FACULTY OF ENGINEERING AND TECHNOLOGY

MINING ENGINEERING DEPARTMENT

APPLICATION OF MACHINE LEARNING TO OPTIMIZE GOLD RECOVERY DURING CYANIDE LEACHING AT WAGAGAI GOLD MINE

CASE STUDY: WAGAGAI MINING (U) LIMITED

By

AYEBALE REYES
2200401601
BU/UG/2022/1601

A final year project report submitted to the department of Mining Engineering as a partial fulfilment of the requirements of the award of Bachelor of Science degree in Mining Engineering

MAY 2026

ABSTRACT
The cyanide leach process, also known as gold cyanidation, is a hydrometallurgical technique used to extract gold from low-grade ore, yet its efficiency is often compromised by the complex, non-linear interactions among various operational parameters (e.g., cyanide concentration, pH, leaching time, pulp density, and dissolved oxygen). Traditional optimization methods heavily rely on subjective operator experience, leading to suboptimal and inconsistent gold recovery rates.This study addressed this challenge by applying machine learning techniques to predict and optimize gold recovery during cyanide leaching at Wagagai Mining (U) Limited. A dataset of 2801 historical plant records was developed using key process variables, including cyanide concentration, pH, leaching time, pulp density, and dissolved oxygen. Three models including Multiple Linear Regression (MLR), Random Forest (RF), and Artificial Neural Network (ANN) were developed and evaluated using statistical performance metrics including the coefficient of determination (R^2), Root Mean Square Error (RMSE), and Mean Squared Error (MSE).

The results showed that the ANN model outperformed the others, achieving the highest predictive accuracy with an R^2 of 0.9444 and RMSE of 1.0209, compared to RF (R^2 = 0.9133) and MLR (R^2 = 0.8509). Hyperparameter optimization further improved ANN performance and confirmed that a single hidden layer with 10 neurons was optimal.

The optimized ANN predicted a maximum gold recovery of 96.42% at a cyanide concentration of 800 mg/L, pH of 10.67, leaching time of 48 hours, pulp density of 35%, and dissolved oxygen of 12 mg/L. Partial Dependence Plots (PDPs) and sensitivity analysis were used to interpret the model, revealing that pulp density and cyanide concentration are the most influential variables affecting recovery, while pH and dissolved oxygen play significant but moderate roles. Sensitivity analysis revealed that pulp density (10.53) and cyanide concentration (8.004) were the most influential variables.

The findings demonstrate that ANN-based models are effective tools for optimizing cyanide leaching and improving operational efficiency in gold mining.

Key words: Machine learning, Random forest, Artificial Neural Network, Multiple Linear Regression, Partial Dependence Plots and performance metrics.

DECLARATION

I **AYEBALE REYES,** registration number BU/UG/2022/1601 declare that this research report was originally written and organized by me and has not been previously published, therefore presented as a requirement for the award of Bachelor of Science in mining engineering.

Signature: …………………………………..

Date: ………………………………………..

Email: reyesganzi50@gmail.com

Contact: 0773133092

APPROVAL

I AYEBALE REYES, submit this report of my final year project to the faculty of engineering in the department of mining for the examinations with approval of my supervisors

DR. NANGENDO JACQUELINE

Signature……………………………………………………….

Date…………08/06/2026………….

MR. KWIKIRIZA MATHIAS

Signature………………………………………………

Date……………………………………………………….

ACKNOWLEDGEMENT

I take this opportunity to thank the almighty God for the good health and the ability to research and gather information in this report. I thank my supervisors for time, support and advice provided to me during my research. I would like to thank the almighty God for this guidance, protection and knowledge. I would to like to thank my guardians for their continuous support both spiritually and financially. My sincere gratitude goes to my project Supervisor Dr. Nangendo Jacqeline and Mr. Kwikiriza Mathias for their guidance, fellow students and Mining Department who have equipped me with the necessary knowledge that has enabled me to excel in academics.

DEDICATION

I dedicate this report to my family for their unwavering love, support, and encouragement have been the driving force behind my academic pursuits, may the almighty God bless them abundantly.

LIST OF ACRONYMS

ANN	Artificial Neural Networks
CNN	Convolutional Neural Networks
LR	Linear Regression
MAE	Mean absolute error
MLR	Multiple Linear Regression
MNLR	Multivariate Nonlinear regression
R^2	Coefficient of determination
RF	Random Forest
RMSE	Root mean square error
SVM	Support Vector Machines
HPO	Hyperparemeter optimization
ML	Machine learning
PDP	Partial Dependence Plots

Table of Contents

CHAPTER ONE: INTRODUCTION

1.1 Background

The cyanide leach process, also known as gold cyanidation, is a hydrometallurgical technique used to extract gold from low-grade ore. In this process, a cyanide solution (usually sodium cyanide) is used to dissolve gold from the ore, forming a water-soluble coordination complex. It involves crushing and grinding of ore to liberate the gold particles, a cyanide solution is applied to the ore, which dissolves the gold and the gold-laden solution is then processed to recover the gold, often using activated carbon or zinc precipitation (Gail et al., 2011).

Globally, gold extraction through cyanide leaching remains the dominant metallurgical process for processing free-milling gold ores due to its high efficiency and cost-effectiveness (Alguacil, 2006). However, the efficiency of this process is notoriously dependent on a complex interplay of chemical and physical parameters, including cyanide concentration, pH, and temperature, particle size, leaching time, and dissolved oxygen levels (Asghar Azizi 1 *, 2012). Suboptimal operation leads to significant financial losses through poor recovery and excessive reagent consumption, alongside severe environmental and safety risks associated with the handling and discharge of toxic cyanide (Han et al., 2024). In response to these challenges, the global mining industry is increasingly turning to data-driven approaches, such as machine learning (ML) and artificial intelligence (AI), to model complex processes, predict outcomes, and optimize parameters beyond the capabilities of traditional empirical methods (Qiu F, 2025). Liu, (2023) introduced a data-based compensation method for optimizing cyanide leaching operations. Recognizing the complexity of gold cyanide reaction mechanisms, they proposed a ML framework capable of adjusting process settings dynamically, outperforming traditional rule-based approaches. Their work demonstrated that ML could stabilize leaching processes under variable ore conditions, reducing uncertainty and reagent waste. However, their work focus on maximizing a single objective like recovery, leaving a gap in comprehensive frameworks that simultaneously optimize for multiple conflicting goals, such as maximizing gold yield while minimizing cyanide consumption for enhanced economic and environmental sustainability at a specific mine site. Flores & Leiva, (2021) show that Artificial Neural Networks (ANNs) have been effectively used for predicting copper recovery in leaching processes, achieving high accuracy and precision, with mean accuracy values around 94.3% and a recall of 99.5% in some studies. They are also noted for their ability to model the recovery rate of uranium in in situ leaching processes with high accuracy compared to traditional methods (Aizhulov et al., 2024). Zhang et al., 2022a Support Vector Machines (SVM) have demonstrated strong predictive performance in various leaching scenarios. For instance, in pyrolusite leaching, SVM achieved a regression index (R^2) of 0.92, indicating excellent prediction accuracy. Random Forest (RF) models have been shown to perform well in predicting gold accessibility, with a coefficient of determination (R^2) of 0.77, indicating a robust predictive capability (Costa et al., 2023).

In Africa, Many operations, particularly small to medium-scale mines, continue to rely on heuristic methods and generalized rules of thumb for their leaching circuits. This often results in recovery rates that are below the industry potential and operational costs that are higher than necessary (Akhabue, 2023). The implementation of sophisticated process control and optimization strategies, like those proposed in this study, presents a significant opportunity to enhance the competitiveness and sustainability of the African gold mining sector. For instance, a study of cyanide vs thiosulfate leaching in artisanal gold mining in Zimbabwe found cyanide gave only ~ 68–72% extraction in some ores due to limited control of leaching parameters (Manzila et al., 2022).

Within East Africa, a burgeoning gold-producing region, the push for technological modernization is gaining momentum. Countries like Tanzania and Kenya are investing in modern mining technologies to improve efficiency and yield (East African Community, 2021). Uganda, in particular, with its developing gold mining industry, stands to benefit immensely from research that directly addresses process optimization. The country's focus on developing its mineral resources, as outlined in its National Development Plan III, highlights the need for efficient and environmentally sound extraction technologies to ensure the sector's contribution to national economic growth is maximized (Government of Uganda, 2020).

The Busia District is a focal point of Uganda's gold mining activities, hosting projects like the Wagagai Gold Mine. The geological setting of the Busia-Budadiri belt presents orogenic gold mineralization that responds uniquely to leaching parameters. While the region holds economic potential, operational challenges such as fluctuating ore grades and the need for precise process control are prevalent. Optimizing the leaching process for these specific ores is therefore not just a technical exercise but a necessity for the economic viability of mining operations in the district (Tuhumwire, 2019).

At the Wagagai Gold Mine, the beneficiation plant utilizes cyanide leaching as its primary extraction method. Current operations are guided by conventional practices. There is a recognized opportunity to enhance recovery efficiency and reduce cyanide consumption through a more scientific, data-oriented approach. This study proposes to bridge this gap by leveraging machine learning to model and optimize the leaching process specifically for

Wagagai ore. By applying Machine Learning algorithms, this research aims to define the optimal parameter combinations that maximize gold recovery while minimizing operational costs and environmental impact, thereby contributing directly to the efficiency and sustainability of gold mining in Busia District and Uganda at large.

1.2 Problem statement

At Wagagai Mining (U) Limited, cyanide leaching process suffers from suboptimal (below the standard) performance with average recovery of 75% due to its reliance on traditional optimization methods (i.e. Rule of thumb sparging and trial and error), which rely on operators' expertise, does not favor inexperienced operators, leading to economic inefficiency from lower-than-potential gold recovery and environmental risk from the overuse of cyanide. This operational gap arises from an inability to capture the complex, non-linear interactions between critical parameters such as cyanide concentration, pH, and dissolved oxygen, resulting in significant revenue loss and heightened safety hazards. This study proposed a supervised machine learning-based optimization framework that provided practical guidelines for the mining industry to enhance gold recovery while reducing costs and minimized chemical waste. This solution involves training predictive models such as artificial neural network, Random Forest and Multilinear regression models on plant data to accurately forecast recovery, systematically tuning these models to identify parameter combinations that maximize yield while minimizing cyanide consumption, and finally generating actionable, data-backed operational guidelines tailored to Wagagai's specific ore.

1.3 Justification

This research is justified by the urgent need to enhance the gold recovery efficiencies and environmental sustainability of gold extraction at Wagagai Mining (U) Limited. The current reliance on empirical methods for controlling the cyanide leaching process leads to suboptimal gold recovery of 75%, representing significant revenue loss of 21%, and excessive cyanide consumption, which unnecessarily inflates operational costs and poses severe environmental and safety risks. Since cyanide is highly toxic, its overuse also poses significant environmental and safety risks to workers and surrounding communities. This project addresses these challenges directly by pioneering a machine learning-based optimization framework, which moves the process from guesswork to data-driven precision. By systematically identifying the ideal parameter combinations, the study aims to provide practical guidelines to both junior and senior operators, minimize ecological impact through reduced cyanide usage, and bridge a crucial technological gap, thereby promoting safer, more efficient, and sustainable mining practices.

1.4 Objectives

1.4.1 Main objective:

To optimize gold recovery during cyanide leaching by application of machine learning.

1.4.2 Objectives:

1. To develop machine learning models for the cyanide leaching process.
2. To optimize the best performing developed model.
3. To utilize the optimized developed model to analyze nonlinear parameter interactions and predicting gold recovery.

1.5 Research questions:

1. What are the critical cyanide leaching parameters?
2. What is the effective machine learning model and how can machine-based analysis improve prediction and optimization of recovery outcomes?
3. What parameter combinations yield the best balance between high gold recovery and low cyanide use?

1.6 Significance

It will provide practical guidelines to the inexperienced mining engineers to enhance gold recovery while promoting responsible cyanide use to minimize chemical waste, and contributes to safer operations by lowering risks of cyanide exposure for workers and nearby communities hence promoting SDG12 (Responsible consumption and production). It also enriches academic research by advancing knowledge on hydrometallurgical optimization through the integration of experimental data and statistical modeling which promotes SDG 9 (Industry, Innovation, and Infrastructure), while offering policymakers' evidence-based recommendations to support safer and more efficient cyanidation standards.

1.7 Scope

1.7.1 Geographical scope

This project is limited to Wagagai Mining (U) Limited in Eastern Uganda, Busia district.

1.7.2 Conceptual framework

This project only focused on applying machine learning techniques to optimize gold recovery during cyanide leaching process. It did not include carrying out laboratory tests. However, ore mineralogy, agitation speed, temperature and particle size were not considered as model input parameters.

The theoretical foundation of this study is based on Elsner's Equation ($4Au + 8CN^- + O_2 + 2H_2O \rightarrow 4Au(CN)_2^- + 4OH^-$), which describes the cyanidation process whereby gold reacts with sodium cyanide, oxygen, and water to form a soluble gold–cyanide complex. This equation highlights that successful gold dissolution is strongly dependent on both cyanide concentration and oxygen availability under alkaline conditions, maintained using lime. Thus, controlling these factors is critical to maximizing leaching efficiency.

Independent variables (input parameters)	**Mediating and controlled factors**	**Dependent variable (output)**
• Cyanide Concentration • Solution pH • Leaching Time • Pulp Density • Dissolved Oxygen Level	Ore Mineralogy (constant ore source) Mixing Efficiency (Agitation speed) Temperature (Held Constant) Particle Size (P80) of the ore	**Gold Recovery (%)**

1.7.3 Time

This project was done within period of 3 months.

2. CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

This chapter explores more about machine learning techniques used during mineral processing particularly during cyanide leaching with specific focus on development of a model, optimization and validation of the ML model to improve gold recovery.

Consequently, this review articulates notable advancements, challenges, and deficiencies that necessitate the present inquiry, thereby positioning the research within the broader framework of ML-driven innovations.

STUDY AREA

Wagagai Mining Company Limited, stands as my case study area. Wagagai Mining (U) Limited is a wholly owned subsidiary of Liaoning Hongda Group, registered in Busia, a town in the Administrative Centre of Busia District located in the Eastern Region of Uganda. The company owns the gold mining right of 9.24 square kilometers in the North of the town, and about 20 million tons of gold ore resources have been controlled in the first mining section. The total gold resource of the mining area can reaches 20 tons, contributing to Uganda's industrial and economic development and it stands as one of the largest and most significant players in the country's mining industry specialized in gold extraction and processing

Figure 1 Arial view of wagagai

Cyanide leaching process for gold extraction.

Cyanide leaching, or cyanidation, is the dominant hydrometallurgical technique for extracting gold from ores, accounting for over 90% of global gold production (Marsden, 2006). Its dominance stems from its high efficiency, selectivity for gold, and relative cost-effectiveness compared to alternative methods. The process was pioneered in the late 19th century by J.S. MacArthur, R.W. Forrest, and W. Forrest, who patented the use of a dilute cyanide solution to dissolve gold, revolutionizing the gold mining industry (Habashi, 2005).

The efficiency of the cyanide leaching process is governed by a complex interplay of chemical and physical parameters, where the stoichiometric relationship between cyanide and oxygen (Elsner's Equation) dictates that dissolution rates are maximized when both reagents are readily available, often requiring cyanide concentrations above stoichiometric levels (200-500 ppm) to overcome diffusion limitations (Azizi et al., 2016). The process must be maintained in an alkaline environment (pH 10-11) using lime to prevent the formation of toxic HCN gas and suppress impurity dissolution (Habashi, 2005), while optimal particle size liberation (e.g., P80 of 75μm) is crucial to expose gold surfaces without creating problematic slimes (Tan Liu; Qingyun Yuan; Lina Wang; Yonggang Wang, 2023). Furthermore, although reaction rates increase with temperature, practical operations balance leaching time (typically 24-48 hours) and agitation intensity to ensure sufficient diffusion and mass transfer while maintaining economic viability (Marsden, 2006).

2.2 Machine Learning in Mineral Recovery

Machine learning (ML) in mineral processing refers to the application of data-driven computational algorithms capable of identifying patterns, modeling complex process dynamics, and making predictions or control decisions without explicit programming (McCoy & Auret, 2019). It is used to enhance predictability, process control, and optimization across unit operations such as comminution, flotation, and leaching.

According to McCoy & Auret, (2019), ML provides "a structured means of modeling mineral processing data for fault detection, performance prediction, and process optimization through algorithms that learn from plant or laboratory datasets rather than relying solely on first-principle equations." Similarly, Jooshaki et al., (2021) describe ML as "a subset of artificial intelligence designed to enable computers to solve complex mineralogical problems by learning from large, multidimensional datasets that represent ore characteristics, process parameters, and product outputs."

2.3 Data collection

A study by Osei et al., 2024 compared ML models for gold recovery prediction using a dataset of 98 laboratory experiments, also Tan Liu; Qingyun Yuan; Lina Wang; Yonggang Wang, 2023 used a data-driven approach with high-frequency measurements, but their initial model was built on a foundational dataset of over 120 controlled batch tests. The quality, variety, and granularity of the data collected directly influence the performance of ML algorithms. Inadequate, inconsistent, or noisy data can lead to biased models that fail under real-world variability (Ahamed, 2025). Thus, the design of a robust data collection strategy, covering experimental setup, parameter measurement, data cleaning, and validation is essential for successful optimization of leaching processes. Tan Liu; Qingyun Yuan; Lina Wang; Yonggang Wang, 2023 highlighted that gold leaching processes involve numerous interacting parameters such as cyanide concentration, pH, temperature, particle size, and dissolved oxygen. They proposed a data-driven compensation approach that uses high-frequency laboratory measurements to adjust process settings dynamically. This approach relied on carefully collected time-series data from agitated tank reactors, providing continuous input-output relationships for ML training.

