substantial amount of data to work with, which can be both advantageous and challenging. The significant sample variance and irregular distribution of geochemical anomalies present the greatest difficulty in working with a large dataset of drillhole samples. Therefore, choosing a data partitioning method carefully is essential since it might have a big impact on how accurate the prediction model is. It is important to ensure that the data partitions are accurate representations of the entire dataset. On the basis of a core sample interval of 1m, individual samples were modeled along the z-axis.

The data was split into two sets, namely the training set and the testing set, using the hold-out method. The training dataset consisted of 14179 datasets while the testing dataset consisted of 115 datasets. The adoption of the Bayesian regularization backpropagation algorithm was implemented to train the ANN model, with the objective of effectively managing uncertainties and improving the overall accuracy and precision. The training of the model involved the use of seven input variables (lithology, alteration, dip, azimuth, northings, eastings and altitude). The results suggest that the artificial neural network model demonstrates a moderate degree of precision in predicting the grade of gold. The ideal model design consists of a total of seven input nodes, one hidden layer, and a singular output node as indicated by Figure 25. The computational training time for the model was 12.05 minutes.

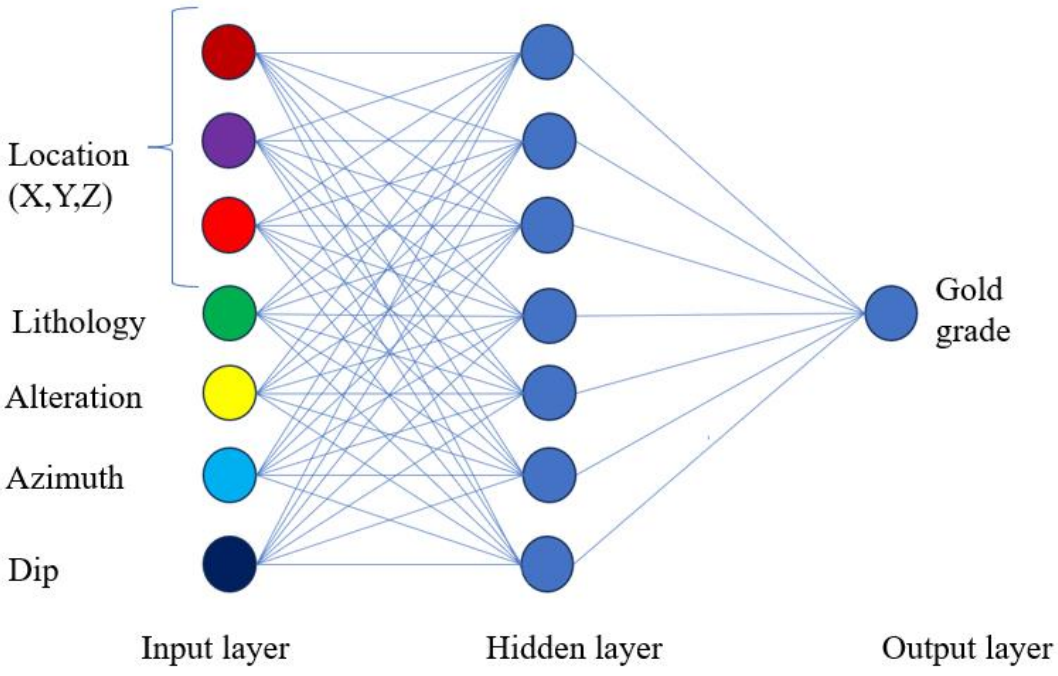


Figure 25: The ANN structure for Gold without feature selection

The correlation coefficient (R) measures the strength and direction of the linear relationship between the input variables and the output variables. Figure 26 shows the ANN regression model. The correlation coefficient, R, was calculated for the training, testing, and overall model. The values obtained for the variable R were 0.754, 0.691, and 0.740, respectively. This observation implies a strong positive correlation between the input variables and the output variables. The errors for the prediction of gold are high because there is a massive variation in gold ore grades. Figure 27 and 28 show the error histogram and the MSE plots correspondingly. Table 7 shows the summary of the results.

Table 7:Summary results for Gold without feature selection

	Results
Algorithm	Bayesian regularization
R for training	0.754
R for Test	0.691
MAE	0.293
MSE	0.502
RMSE	0.694

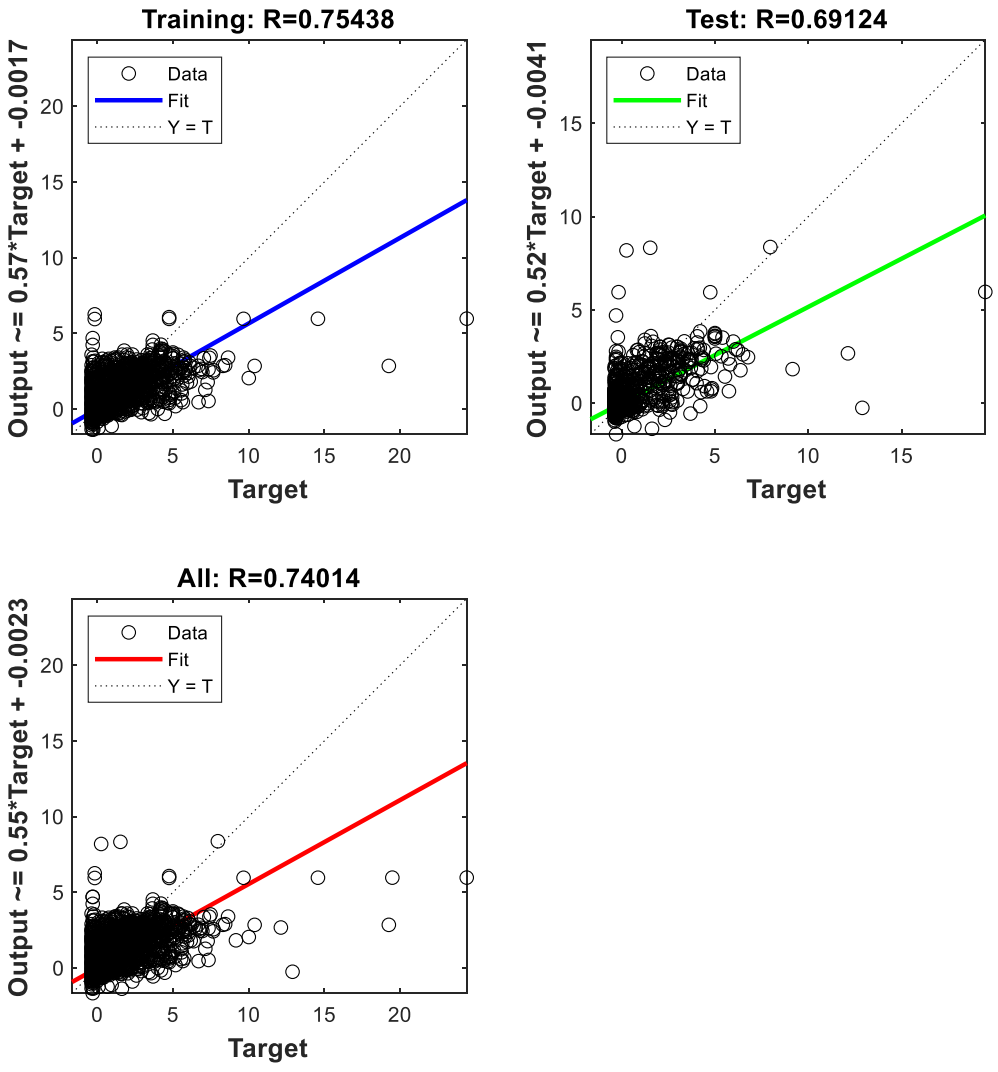


Figure 26:ANN regression models with all features showing actual and predicted data distribution with the blue, green, and red showing best line fit for training, test, and the overall model data, respectively. The white circles represent the data set.

