National College of Ireland

Attention-based Deep Learning Prediction of Silica Concentrate for Quality Optimization in Mineral

MSc in Data Analytics

Fardeen Mohammad
Student ID: 24109878

School of Computing
National College of Ireland

Supervisor: Vikas Tomer

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Fardeen Mohammad

Student ID: 24109878

Programme: MSc in Data Analytics **Year:** 2025

Module: MSc Research Practicum Part-2

Supervisor: Vikas Tomer

Submission Due Date: 27-01-2026

Project Title: Attention-Based Deep Learning Prediction of Silica Concentrate

Word Count: 8,914 **Page Count** 21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Fardeen Mohammad

Date: 27-01-2026

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Attention-based Deep Learning Prediction of Silica Concentrate for Quality Optimization in Mineral Processing

Fardeen Mohammad
x24109878@student.ncirl.ie
MSc in Data Analytics

Abstract

The prediction of silica and impurity concentration is essential for controlling mineral processing operations, where fluctuations in feed composition and process conditions significantly affect product quality and energy efficiency. Traditional laboratory-based measurement methods are accurate but slow, creating delays that hinder real-time decision-making and operational responsiveness. Although machine learning and deep learning models have been increasingly applied to address these challenges, many existing approaches rely on predefined feature sets or dimensionality reduction techniques, limiting their ability to capture high-order and dynamic feature interactions. Furthermore, conventional models often lack transparency and struggle to generalize across variable industrial conditions. This study introduces an attention-based AutoInt framework designed to overcome these limitations by automatically learning low- and high-order interactions among process variables using multi-head self-attention. The leveraging approach transforms each operational parameter such as reagent dosage, pulp density, air flows, and flotation column indicators into learnable embeddings, enabling adaptive modelling of complex industrial behaviour. In addition, the research a comprehensive pipeline that compares AutoInt with traditional machine learning models and a deep learning architecture using real iron ore flotation plant data. Experimental results show that AutoInt significantly outperforms all baseline models, achieving the lowest MAE value 0.3601, MSE value 0.2774, and RMSE value 0.5267, along with the highest R^2 score of 0.7777. These findings confirm AutoInt's superior ability to capture nonlinear and high-order dependencies, making it more robust under fluctuating industrial conditions. The model's attention mechanisms also provide, revealing which variables most strongly influence silica concentration.

1 Introduction

Silicon and impurity concentration prediction is a major problem to be faced in mineral processing and metallurgical industries due to the fact that they directly influence product quality, metallurgical productivity and the amount of energy used in the downstream refining processes. Conventional measurements which are accurate but slow and require delays that do not make it responsive to real-time operations (Nagarjuna and Ananthan, 2022). Such delays are problematic in dynamic flotation systems where feed composition, aeration conditions, and

reagent dosages vary continuously and timely prediction of impurities is necessary to achieve the consistency in the products and to sustain a good performance of the plants.

In this case, the application of smart and analytic systems capable of providing fast, accurate, and dynamic predictions in various industrial environments is becoming a necessity. The use of machine learning and deep learning models has become an appealing option because it can learn the nonlinear relation on the complex process variables and respond to the plant dynamic conditions (Chen et al., 2025). In addition, the current increasing availability of high frequency sensor data in the mining plant provides an ideal foundation on which predictive models can be implemented to reduce reliance on the slow laboratory tests and focus the real-time decision-making process.

The previous two papers have talked about the diversity of machine learning and deep learning techniques, which can be used to predict the impurity and optimize the process. Machine learning random Forest, XGBoost, and ensemble feature ranking models have shown positive outcomes in the identification of nonlinear trends in flotation and smelting data (Nagarjuna and Ananthan, 2022; Anjum et al., 2022; Nistala et al., 2024). LSTMs, CNNs, and hybrid MI-PCA architectures have been studied in deep learning to enhance the ability to predict the future and adjust to the changing conditions of the process (Chen et al., 2025; Hao et al., 2025). Metallurgical systems have also been implemented into hybrid intelligence frameworks that combine dimensionality reduction, regularization, and optimization algorithms (Cherkaev et al., 2024; Kolodziej et al., 2025; Kraiem et al., 2024). Nevertheless, a lot of these approaches are based on predefined sets of features, they lack interpretability or ability to learn high-order interaction, and more flexible and transparent modelling frameworks are possible.

Research Question:

How can attention-based models such as AutoInt be used to automatically learn high-order feature interactions and improve the prediction accuracy of silica concentration?

Justification:
The research question is specifically related to the limitations that have been found in current machine learning and deep learning approaches, especially their inability to capture complex feature interactions and give explanations in real-time industrial contexts.

This study has a number of contributions. First, it presents an attention-based AutoInt framework to the mineral processing field, which can automatically discover both low- and high-order interactions of features that have an impact on the silica concentration. Second, it gives an extensive comparison of the traditional machine learning, deep learning, and AutoInt models on real flotation industrial data. Lastly, the paper shows that AutoInt is much more effective than current methods in terms of predictive accuracy that provides a more dynamic and predictable framework of predicting impurity real-time and controlling processes.