2.3.1 Data preprocessing

Before data can be used for training machine learning models, it must undergo rigorous preprocessing. Kavoni et al., 2025 pointed out that many industrial and laboratory datasets are incomplete, inconsistent, or contain outliers, especially in complex bioprocessing and metallurgical environments. In their work on optimizing monoclonal antibody production, a process analogous to leaching in terms of nonlinear kinetics, they implemented data normalization and dimensionality reduction to improve model stability. Translating this to cyanide leaching, similar preprocessing steps such as standardization, outlier removal, and normalization are necessary to prevent model bias.

Furthermore, Qiu et al., (2025b) proposed a multi-objective optimization framework that integrates heterogeneous data from laboratory and operational settings. They demonstrated that ML models trained on multi-source datasets, combining experimental, environmental, and economic variables can achieve better generalization. In the context of gold cyanidation, integrating laboratory test data with operational plant data can help models learn both micro-scale reaction behaviors and macro-scale production trends.

2.4 Types of models used.

2.4.1 Multiple Linear Regression (MLR).

In contemporary hydrometallurgical research, Multiple Linear Regression (MLR) serves primarily as a performance baseline, as demonstrated by Osei et al., 2024, who reported that while MLR provided a reasonable fit ($R^2 = 0.76$) for gold cyanide leaching prediction, it was significantly outperformed by non-linear ensemble methods like Random Forest and ANNs due to its inability to capture complex parameter interactions. Despite this limitation, MLR remains valuable for preliminary parameter screening, exemplified by Mendez et al. (2022), who used it to identify statistically significant factors in copper leaching, thereby streamlining subsequent complex modeling efforts. Furthermore, in hybrid approaches, MLR's explanatory clarity complements advanced models, as shown by Qiu et al. (2025), who paired MLR-based sensitivity analysis with high-accuracy ML predictors to interpret variable influences. However, MLR's utility is constrained by inherent limitations: its assumption of linearity and additivity fails to represent logarithmic relationships (e.g., cyanide concentration effects) or parameter interactions (e.g., pH-cyanide interdependence) without explicit transformation (Marsden & House, 2006); its sensitivity to multicollinearity inflates coefficient variances, destabilizing interpretations (Farrar & Glauber, 1967); and its propensity for oversimplification renders it inadequate for high-dimensional, non-linear data in modern mineral processing contexts (Aldrich & Feng, 2020).

2.4.2 Support Vector Regression (SVR).

Support Vector Regression (SVR) is a powerful ML algorithm based on the principles of Support Vector Machines. It works by mapping input data into a high-dimensional feature space to find a hyperplane that best fits the data with a minimal margin of error, making it effective for handling non-linear relationships.

Scholars have successfully applied SVR to metallurgical challenges. Hapurima et al. (2023) demonstrated that SVR outperformed traditional response surface methodology in predicting gold recovery, particularly due to its robustness against overfitting. Furthermore, a study by Aldrich and Feng (2020) highlighted that SVR's performance is highly dependent on the careful selection of its kernel function and hyperparameters (e.g., C, gamma), but when tuned correctly, it can model complex chemical processes with high precision, even with a limited number of data points.

2.4.3 Random Forest Regressor (RFR).

Random Forest is an ensemble learning method that operates by constructing a multitude of decision trees during training and outputting the mean prediction of the individual trees. It is highly effective for tabular data, resistant to overfitting, and provides native feature importance rankings.

The application of RFR in mineral processing is well-documented. Osei et al. (2024) utilized a Random Forest model as part of their comparative analysis for gold recovery prediction and found it to be highly accurate, noting its ability to handle noisy data from industrial processes. In a broader context, Qiu et al. (2025) emphasized that tree-based models like RFR are particularly valuable for identifying the most influential leaching parameters (e.g., cyanide concentration and dissolved oxygen), thereby providing actionable insights for process engineers beyond mere prediction.

2.4.4 Gradient Boosting Machines (GBM).

Gradient Boosting Machines (GBM) are another class of ensemble methods that build models sequentially, where each new model corrects the errors of the previous one. Advanced implementations like XGBoost, LightGBM, and CatBoost have become state-of-the-art for many predictive tasks due to their superior speed and accuracy.

Recent literature shows a strong preference for these advanced GBM models. Qiu et al. (2025) explicitly used a hybrid approach with a GBM model (CatBoost) to predict key performance indicators in gold mining, achieving high accuracy for cost and emission outputs. Another study by Zhang and Yang (2022) compared various algorithms for a flotation process and concluded that XGBoost consistently delivered the best predictive performance, attributing this to its efficient handling of complex variable interactions and its built-in regularization which prevents overfitting.

2.4.5 Artificial Neural Networks (ANN).

Artificial Neural Networks (ANNs) are computational models inspired by the human brain, consisting of interconnected layers of nodes (neurons). They are universal function approximators, capable of learning extremely complex, non-linear relationships, making them ideal for modeling multifaceted processes like cyanide leaching.

ANNs have a long history of application in hydrometallurgy. Seminal work by Apling et al. (2020) demonstrated that a well-structured multi-layer perceptron (MLP) could model a full gold processing plant, integrating data from crushing through to leaching. More recently, Bascur and Soudek (2021) argued that deep learning architectures, while data-hungry, offer the highest potential for capturing the dynamic and transient behaviors in leaching circuits that simpler models miss, paving the way for real-time adaptive control.

2.4.6 Empirical Models.

Empirical models are based on observed relationships and experimental data rather than theoretical first principles. In gold cyanidation, the most foundational empirical model is Elsner's Equation (1846), which describes the stoichiometry of the core reaction. While essential for fundamental understanding, these models often lack the precision for operational optimization under variable conditions.

Scholars have recognized the limitations of standalone empirical models. Marsden and House (2006), in their authoritative text, detail various kinetic models for leaching but note that their accuracy is constrained by assumptions of ideal conditions. Modern research, such as that by Tan et al. (2023), uses these empirical kinetics not as predictive tools, but as a physical basis to inform and constrain data-driven models, creating a powerful hybrid approach that leverages both domain knowledge and ML's predictive power.

2.5 Optimization of the Machine Learning Model

The integration of ML with multi-objective optimization in mining has been explored in various applications to show its potential for optimizing operational efficiency. Numerous studies have been performed to examine the application of ML for ore grade prediction, energy consumption forecasting, and process optimization. Little work has been done on a comprehensive framework to optimize several conflicting objectives in gold mining operations simultaneously (Qiu et al., 2025b). -, 2024 and (Thanammal Indu et al., 2025) focuses on automated validation frameworks, domain-specific feature development, systematic hyper parameter optimization, and leveraging cloud-native architectures. Implementing strong data engineering techniques significantly enhanced model performance by 42% and prediction accuracy by 40%, reduces error rates, and accelerates development timelines.

2.5.1 Optimization algorithm

In hyperparameter optimization, Bayesian Optimization (BO) has emerged as a superior alternative to exhaustive methods like Grid Search. By constructing a probabilistic model to guide the search for optimal parameters, BO efficiently minimizes expensive function evaluations, making it ideal for tuning complex machine learning models (Shahriari et al., 2016). Modern implementations, such as the Tree-structured Parzen Estimator (TPE) within frameworks like Optuna, have further advanced this approach by effectively handling high-dimensional and conditional parameter spaces (Takuya Akiba, 2019).

For process optimization involving multiple conflicting objectives, such as maximizing gold recovery while minimizing reagent consumption, metaheuristic algorithms are particularly effective. The Non-dominated Sorting Genetic Algorithm (NSGA-II) is a prominent method that evolves a population of solutions to identify the Pareto-optimal front, representing the best possible trade-offs (Deb et al., 2002). Recent research demonstrates the power of using machine learning models as fast surrogates for the actual process, which are then optimized with these algorithms to rapidly identify optimal operational parameters, a key methodology in modern mineral processing (Qiu et al., 2025b).

2.5.2 Model evaluation

A thorough evaluation of machine learning models requires multiple statistical metrics to reliably measure predictive accuracy. The coefficient of determination (R^2) is commonly used to quantify the proportion of variance in the dependent variable explained by the model, although researchers warn against relying on it alone as it can be artificially inflated and does not reveal prediction bias (Chicco et al., 2021). Alongside R^2, both Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) offer essential insights into the size of prediction errors, with RMSE's sensitivity to large errors making it especially useful for cost-sensitive applications like gold recovery, while MAE provides a more robust measure of average error.

In addition to metric reporting, robust validation techniques are crucial to ensure model generalizability. k-fold cross-validation has emerged as a gold standard method, especially with limited datasets, as it provides a nearly unbiased estimate of generalization error by repeatedly partitioning data into training and validation subsets (Hastie, 2009). This approach significantly reduces performance estimate variance compared to single train-test splits, thereby enabling more confident model selection for scientific and industrial applications (Payam Refaeilzadeh, 2016).

2.6 Partial Dependence Plots (PDPs)

Partial Dependence Plots (PDPs) remain a cornerstone technique for interpreting complex machine learning models, particularly for understanding the average marginal effect of one or two features on the predicted outcome. Introduced by Friedman, 2001in the context of gradient boosting, PDPs work by marginalizing the model's predictions over the distribution of all other features, thus showing the relationship between a target feature and the prediction while accounting for the average effect of the other variables. This provides a highly intuitive visual summary that is model-agnostic, making it applicable to everything from linear models to sophisticated ensembles and deep neural networks (Molnar, 2022).

Despite their widespread use, a significant limitation of PDPs is their reliance on the assumption of feature independence. When features are correlated, the PDP can create and display unrealistic data points, leading to misleading interpretations by extrapolating in regions of the feature space without actual data (Apley, 2020). To address this, Accumulated Local

Effects (ALE) plots have been proposed as a robust alternative. ALE plots compute the difference in predictions within local intervals of the feature, thereby avoiding the extrapolation issue inherent in PDPs and providing a more reliable estimate of the feature effect in the presence of correlation (Apley, 2020). Furthermore, PDPs only show the average global relationship, which can obscure heterogeneous effects. To uncover individual instance-level behavior, Individual Conditional Expectation (ICE) plots are often used in tandem with PDPs. ICE plots disaggregate the average PDP by displaying the prediction path for each individual instance, revealing interaction effects and subgroups within the data that might cancel each other out on the global plot (Goldstein, 2015). The integration of PDPs with SHapley Additive exPlanations (SHAP) has also gained traction, as SHAP dependency plots can provide a similar view to PDPs but with the added benefit of explaining individual predictions and visualizing the distribution of other interacting features (Lundberg, 2017). In applied domains like mineral processing, PDPs have been effectively used to visualize the non-linear relationships and interaction "sweet spots" between critical leaching parameters, such as cyanide concentration and dissolved oxygen, directly informing operational guidelines (Qiu et al., 2025b).

2.7 Summary of the review

While the literature convincingly shows that ML models like Random Forest and ANN are very effective for predicting gold recovery (Osei et al., 2024; Osvaldo A. Bascur, 2021)), these models are highly sensitive to the data they are trained on. There is a significant lack of studies that develop and validate these models specifically for the unique ore mineralogy of the Busia-Budadiri belt in Uganda (Tuhumwire, 2019). A model trained on data from a North American mine might not perform well at the Wagagai Gold Mine, highlighting the need for a localized, site-specific predictive model.

Recent research by Qiu et al., 2025c explicitly notes that "little work has been done on a comprehensive framework to optimize several conflicting objectives in gold mining operations simultaneously." The literature focuses on maximizing a single output (e.g., recovery). However, for operations in Uganda, where economic viability and environmental sustainability are equally pressing, there is a clear gap in frameworks that simultaneously optimize for both high gold recovery and minimal cyanide consumption.

The review highlights a global trend towards advanced process control (Liu, 2023) but also notes that many African mines still rely on heuristic methods (Akhabue, 2023). This indicates a significant technology adoption gap. Therefore, there is a need for research that not only develops a sophisticated ML model but also bridges the "last mile" by translating the model into practical, actionable operational guidelines that can be readily adopted by a mine like Wagagai, which currently uses conventional practices.

3. CHAPTER THREE: METHODOLOGY

This study adopted a quantitative, data-driven modeling approach to investigate and optimize gold recovery during cyanide leaching. Machine learning techniques were employed due to their ability to capture complex nonlinear relationships between process variables and metallurgical performance. Recent studies have demonstrated that artificial intelligence models significantly improve predictive accuracy in mineral processing applications (Herrera, 2023; Osei et al., 2024). The methodology consisted of three stages: model development, model optimization, and process optimization.

3.1 Development of machine learning models for the cyanide leaching process for gold recovery.

To accomplish this objective, a structured approach was adopted involving data acquisition, preprocessing, model development, and performance evaluation. The focus was on developing accurate machine learning models capable of predicting gold recovery based on key leaching parameters.

3.1.1 Data collection

The data collection process adopted a data-driven approach to collect data (existing plant records) necessary for modeling and optimizing the cyanide leaching process (Qiu et al., 2025a). A total of 2801 historical leaching records were obtained from Wagagai Mining (U) Limited. These records spanned a full year and half of operations, capturing daily performance data under varying process conditions. Each record included measurements for the following initial process parameters: cyanide concentration (mg/L), leaching Time (hrs.), pH, Pulp Density (%), Dissolved oxygen (mg/L) and Gold Recovery (%).

3.1.2 Data cleaning and processing

The dataset was normalized using z-score standardization to ensure consistent scaling of input variables. (Goodfellow et al., 2016).

$$z = [x - \mu] / \boldsymbol{\sigma}$$

Where: **x** = original value, **μ** = mean of the dataset, **σ** = standard deviation of the dataset.

3.1.3 Data partitioning

The dataset was then divided into training (80%) and testing (20%) subsets using MATLAB's cvpartition (c = cvpartition (n, 'Holdout', p) ;). Normalization improves convergence and enhances model performance, particularly for neural networks (B. Ghobadi1, 2014). A stratified splitting strategy was employed to ensure that the distribution of the target variable is consistent across both sets. Although there is no universal rule for partitioning ratios, Szmigiel et al.(2024) the rule of thumb is that the training dataset should be significantly larger than the validation dataset to capture the full characteristics of the data The remaining 20% was used as the testing dataset to validate the best-performing model. Each model was trained using the training dataset and then evaluated using the test dataset for performance assessment (Durmus et al., 2024).

3.1.4 Feature Scaling

The input variables were standardized to possess a zero mean and unit variance by employing MATLAB's normalization functions. This preprocessing action improved the stability and convergence of the model, particularly for algorithms like Random Forests (RF), which are influenced by feature scaling (Lapshyn et al., 2020). Normalization (Min-Max scaling) was used to rescale all features to a fixed range, typically [0, 1]. This is particularly important for the Artificial Neural Network and improves the convergence of many other algorithms (Wesam S. Bhaya, 2017). The scaling parameters (e.g., min, max) was calculated on the training set and then applied to the test set to prevent data leakage.

3.1.5 Model development

Three machine learning models were developed: Multiple Linear Regression (MLR), Random Forest (RF), and Artificial Neural Network (ANN). MLR served as a baseline model, RF captured nonlinear relationships, and ANN modeled complex interactions between variables.

3.1.5.1 Multiple Linear Regression (MLR)

An MLR model establishes the relationship between one output variable y and one or more input variables x. It was developed using the Ordinary Least Squares (OLS) method because it estimates the relationship between a dependent variable and one or more independent variables. It achieves this by minimizing the sum of squared residuals, which are the differences between observed and predicted values. As noted by Zhang et al., 2022, while MLR often provides a reasonable preliminary fit, it is frequently outperformed by non-linear models due to its inability to capture complex parameter interactions inherent in hydrometallurgical processes.

The formulation of the MLR model to be used in this work is shown in Equation (1) (Zhang et al., 2022)

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n \tag{1}$$

y represents the output of the model, x_i (i=1,2, ...n) represents the input features, and θ is the coefficient term of the input features of the equation.The optimization objective of multiple linear regression is to obtain the optimal coefficient θ vector set corresponding to the minimum value of the sum of distances of the equation to all input data. In this study, a multiple linear regression model containing a constant term was developed for the importance analysis of leaching conditions based on the leaching data (Zhang et al., 2022).

3.1.5.2 Random Forest Regressor (RFR)

A random forest (RF) is a machine learning model, an ensemble of decision trees, which handles classification and regression problems efficiently. For Random Forest regression, MATLAB's TreeBagger was utilized to construct an ensemble of 200 decision trees, integrating out-of-bag (OOB) error evaluation. The model's parameters, including feature subsampling (mtry) and tree depth, were adjusted to balance the bias–variance trade-off. RF was selected for its resilience to noisy data, its ability to mitigate overfitting through bagging, and its provision of variable importance assessments (Zhang et al., 2022).

An RFR model was developed due to its robustness and high performance with tabular data (Ma et al., 2025). Initially, the model was trained with default hyperparameters as a starting point. Its native ability to provide feature importance scores was leveraged for later interpretation, helping to identify the most critical leaching parameters at Wagagai.

Samples were randomly taken from training samples (n*samples) for n times to form a training set. Repeat r times to obtain training sets: D_1, $D_{2..}$,,D_3.

For each training set, k attributes are randomly selected from the attributes set (m×attribute), k=log2m, and then cart trees are established: $f_1(x)$, $f_2(x)$…, $f_r(x)$.