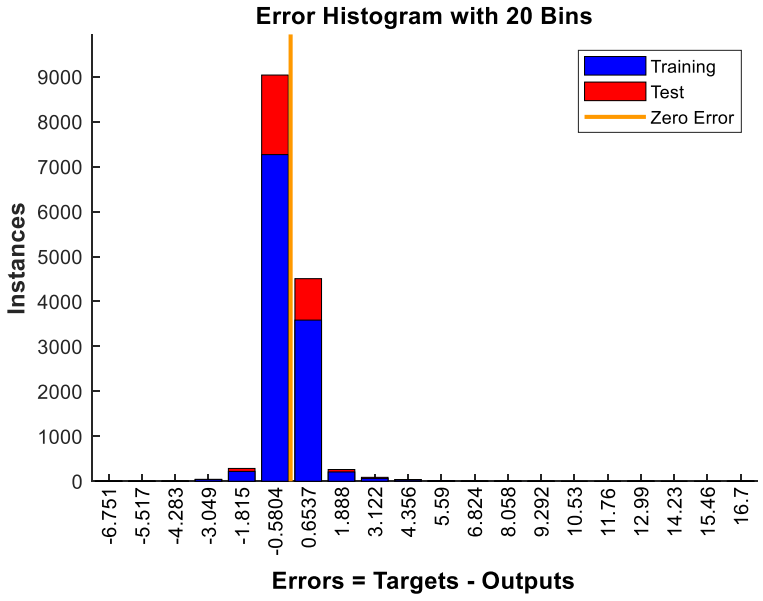


Figure 27: Error histogram for model with all features.

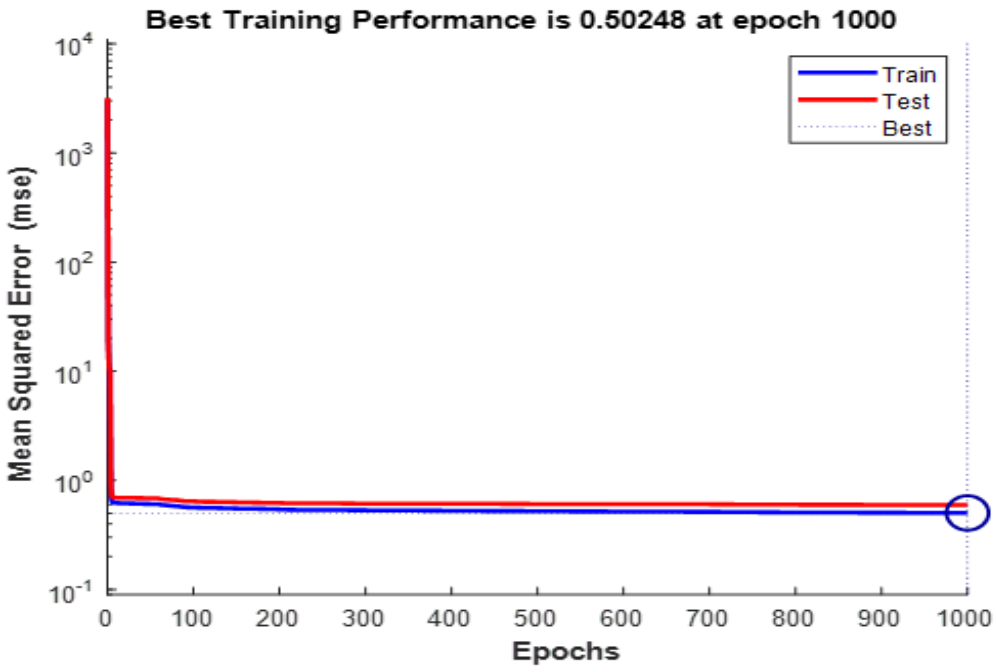
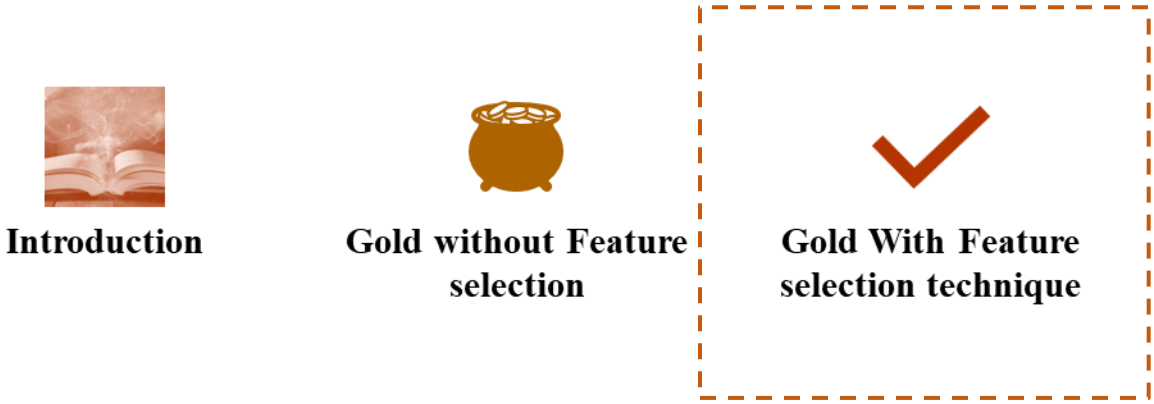


Figure 28: MSE Plot for model with all features

5.3 Gold with Feature Selection



5.3.1. Data and Materials

A comparable methodology was employed to build an ANN model incorporating correlation features, with the exception that during the feature selection phase as shown by the flowchart on Figure 29. The correlation coefficient feature based technique was used to evaluate the relationship between each feature and the gold grade. However, after the data preprocessing phase, the correlation coefficient-based feature selection was implemented. Correlation based feature selection technique was used to identify the critical features that are relevant for gold ore grade prediction and at the very least to remove some of the redundant input features. The most popular filter methods are correlation-based feature selection. This method calculates the correlation between each feature and the target variable (or among features) and selects the features with the highest correlation scores. Features that are highly correlated with the target variable are considered more important. The study investigated the

interrelationships among features, assessed the correlations between features and the target, and identified potential outliers. The statistical methods were used to rank the importance of each feature. The probability value(p-value) was used as a measure of the significance of each feature's contribution to the model. The correlation between each feature and the output was computed. The ANN model was then trained using these features and its performance was evaluated using correlation coefficient(R), mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE). Variables with higher absolute correlation coefficients were deemed to have greater relevance in predicting the gold grade. Consequently, these variables were further refined to include six input variables, specifically sample location (X and Z), alteration, lithology, and dip based on the highest absolute correlation scores. Figure 30 shows the correlation matrix based on correlation coefficient of the relationship between the input variables and the output variable (gold ore grade).