2 Related Work

Silicon and impurity prediction is a necessary procedure in the efforts to simplify the process of mineral processing and metallurgical operations because it directly influences the quality of products, energy use, and environmental results. Traditional silicon concentration measurement methods, which are carried out in laboratories, are not only time-consuming, but also, they slow down process control and decision-making. To overcome these limitations, current studies have taken into account machine learning (ML) and deep learning (DL) techniques of real-time estimation and prediction of process parameters. These intelligent models can acquire complex non-linear association among variables and react to the evolving industrial variables. However, along with significant successes, the adaptability remains to be enhanced, and predictive accuracy with the help of more advanced system designs such as attention-based models.

2.1 Machine Learning Models for Process Prediction

Machine learning is one of the essential tools being applied in the mineral processing and metallurgical industries to increase the efficiency and product quality. In their estimation of silicon concentration application, Nagarjuna and Ananthan (2022) employed the Random Forest Regressor, Support Vector Regressor, and Extreme Gradient Boosting models to estimate silicon concentration in the froth flotation process. Their article demonstrated that complex, multidimensional data may be effectively dealt with by ensemble-based approaches. Principal Component Analysis was also incorporated to reduce the redundancy in the data, as well as, enhanced the generalization capability of the model. The present research paper has concluded that the conventional methods of laboratory analysis can be substituted by the ML models that provide fast, accurate, and cost-effective estimates of the silicon concentration in industrial processes.

The role of ensemble machine learning techniques has also been widely found to be used in other material and process prediction tasks. Anjum et al. (2022) also made use of diverse ensemble models, including Gradient Boosting, Random Forest, Bagging Regressor, and AdaBoost Regressor, to calculate the compressive strength of fibre-reinforced nano-silica concrete. Their results showed that ensemble learning techniques could effectively bring out nonlinear relationships between material parameters and performance results. Cherkaev et al. (2024) have also developed a hybrid machine learning model to forecast silicon content of ferrochrome smelting. The dimensionality reduction combined with regularized regression boosting helped the model to identify the most meaningful process parameters such as the concentration of carbon and titanium, and the resistance of the electrode. These findings confirm that the combination and the hybrid ML methods are able to improve the accuracy of the prediction in the case of metallurgical modelling.

Other advancements in the field have been to streamline the model performance and feature selection. Wang, Hu, and Tang (2021) came up with a Mult objective evolutionary nonlinear ensemble learning framework in which an evolutionary feature selection mechanism was

employed to forecast silicon content in blast furnaces. Their optimization algorithm considered the accuracy of the model and its diversity that made possible the construction of the strong predictive models that can be applicable to the various conditions of the process. On the basis of this concept, Nistala et al. (2024) proposed an ensemble method of feature selection in which hundreds of blast furnace variables were ranked by their influence on silicon concentration. Their models made reliable and stable predictions that were possible by selecting the most relevant features and they were used to control real-time processes. All these studies demonstrate that the classical and ensemble-based ML models with the advanced feature selection and optimization prove highly efficient in predictive modelling in industrial and mining tasks.

2.2 Deep Learning Models for Process Prediction

The increased complexity of industrial mineral processes has provided researchers with the incentive to resort to deep learning techniques that can be employed to model nonlinearities and temporal correlations in the system. A time-series prediction model has been suggested by Chen, Yang, Li, and Song (2025), and its application is premised on the application of a combination of a Mutual Information and Principal Component Analysis (MI-PCA) and a superior Long Short-Term Memory (LSTM) network when predicting silicon content in froth flotation. Their model has been concerned with smart selection of features and adaptability to change with time. Similarly, Hao, Liu, Gao, and Wang (2025) proposed a multi-scale fusion convolutional neural network (MSF-CNN) that is capable of obtaining local and global time-related information of blast furnace data to improve time-dependent process behaviour. All these studies emphasize the potential of deep learning hybrid models to help in the integration of time dynamics and feature relevance to enhance predictive accuracy of metallurgical systems.

It was noted that image-based deep learning models have a high potential in streamlining the procedures in the broader case of flotation and mineral concentration. The visual froth was considered a predictor in a convolutional and backpropagation neural network developed by Nakhaei, Rahimi, and Fathi (2022) that predicts the sulphur removal rate in the column flotation process using visual froth as a predictor variable, depending on the texture, bubble size, and colour characteristics. Their results confirmed that deep neural networks can be trained on complicated visual patterns that cannot be trained by the conventional techniques of statistics. Similarly, Kukartsev and Degtyareva (2024) applied a neural network model to predict silica concentration in iron ore concentrate based on the process data. Their study found out that the neural architecture can be trained to take effective decisions in the mining industry to regulate the quality and the manipulation of the procedure with the assistance of neural architecture.

In addition to the prediction of individual processes, deep learning has been demonstrated to be promising in solving larger scale modeling problems in heterogeneous mineral systems. The article by Kraiem, Maalaoui, Bouhlel, Sadik, and Sdiri (2024) demonstrates the value of automated learning of features in predicting iron content in complex geological formation by

a deep neural network with the focus on high-dimensional compositional data. On the same note, Cherkaev, Erwee, Reynolds, and Swanepoel (2024) used data-driven modelling that utilised the concept of deep learning to predict silicon behaviour in ferrochrome smelting. Their results highlighted that deep architectures complemented with domain-specific feature engineering has the potential to facilitate adaptive modeling of complex chemical and thermal processes in smelting processes. Taken together, these papers reveal that deep learning systems have become a part of the toolkit to create intelligent and data-driven control systems in mineral and metallurgical industries.