The final prediction value of Random Forest is determined by the average method: (Amankwaa-Kyeremeh et al., 2024)

$$f(x) = \frac{1}{r}\sum_{i=1}^{r} f1(x) \tag{2}$$

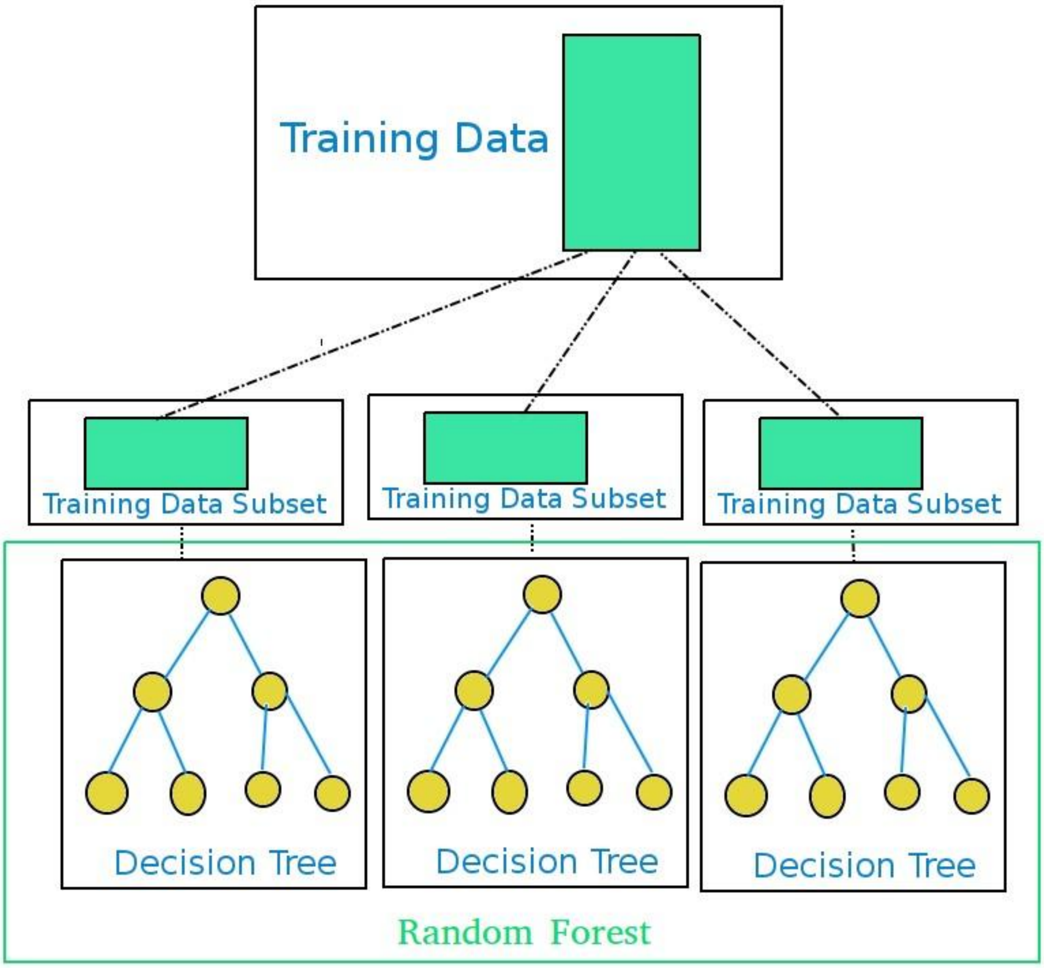


Figure 2 Random Forest regressor

3.1.5.3 Artificial Neural Network (ANN)

ANN is a powerful modeling tool due to its capability to establish a correlation among input and output variables of complex nonlinear system. ANN model is similar relationship with the function of the human nervous system. In its basic form, several layers of neurons are used for the training of ANN that also comprises one or more hidden layers linked with the output layer. These neurons are associated with weights. The network works by transferring the signal from one layer to another till the output layer reached (Madjid et al., 2024).

The Artificial Neural Network (ANN) model for predicting gold recovery was developed using a feed-forward architecture with a 8-10-1 arrangement, meaning it comprises 8 input neurons, 10 hidden layer neurons, and 1 output neuron. This model was trained using the back-propagation learning algorithm (Azizi et al., 2016).

The output of a neuron in an Artificial Neural Network (ANN) was determined by a two-step mathematical expression (Azizi et al., 2016):

1. Weighted Sum (Net Input): First, the neuron calculates a weighted sum of its inputs. This is represented as: z = Σ (w_i * x_i) + b Where:

z is the net input to the neuron.
x_i are the inputs from the previous layer or features.
w_i are the weights associated with each input x_i.
b is the bias term, which allows the activation function to be shifted.
Σ denotes the sum over all inputs.

2. Activation Function: Next, this weighted sum (z) is passed through a non-linear activation function. This introduces non-linearity into the model, allowing it to learn complex patterns. The output of the neuron (a) is then: a = f(z) where:

a is the output (activation) of the neuron.
f is the activation function (e.g., sigmoid, ReLU, tanh, softmax).

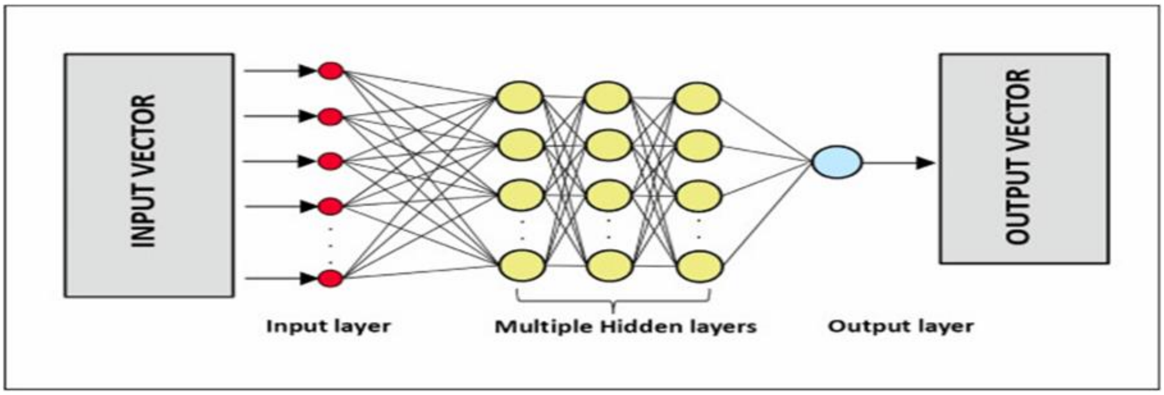


Figure 3 Layout of ANNs

3.1.6 Model Evaluation

The coefficient of determination (R^2), Root Mean Squared Error (RMSE) and mean square error (MSE) were selected as the evaluation criteria to find out the excellent prediction models among the machine learning models (Zhang et al., 2022b).

$$R^2 = 1 - \frac{\Sigma(yi - \hat{y}i)}{\Sigma(yi - \mu y)} \quad (3)$$

Where yi denotes the model output value, ŷi denotes the actual value of the model, μy denotes that the y is the mean value of the model, and i is the ordinal number of the observation, such as 1, 2, 3, ..., n.

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

This measured the average magnitude of the prediction errors, giving a higher weight to large errors. It is expressed in the same units as the target variable (percentage gold recovery), making it highly interpretable. A lower RMSE indicates a better fit.

$$\text{MSE} = \frac{1}{n}\sum_{i=n}^{n}(yi - \hat{y}\text{i}) \tag{5}$$

R^2 was used to represent the performance of the model to evaluate the degree of conformity between the predicted value of the model and the actual value. MSE was used to explain the magnitude of the error. The above three performance indicators effectively evaluated the prediction effect of the models (Zhang et al., 2022b).

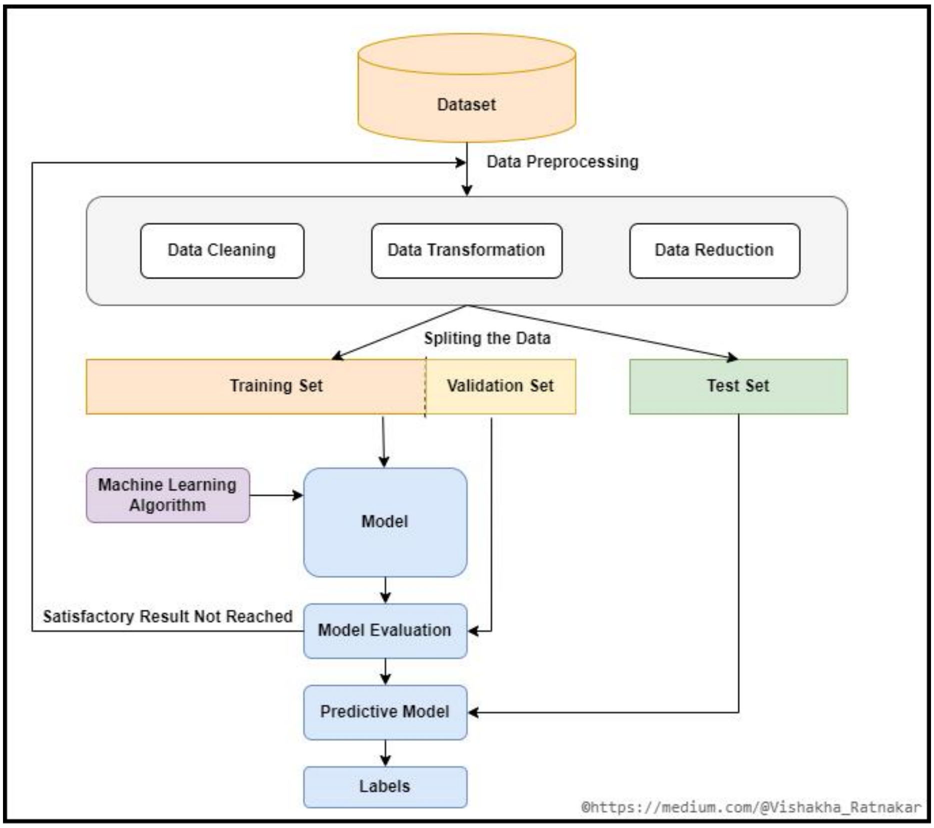


Figure 4 Flowchart for objective one

3.2 Optimization of the best performing developed model.

3.2.1 Selection of the Champion Model for Optimization

The best-performing model, as determined by the highest R^2 and lowest RMSE/MAE on the test set, was selected as the base for optimization. This ensures that computational resources are dedicated to improving the most promising algorithm (Zhang et al., 2022b).

3.2.2 Hyperparameter Tuning Strategy

Hyperparameter tuning was conducted using a grid search approach and different network architectures were tested by, varying the number of hidden layers and neurons in the ANN. This improves model generalization and avoids overfitting. A systematic search for the optimal combination of these hyperparameters was conducted (Matthias Feurer & Frank Hutter, 2019). Hyperparameter spaces were defined for an ANN (e.g., number of hidden layers, neurons per layer, learning rate, activation function).

3.2.2.1 Optimization Algorithm

A Randomized Search Cross-Validation (Randomized Search CV) method was used for hyperparameter tuning. This technique tests a set number of random parameter combinations from the specified search space. Compared to an exhaustive Grid Search, Randomized Search CV is more efficient in terms of computation and has been shown to find high-performing hyperparameters more quickly, with only a minor impact on the final model performance (Fatihah Rahmadayana & Yuliant Sibaroni, 2021).

A predefined number of iterations were set. For each iteration, a random combination of hyperparameters were selected from the search space. The model was trained and evaluated using 5-Fold Cross-Validation on the training set only (eongsoo Han 1, 2021). The average performance across the 5 folds (using RMSE as the primary scoring metric) was recorded for that parameter set.

3.2.3 Evaluation and Selection of the Optimized Model

After the randomized search completes its iterations:

- **Identification of Best Parameters:** The hyperparameter combination that yielded the best average cross-validation score (lowest RMSE) was identified.
- **Retraining the Final Model:** The model was retrained on the entire training dataset (80% of the original data) using this single best set of hyperparameters. This final model incorporates all available training data at its optimal configuration.
- **Final Evaluation:** The performance of this retrained, optimized model was rigorously evaluated on the held-out test set (20% of the original data) that was not used in any part of the tuning process. This provided an unbiased estimate of its real-world performance (Matthias Feurer & Frank Hutter, 2019).

The performance was assessed using the same metrics: R^2, RMSE, and MAE. A successful optimization was demonstrated by a significant improvement in these metrics on the test set compared to the base model.

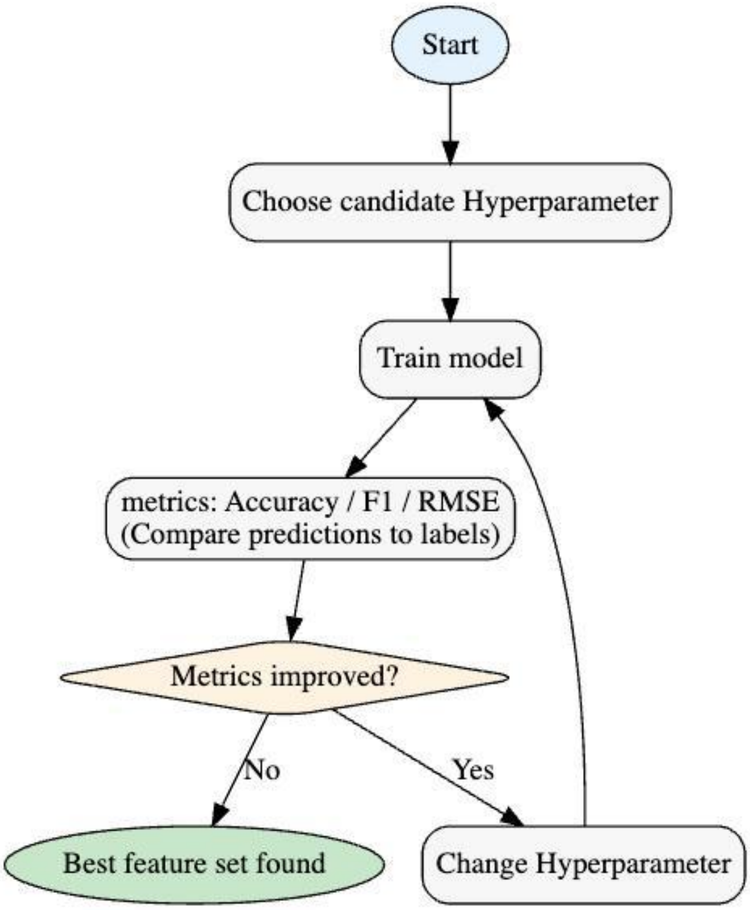


Figure 5 Hyperparameter optimization

3.2.4 Sensitivity Analysis.

The sensitivity analysis was conducted to evaluate the influence of individual cyanide leaching process parameters on the predicted gold recovery Saim, A. K., et al. (2025). The analysis was performed using the optimized Artificial Neural Network (ANN) model, which was developed and validated in the earlier objective. The main goal was to determine the extent to which each input parameter contributes to fluctuations in gold recovery when varied within its operational bounds. To achieve this, a one-at-a-time (OAT) approach was employed. Each parameter was independently varied from its minimum to maximum operational value, while all other parameters were held constant at their respective optimal values Ullah et al., (2023). The ANN model was then used to predict gold recovery for both the minimum and maximum values of each parameter. The difference in predicted recovery was recorded as the maximum recovery change for that parameter. This approach allows for ranking the relative importance of variables and identifying the most influential factors affecting gold recovery (Zhang et al., 2022b).

Sensitivity=Ymax−Ymin

Where:

- Ymax= maximum predicted recovery
- Ymin = minimum predicted recovery

3.3 Application of the optimized developed model in modeling parameter interactions and predicting gold recovery outcomes.

3.3.1 Modeling Parameter Interactions

To understand the complex, non-linear relationships between input variables and the gold recovery output, the model was used as a digital twin of the leaching process. Key techniques included:

3.3.2 Partial Dependence Plots (PDPs)

Partial Dependence Plots (PDPs) were generated using Matlab scripts to visualize the marginal effect of key leaching parameters on the predicted gold recovery, illustrating how the model's output changes as one or two features are varied across their ranges while all other parameters are held constant at their average values (Greenwell, 2017). This technique included one-dimensional PDPs, for instance, to show the relationship between cyanide concentration and recovery, as well as two-dimensional PDPs to reveal the interaction between critical pairs of parameters, such as cyanide concentration and dissolved oxygen, thereby mapping their combined impact on recovery outcomes.

3.3.3 Predicting Optimal Operational Set points

The optimized ANN model was used to simulate various operating conditions in order to determine the combination of input variables that maximizes gold recovery. A grid-based search approach was used, where ranges of input variables were systematically varied. For each combination of parameters, the trained ANN model was used to predict gold recovery. The optimal conditions were identified as those corresponding to the maximum predicted recovery.

A comprehensive synthetic dataset was created using Matlab grid-based search function, encompassing a wide but realistic range of all input parameters (CN^- concentration, pH, pulp density, dissolved oxygen and time). This dataset represents thousands of potential operational scenarios.

The optimized ANN model was used to predict the gold recovery for every scenario in the synthetic dataset. Subsequently, the results were analyzed to identify scenarios that achieve a multi-objective optimum (Qiu et al., 2025b).

3.3.4 Derivation of Operational Recommendations

From this analysis, specific, quantifiable operational recommendations was formulated. For instance:

"The model predicts that for a target recovery of 95%, the optimal parameter range is Cyanide Concentration: 450-550 ppm, pH: 10.5-11.0, and Dissolved Oxygen: 6-8 ppm, given a P80 particle size of 75µm."

4. CHAPTER FOUR: RESULTS AND DISCUSSION

4.1 Development of predictive models for gold recovery

The performance of three models, Multiple Linear Regression (MLR), Artificial Neural Network (ANN), and Random Forest (RF) was evaluated for predicting gold recovery using five selected cyanide leaching variables. Each model was trained on 80% of the dataset and validated using the remaining 20%. Their performance was assessed using R-squared (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) (Zhang et al., 2022b).