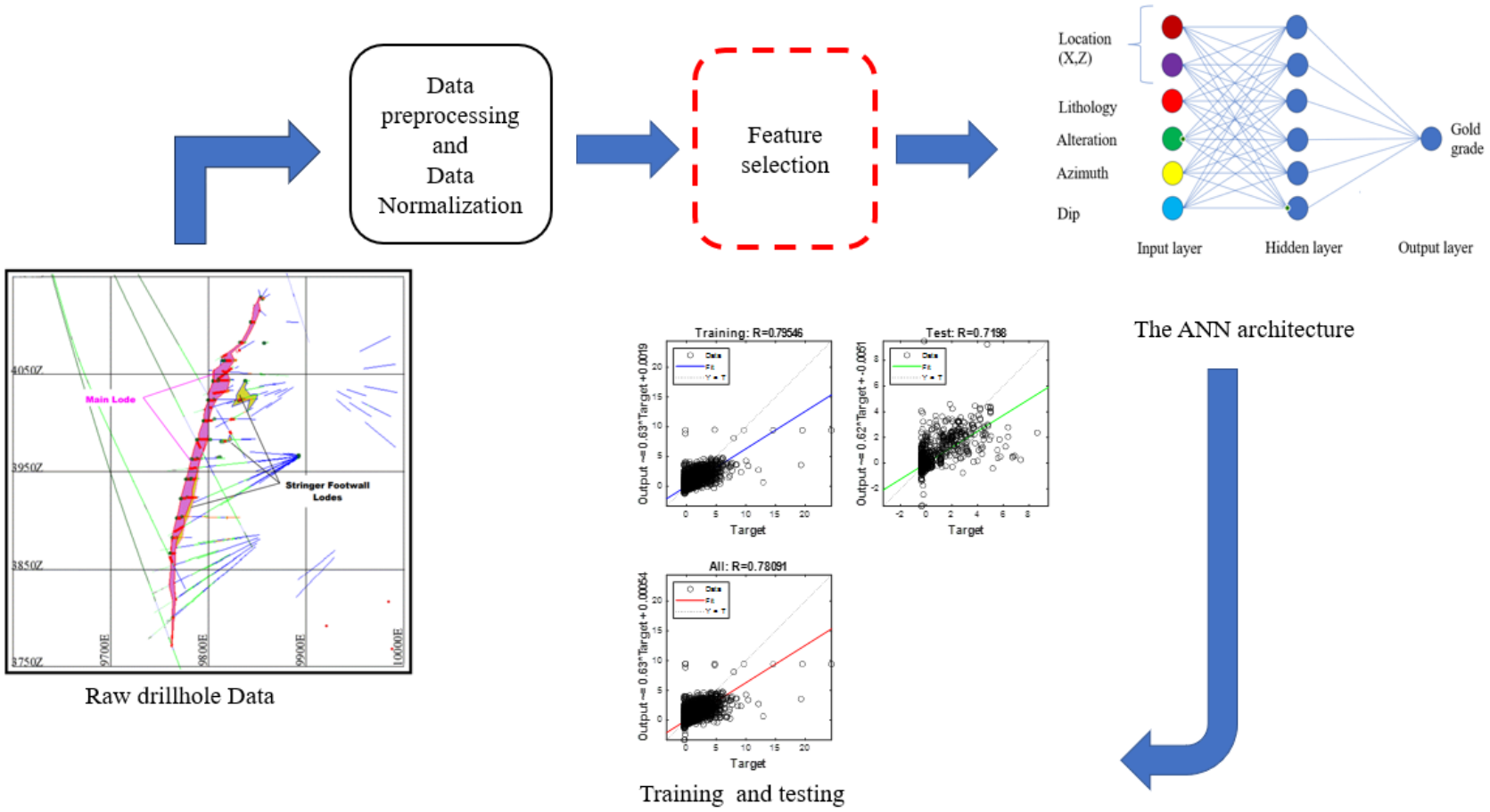


Figure 29: Flowchart of the overall analysis of Gold With feature selection

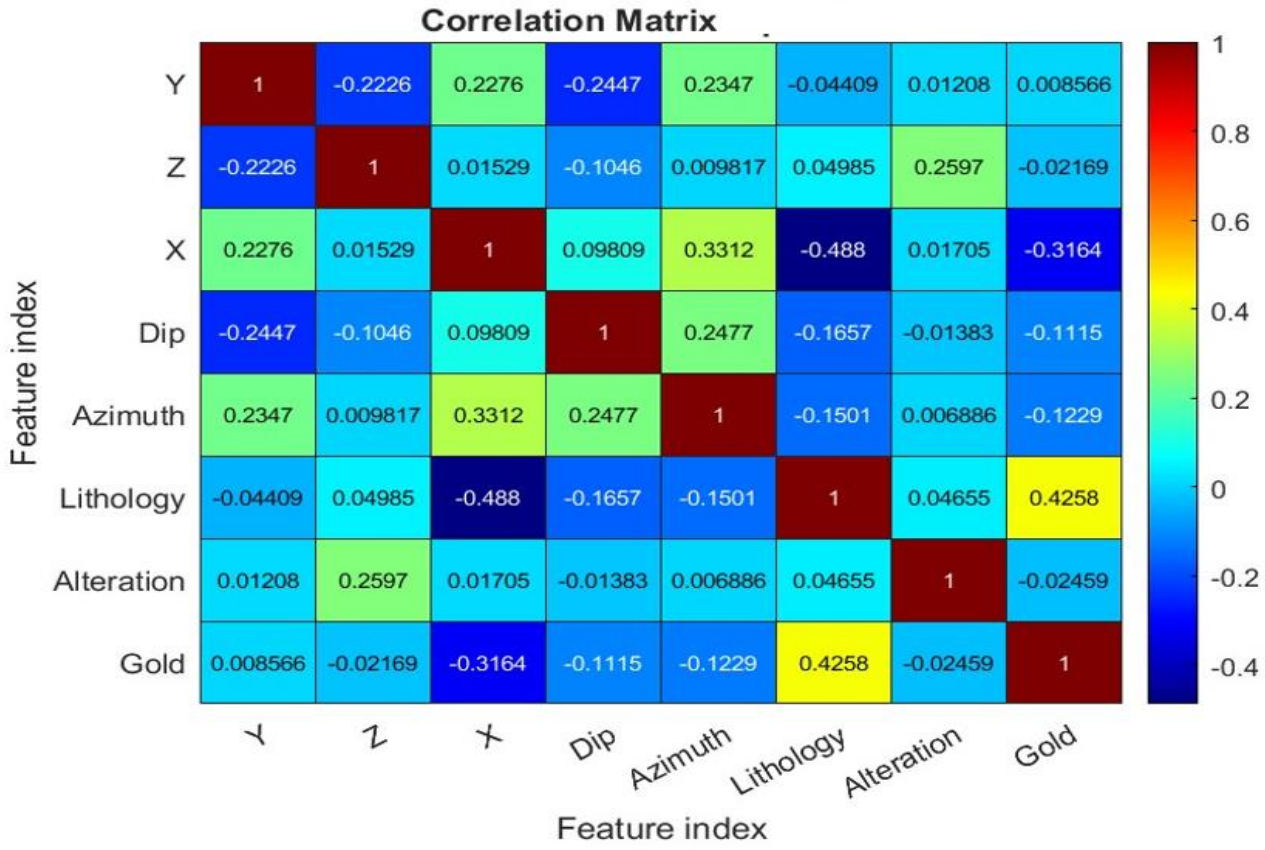


Figure 30:Correlation matrix based on predictors and output variable with dark red showing strong positive correlation and blue showing strong negative correlation between variables.

- Features with positive correlations (R > 0) indicate that as the feature's value increases, the target variable tends to increase as well. These features are positively associated with the target and may be useful for prediction.
- Features with negative correlation (R< 0) indicate that as the feature's value increases, the target variable tends to decrease. These features are negatively associated with the target and may also be useful for prediction.

The heatmap shows that alteration positively correlates with northings, altitude, eastings, azimuth, and lithology while it negatively correlates with dip.

To establish the statistical validity of the correlation scores, a p-value was calculated to evaluate the statistical significance of the correlation between the input and output variable . Figure 31 displays a statistical correlation heat map which illustrates the relationships between