2.3 Hybrid Intelligence and Optimization in Mineral Processing

The greater convergence between machine learning and deep learning has led to the development of hybrid intelligent architectures that have the flexibility of neural architectures and predictability of traditional algorithms. Nistala et al. (2024) proposed an ensemble feature selection and modeling system, in which machine learning is used to forecast the content of molten iron in blast furnaces in terms of silicon. The study involved an interest in the applicability of the combination of different methods of ranking features to be able to estimate the most important variables in the process, thereby ensuring the more effective generalization in terms of operating conditions. Likewise, Wang, Hu, and Tang (2021) suggested a Mult objective evolutionary nonlinear ensemble learning framework, as per which the evolutionary optimization was used to balance the accuracy and the diversity of the model. These hybrid methods emphasize a shift of paradigm in standalone modeling to integrated systems of learning that can be designed to adapt to the shifting industrial conditions and dynamic factors of the process.

The optimization approach to hybrid modeling has also been applied in solving the energy efficiency and process control in the mineral industries. Kolodziej et al. (2025) applied the artificial neural networks optimized with Bayesian and genetic optimisation algorithms in their research to maximize the energy usage of the industrial limestone grinding. This approach portrayed how it is possible to use soft computing successfully to guide the process settings to be geared towards sustainability objectives by identifying the optimal working conditions. Cherkaev et al. (2024) have explored the potential of predicting the content of silicon in ferrochrome smelting using the data management approach built on a hybrid model, which is the combination of dimensionality reduction and boosting, to complement it. They demonstrated in their model that it is possible to attain a significant improvement in the stability of the prediction as well as give valuable data on the chemical and operational processes of the smelting systems through feature compression and optimization of the ensembles. These combined methods point to the increased significance of the optimization-based learning to achieve the predictive accuracy and process efficiency.

The application of hybrid intelligence has not only been restricted to prediction over the last several years, but has also been applied to service and control more processes and decisions. Subsequently, Saavedra et al. (2025) pointed out that machine learning will be applicable in managing the ore blending process to enhance consistency in the feeds and reduce variability

in downstream operations. Their article emphasized the significance of the interaction of predictive analytics and automation systems in changing the operational strategies with the capability of making real-time decisions and their foundation on data. Similarly, Kraiem et al. (2024) also showed the versatility of the deep neural network combined with unsupervised methods of feature reduction, such as PCA and hierarchical clustering, in predicting the iron content of the heterogeneous mining sites. These papers, when combined, indicate that the joint efforts of deep learning, optimization algorithms, and ensemble intelligence are establishing the road to the next generation of flexible, sustainable, and autonomous mineral processing systems.

2.4 Research Gaps

Although it is evident that machine learning and deep learning models have made significant advancements in predicting mineral and metallurgical processes, there are a number of limitations still present in the literature. In the majority of the existing models, the input features are highly predetermined or predictable by use of methods of dimensionality reduction like PCA or MI-PCA, potentially missing important nonlinear interactions between process variables (Nagarjuna and Ananthan, 2022; Nistala et al., 2024). Moreover, even though LSTM and CNN-based models have enhanced their capabilities in recognizing temporal and spatial patterns, they fail to adapt well to long-term dependencies and experience poor performance in dynamic industries (Chen et al., 2025; Hao et al., 2025). Most of the models are trained using small or static datasets resulting to overfitting and poor generalization in different process conditions. Also, hybrid optimization models, like evolutionary and ensemble (Wang et al., 2021; Kolodziej et al., 2025), are based predominantly on the accuracy metrics but do not consider the explainability and adaptability needed to implement them in real-time industries. Such constraints underscore the importance of more relaxed, and context-sensitive modelling models.

Attention-based models provide a good solution to these limitations by selectively emphasizing the most useful items in the process data. Attention mechanisms, in contrast to traditional LSTM or CNN models, dynamically assign weights to important variables in the input, thus enhancing feature and long-term dependency capturing. Using the example of a self-attention module, correlations between temperature, the feed composition, and gas flow in blast furnace work can be considered, such that the model can prioritize the important signals when making prediction. It is possible to incorporate attention layers into recurring or convolutional networks and, therefore, achieve a higher level of prediction accuracy by directing the computational resources to the most important temporal or spatial patterns (Hao et al., 2025). In addition, the multi-head attention architectures can capture a wide range of contextual representations of process data, which give a better vision of underlying physical and chemical interactions which control silicon or impurity variation.

In addition to making features more relevant and providing contextual insights, attention-based architectures also solve the problems of model transparency and real-time flexibility, which are frequently neglected in the traditional ML and DL methods. Attention maps and the learned

weight distributions give clear indications of the process variables that significantly affect impurity or silicon predictions at a given time interval, which allows the engineers to correlate the model outputs with the actual operational behaviours. Such transparency assists operators to correlate the results of prediction with other phenomena like feeding composition changes or alterations in aeration conditions in flotation systems. Moreover, the plastic nature of attention systems enables them to dynamically reorient their attention towards incoming data streams so that the model does not tend to stabilize, but is dynamic enough to react to the changing industrial conditions. Attention-based models are especially well applied in the real-time monitoring, predictive control, and intelligent decision-support systems of mineral processing environments through this dynamic capability.