Training results table

Model	R2	RMSE	MSE
Linear Regression	0.850929498	1.56765311	2.45753628
Random Forest	0.913271842	1.19573288	1.42977712
Artificial Neural Network	0.931590403	1.06197093	1.12778225

Test results table

Model	R2	RMSE	MSE
Linear Regression	0.870608423	1.54688863	2.3928644
Random Forest	0.919411955	1.22079109	1.4903309
Artificial Neural Network	0.944387266	1.01412892	1.0284575

4.1.1 Multiple Linear Regression (MLR)

The MLR model (Figure 6) exhibited the weakest performance among the three models (Osei et al., 2024).During training, the model achieved an R^2 of 0.851, indicating that it could only explain approximately 85.1% of the variance in the gold recovery data. The associated RMSE and MAE values were 1.568 and 2.458, respectively, reflecting significant deviations between predicted and actual recovery values. On the testing dataset, the MLR model continued to underperform compared to RF and ANN models, achieving an R^2 of 0.87, RMSE of 1.55, and MAE of 2.39. These test results confirm the model's limited predictive capacity and reinforce the conclusion that linear modeling is inadequate for capturing the multi-variable, non-linear dynamics of the cyanide leaching process.

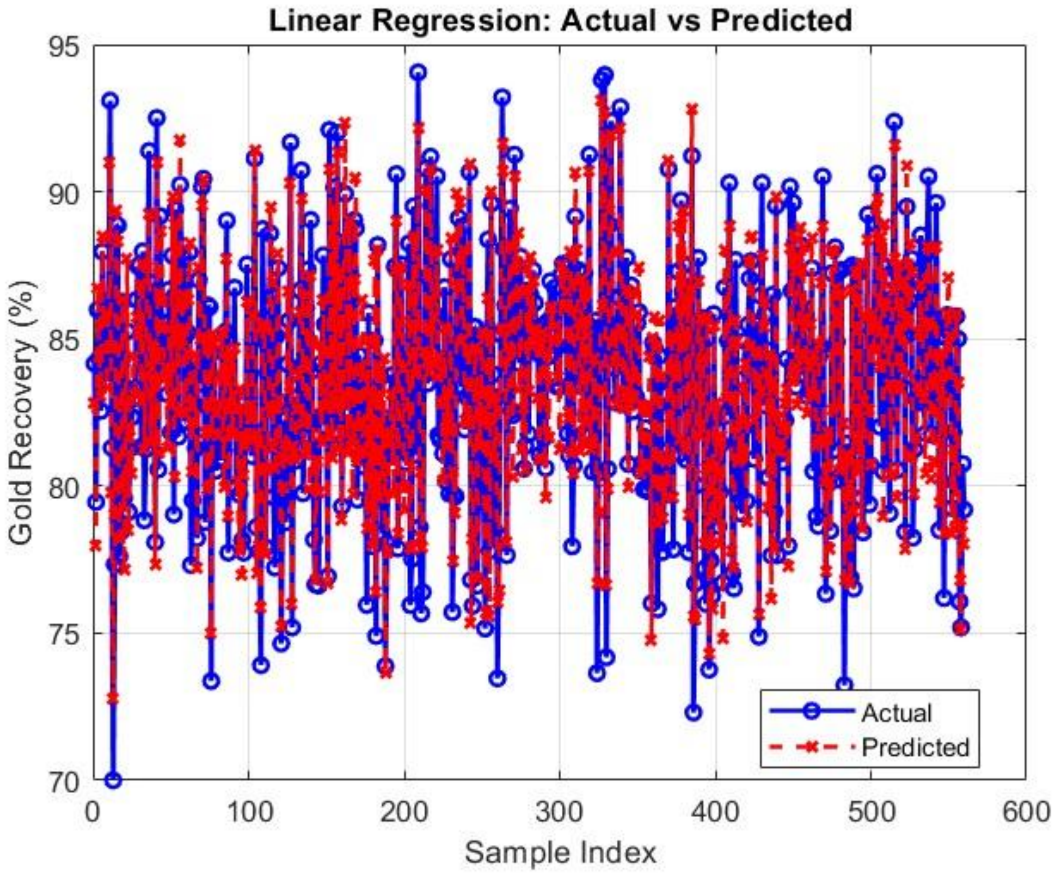


Figure 6 Graph of multiple linear Regression

4.1.2 Random Forest

The Random Forest model (Figure 7) showed improved predictive performance with an R^2 of 0.913 and reduced error metrics of RMSE AND MAE of 1.196 and 1.430 respectively. This indicates that nonlinear relationships and interactions among variables such as cyanide concentration, pH, and dissolved oxygen significantly influence gold recovery Osei et al. (2024). on the testing dataset, the RF model produced an R^2 of 0.92, RMSE of 1.22, and MAE of 0.49. These results confirm its stability and consistent performance across all datasets, reinforcing its capability to generalize well to unseen data. While not as precise as the ANN model, the RF model still demonstrated reliable predictive ability and was notably superior to the MLR model.

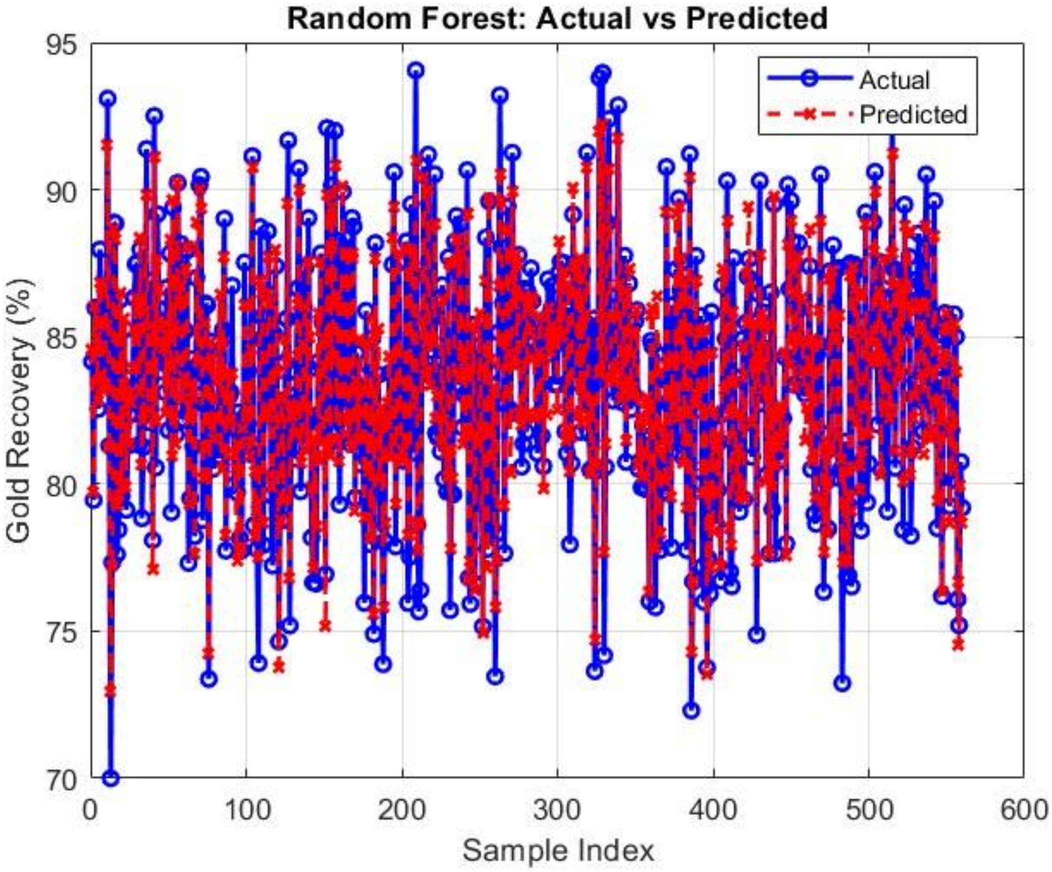


Figure 7 Graph of actual Vs predicted for RF

4.1.3 Artificial Neural Network (ANN)

The Artificial Neural Network (Figure 8) outperformed both Linear Regression and Random Forest and demonstrated the highest accuracy and consistency , achieving the highest R^2 value of 0.932 and the lowest prediction errors of 1.062 (RMSE) AND 1.128 (MAE). This demonstrates its superior ability to capture complex nonlinear relationships within the leaching process (Bascur and Soudek (2021). On the test dataset, the ANN model maintained its performance with an R^2 of 0.944, RMSE of 1.014, and MAE of 1.028. These results confirm the ANN model's robustness and generalization ability, making it the most effective model for predicting gold recovery in this study Monyake et al. (2020).

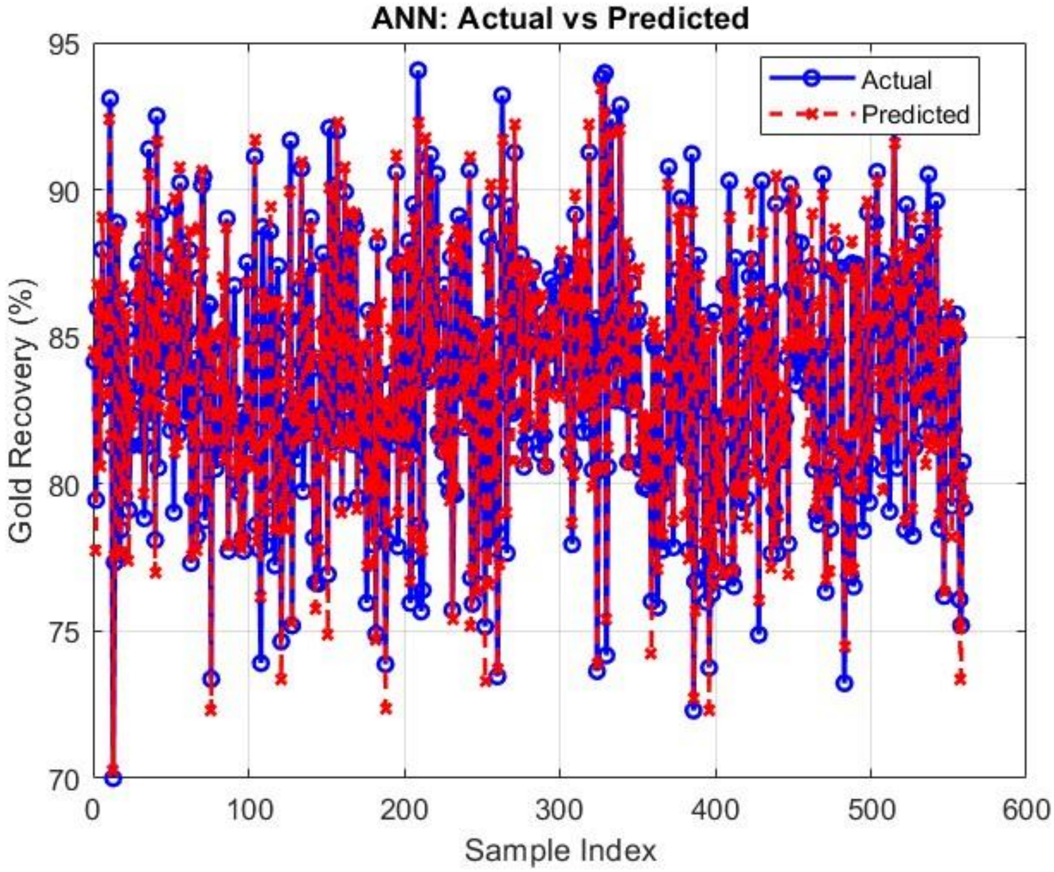


Figure 8 Graph for ANN model showing actual vs predicted gold recovery

4.1.4 Discussion of results for objective one

The results obtained in objective one show a clear performance difference among the three machine learning models developed for predicting gold recovery during cyanide leaching. On the test dataset, the Multiple Linear Regression model achieved an R^2 of 0.8706, the Random Forest model achieved an R^2 of 0.9194, and the Artificial Neural Network achieved the highest R^2 of 0.9444, with the ANN also producing the lowest prediction errors among the three models. This ranking indicates that the ANN provided the most accurate representation of the relationship between cyanide concentration, pH, leaching time, pulp density, dissolved oxygen, and gold recovery in this study.

The relatively weaker performance of the Multiple Linear Regression model is consistent with the nature of cyanide leaching, which is controlled by complex and interdependent physicochemical variables rather than simple additive linear effects. In practical leaching systems, variables such as pH and cyanide concentration interact with dissolved oxygen availability and residence time, while pulp density influences slurry rheology and mass transfer. Because of this, a linear model can provide a useful baseline, but it is less capable of capturing curved responses, threshold effects, and variable interactions. This interpretation agrees with

Osei, (2024), who reported that simpler models were less effective than advanced AI-based approaches for gold cyanide leaching recovery prediction, and with Saldaña et al., (2022), who noted in their review that leaching systems are typically nonlinear and therefore benefit from machine learning approaches that can accommodate complex relationships.

The Random Forest model performed substantially better than Multiple Linear Regression, which suggests that ensemble tree methods are more suitable for this type of hydrometallurgical dataset. Random Forest is able to model nonlinear trends and interaction effects without requiring the strict assumptions of linear regression, and this likely explains its stronger predictive ability in the present study. The improved performance of Random Forest also aligns with recent literature showing that tree-based and other nonlinear machine learning models are effective in mineral-processing prediction tasks. For example, Osei, (2024) developed AI-based models for gold cyanide leaching recovery prediction, while Zhang et al., (2022) showed in a mineral leaching case study that machine learning models can successfully capture nonlinear process behavior in hydrometallurgical systems. The present findings therefore support the argument that nonlinear ensemble methods are more appropriate than purely linear approaches when modeling gold recovery.

Among the three models, the Artificial Neural Network gave the best predictive performance, which indicates that the relationship between the selected leaching variables and gold recovery is highly nonlinear and is best represented by a flexible function approximator. ANNs are especially suitable for problems where several variables act simultaneously and where the effect of one parameter depends on the level of another. In this study, the superior ANN performance suggests that gold recovery at Wagagai-type conditions depends on interaction-rich process behavior rather than isolated variable effects. This result is in agreement with Osei, (2024), whose study on gold cyanide leaching recovery prediction also found strong performance from AI-based modeling approaches, and with the wider review by Saldaña et al., (2022), which identified ANN as one of the most promising tools for modeling leaching systems. It is also consistent with Zhang et al., (2022), who concluded that machine learning methods are well suited to mineral leaching prediction because they can learn complex input–output relationships directly from data.

Based on the comparative evaluation of the three developed models, the Artificial Neural Network was identified as the most effective predictor of gold recovery, followed by Random Forest and Multiple Linear Regression. The superior performance of ANN confirms that the cyanide leaching process is governed by complex nonlinear interactions among process variables, and this agrees with recent studies in leaching and mineral-processing machine learning literature. Therefore, the ANN model was selected as the champion model for optimization in Objective Two.

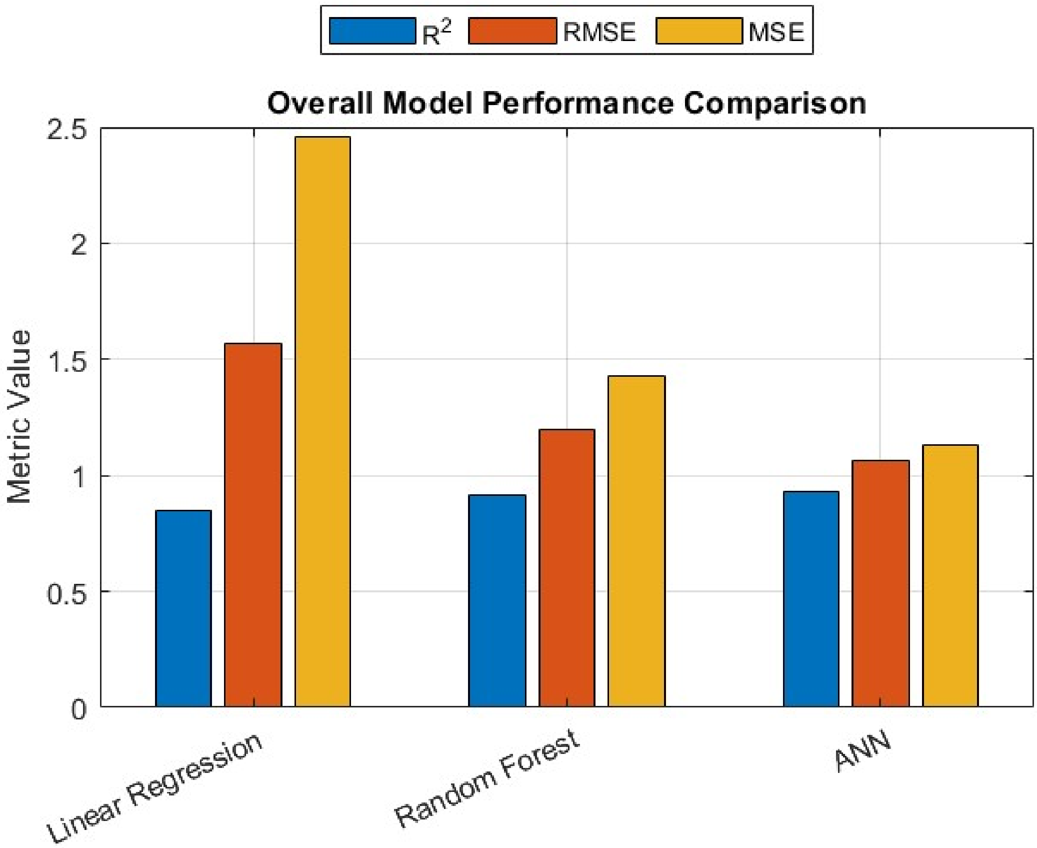


Figure 9 Performance of the models (r2,rmse and mse)

4.2 Optimization of the Best model developed

Based on the comparative evaluation of the three developed models, the Artificial Neural Network was identified as the most effective predictor of gold recovery which achieved the highest R^2 value of 0.932 and the lowest prediction errors of 1.062 (RMSE) AND 1.128 (MAE).