input and output variables. The p-value, short for "probability value," is a statistical measure used in hypothesis testing to assess the strength of evidence against a null hypothesis. It quantifies the probability of observing a test statistic as extreme as, or more extreme than, the one calculated from the sample data, assuming that the null hypothesis is true A small p-value (typically less than 0.05) suggests that the observed results are unlikely to have occurred by random chance alone, leading to the rejection of the null hypothesis. Conversely, a large p-value suggests that the observed results are consistent with the null hypothesis, and there is insufficient evidence to reject it.

It is important to note that a p-value does not provide information about the size or practical significance of an effect; it only indicates whether there is evidence to reject the null hypothesis. Additionally, p-values should be interpreted carefully, and their significance should be considered in the context of the proposed mineral estimation model and its design.

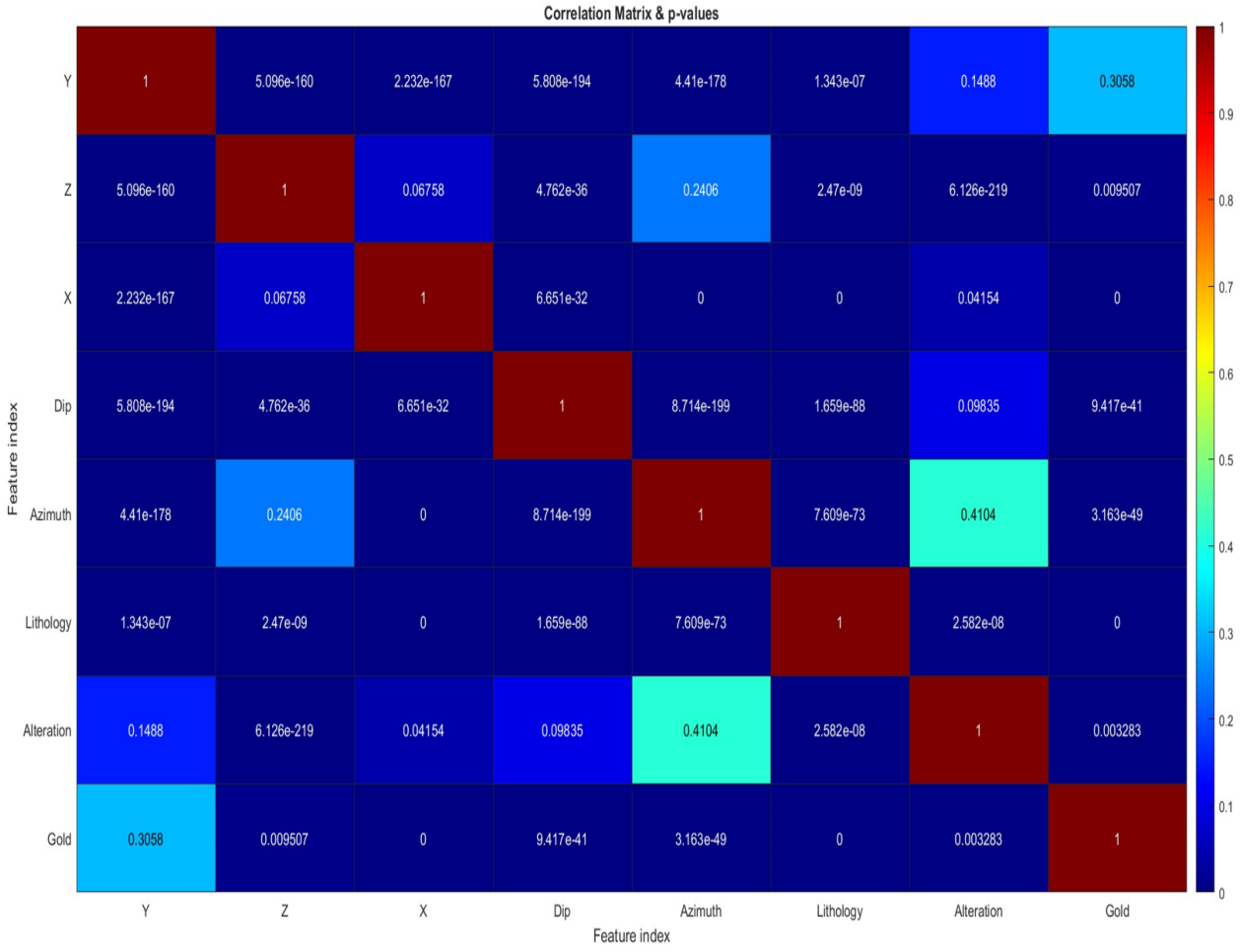


Figure 31:Correlation matrix showing p-values with dark blue showing the least contribution and dark red showing the most contribution.