Author(s) & Year	Model / Technique Used	Key Contribution	Results / Findings	Limitations / Gaps
Nagarjuna & Ananthan (2022)	RFR, SVR, XGB, PCA	Predicted silicon concentration in froth flotation; PCA reduced dimensionality.	Low MSE, MAE, and RMSE, Random Forest best performer.	Static model, lacked temporal analysis.
Anjum et al. (2022)	Gradient Boosting, Random Forest, Bagging, AdaBoost	Predicted compressive strength of fiber-reinforced nano-silica concrete using ensemble ML.	AdaBoost highest precision (R^2=0.92, MAE=3.73 MPa).	No temporal modeling, limited data-specific results.
Chen et al. (2025)	MI-PCA + Improved LSTM	Combined MI-PCA with LSTM for time-series silicon forecasting.	Outperformed LSTM, RNN, and XGBoost with lowest RMSE and MAE.	Limited long-term dependency handling.
Nistala et al. (2024)	Ensemble ML + Feature Ranking	Ranked 374 variables for silicon prediction in blast furnaces.	90% hit rate, identified new influential variables.	No temporal dynamics, lacked adaptive learning.
Cherkaev et al. (2024)	Hybrid ML (Dimensionality Reduction + Boosting + Regression)	Predicted silicon in ferrochrome smelting; linked data to thermodynamics.	R^2=0.63, identified key predictors (C, Ti, CaO, resistance).	Moderate accuracy, non-dynamic, single-site data.
KoÅ‚odziej et al. (2025)	ANN + Bayesian & Genetic Optimization	Optimized energy efficiency in limestone grinding using ANN.	Energy savings up to 48%.	Static model, not generalizable.
Nakhaei, Rahimi & Fathi (2022)	CNN, BPNN, MLR, K-Means	Used froth image features to predict sulfur removal.	CNN error=10%, BPNN R=0.97, MLR error=15%.	Small dataset, lacks sensor data integration.

Saavedra et al. (2025)	ML for Ore Blending	Reviewed ML for ore blending and feed consistency management.	Synthesized real-time blending methodologies.	Conceptual only, no model implementation.
Harsha et al. (2021)	MLR, RF, DT, LSTM	Predicted silica impurity in mining; built real-time web app.	LSTM best performer for real-time prediction.	Overfitting risk; lacked adaptive attention.
Deng et al. (2022)	Mechanism-based Semi-supervised LSTM (MSLSTM)	Integrated mechanism model with LSTM for copper smelting matte grade.	Outperformed data-only models, improved nonlinear mapping.	Domain-specific, no feature prioritization.

3 Research Methodology

3.1 CRISP-DM Methodology

CRISP-DM (Cross-Industry Standard Process for Data Mining) is a structured and systematic method of performing data-driven studies and is therefore very applicable to industrial use, e.g., silicon and impurity prediction in mineral and metallurgical processes. In this piece of work, the methodology commences with the business and process understanding, in which the necessity of real-time silicon prediction is developed. Conventional laboratory measurements add delays to a process, decreasing the responsiveness of the process, and there is an obvious need to have intelligent models that can be used to describe nonlinear interactions among operating variables. The second stage, data understanding, is the analysis of the existing process data including sensor data, operational logs, and time-series data to find the corresponding patterns, anomalies, and correlations that can influence predictive results. This guarantees a knowledgeable base of the further steps of the job process.

After this, there is the data preparation phase, which aims at preparing and streamlining the data to make it machine and deep learning model friendly. In the current study, this involves the processing of missing values, normalization of numerical data, aligning time-dependent factors and improving the dataset, such that sophisticated architectures like attention-based models can effectively discover influential variables directly out of the data instead of using manual dimensionality reduction. During the modeling stage, different ML, DL, and attention-enhanced networks are built and trained to predict silicon concentration at dynamic industrial conditions. The evaluation stage then analyses the performance of the model through RMSE, MAE, and hit rate indicators to ascertain the accuracy, stability and robustness of the model. The last stage of deployment defines how the attention-based prediction system may be incorporated into the operating workflow of real-time plants to allow the engineers to make timely changes in the process, decrease the level of impurities, increase the quality of the product, and make the energy consumption more efficient. In this hierarchical CRISP-DM framework, the methodology guarantees that the resultant models are robust, practical and versatile in diverse industrial settings.

3.2 Dataset Description

The data applied in the present study is the measurements of industrial processes that are taken in a flotation plant, which is functioning, which is the collection of various variables that affect the quality of the feed and the composition of the final concentrate. Each record contains timed observations and process parameters of 100% Iron Feed, 100% Silica Feed, reagent flows (Starch Flow and Amina Flow), Ore Pulp Flow, Ore Pulp pH, and Ore Pulp Density. Moreover, the data set also has operational indicators in detail, such as air flow rates and pulp levels on seven flotation columns, which capture the dynamic behaviour of the separation process. The target variables will be the percent Iron Concentrate and percent Silica Concentrate which is the final product quality. The common features of the data set are that it has outliers, non-uniform values, and potential sensor noise which are typical of the actual operation. These features make it suitable in evaluating the power of more advanced prediction frameworks, particularly attention-based frameworks, which are able to learn important features in different conditions of the process variables. Overall, the data set is a rich multivariate time-series setting that can be applied to assist of accurate and context sensitive silicon and impurity forecasting systems in mineral processing processes.

3.3 Data Preprocessing

The preprocessing phase started with loading the flotation plant dataset and carrying out a preliminary analysis to get familiar with its structure and content. This involved inspection of the attributes that were available, the extent to which they were numerical, categorical and time related fields, and the general consistency of data. The check-up proved that the dataset consisted of numerous parameters of operations, chemicals, and flotation that were recorded over a long industrial term, which is an excellent foundation of predictive modeling.