4.2.1 Hyperparameter optimization.

The optimal Artificial Neural Network architecture consisted of a single hidden layer with 10 neurons, achieving the highest R^2 of 0.9436 and the lowest prediction error, RMSE of 1.0209. The optimized ANN model generated trained connection weights and biases which define the mathematical relationship between the input variables and gold recovery prediction. The input-to-hidden layer weights (W_1) represent the influence of cyanide concentration, pH, leaching time, pulp density, and dissolved oxygen on the hidden neurons of the network. Large positive and negative weight values indicate strong nonlinear relationships between process variables and gold recovery. The hidden-layer biases (b_1) shifted neuron activation thresholds, enabling the model to capture complex process interactions.

The hidden-to-output weights (W_2) and output bias (b_2) were then used to combine the hidden neuron outputs into the final predicted gold recovery. The final ANN predictive equation obtained from the optimized network was:

$$Y = W_2 \times \text{tansig}(W_1 X + b_1) + b_2$$

Where;

X = normalized input variables

W_1 = Input to hidden layer weights

b_1 = hidden layer biases

W_2 = hidden to output weights

b_2 = output bias

tansig = nonlinear sigmoid activation function

Y = predicted gold recovery

The use of the tansig activation function enabled the ANN model to effectively capture nonlinear relationships among cyanide concentration, pH, leaching time, pulp density, dissolved oxygen, and gold recovery. This explains the superior predictive performance of the ANN model compared to conventional linear approaches.

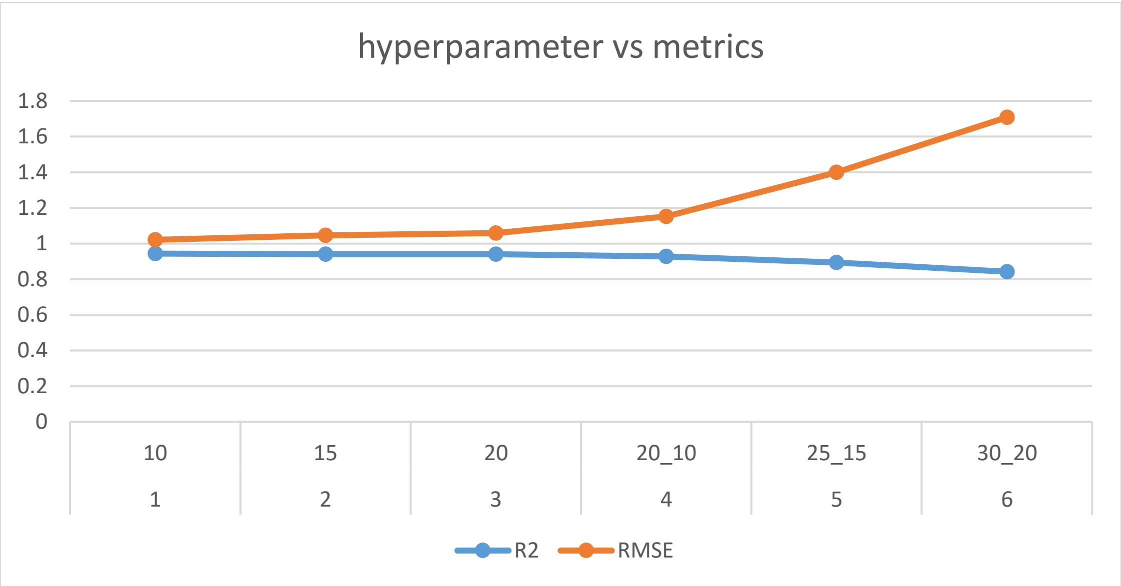


Figure 10 A graph showing hyperparameter tuning.

4.2.2 Sensitivity analysis of process variables.

Variable	Sensitivity
Pulp Density	10.5297
Cyanide Concentration	8.0040
pH	5.5116
Dissolved Oxygen	4.0357
Leaching Time	3.4529

Figure 11 Table showing sensitivity results

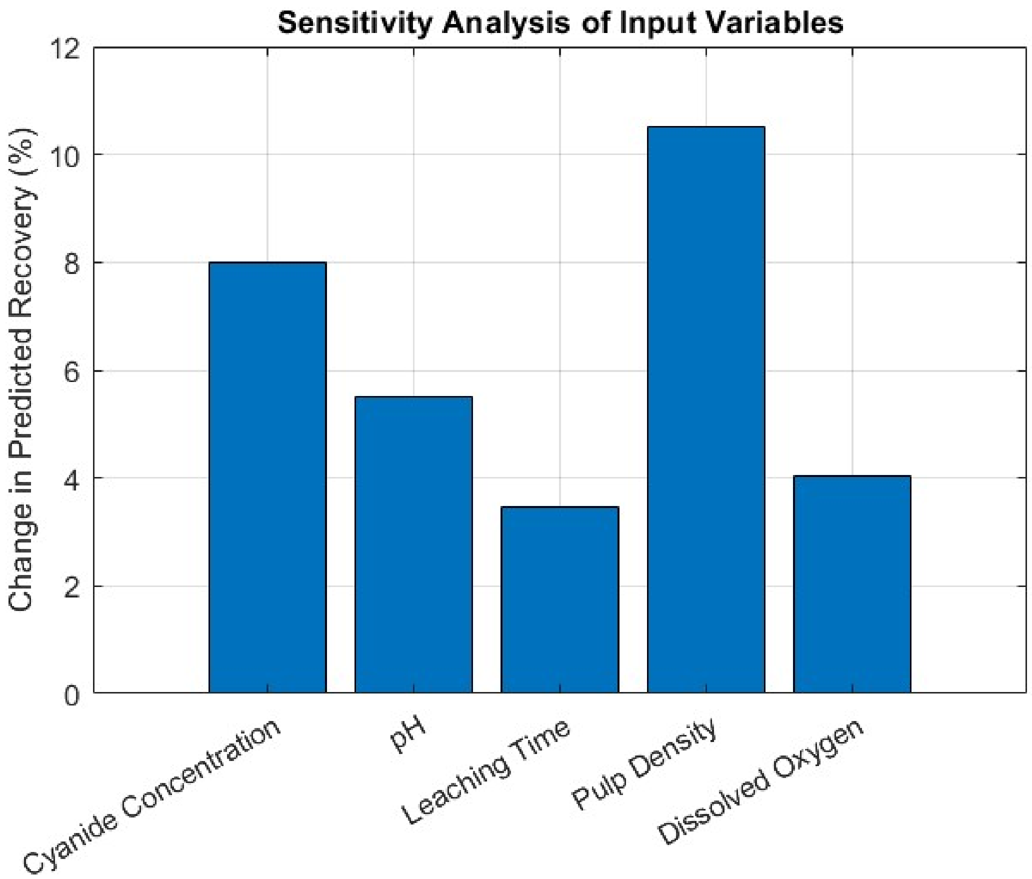


Figure 12 A bar chart for sensitivity analysis

Sensitivity analysis was conducted to quantify the relative influence of each input variable on gold recovery. The results showed that pulp density was the most influential parameter, followed by cyanide concentration, pH, dissolved oxygen, and leaching time.

The high sensitivity of pulp density indicates that it has the greatest impact on recovery, highlighting the importance of slurry characteristics in controlling leaching efficiency. This result is particularly significant because it demonstrates that process efficiency is not governed solely by chemical variables, but also by physical factors such as mass transfer and mixing conditions.

Cyanide concentration was identified as the second most sensitive parameter, confirming its critical role in gold dissolution. The moderate sensitivity of pH and dissolved oxygen further supports their importance in maintaining favorable chemical conditions for cyanidation. Leaching time showed the lowest sensitivity, indicating that while it influences recovery, its effect is less pronounced compared to other variables.

These findings are consistent with recent studies in mineral processing, where sensitivity analysis has been used to identify key parameters influencing recovery. For example, studies on machine learning modeling of gold recovery systems have highlighted the importance of slurry conditions, reagent concentration, and process control variables in determining recovery performance (Saldaña et al., 2022).

4.2.3 Discussion of results for objective two

The optimization of the Artificial Neural Network (ANN) model resulted in a significant improvement in predictive performance, confirming the importance of hyperparameter tuning in machine learning applications for mineral processing. The optimal ANN configuration identified in this study consisted of a single hidden layer with 10 neurons, which achieved an R^2 value of 0.9436 and RMSE of 1.0209. This represents an improvement over the baseline ANN model, demonstrating that proper tuning of network architecture enhances the model's ability to capture complex relationships in cyanide leaching systems.

The improvement observed after optimization is consistent with the fundamental principle that machine learning models are highly sensitive to hyperparameter selection. Parameters such as the number of neurons, network depth, and training configuration directly influence the model's capacity to learn nonlinear relationships. Studies have shown that poorly tuned models may underfit or overfit data, while optimized models achieve better generalization and predictive accuracy (Franceschi, 2024). In this study, increasing the network complexity beyond the optimal configuration (e.g., deeper architectures such as 25–15 or 30–20 neurons) produced lower R^2 values of 0.8940 and 0.8421 respectively, with higher RMSE values, indicating overfitting and loss of generalization. This observation highlights the importance of balancing model complexity with dataset characteristics.

The results of this study are in agreement with recent research in mineral processing and related engineering systems. For example, a study involving machine learning optimization for catalytic and environmental processes demonstrated that optimized models significantly outperform baseline configurations due to improved parameter tuning and model calibration

(Easwaramoorthy, 2023). Similarly, recent hydrological and environmental modeling studies have shown that ANN performance improves substantially after optimization, often outperforming other models when properly tuned (ScienceDirect., 2025). These findings reinforce the conclusion that optimization is not optional but essential for achieving high-performance predictive models.

Another important observation from this study is that optimization improved not only accuracy but also model stability. The optimized ANN maintained high performance across both training and testing datasets, indicating good generalization. This is critical in industrial applications such as cyanide leaching, where models must perform reliably under varying process conditions. Recent studies emphasize that optimized models are more robust and less sensitive to noise, making them more suitable for real-world deployment in mining operations (ResearchGate., 2025).

The optimization results clearly demonstrate that the performance of machine learning models is highly dependent on proper hyperparameter tuning. The optimized ANN model outperformed all other configurations and baseline models, confirming its suitability for modeling the cyanide leaching process. This aligns with recent studies that emphasize the importance of model optimization in achieving high predictive accuracy and reliability in mineral processing applications. Therefore, the optimized ANN was selected as the final model for further analysis in Objective Three.

4.3 Application of the optimized ANN model in modeling parameter interactions and predicting gold recovery outcomes.

4.3.1 Predicting Optimal Operational Set points

The optimized Artificial Neural Network (ANN) model was applied to simulate various operating conditions in order to determine the combination of process variables that maximizes gold recovery. The model predicted a maximum gold recovery of 96.42%, achieved at a cyanide concentration of 800 mg/L, pH of 10.67, leaching time of 48 hours, pulp density of 35%, and dissolved oxygen concentration of 12 mg/L. The ANN recommendation of 800 mg/L cyanide concentration compares favorably with recent experimental investigations. A study by Putri, 2023 on cyanide concentration variation reported gold recoveries of approximately 96.5–97.9% within the range of 700–800 ppm cyanide concentration during 48-hour leaching tests. The predicted leaching time of 48 hours is also supported by previous findings indicating that gold recovery increases rapidly during the initial leaching stages and gradually approaches equilibrium after prolonged contact time. Likewise, the dissolved oxygen concentration of 12 mg/L aligns with established cyanidation theory, which states that oxygen is an essential reactant in the Elsner equation governing gold dissolution. Studies have demonstrated that increased dissolved oxygen availability improves cyanidation efficiency and accelerates gold recovery (Behera, 2025). Similarly, recent cyanidation studies have shown that maintaining pH values above 10.5 enhances gold leaching kinetics and process stability (Umarova, 2025).

This is consistent with the chemistry of cyanide leaching, where gold dissolution occurs through the formation of a soluble aurocyanide complex in the presence of cyanide ions and oxygen. The role of oxygen as an oxidizing agent and the importance of maintaining alkaline conditions to prevent cyanide loss are well established in hydrometallurgical systems (Marsden, 2006).

4.3.2 Partial Dependence Plots (PDPs)

Partial Dependence Plots (PDPs) were used to interpret the behavior of the optimized Artificial Neural Network model by illustrating the relationship between individual cyanide leaching parameters and the predicted gold recovery while holding other variables constant at their mean values. This approach allows for a clear understanding of how each parameter independently influences the recovery process, despite the complexity of the underlying model.

Pulp density PDP

The PDP for pulp density (Figure 13) shows a negative relationship with gold recovery, where recovery decreases as pulp density increases. This indicates that higher solid concentrations adversely affect the leaching process.

The reduction in recovery at higher pulp densities can be attributed to:

- Reduced mass transfer efficiency
- Increased slurry viscosity
- Limited interaction between cyanide solution and gold particles

This finding highlights the importance of maintaining optimal slurry conditions for efficient leaching and is consistent with previous studies emphasizing the role of physical parameters in hydrometallurgical processes Saldaña et al., 2022.

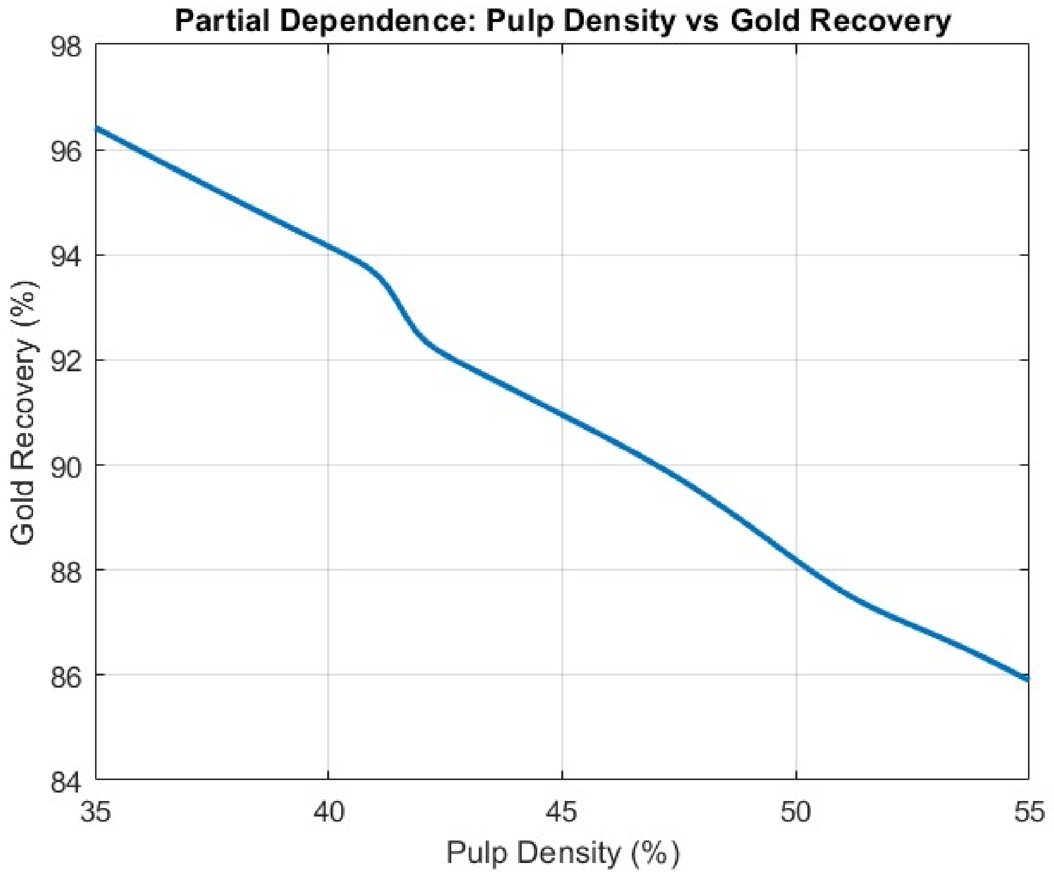


Figure 13 Graph of pulp density vs gold recovery

Cyanide concentration PDP

The PDP for cyanide concentration (Figure 14) shows a progressive increase in gold recovery with increasing cyanide concentration, followed by a gradual leveling off at higher concentrations. This indicates that cyanide plays a critical role in enhancing gold dissolution, as it directly participates in the formation of soluble gold-cyanide complexes. However, the observed plateau suggests that beyond a certain concentration, the system reaches a saturation point where additional cyanide does not significantly improve recovery.

This behavior reflects the fundamental chemistry of cyanidation and aligns with previous studies, which report diminishing returns at higher cyanide dosages due to mass transfer limitations and reagent inefficiency (Marsden, 2006; Zhang et al., 2022).