Table 7 shows the interpretation of the p-values on the variables used in gold grade estimation. The significant level of p-value is as follows,

$p > 0.05$; not significant

$p \leq 0.05$ (5%); significant

$p \leq 0.01$ (1%); very significant

$p \leq 0.001$ (0.1%); highly significant

Table 8: Interpretation of p-values

Input Feature	Values of p	Inference
Y	0.3058	Not significant
Z	0.0095	Very significant
X	0	Highly significant
Dip	$9.417e^{-41}$	Highly significant
Azimuth	$3.163e^{-49}$	Highly significant
Lithology	0	Highly significant
Alteration	0.0033	Very significant

Based on the findings presented in Table 8, it can be concluded that the northings did not exhibit a statistically significant influence on the model. This is evident from its correlation score falling below the threshold that was established, leading to its exclusion from the model training process. However, the variables of alteration, lithology, dip, azimuth, eastings, and altitude displayed notable contributions. The alteration and lithology are important because they show the chemical composition of the mineral deposit.

Figure 32 displays a feature importance bar plot where each feature is represented by a bar whose height corresponds to its importance score. It demonstrates the most influential features of the predictive model. As mentioned earlier lithology, alteration, dip, azimuth and coordinates X and Z are very influential in this case.

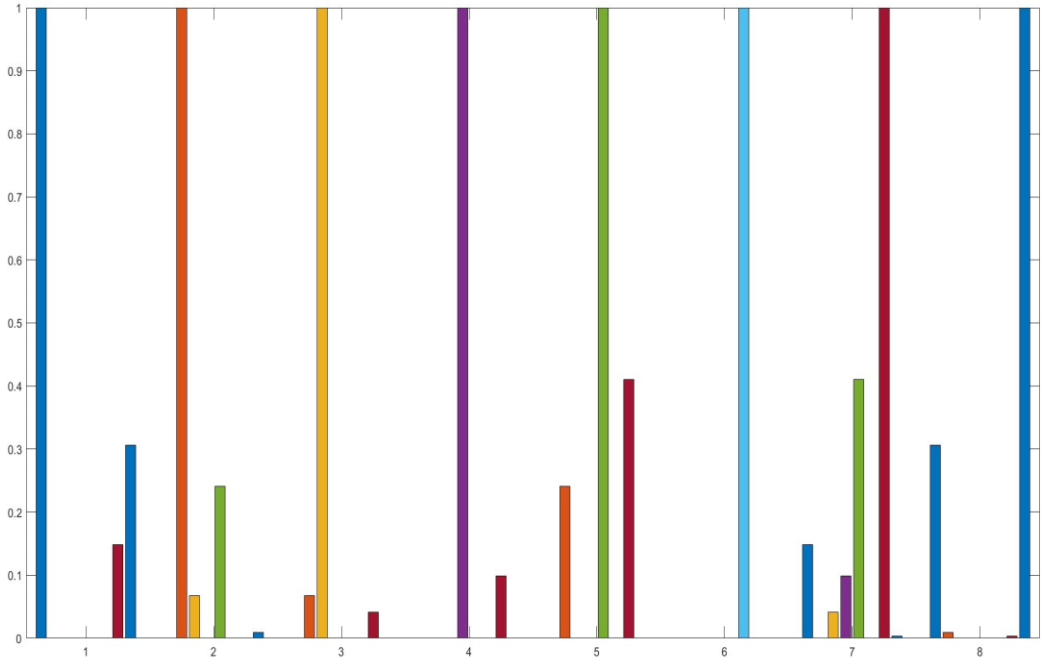


Figure 32:Feature importance bar plot indicating the importance of each variable on the prediction of gold ore grade.

The optimal model architecture has a set of six input nodes, a single hidden layer, and a solitary output node.

5.3.2 The Proposed ANN model for gold grade prediction

Artificial neural networks are a resilient machine learning methodology that may be utilized to represent intricate patterns, address prediction challenges by identifying established correlations within a dataset and predict the output values for a provided input dataset. An iterative process involving the exploration of various neural network configurations produced the architecture of the artificial neural network. The configuration with the lowest error rate

was ultimately selected as the optimal architecture. The proposed ANN structure consists of a single input layer with six neurons, a hidden layer, and an output layer, as seen in Figure 14.

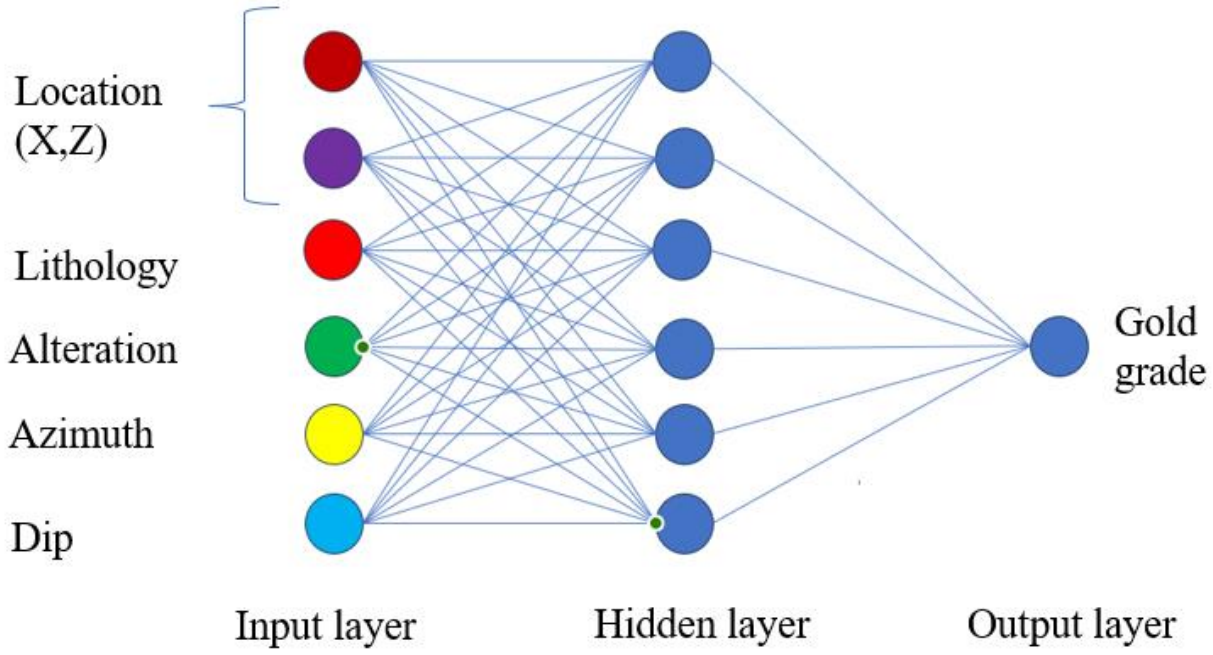


Figure 33: The proposed ANN structure for gold ore grade estimation

These selected variables were used in training the ANN model. The correlation coefficient, R, was computed for the training, testing, and overall model and the obtained values for R were 0.795, 0.720, and 0.781, respectively. Figure 34 shows the regression analysis diagram for training, testing and overall data. This observation suggests a high correlation between the input variables and the output variables. The mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE) were calculated to be 0.264, 0.374, and

0.653, respectively. MSE measures the average squared difference between the predicted values and the actual values. It provides a measure of the variance of errors. The MSE of 0.374 suggests that the model's errors have gold grade variability. Figure 35 illustrates the mean square error while Figure 36 shows the error histogram. The error is concentrated between -0.5 and 0.8. The findings indicate that the artificial neural network model exhibits a moderate level of accuracy in estimating the grade of gold. Not only did the accuracy of the model improve but the computational efficiency improved as well as the total training time for the model was recorded as 10.02 minutes.