Many of the numerical measurements were initially represented as text because of the use of commas as decimals separators, as commonly used in industrial data collection systems. The standardization of these values was done by converting the comma-formatting to the appropriate decimal notation and changing the fields into numerical values. This standardization made all the variables of the processes appropriate to machine learning models and removed the problems of inconsistency due to mixed data formats.

Some variables that were irrelevant to the prediction of silica concentration were eliminated to make the dataset relevant to the prediction goal of the study. The variables that represented the feed composition or other indicators of product quality were omitted to avoid duplicity and the chances of information leakage. This optimization made sure that the model is only trained on meaningful and independent values of process that affect the target outcome.

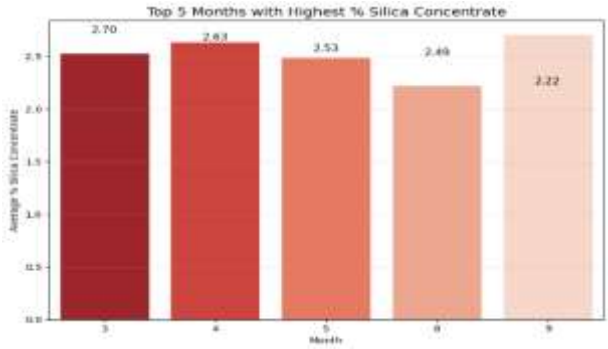
A line-by-line examination was made to determine the existence of missing values in all the columns, and the data was determined complete with no gaps that needed to be imputed. Nevertheless, some duplicate records were also identified, probably due to duplicate sensor

records or system logging overlaps. Such redundant observations were discarded to prevent the model learning to be skewed and the integrity of the data.

The column that contained the timestamp was altered to a normalized date time so that one could extract the meaningful time features. Using this reformed timestamp, other characteristics such as hour, day and month number were calculated to indicate the patterns of operations, day to day variation and bigger time trends that dominate in the mining industry. These fictitious time-based attributes complement the data provided by the dataset by providing it a contextual knowledge of the dynamic behaviour of the plant. The field of original timestamp was removed when it was extracted and redundancy was removed. The data set was organized in a time sequence in order not to interfere with the sequence of the events of the process. Time consistency is particularly critical in the process of modeling industrial processes which are prone to a time-varying response and a characteristic operational cycle.

The processed data was split into input features and target variable where silica concentration was chosen as the predictive output. A section of the data was kept aside as the training and the rest was used as the testing part. This division allowed objective performance evaluation and also ensured that predictive performance of the model was tested on new data. A random seed was fixed to ensure that there was reproducibility in data split. The dataset has variables of different magnitudes ranging between physical flows and reagent dosages to pH values and flotation levels so the feature scaling was used to normalize all the variables. Standardization enabled every variable to make a proportional contribution in training and enhanced stability and convergence behaviour of neural network models. The scaling parameters were estimated using the training data and were used uniformly to the training and the test data.

3.4 EDA (Exploratory Data Analysis)



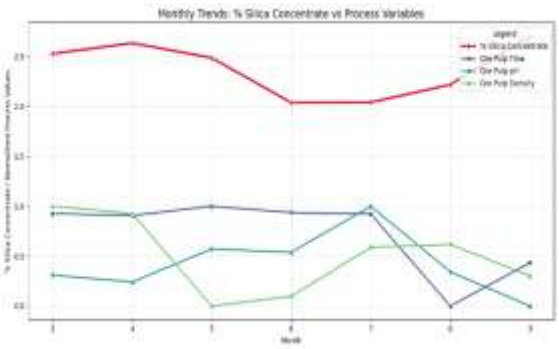


Figure 1. Top Months Highest Silica Concentrate **Figure 2. Silica & Process Variables**

Figure 1, chart demonstrates the five months that had the highest average silica concentrate with top ratings of March, April, and September. The chart in figure 2 compares monthly trends in silica concentrate to normalized ore pulp flow, pH, and density indicating how these process variables vary with respect to silica. In general, the concentration peaks of silica are not exactly correlated with any of the process variables indicating that there are several interacting factors.

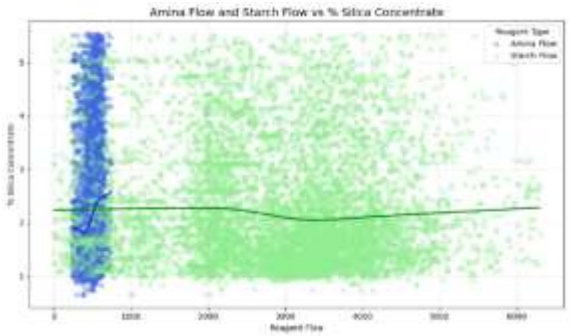


Figure 3. Reagent Flow vs. Silica Concentrate

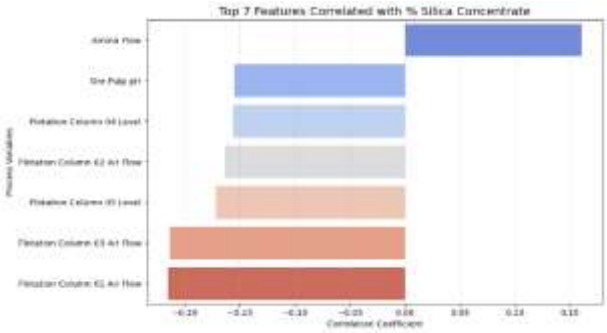


Figure 4. Top Correlated Variables with Silica

Figure 3, plot demonstrates that Amina Flow is increasing very slightly with the percentage Silica Concentrate whereas Starch Flow is increasing by almost no significant percentage. A better trend is shown by the Aminas formations, which implies that silica is more susceptible to Amina dosage. Figure 4 shows that plot ranks the best correlated variables with Amina Flow being the strongest positive predictor of higher silica. There are moderate negative relationships between several flotation air flows and levels, which implies that they tend to contribute to silica reduction. In general, the visuals demonstrate that the most influential variable is Amina Flow as other variables have weaker or opposite impact.