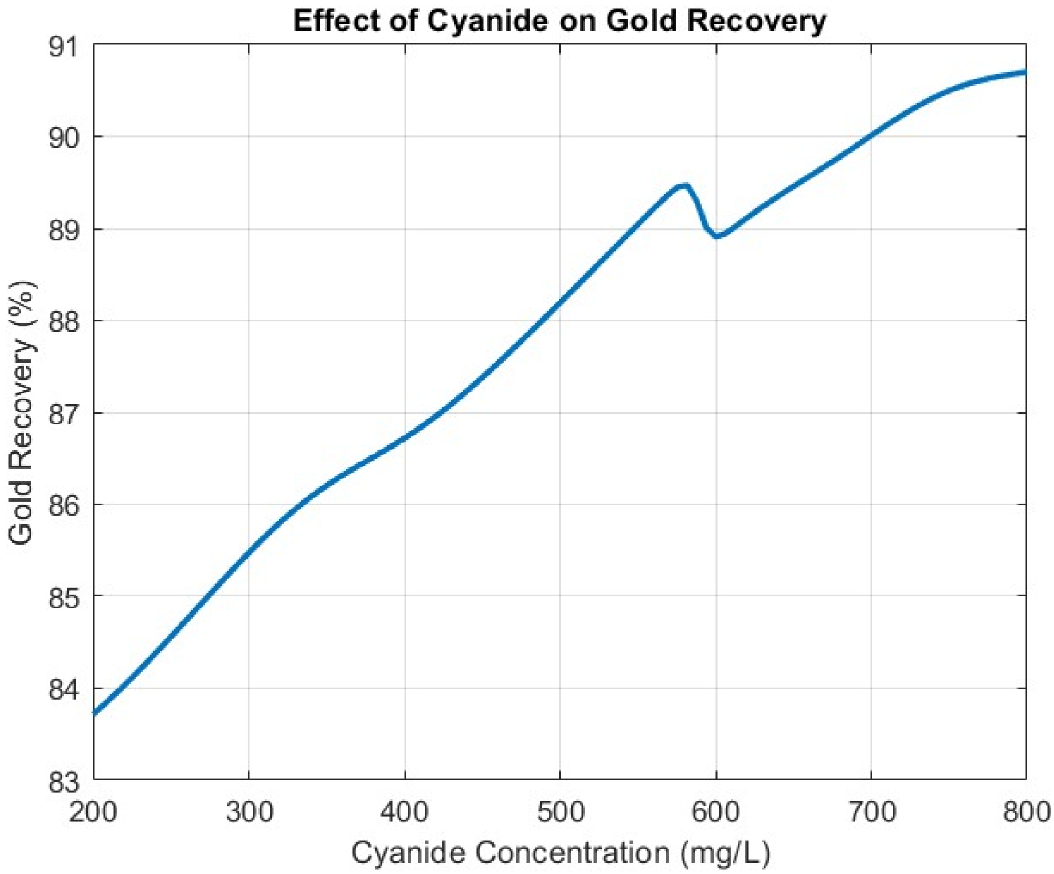


Figure 14 Graph of cyanide effect on recovery

Dissolved oxygen PDP

The PDP for dissolved oxygen (Figure 15) shows a positive relationship between oxygen concentration and gold recovery, confirming its essential role as an oxidizing agent in the cyanidation process. As oxygen concentration increases, recovery improves due to enhanced oxidation of gold.

However, the curve begins to flatten at higher oxygen levels, indicating that once sufficient oxygen is available, further increases do not significantly enhance the reaction. This suggests that oxygen availability becomes non-limiting beyond a certain threshold, which is consistent with cyanide leaching kinetics (Marsden & House, 2006).

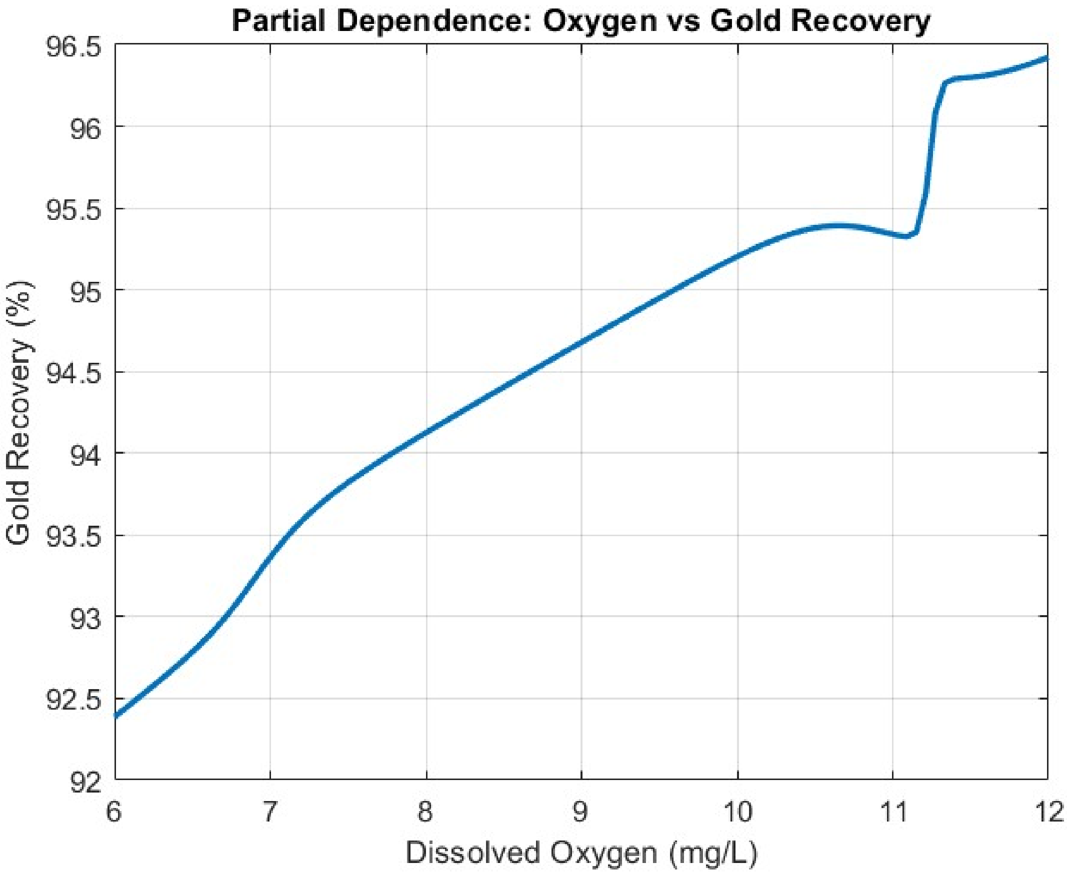


Figure 15 Graph of oxygen effect on recovery

pH PDP

The PDP for pH (Figure 16) demonstrates that gold recovery increases as the pH approaches an optimal alkaline range, after which the recovery stabilizes or slightly declines. The optimal pH range observed in this study is approximately 10.5–10.8, which is consistent with industrial cyanidation practices.

This trend can be explained by the need to maintain alkaline conditions to prevent the formation of hydrogen cyanide gas and to ensure the stability of cyanide ions in solution. At lower pH values, cyanide loss occurs, reducing recovery efficiency, while excessively high pH levels may reduce reaction kinetics. Similar optimal pH ranges have been reported in machine learning-based leaching studies (Osei et al., 2024).

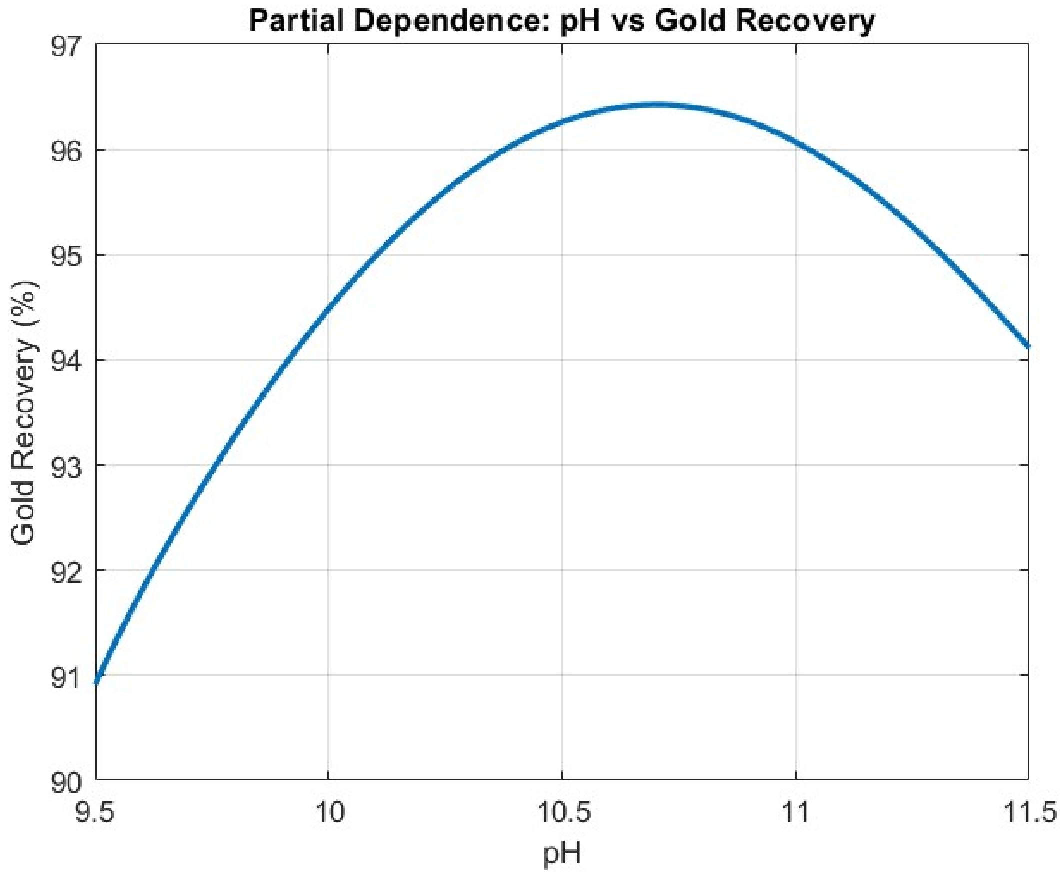


Figure 16 Graph of pH vs recovery

Leaching time PDP

The PDP for leaching time (Figure 17) reveals a gradual increase in gold recovery with time, followed by a plateau. This indicates that longer leaching durations allow more gold to dissolve, but after a certain period, the rate of dissolution decreases as most of the available gold has already reacted.

This behavior reflects typical leaching kinetics, where reaction rates are initially high and then decline as the system approaches equilibrium. The results suggest that excessively long leaching times may not be economically justified, as they yield minimal additional recovery.

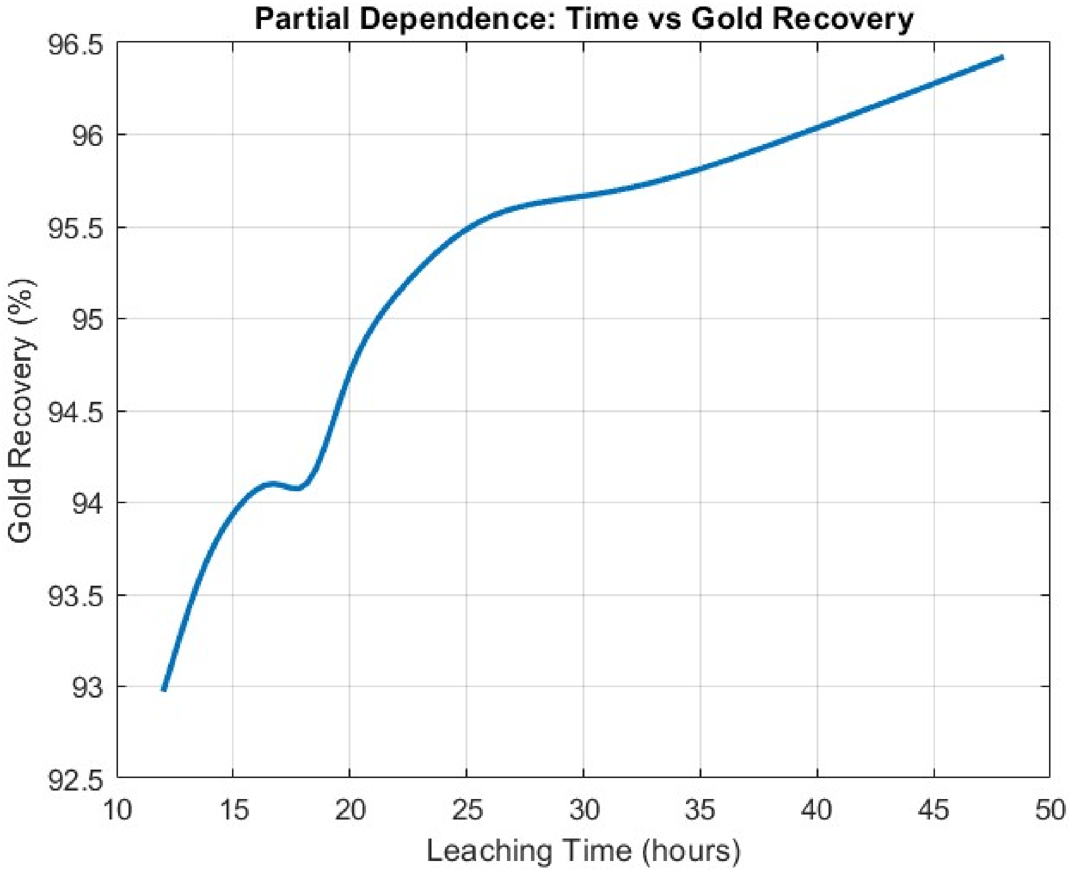


Figure 17 Graph of leaching time on recovery

The PDP analysis provided valuable interpretability to the optimized ANN model, enabling the identification of optimal parameter ranges and reinforcing the understanding of cyanide leaching behavior, thereby supporting data-driven decision-making for improved gold recovery.

4.3.3 Discussion of Results for Objective Three.

Optimization of Gold Recovery using ANN

The optimized Artificial Neural Network (ANN) model was applied to simulate various operating conditions in order to determine the combination of process variables that maximizes gold recovery. The model predicted a maximum gold recovery of 96.42%, achieved at a cyanide concentration of 800 mg/L, pH of 10.67, leaching time of 48 hours, pulp density of 35%, and dissolved oxygen concentration of 12 mg/L.

These results indicate that gold recovery is favored under conditions of high cyanide concentration, optimal alkaline pH, extended leaching time, lower pulp density, and increased oxygen availability. This is consistent with the chemistry of cyanide leaching, where gold dissolution occurs through the formation of a soluble aurocyanide complex in the presence of cyanide ions and oxygen. The role of oxygen as an oxidizing agent and the importance of maintaining alkaline conditions to prevent cyanide loss are well established in hydrometallurgical systems (Marsden & House, 2006).

An important observation is that the optimal cyanide concentration occurred at the upper boundary of the tested range, suggesting that cyanide concentration is a highly influential parameter. However, this also indicates that further increases may not be economically or environmentally sustainable, highlighting the need to balance recovery efficiency with reagent consumption and environmental compliance. Similar findings have been reported in recent studies, where optimization using machine learning models identified high reagent concentration as a key driver of recovery, while also emphasizing cost and sustainability trade-offs (Qiu et al., 2025).

An important observation is that the optimal cyanide concentration occurred at the upper boundary of the tested range, suggesting that cyanide concentration is a highly influential parameter. However, this also indicates that further increases may not be economically or environmentally sustainable, highlighting the need to balance recovery efficiency with reagent consumption and environmental compliance. Similar findings have been reported in recent studies, where optimization using machine learning models identified high reagent concentration as a key driver of recovery, while also emphasizing cost and sustainability trade-offs (Qiu et al., 2025).

Interpretation Partial Dependence Plots (PDPs)

The PDP for cyanide concentration showed an increasing trend followed by a plateau, indicating that recovery improves with increasing cyanide concentration up to a certain point, beyond which additional cyanide yields diminishing returns. This behavior reflects the saturation of reactive sites and aligns with findings reported in machine learning studies on mineral leaching systems, where nonlinear saturation effects are commonly observed (Zhang et al., 2022).

The pH PDP revealed a clear optimal range around 10.5–10.8, beyond which recovery decreases slightly. This confirms that maintaining a controlled alkaline environment is critical for efficient cyanidation. At lower pH values, cyanide can convert to hydrogen cyanide gas, reducing leaching efficiency, while excessively high pH may reduce reaction kinetics. Similar optimal pH ranges have been reported in gold leaching studies and AI-based modeling of metallurgical processes (Osei et al.,, 2024).

The PDP for dissolved oxygen showed a positive relationship with recovery, indicating that increased oxygen availability enhances gold dissolution. However, the rate of increase slows at higher concentrations, suggesting diminishing marginal gains. This is consistent with the role of oxygen as a reactant in the oxidation of gold, as described in cyanidation chemistry.

The leaching time PDP demonstrated a gradual increase in recovery with time, followed by a plateau, indicating that prolonged leaching improves recovery up to a saturation point. This reflects the completion of most dissolution reactions over time, after which additional residence time contributes minimally to recovery.

The pulp density PDP showed a negative relationship with recovery, indicating that higher pulp densities reduce gold recovery. This is attributed to reduced mass transfer efficiency and increased slurry viscosity, which limit the interaction between cyanide solution and gold particles. Similar trends have been observed in mineral processing studies, where high solids concentration negatively affects leaching performance.

5. CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS

5.1 CONCLUSIONS

Three machine learning models (Multiple Linear Regression (MLR), Random Forest (RF), and Artificial Neural Network (ANN)) were developed and evaluated. The results showed a clear performance hierarchy, with the ANN model achieving the highest predictive accuracy (R^2 = 0.932) and the lowest error values on the test dataset. This confirmed that gold recovery in cyanide leaching is governed by complex nonlinear interactions among process variables, which are better captured by advanced machine learning models than by traditional linear approaches.