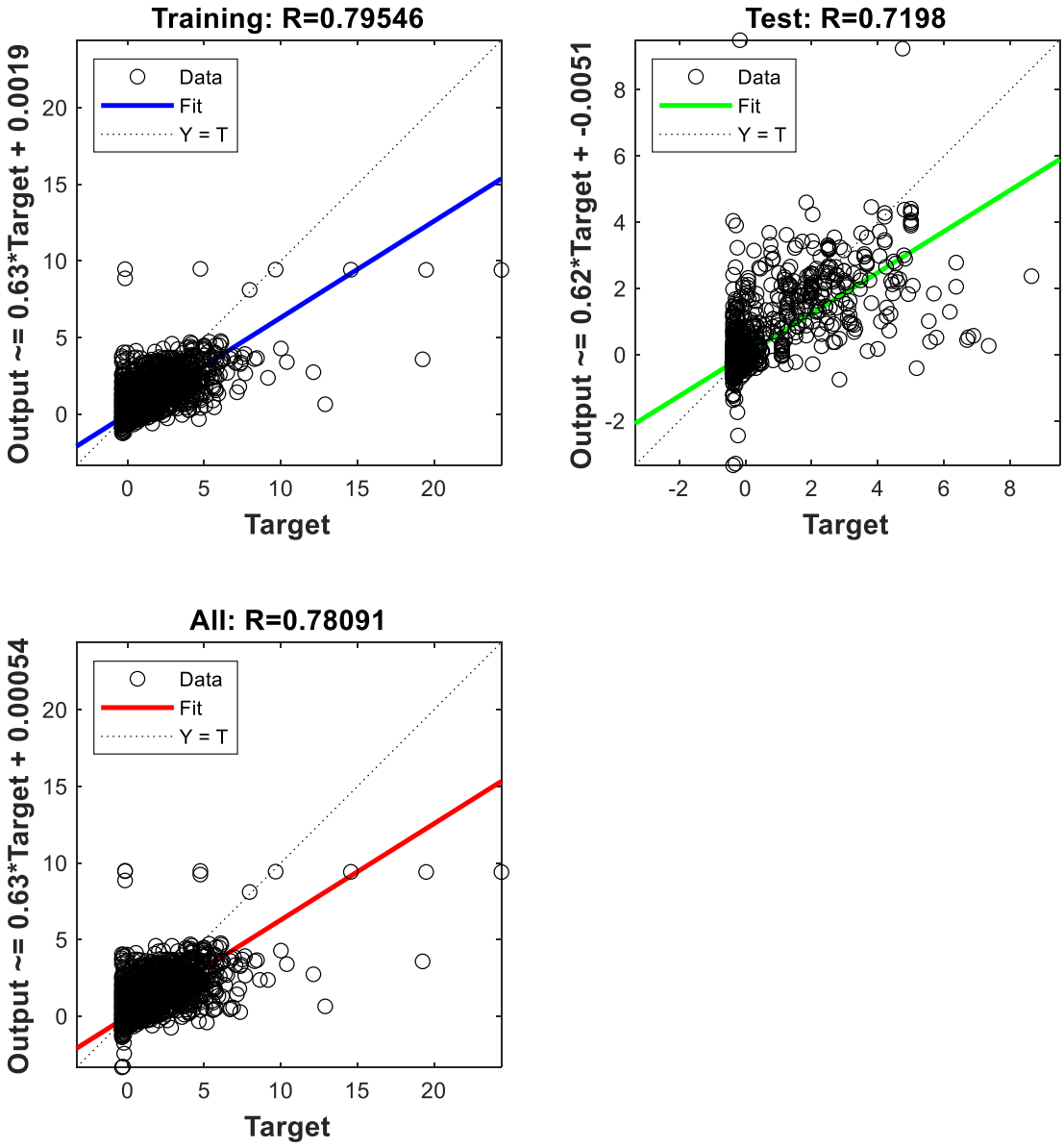


Figure 34:ANN regression model with correlation features showing actual and predicted data distribution with the blue, green, and red showing best line fit for training, test, and the overall model data, respectively. The white circles represent the data set.