3.5 Model Architecture: AutoInt for Feature Interaction Learning

In order to overcome the limitations of the existing works, the paper will utilize the idea of AutoInt (Automatic Feature Interaction Transformer) to overcome the limitations that were observed in the previous research on machine learning and hybrid models. The earlier methods tended to utilize a combination of predetermined interaction of features, manual selection of variables, or rudimentary ensemble models and this limited their capability to represent the dynamic and multifaceted association between the process parameters in industrial flotation systems. Such techniques usually consider individual features separately or as a combination in fixed, superficial ways, and hence limited generalization in dynamic operating conditions. AutoInt manages to overcome these challenges by its multi-head self-attention mechanism which learns automatically both low-order as well as high-order interactions in the raw input space. The model can put dynamic attention weights on the most significant variables including reagent dosage, air flows, pulp characteristics, and column-level indicators and still retain delicate dependencies that are often not reflected in other traditional models. This allows AutoInt to capture nonlinear, context-dependent interactions in a better way, with better predictive qualities.

Here, AutoInt is used on the multivariate time-series data of the flotation plant by embedding each process variable, which represents its underlying behaviour. The architecture determines the interactions and effects of various operational features on silica concentration under varying temporal and process conditions through the use of stacked attention layers. The learned feature interactions are then combined with the feed-forward layers that are created to provide stable and efficient prediction, and thus the model is suitable in real-time industrial application. Another benefit of AutoInt is that transparency attention patterns can be visualized giving

insightful information to process engineers as to which variables are the most significant contributors to impurity fluctuations. On the whole, the inclusion of AutoInt brings in a more flexible, and context-sensitive modeling structure, which is congruent with the complexity and multi-factorial nature of mineral processing processes.

3.6 Model Training and AutoInt Working Mechanism

The architecture of AutoInt that is implemented in this paper is developed based on the concept of learning meaningful feature interactions automatically by focusing attention based on attention-guided learning. AutoInt instead of using manual feature engineering or the application of predefined mathematical transformations, maps each process variable: reagent flows, air rates, pulp characteristics, and flotation column indicators into a continuous representation space. These embeddings enable the model to encode raw industrial measurements into the structured vectors where the numerical magnitude and the contextual behaviour are retained. The features are then fed out to stacked multi-head self-attention layers, where the model determines the strength with which each variable interacts with all the rest of the variables. This interaction learning is vital in mineral processing as levels of impurity e.g. silica concentration is normally affected by various factors operating simultaneously as opposed to single input.

The best thing with AutoInt is that it can identify high-order features and nonlinear features interactions which are not easily identified by the traditional machine learning models. Amina dosage, pulp density, and air flow are not independent variables in flotation plants, and their level of interactions is dynamic and complex depending on the operating conditions. The attention mechanism of AutoInt learns to weight these interactions, which allows the model to concentrate more weight on variables that have the highest impact in a particular operation context and reduces the effects of less significant or noisy signals. This adaptive weighting is more predictive and robust especially in dynamic industrial settings where the process dynamics change with time. In addition, the attention structure is automatically, i.e. engineers are able to study what variables the model depended on the most when predicting the silica concentration.

In the course of training, the model is taught to optimize these interaction weights by employing backpropagation, and slowly increases its capacity to find meaningful relationships in the multivariate time-series data. The results of each attention layer are combined and fed to fully connected prediction layers, which convert the learned interactions to a final estimate of silica concentration. This hierarchical structure will make sure that the model takes into account either direct or indirect effects between the process variables at various levels of abstraction. Due to this, AutoInt can effectively model the complex behaviour of flotation systems compared to traditional methods and also make decisions about processes in real-time. The AutoInt model is oriented towards the practical requirements of the mineral processing operations because it combines automatic feature interaction learning with a transparent and flexible architecture and allows predicting impurities more closely and enhancing the control of the process on the whole.

3.7 Model Evaluation Metrics

The evaluation of the prediction model is done with four major regression measures; Mean Absolute Error (MAE), which calculates the average difference between the predicted and actual values of silica; the Mean Squared Error (MSE) which is more stringent in penalizing larger errors, by squaring them; the Root Mean Squared Error (RMSE) which is on the same scale as the target variable and the R2 Score, which is used to determine the extent to which the model can explain the variation in silica concentration. All these metrics can be used to evaluate the model in terms of its accuracy, consistency and generalizability under various conditions of processes.