The ANN model was optimized through hyperparameter tuning. The optimal configuration (a single hidden layer with 10 neurons) achieved improved performance (R^2 = 0.9436 and RMSE = 1.0209), demonstrating that proper model tuning significantly enhances predictive accuracy and generalization capability. The optimization process also revealed that increasing model complexity beyond an optimal level leads to overfitting, emphasizing the importance of balancing model structure with dataset characteristics. Sensitivity analysis further showed that pulp density and cyanide concentration are the most influential parameters, highlighting the combined importance of both physical and chemical factors in the cyanide leaching process.

The optimized ANN model was successfully applied to simulate process conditions, analyze parameter interactions, and identify optimal operating conditions. The model predicted a maximum gold recovery of 96.42%, achieved under specific conditions of cyanide concentration, pH, leaching time, pulp density, and dissolved oxygen. Partial Dependence Plot (PDP) analysis revealed clear nonlinear relationships, optimal operating ranges, and diminishing returns for key variables.

5.2 RECOMMENDATIONS

Future studies should incorporate additional process variables such as particle size (P80), temperature, and ore mineralogy to improve the robustness and applicability of the predictive model. These variables play a critical role in gold cyanidation, as particle size influences surface area for reaction, temperature affects reaction kinetics, and mineralogical composition determines gold accessibility and reagent consumption. Including these parameters would enhance the model's ability to represent real plant conditions more accurately and improve prediction reliability.

Advanced interpretability techniques such as SHAP (SHapley Additive Explanations) should be applied to provide deeper insights into the contribution of each variable to model predictions. Unlike traditional sensitivity analysis, SHAP offers a more detailed, instance-based explanation of model behavior, enabling better understanding of complex nonlinear interactions. Furthermore, future work could explore hybrid modeling approaches that combine machine learning models with process-based or mechanistic models to improve both prediction accuracy and physical interpretability.

It is also recommended that the developed model be validated using controlled laboratory experiments. Validation with laboratory experiments would provide controlled environments to verify specific model predictions. This step is essential to bridge the gap between theoretical modeling and industrial application, ensuring the model's credibility and usability in decision-making.

Finally, future research should explore multi-objective optimization frameworks, such as the Non-dominated Sorting Genetic Algorithm II (NSGA-II), to simultaneously optimize gold recovery and reagent consumption. While this study focused on maximizing recovery, practical mining operations require a balance between efficiency, cost, and environmental impact. Multi-objective optimization would enable the identification of optimal trade-offs, leading to more sustainable and economically viable process optimization strategies.

REFERENCES

1. Ahamed, F. I. (2025). A Comprehensive Guide to Optimizing Machine Learning and Deep Learning Models. *European Journal of Computer Science and Information Technology*, *13*(11), 99–113. https://doi.org/10.37745/ejcsit.2013/vol13n1199113
2. Aizhulov, D., Tungatarova, M., Kurmanseiit, M., & Shayakhmetov, N. (2024). Artificial Neural Networks for Mineral Production Forecasting in the In Situ Leaching Process: Uranium Case Study. *Processes*, *12*(10), 2285. https://doi.org/10.3390/pr12102285
3. Akhabue, C. E. , & O. F. N. (2023). *Optimization of Cyanidation Parameters for Improved Gold Recovery: A Review of Techniques and Challenges in the African Context. Journal of Sustainable Mining*.
4. Alguacil, F. (2006). The Chemistry of Gold Extraction (2nd edition) John O Marsden and C Iain House SME. *Gold Bulletin*, *39*(3), 138–138. https://doi.org/10.1007/BF03215543
5. Amankwaa-Kyeremeh, B., McCamley, C., Ehrig, K., & Asamoah, R. K. (2024). Multiobjective Optimisation of Flotation Variables Using Controlled-NSGA-II and Paretosearch. *Resources*, *13*(11), 157. https://doi.org/10.3390/resources13110157
6. Apley, D. W. , & Z. J. (2020). *Visualizing the effects of predictor variables in black box supervised learning models. Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.
7. Asghar Azizi 1 *, S. Z. S. 1, R. R. 2, A. H. 3, M. P. (2012). *Estimating of gold recovery by using back propagation neural network and multiple linear regression methods in cyanide leaching process*.
8. Azizi, A., Ghaedrahmati, R., Ghahramani, N., & Rooki, R. (2016). Modelling and simulation of the cyanidation process of Aghdareh gold ore using artificial neural network and multiple linear regression. *International Journal of Mining and Mineral Engineering*, *7*(2), 139. https://doi.org/10.1504/IJMME.2016.076497
9. B. Ghobadi1, M. N. S. Z. S. M. U. (2014). *Optimization of cyanidation parameters to increase the capacity of Aghdarre gold mill* .
10. Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
11. Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623
12. Costa, F., de Carvalho Carneiro, C., & Ulsen, C. (2023). *Predicting Gold Accessibility from Mineralogical Characterization Using Machine Learning Algorithms*. https://doi.org/10.2139/ssrn.4591763
13. Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, *6*(2), 182–197. https://doi.org/10.1109/4235.996017
14. East African Community. (2021). *EAC Strategy for the Development of the Mineral Sector. Arusha, Tanzania: EAC Secretariat.*

15. eongsoo Han 1, †, Minuk Jung 2, 3, †, Wonjae Lee 2, Seongmin Kim 3ORCID, Kyoungmun Lee 1ORCID, Geun-tae Lim 4, Ho-Seok Jeon 3, 5, Siyoung Q. Choi 1, * andYosep Han 3, 5. (2021). *Diagnosis and Optimization of Gold Ore Flotation Circuit via Linear Circuit Analysis and Mass Balance Simulation.*
16. Flores, V., & Leiva, C. (2021). A Comparative Study on Supervised Machine Learning Algorithms for Copper Recovery Quality Prediction in a Leaching Process. *Sensors*, *21*(6), 2119. https://doi.org/10.3390/s21062119
17. Friedman, J. H. (2001). *Greedy function approximation: A gradient boosting machine. The Annals of Statistics*.
18. Gail, E., Gos, S., Kulzer, R., Lorösch, J., Rubo, A., Sauer, M., Kellens, R., Reddy, J., Steier, N., & Hasenpusch, W. (2011). Cyano Compounds, Inorganic. In *Ullmann's Encyclopedia of Industrial Chemistry*. Wiley. https://doi.org/10.1002/14356007.a08_159.pub3
19. Goldstein, A. , K. A. , B. J. , & P. E. (2015). *Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. Journal of Computational and Graphical Statistics*.
20. Government of Uganda. (2020). *Third National Development Plan (NDPIII) 2020/21 - 2024/25*. Kampala: National Planning Authority.*
21. Greenwell, B. M. (2017). pdp: An R Package for Constructing Partial Dependence Plots. *The R Journal*, *9*(1), 421. https://doi.org/10.32614/RJ-2017-016
22. Habashi, F. (2005). *A short history of hydrometallurgy. Hydrometallurgy,*.
23. Han, J., Dai, S., Deng, J., Que, S., & Zhou, Y. (2024). Technology for Aiding the Cyanide Leaching of Gold Ores. *Separations*, *11*(8), 228. https://doi.org/10.3390/separations11080228
24. Hastie, T. , T. R. , & F. J. (2009). *The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer*.
25. Jooshaki, M., Nad, A., & Michaux, S. (2021). A Systematic Review on the Application of Machine Learning in Exploiting Mineralogical Data in Mining and Mineral Industry. *Minerals*, *11*(8), 816. https://doi.org/10.3390/min11080816
26. Kavoni, H., Shahidi Pour Savizi, I., Gopalakrishnan, S., Lewis, N. E., & Shojaosadati, S. A. (2025). Machine learning-driven optimization of culture conditions and media components to mitigate charge heterogeneity in monoclonal antibody production: current advances and future perspectives. *MAbs*, *17*(1). https://doi.org/10.1080/19420862.2025.2547084
27. Liu, T. , Y. Q. , W. L. , & W. Y. (2023). *Data-based compensation method for optimal operation setting of gold cyanide leaching process. Journal of Mining Science, 59(4), 721–734*.
28. Lundberg, S. M. , & L. S. I. (2017). *A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems*.
29. Madjid, M. K. A., Noureddine, S., & Amina, B. (2024). *An Introductory Review of Artificial Neural Networks (ANNs)* (pp. 394–406). https://doi.org/10.1007/978-3-031-60632-8_34
30. Ma, G., Chen, J., Li, J., Shi, H., & Chen, Y. (2025). Random Forest Regression May Become the Optimal Regression Model for Osteoarthritis of the Knee in Elderly, in the Context of Embodied Cognition and Psychosomatic Medicine.

Journal of Multidisciplinary Healthcare, *Volume 18*, 4219–4232. https://doi.org/10.2147/JMDH.S519195

31. Manzila, A. N., Moyo, T., & Petersen, J. (2022). A Study on the Applicability of Agitated Cyanide Leaching and Thiosulphate Leaching for Gold Extraction in Artisanal and Small-Scale Gold Mining. *Minerals*, *12*(10), 1291. https://doi.org/10.3390/min12101291
32. Marsden, J. O. , & H. C. I. (2006). *The Chemistry of Gold Extraction (2nd ed.). Society for Mining, Metallurgy, and Exploration.*
33. Matthias Feurer & Frank Hutter. (2019). *Hyperparameter Optimization*.
34. McCoy, J. T., & Auret, L. (2019). Machine learning applications in minerals processing: A review. *Minerals Engineering*, *132*, 95–109. https://doi.org/10.1016/j.mineng.2018.12.004
35. -, M. D. (2024). Optimizing Machine Learning Models: A Data Engineering Perspective. *International Journal For Multidisciplinary Research*, *6*(6). https://doi.org/10.36948/ijfmr.2024.v06i06.32242
36. Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable.*
37. Osei, P. A., Brew, L., Amankwah, R. K., Ziggah, Y. Y., & Owusu, C. (2024). Gold cyanide leaching recovery prediction model based on neighbourhood component analysis and artificial intelligence technique. *Modeling Earth Systems and Environment*, *10*(3), 3865–3880. https://doi.org/10.1007/s40808-024-01970-z
38. Osvaldo A. Bascur. (2021). *Digital Transformation for the Process Industries A Roadmap*.
39. Payam Refaeilzadeh, L. T. & H. L. (2016). *Cross-Validation*.
40. Qiu F, L. W. Z. Y. L. H. C. X. N. J. L. X. S. B. (2025). *Effect of Saccharomyces cerevisiae inoculation on the co-fermentation of Clostridium kluyveri and Clostridium tyrobutyricum: A strategy for controlling acidity and enhancing aroma in strong-flavor Baijiu. Int J Food Microbiol 435:111172*.
41. Qiu, L., Yang, X., Tang, J., & Fan, L. (2025a). Machine learning-driven multi-objective optimization for sustainable, cost-effective, and low-emission gold mining. *Journal of Cleaner Production*, *511*, 145621. https://doi.org/10.1016/j.jclepro.2025.145621
42. Qiu, L., Yang, X., Tang, J., & Fan, L. (2025b). Machine learning-driven multi-objective optimization for sustainable, cost-effective, and low-emission gold mining. *Journal of Cleaner Production*, *511*, 145621. https://doi.org/10.1016/j.jclepro.2025.145621
43. Qiu, L., Yang, X., Tang, J., & Fan, L. (2025c). Machine learning-driven multi-objective optimization for sustainable, cost-effective, and low-emission gold mining. *Journal of Cleaner Production*, *511*, 145621. https://doi.org/10.1016/j.jclepro.2025.145621
44. Saldaña, M., Neira, P., Gallegos, S., Salinas-Rodríguez, E., Pérez-Rey, I., & Toro, N. (2022). Mineral Leaching Modeling Through Machine Learning Algorithms − A Review. *Frontiers in Earth Science*, *10*. https://doi.org/10.3389/feart.2022.816751
45. Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & de Freitas, N. (2016). Taking the Human Out of the Loop: A Review of Bayesian Optimization.

Proceedings of the IEEE, *104*(1), 148–175. https://doi.org/10.1109/JPROC.2015.2494218
46. Takuya Akiba, S. S. T. Y. T. O. M. K. (2019). *Optuna: A Next-generation Hyperparameter Optimization Framework*.
47. Tan Liu; Qingyun Yuan; Lina Wang; Yonggang Wang. (2023). *Data-based Compensation Method for Optimal Operation Setting of Gold Cyanide Leaching Process.*
48. Thanammal Indu, V., Elumalai, M., Anto Gracious, L. A., & R., S. S. (2025). *Optimization Algorithms for Machine Learning Model Enhancement in Healthcare* (pp. 159–192). https://doi.org/10.4018/979-8-3373-5137-7.ch006
49. Tuhumwire, J. T. , & M. F. (2019). *Characterization and Gold Recovery Response of Oxidized Ores from the Busia-Budadiri Belt, Eastern Uganda. Uganda Journal of Geological and Mineral Sciences*.
50. Wesam S. Bhaya. (2017). *Review of Data Preprocessing Techniques in Data Mining*.
51. Zhang, Z., Zhang, X., Zhang, D., Zhang, X., Qiu, F., Li, W., Liu, Z., Shu, J., & Tang, C. (2022a). Application of Machine Learning in a Mineral Leaching Process─Taking Pyrolusite Leaching as an Example. *ACS Omega*, *7*(51), 48130–48138. https://doi.org/10.1021/acsomega.2c06129
52. Zhang, Z., Zhang, X., Zhang, D., Zhang, X., Qiu, F., Li, W., Liu, Z., Shu, J., & Tang, C. (2022b). Application of Machine Learning in a Mineral Leaching Process─Taking Pyrolusite Leaching as an Example. *ACS Omega*, *7*(51), 48130–48138. https://doi.org/10.1021/acsomega.2c06129
53. Behera, S. (2025). *Gold cyanidation process: Leaching and recovery*.
54. Easwaramoorthy, V. S. (2023). *Machine learning algorithms for catalytic performance prediction*.
55. Franceschi, L. (2024). *Hyperparameter optimization in machine learning*.
56. Marsden, J. O. , & H. C. I. (2006). *The Chemistry of Gold Extraction (2nd ed.). Society for Mining, Metallurgy, and Exploration.*
57. Osei, P. A. , B. L. , A. R. K. , Z. Y. Y. , & O. (2024). *Gold cyanide leaching recovery prediction model based on neighbourhood component analysis and artificial intelligence technique. Modeling Earth Systems and Environment*.
58. Putri, R. (2023). *Study on the effect of cyanide concentration on gold and silver recovery during leaching process. Equilibrium Journal of Chemical Engineering*.
59. Qiu, L., Yang, X., Tang, J., & Fan, L. (2025). Machine learning-driven multi-objective optimization for sustainable, cost-effective, and low-emission gold mining. *Journal of Cleaner Production*, *511*, 145621. https://doi.org/10.1016/j.jclepro.2025.145621
60. ResearchGate. (2025). *Hybrid deep learning in mining systems*.
61. Saldaña, M., Neira, P., Gallegos, S., Salinas-Rodríguez, E., Pérez-Rey, I., & Toro, N. (2022). Mineral Leaching Modeling Through Machine Learning Algorithms – A Review. *Frontiers in Earth Science*, *10*. https://doi.org/10.3389/feart.2022.816751
62. ScienceDirect. (2025). *Recent groundwater ML study*.
63. Umarova, I. K. (2025). *Investigation of factors affecting the kinetics of gold heap leachin*.

64. Zhang, Z., Zhang, X., Zhang, D., Zhang, X., Qiu, F., Li, W., Liu, Z., Shu, J., & Tang, C. (2022). Application of Machine Learning in a Mineral Leaching Process─Taking Pyrolusite Leaching as an Example. *ACS Omega*, *7*(51), 48130–48138. https://doi.org/10.1021/acsomega.2c06129

APPENDIX: CODE REPOSITORY

All MATLAB scripts, tools for data preprocessing, and machine learning models utilized in this research can be accessed through the following link:

project codes - Google Docs

This repository contains:

Scripts for data cleaning, partition and normalization

Code for training and evaluating Multiple Linear Regression, Random Forest and Artificial Neural Network (ANN) models.

Scripts for ANN optimization and gold recovery optimization

Scripts for sensitivity analysis and visualization.

If you encounter any access difficulties or need further information, please reach out to the author at reyesganzi50@gmail.com

cyanide_concentration	pH	leaching_time	pulp_density	dissolved_oxygen	gold_recovery
424.7241	11.0852	21.6532	44.9807	7.3629	81.3671
770.4286	11.3158	23.3641	53.6307	8.3378	80.5218
639.1964	11.3874	16.5090	46.6667	9.6387	84.2497
559.1951	11.4203	35.5984	45.7677	10.5320	84.7649
293.6112	10.5429	13.4327	44.0897	11.4099	83.1562
293.5967	11.4546	33.5166	39.1858	11.4928	84.9175
234.8502	11.0146	43.3091	35.0744	10.6598	86.8666
719.7057	9.8233	44.0700	39.8301	7.3700	87.6196
560.6690	10.4538	28.1654	45.1945	6.1489	84.1581
624.8435	10.9367	28.5349	42.5992	8.3026	90.5159
212.3507	9.9947	14.7709	43.9697	6.8493	79.4571

DEPARTMENT OF MINING ENGINEERING

COMPLIANCE REPORT FOLLOWING COMMENTS RAISED ON THE PANEL, AFTER THE CONSULTATION BY BU/UG/2022/1601 ON May 14, 2026 AT 11:00 AM.

NO.	**Comment Raised**	**Response**	**Location**
1.	What data was used and how was it collected?	Methodology section 3.1.1 clearly shows how data was collected.	Page 24
2.	Show the splitting method used.	Methodology section 3.1.3 has been revised to show methods used in splitting the data.	Page 24
3.	Show the validation method used.	Methodology section 3.2.2.1 clearly shows that 5-fold cross validation method was used on the ANN model.	Page 30
4.	The background should provide justification to support your problem.	In chapter one, the background has been revised to justify the problem solved.	Page 10-12
5.	Revise your justification.	In chapter 1.3, justification section has been critically reviewed for better	Page 12-13

		understanding of the study.	