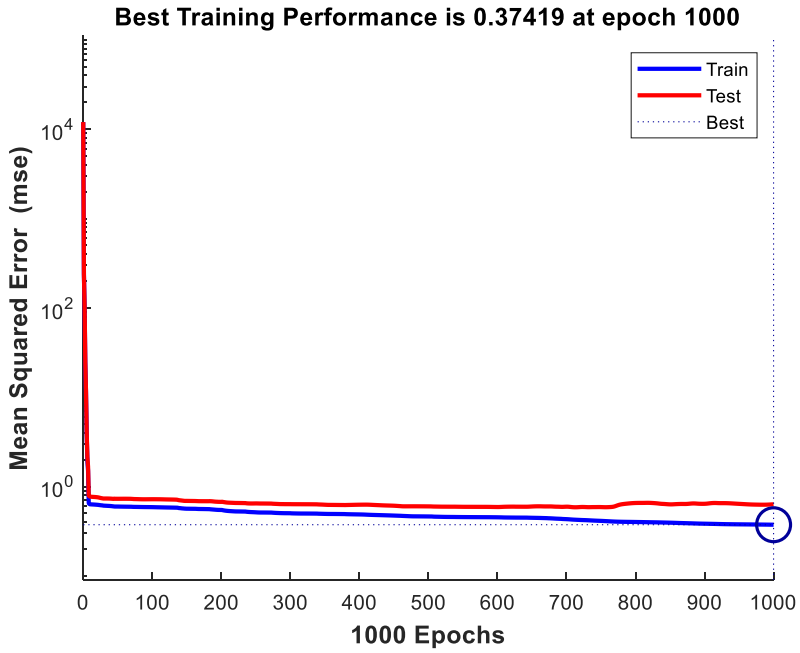


Figure 35: MSE plot for model with correlation features

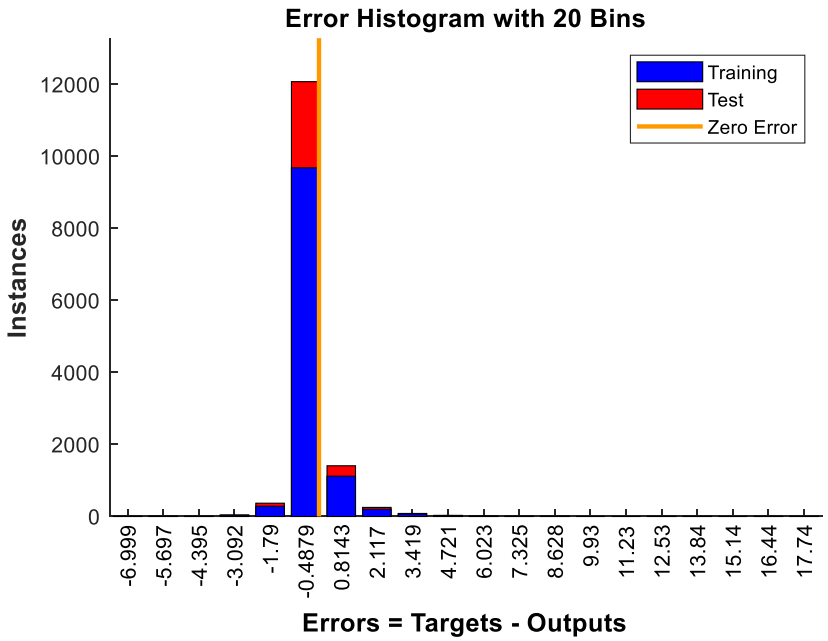


Figure 36:Error histogram for model with correlation features

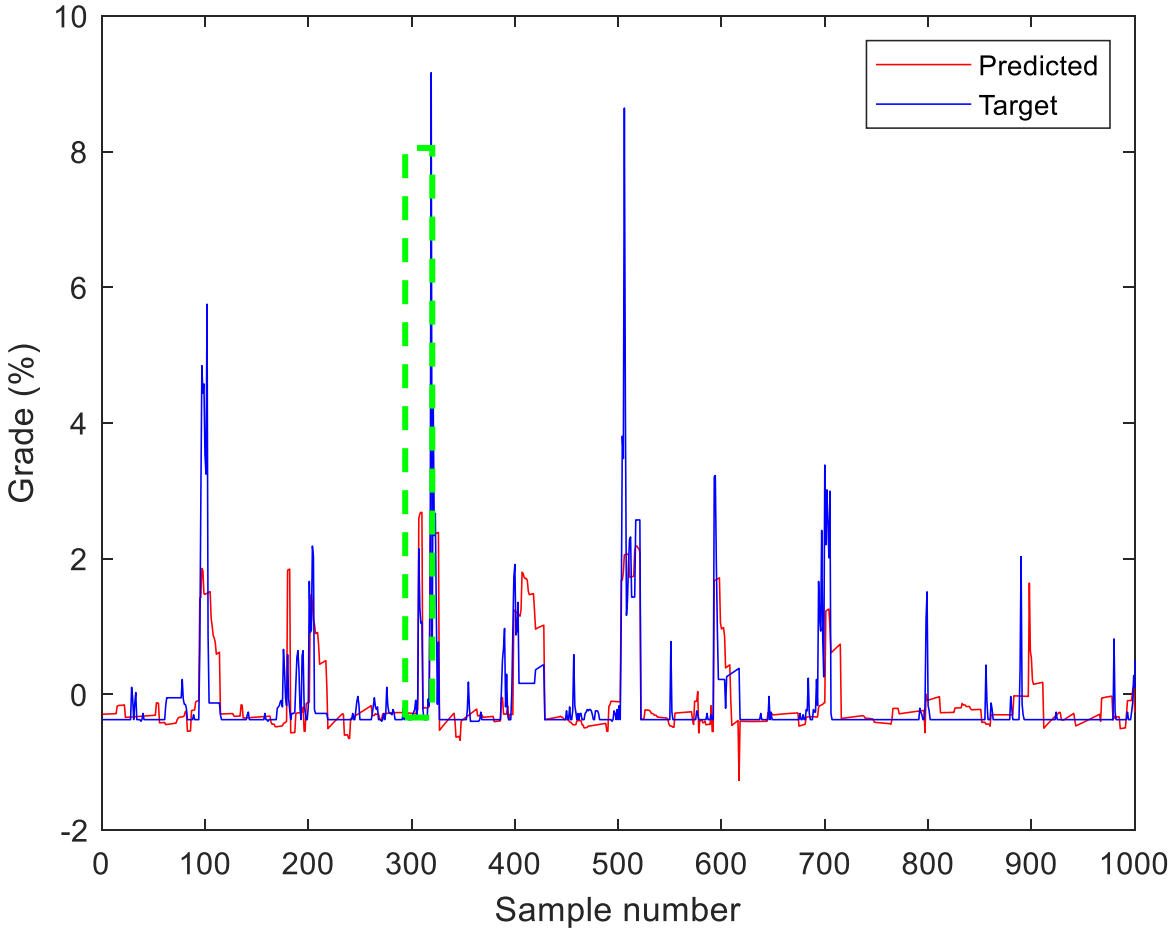


Figure 37: Data Distribution of actual and predicted values for gold with feature selection.

Figure 367showcases the performance of the ANN model in predicting gold grades. This is a visual representation that illustrates how close the model's predictions were to the actual grades. The model predicted the model moderately at sample 300 while the other predictions were fairly predicted.

Comparison of the models with and without feature selection

Table 9 shows the results of the models with or without feature selection based on correlation coefficient, mean absolute error, mean squared error, root mean squared error and computational time. Though both models made reasonable predictions on the gold grade, it was seen that the model using feature selection demonstrated superior performance in estimating the gold grade compared to the model without feature selection. The model that used feature selection exhibited a reduced computational training time of 10.02 minutes in comparison to the model that did not incorporate feature selection. It is important to acknowledge that the proposed methodology has the potential to be applied to many mineral resources. However, it is crucial to recognize that the selection of features is contingent upon the specific characteristics of each mineral. Although correlation feature selection approaches have demonstrated effective outcomes, it is imperative to examine other methods to provide a full study.

Table 9: Model comparison with and without feature selection

Mineral	**R**	**MAE**	**MSE**	**RMSE**	**Computational time(minutes)**
Gold (with feature selection)	0.720	0.264	0.374	0.653	10.02
Gold (without feature selection)	0.691	0.293	0.502	0.694	12.05

5.4 Summary

In summary, the correlation coefficient-based feature selection technique was applied to identify the most significant features with strong correlation with the gold ore grade. Furthermore, the study conducted a comparison of the performance of the ANN models with and without feature selection, thereby offering useful insights into the real-world benefits of employing these strategies in mineral grade prediction tasks. The findings indicate that the inclusion of feature selection strategies in the gold ore grade estimate resulted in improved performance of the model, as evidenced by correlation coefficient values of 0.720, a mean absolute error of 0.264, a mean square error of 0.374, and a root mean square error of 0.653. filter methods can be adapted to various mining operations and applications where feature selection is critical for improving the efficiency, interpretability, and generalization of machine learning models or data analysis tasks. This work highlights the importance of feature selection in predicting gold grades. However, future research can explore alternative feature selection techniques to enhance the accuracy and precision of the model and to ensure comprehensive model analysis. This study did not incorporate faults and geological discontinuities into the predictive model for gold ore grade.