4 Design Specification

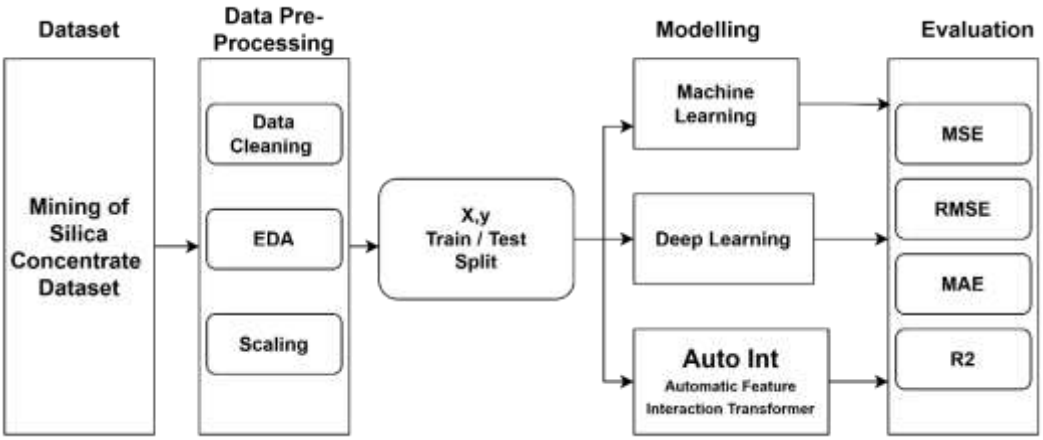


Figure 5. End-to-End Data Mining and Modeling Workflow Diagram

4.1 System Architecture Overview

The system architecture in figure 5, is a full end-to-end silica prediction workflow, starting with the mining data, which has actual data of flotation plants. The initial stage in processing this data involves a preprocessing pipeline which involves data cleaning to eliminate duplicates and false values, EDA and scaling to have a feel of the trends and normalise the variables and lastly a train/test split to have an objective assessment. Following preprocessing, the modeling phase uses three parallel methods that are traditional Machine Learning techniques, more complex Deep Learning networks, and the AutoInt architecture which automatically learns both low- and high-order feature interactions through self-attention mechanisms. These models examine the effects of the variables like reagent dosage, pulp density, and air flows on silica levels. The last step, evaluation, applies MAE, MSE, RMSE, and R2 to assess the accuracy, strength and predictive ability of every model and finally the most trusted method of predicting silica in real-time in industry is identified.

4.2 Workflow and Pipeline Implementation

The silica concentration prediction system workflow/pipeline implementation is based on a structured and modular design specification that is aimed at facilitating the robust data processing and model training in an industrial mineral processing environment. The system architecture is structured into interrelated levels that handle data intake, preprocessing, model implement and assessment, to maintain the flowing of information in a smooth manner between the raw operations inputs and the end predictive outputs. The pipeline starts with the mining data being loaded using sensors on the flotation plant and then passes through a thorough preprocessing step as the data is cleansed to eliminate any duplicates or formatting errors, coded and normalized to be numerically stable and then split into training and testing data sets whilst preserving the temporal order. These clean datasets are further transformed as per model needs to flat tabs to machine learning regressors, sequence tensors to deep learning structures and embedded feature vectors to the AutoInt model. The model training pipeline combines three modeling paradigms traditional machine learning algorithms to learn statistical relationships between operational parameters; deep learning models to leverage temporal dependencies and nonlinear dynamics; and the AutoInt architecture to learn high-order relationships between feature interactions germane to silica behaviour by embedding each input variable as a token. Each and every model is trained on the pre-prepared datasets and evaluated on standardized evaluation measures, to guarantee that they are accurate, have generalization and can be applied in the industry.

5 Implementation

5.1 Machine Learning Models Implementation

This paper used hyperparameter tuning with both grid search and randomized search to explore parameter combinations in a systematic way and find the most effective parameter settings of the model. These are cross-validation methods based on R2 scoring, which is meant to guarantee good generalization and avoid overfitting. In all machine learning models, the most important parameter that was tuned was the n number of estimators which defines the number of trees or boosting rounds and has a direct influence on the complexity of the model, the depth of learning and stability.

In the case of the Random Forest Regressor, tuning was restricted to the optimal value of number of estimators that was experimented between 50,100,150,200. This parameter is important in enhancing the variance reduction ability of the ensemble as it grows the number of decision trees making a collective contribution in the final prediction. Having compared the values of the candidate, matching n estimators was found to be the most successful of the configuration, namely n 200, meaning that a larger forest was needed to capture the nonlinear interactions of the flotation plant data.

In the case of the XGBoost Regressor, the tuning of the number of estimators parameter was done by testing values of 50,100,150,200 to determine the effect that the number of boosting

rounds has on learning. Given that XGBoost is more accurate as the boosting cycle is repeated, it is possible to capture more intricate trends in industrial data with larger boosting cycles. The number of estimators value that was found as the best is 200, which shows that long boosting was required to make silica concentration a nonlinear behavioural model.

In the case of the LightGBM Regressor, number of estimators (50,100,150,200) were also considered in order to find the number of boosting iterations to achieve stable learning. LightGBM is effective in high-dimensional interaction of features, and additional boosting rounds usually enhance generalization in multivariate process data. The tuning process also picked a best configuration of n estimators as 200, which attests to the fact that increased number of iterations was needed in order to capture the fluctuations and nonlinearities that exist in the flotation plant data.