MANUSCRIPT

APPLICATION OF MACHINE LEARNING TO OPTIMIZE GOLD RECOVERY DURING CYANIDE LEACHING AT WAGAGAI GOLD MINE.

CASE STUDY: WAGAGAI MINING (U) LIMITED, BUSIA DISTRICT

Ayebale Reyes[1.1], Dr. Nangendo Jacqueline [1.2], Mr. Kwikiriza Mathias[1.3]

[1.1]Department of Mining Engineering, Busitema University, P.O. Box, 236 Tororo- Uganda; reyesganzi50@gmail.com

[1.2] Department of Mining Engineering, Busitema University, P.O. Box, 236 Tororo- Uganda; jackie.nangendo@gmail.com

[1.3] Department of Mining Engineering, Busitema University, P.O. Box, 236 Tororo- Uganda; kwikirizamathias616@gmail.com

ABSTRACT

This study applied machine learning techniques to predict and optimize gold recovery during cyanide leaching at Wagagai Mining (U) Limited. A dataset of 2801 historical plant records was developed using key process variables including cyanide concentration, pH, leaching time, pulp density, and dissolved oxygen. Three machine learning models, namely Multiple Linear Regression (MLR), Random Forest (RF), and Artificial Neural Network (ANN), were developed and evaluated using the coefficient of determination (R^2), Root Mean Square Error (RMSE), and Mean Squared Error (MSE).

The ANN model achieved the best predictive performance with an R^2 of 0.9444 and RMSE of 1.0209, outperforming RF ($R^2 = 0.9194$) and MLR ($R^2 = 0.8706$). Hyperparameter optimization revealed that a single hidden layer with 10 neurons provided the optimal ANN architecture. The optimized ANN predicted a maximum gold recovery of 96.42% at a cyanide concentration of 800 mg/L, pH of 10.67, leaching time of 48 hours, pulp density of 35%, and dissolved oxygen concentration of 12 mg/L. Partial Dependence Plots and sensitivity analysis showed that pulp density and cyanide concentration were the most influential variables affecting recovery.

The findings demonstrate that ANN-based models provide effective tools for optimizing cyanide leaching processes and improving operational efficiency in gold mining.

Keywords: Gold recovery, cyanide leaching, machine learning, ANN, Random Forest, process optimization.

1. INTRODUCTION

Gold cyanidation remains the dominant hydrometallurgical process used for extracting gold from low-grade ores due to its efficiency and economic viability (Marsden, 2006). However, the cyanide leaching process is governed by complex nonlinear interactions among operational variables such as cyanide concentration, pH, dissolved oxygen, pulp density, and leaching time (Azizi et al., 2016). Traditional optimization methods used in many gold-processing plants rely heavily on operator experience and trial-and-error approaches, which often result in reduced recovery efficiency, increased cyanide consumption, and higher operational costs (Han et al., 2025).

Recent advances in machine learning have demonstrated strong potential in modeling and optimizing complex mineral processing systems (McCoy & Auret, 2019). Machine learning techniques can capture nonlinear relationships and hidden interactions among process variables more effectively than conventional empirical methods. Artificial Neural Networks (ANNs), Random Forests (RFs), and other machine learning algorithms have been successfully applied in mineral processing for predictive modeling and optimization (Flores & Leiva, 2021; Aizhulov et al., 2024). Yet, limited research has focused on gold recovery optimization in East Africa, particularly Uganda, where artisanal and small-scale mining often relies on heuristic methods (Akhabue, 2023).

At the Wagagai Gold Mine, the beneficiation plant utilizes cyanide leaching as its primary extraction method. Current operations are guided by conventional practices. There is a recognized opportunity to enhance recovery efficiency and reduce cyanide consumption through a more scientific, data-oriented approach. This study aimed to develop machine learning models for predicting gold recovery, optimize the best-performing model, and apply the optimized model to identify optimal operational conditions for cyanide leaching.

2. MATERIALS AND METHODS

2.1 Data Collection

A total of 2801 historical cyanide leaching records were obtained from Wagagai Mining (U) Limited. Each record included measurements for the following initial process parameters: cyanide concentration (mg/L), leaching Time (hrs.), pH, Pulp Density (%), Dissolved oxygen (mg/L) and Gold Recovery (%).

2.2 Data Preprocessing

The dataset was cleaned and normalized using z-score standardization to improve model stability and convergence (Kavoni et al., 2025). The data were partitioned into training (80%) and testing (20%) subsets using MATLAB's cvpartition function.

2.3 Model Development

Three machine learning models were developed: Multiple Linear Regression (MLR), Random Forest (RF), and Artificial Neural Network (ANN). MLR served as a baseline model, RF captured nonlinear relationships, and ANN modeled complex interactions between variables.

2.4 Model Evaluation

The models were evaluated using Coefficient of determination (R^2), Root Mean Square Error (RMSE) and Mean Squared Error (MSE).

2.5 Hyperparameter Optimization

The ANN model was optimized by varying Number of hidden neurons, Hidden layer configurations and Training parameters.

The optimal ANN architecture consisted of one hidden layer with 10 neurons.

2.6 Sensitivity Analysis

Sensitivity analysis was performed using a one-at-a-time approach by varying each process parameter independently while keeping other variables constant.

2.7 Partial Dependence Plots (PDPs)

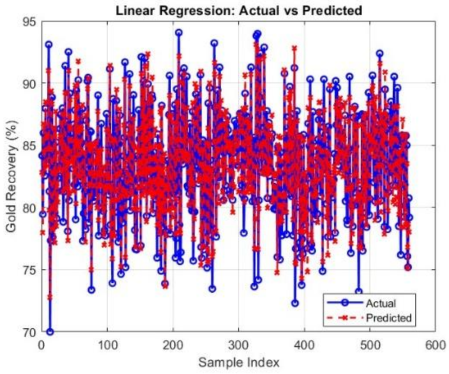
PDPs were generated to analyze the marginal influence of each process variable on predicted gold recovery.

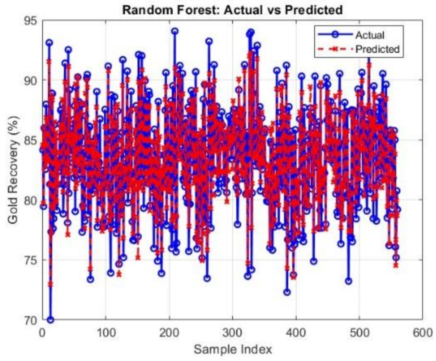
3. RESULTS AND DISCUSSION

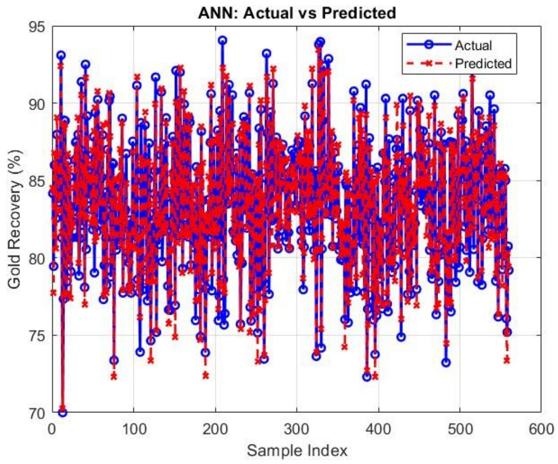
3.1 Model Performance

The ANN model achieved the highest predictive performance compared to RF and MLR.

Model	**R^2**	**RMSE**
MLR	0.8706	1.5469
RF	0.9194	1.2208
ANN	0.9444	1.0141







The superior performance of ANN demonstrates its ability to capture nonlinear interactions among cyanide leaching variables.

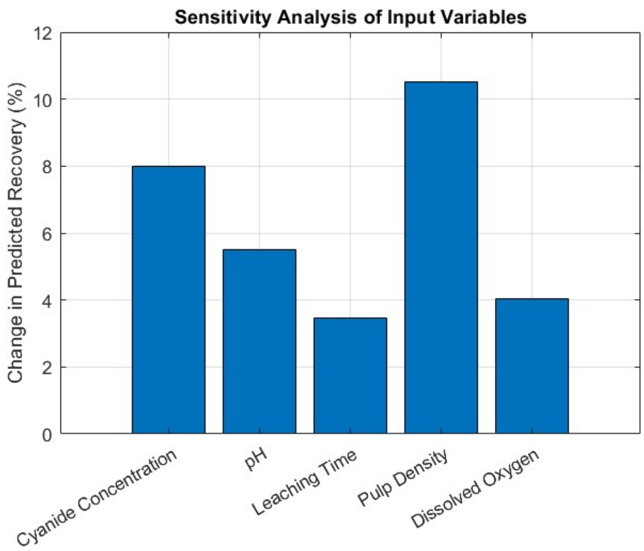
3.2 Hyperparameter Optimization

The optimized ANN architecture consisted of 10 neurons in a single hidden layer and achieved $R^2 = 0.9436$ and RMSE = 1.0209.

Increasing network complexity beyond this configuration resulted in overfitting and reduced model generalization.

3.3 Sensitivity Analysis

Sensitivity analysis showed that Pulp density was the most influential variable, Cyanide concentration was the second most sensitive parameter, PH and dissolved oxygen showed moderate influence and Leaching time had the least influence.



These findings confirm the combined importance of physical and chemical process variables in cyanidation.

3.4 Optimization Results

The optimized ANN predicted a maximum gold recovery of 96.42% under the following conditions:

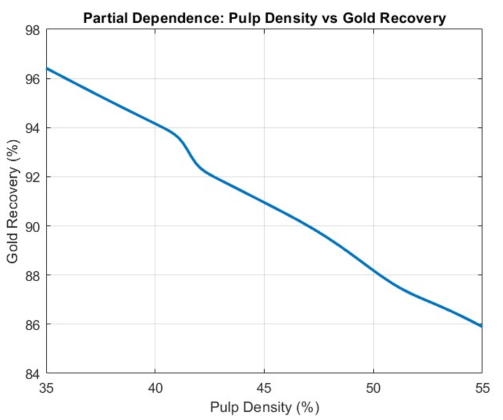
Parameter	Optimal Value
Cyanide concentration	800 mg/L
pH	10.67
Leaching time	48 hours
Pulp density	35%
Dissolved oxygen	12 mg/L

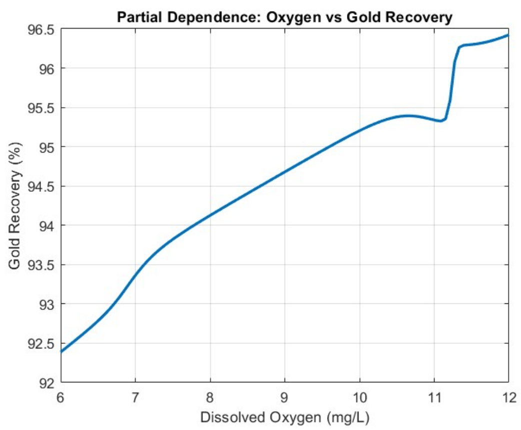
The results demonstrate that maintaining optimal alkaline conditions and sufficient oxygen availability significantly improves cyanidation efficiency.

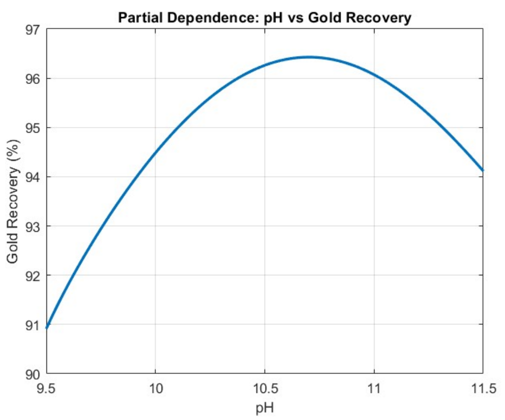
3.5 Partial Dependence Plot Interpretation

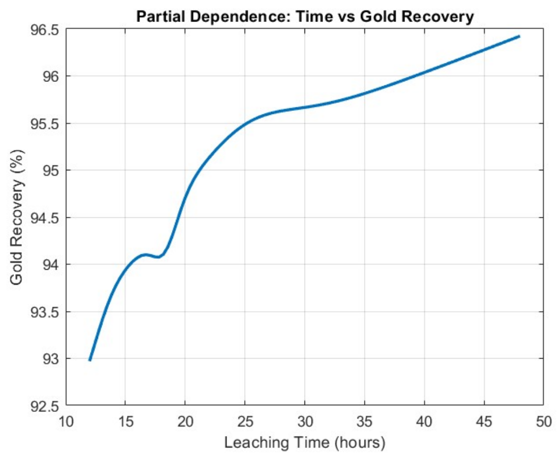
The PDP analysis revealed that:

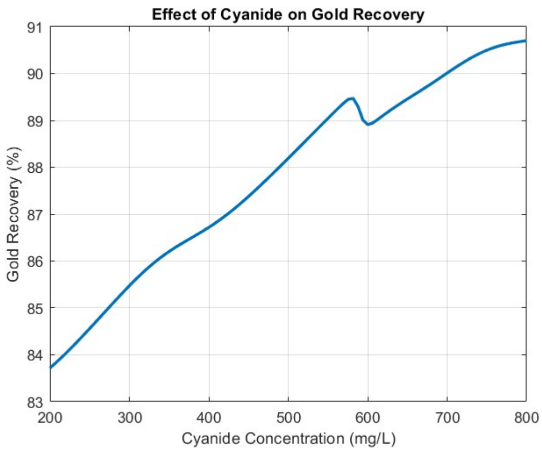
- Gold recovery increases with cyanide concentration before reaching saturation.
- Optimal pH lies between 10.5 and 10.8.
- High pulp density negatively affects recovery.
- Dissolved oxygen positively enhances recovery.
- Excessively long leaching times yield diminishing returns.











4. CONCLUSION

This study successfully demonstrated the application of machine learning techniques for predicting and optimizing gold recovery during cyanide leaching. Among the developed models, ANN achieved the best predictive performance with an R^2 of 0.9444.

The optimized ANN predicted a maximum recovery of 96.42% and identified pulp density and cyanide concentration as the most influential process variables. Findings confirm that ANN models are superior for modeling nonlinear interactions in cyanide leaching. Compared to traditional methods, ML provides actionable insights for operators, reducing reliance on trial-and-error approaches. This aligns with global trends toward AI-driven process optimization (Qiu F, 2025; Liu, 2023).

For Wagagai Mine, adopting ML-based optimization could increase recovery rates from 75% to over 95%, reduce cyanide consumption, and minimize environmental risks.

5. RECOMMENDATIONS

Future studies should:

1. Incorporate additional variables such as temperature, particle size, and ore mineralogy.
2. Validate the developed model using real-time plant data.
3. Apply advanced interpretability techniques such as SHAP.
4. Explore multi-objective optimization methods such as NSGA-II.

REFERENCES

Aizhulov, D., Tungatarova, M., Kurmanseiit, M., & Shayakhmetov, N. (2024). Artificial Neural Networks for Mineral Production Forecasting in the In Situ Leaching Process: Uranium Case Study. *Processes*, *12*(10), 2285. https://doi.org/10.3390/pr12102285

Akhabue, C. E. , & O. F. N. (2023). *Optimization of Cyanidation Parameters for Improved Gold Recovery: A Review of Techniques and Challenges in the African Context. Journal of Sustainable Mining*.

Azizi, A., Ghaedrahmati, R., Ghahramani, N., & Rooki, R. (2016). Modelling and simulation of the cyanidation process of Aghdareh gold ore using artificial neural network and multiple linear regression. *International Journal of Mining and Mineral Engineering*, *7*(2), 139. https://doi.org/10.1504/IJMME.2016.076497

Flores, V., & Leiva, C. (2021). A Comparative Study on Supervised Machine Learning Algorithms for Copper Recovery Quality Prediction in a Leaching Process. *Sensors*, *21*(6), 2119. https://doi.org/10.3390/s21062119

Han, X., Yao, H., Zhan, F., Song, X., Zhao, J., & Zhao, H. (2025). A new framework for X-ray absorption spectroscopy data analysis based on machine learning: *XASDAML*. *Journal of Synchrotron Radiation*, *32*(5), 1244–1256. https://doi.org/10.1107/S1600577525005351

Kavoni, H., Shahidi Pour Savizi, I., Gopalakrishnan, S., Lewis, N. E., & Shojaosadati, S. A. (2025). Machine learning-driven optimization of culture conditions and media components to mitigate charge heterogeneity in monoclonal antibody production: current advances and future perspectives. *MAbs*, *17*(1). https://doi.org/10.1080/19420862.2025.2547084

Liu, T. , Y. Q. , W. L. , & W. Y. (2023). *Data-based compensation method for optimal operation setting of gold cyanide leaching process. Journal of Mining Science, 59(4), 721–734*.

Marsden, J. O. , & H. C. I. (2006). *The Chemistry of Gold Extraction (2nd ed.). Society for Mining, Metallurgy, and Exploration.*

McCoy, J. T., & Auret, L. (2019). Machine learning applications in minerals processing: A review. *Minerals Engineering*, *132*, 95–109. https://doi.org/10.1016/j.mineng.2018.12.004

Qiu F, L. W. Z. Y. L. H. C. X. N. J. L. X. S. B. (2025). *Effect of Saccharomyces cerevisiae inoculation on the co-fermentation of Clostridium kluyveri and Clostridium tyrobutyricum: A* *strategy for controlling acidity and enhancing aroma in strong-flavor Baijiu. Int J Food Microbiol 435:111172.*