5.5 Comparative analysis of the predictive ANN models for Copper and Gold Ore grades

The ANN model was used to predict both copper and gold grades, and the results are presented in Table 10. The ANN model yielded better results based on correlation coefficient, mean absolute error, mean squared error and root mean squared error when predicting copper grades compared to gold grade. The difference in the model performance may primarily from

the significant variation in the grade of gold compared to copper. This suggests that the variability in the gold grades in the dataset is higher than that in copper grades, which made the prediction more challenging. Feature selection improved the gold grade predictions, but the ANN model's performance for copper grades remained superior. Additional features could be beneficial for enhancing the god grade predictions.

Table 10: Comparative Analysis of the Mineral Ore grades

Mineral	**R**	**MAE**	**MSE**	**RMSE**
Copper (without feature selection)	0.765	0.018	0.0016	0.041
Gold (with feature selection)	0.720	0.264	0.374	0.653
Gold (without feature selection)	0.691	0.293	0.502	0.694

REFERENCES

Akbar, D.A. Reserve estimation of central part of Choghart north anomaly iron ore deposit through ordinary kriging method. Int. J. Min. Sci. Technol., 22, pp.573–577(2012)

Haldar, S.K. Mineral Exploration: Principles and Applications; Elsevier: Amsterdam, The Netherlands; ISBN 0128140232(2018)

Brown, W.M., Gedeon, T.D., Groves, D.I., Barnes, R.G., Artificial neural networks: a new method for mineral prospectivity mapping. Aust. J. Earth Sci. 47, pp.757–770. (2000)

Ewusi, A., Ahenkorah, I., Aikins, D., 2021. Modelling of total dissolved solids in water supply systems using regression and supervised machine learning approaches. Appl. Water Sci. 11, pp. 1–6. (2021) doi:10.1007/s13201-020-01352-7

Bellman, R., Dynamic programming. Science 153, pp34–37. (1966) doi:10.1126/science.153.3731.34

Zhang, Z., Zuo, R., Xiong, Y., A comparative study of fuzzy weights of evidence and random forests for mapping mineral prospectivity for skarn-type Fe deposits in the southwestern Fujian metallogenic belt, China. Sci. China Earth Sci. 59, pp. 556–572.(2016) doi:10.1007/s11430-015-5178-3.

Carranza,E.J.M., A. G. Laborte,A.G., "Data-Driven Predictive Modeling of Mineral Prospectivity Using Random Forests: A Case Study in Catanduanes Island (Philippines)," Nat Resour Res, 25(1), pp. 35–50,(2016) doi: 10.1007/s11053-015-9268-x.

Kaplan, U.E., Topal, E. A New Ore Grade Estimation Using Combine Machine Learning Algorithms. Minerals,10,847. (2020)

Samanta, B., Ganguli, R., Bandopadhyay, S. Comparing the predictive performance of neural networks with ordinary kriging in a bauxite deposit. Trans. Inst. Min. Metall. Sect. A Min. Technol.114, pp.129–140. (2005)

Chatterjee, S.; Bhattacherjee, A.; Samanta, B.; Pal, S.K. Ore grade estimation of a limestone deposit in India using an artificial neural network. Appl. GIS, 2,

pp.1–2, (2006)

Li, X.L., Xie, Y.L., Guo, Q.J., Li, L.H. Adaptive ore grade estimation method for the mineral deposit evaluation. Math. Comput. Model., 52, pp.1947–1956. (2010)

Mlynarczuk, M., Skiba, M. The application of artificial intelligence for the identification of the maceral groups and mineral components of coal. Comput. Geosci., (103) pp. 133–141(2017)

Singh, R.K.; Ray, D.; Sarkar, B.C. Recurrent neural network approach to mineral deposit modelling. In

Proceedings of the 2018 4th International Conference on Recent Advances in Information Technology (RAIT), Dhanbad, India, 15(17), pp.1–5, (2018)

Tsae, N.B., Adachi, T., Kawamura, Y., “Application of Artificial Neural Network for the Prediction of Copper Ore Grade,” Minerals, 13(5), pp. 658.

Sánchez-Maroño, N., Alonso-Betanzos. A., M. Tombilla-Sanromán, M., “Filter Methods for Feature Selection – A Comparative Study,” in Intelligent Data Engineering and Automated Learning – IDEAL (2007)

Yin, H., Tino, P.; Corchado, E., Byrne, W., Yao., Eds., Lecture Notes in Computer Science, Berlin, Heidelberg: Springer Berlin Heidelberg. 4881, pp.178–187. (2007)

Dumakor-Dupey, K., Arya, S. Machine Learning—A Review of Applications in Mineral Resource Estimation. Energies, 14, 4079. (2021)

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

CHAPTER 6: STUDY CONCLUSION

The study investigated the application of ANN on copper and gold ore deposits. Accurately predicting ore grade in mining is challenging due to the complex geological processes involved in ore formation and deposition. Precise grade prediction is crucial for mine planning, grade control, and feasibility studies. The ANN model combined seven input variables sample locations (X, Y, and Z coordinates), geological attributes (alteration and lithology), and dip and azimuth in estimating copper ore grade. The performance metrics, R^2, R, MAE, MSE, and RMSE, were used to evaluate the copper ore grade 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 in terms of prediction accuracy. Shapley values are used to assess the importance of input variables in predicting copper ore grade. The study demonstrated that lithology had the greatest influence on copper grade prediction, as it is closely related to the mineral composition of the orebody.

This research work emphasized the importance of feature selection techniques in machine learning. These techniques are crucial for identifying the most relevant features or variables that impact the model's ability to predict gold grade accurately. Feature selection has several goals, including improving model performance, reducing overfitting, enhancing model interpretability, and reducing computational complexity. The choice of a specific feature selection technique depends on the characteristics of the data and the machine learning algorithm used.

The research employed a correlation feature selection technique to identify the crucial features or variables that can improve the accuracy of predicting the grade of gold ore in the framework of artificial neural networks. By pinpointing these important features, the research seeks to reduce prediction errors and improve the overall performance of the model.

A comparative analysis was conducted to evaluate the performance of predictive models with and without feature selection. This comparison provided valuable insights into the actual benefits of using feature selection techniques in predicting mineral grades. In summary, the ANN model yielded better results when predicting copper grades compared to gold grades, primarily due to the larger variation in gold grades. Feature selection improved the gold grade predictions, but the ANN model's performance for copper grades remained superior.

The proposed model is considered innovative because it integrates various types of data that were not commonly combined in existing literature. This innovative approach has the potential to advance techniques for mining exploration and ore grade estimation, suggesting that it may contribute to improving the efficiency and accuracy of mining operations. It can also aid in mine planning, design, and mineral resource evaluation, leading to cost savings.

This study did not incorporate faults and geological discontinuities into the predictive model for the mineral ore grades. The results are promising and open the door for future research to address these limitations and further enhance the model's performance. It would be worthwhile to compare the proposed ANN model with established geostatistical methods like kriging in future directions. Future research can explore alternative feature selection

techniques to enhance the accuracy and precision of the model and to ensure comprehensive analysis.