5.2 Deep Learning Models Implementation

The deep learning aspect of the current study uses a custom Multi-Layer Perceptron (MLP) Regressor to learn nonlinear interactions in the flotation plant data. Once the reproducibility of the seeds of NumPy, Torch, and CUDA was ensured by repairing the seeds, the scaled training and test data were then turned into tensors and loaded with mini-batch Data Loaders, to be effectively calculated. The MLP architecture is designed to have three hidden layers with 128, 64, 32 neurons with a dropout rate and ReLU activation to enable the model to extract higher level feature representations as the architectural levels increase, with the tendency to overfit diminishing. This model was trained with 100 epochs and the AdamW optimizer with the learning rate and MSE used as the loss function to direct the weight changes. The training and validation losses were tracked during training to make sure that the convergence was stable and the optimized network was subsequently applied to predict on unseen data. This MLP architecture gives a nonlinear impurity behaviour that can be represented by a flexible deep learning baseline as a complementary to the conventional machine learning models that will be employed in this study.

5.3 AutoInt Implementation

The AutoInt application in the present research uses an attention-based architecture that has the potential to acquire complex feature interaction automatically based on the flotation plant data. After seeding the seeds, the transformation of each numerical feature into a learnable embedding is done in a linear transformation, which allows the model to express raw process variables in a more expressive vector space. These embeddings are then subjected to several Interacting Layers, in which multi-head self-attention compares variables between reagent dosage, air flows and pulp properties and is able to capture both simple and high-order interactions which other traditional models might not capture. The dropout, residual connections, and layer normalization are used to improve stability and avoid overfitting. After attention layers extract interaction-aware representations, the output is flattened and passed into the MLP prediction head, which is a small feed-forward neural network composed of fully connected layers (128 and 64 units) with ReLU activation. This prediction head serves as the

final stage that aggregates all learned interactions and transforms them into a single numerical output the predicted silica concentration. Training is performed using the AdamW optimizer with MSE loss, and an Early Stopping mechanism with patience of 178 epochs ensures that the best model is retained without unnecessary overtraining. Overall, the AutoInt architecture provides a powerful end-to-end framework for modeling impurity behaviour, as it can automatically discover meaningful interactions among process variables while maintaining through attention mechanisms.

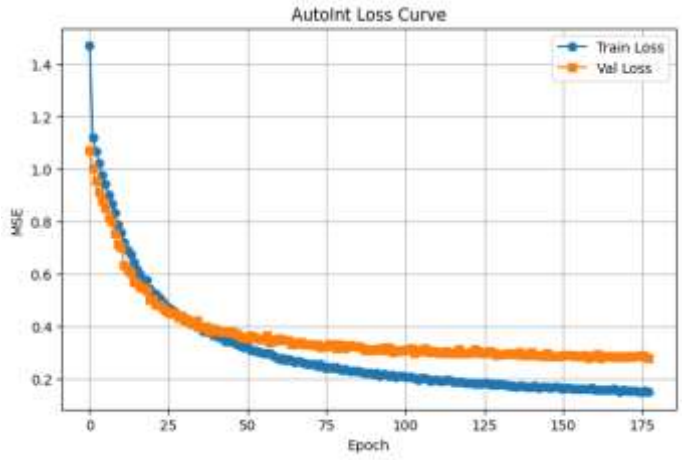


Figure 6. Autoencoder Training vs Validation Loss (MSE)

In figure 6, the training and validation MSE drop sharply in the first few epochs, showing the model learns quickly at the start. After that, both curves decrease more gradually and start to flatten, which suggests the improvements become smaller as training continues. The validation loss remains higher than the training loss throughout, meaning the model performs better on the training data than on the validation data.

6 Evaluation

Table 1. Performance Comparison of all Models

Model	MAE	MSE	RMSE	R2 Score
Random Forest	0.4604	0.3993	0.6319	0.6799
XGBoost	0.4474	0.3704	0.6086	0.7031
Light GBM	0.4571	0.3732	0.6109	0.7009
MLP	0.5001	0.4963	0.7045	0.6022
Auto Int	0.3601	0.2774	0.5267	0.7777

This section evaluates and compares the predictive performance of all models for silica concentration estimation using standard regression metrics (MAE, MSE, RMSE, and R^2). summarizes the results across traditional machine learning, deep learning, and attention-based architectures to highlight differences in error magnitude and variance explanation. The following subsections discuss each model's outcomes in detail and identify the most effective approach for real-time flotation process prediction as shown in table 1.

6.1 Machine Learning Model Results

The Random Forest Regressor achieved an MAE of 0.4604, MSE of 0.3993, RMSE of 0.6319, and an R^2 score of 0.6799, indicating moderately strong predictive accuracy. The model captures key nonlinear relationships in the flotation process but leaves some variance unexplained due to its averaging behaviour. The fact that it has a slightly lower R2 than boosting models implies that it is not very good at learning complex interactions between features. All in all, Random Forest is a good choice but it is outperformed by more sophisticated methods of gradient boosting.

XGBoost performed best in all the machine learning models with a MAE of 0.4474, MSE of 0.3704, RMSE of 0.6086, and highest R2 of 0.7031. These measurements indicate that the model is not only less predictive, but also explains more of the variance in silica concentration. It can learn complex, high-order interaction of features and correct residual error of the earlier trees, so it is especially useful with the nonlinear and dynamic nature of industrial flotation data. The enhanced precision and consistency points to the applicability of XGBoost in the context of modeling impurity behaviour in real-time process settings.

LightGBM had an MAE of 0.4571, MSE of 0.3732, an RMSE of 0.6109 and an R2 of 0.7009, which ranks it near XGBoost. It is efficient in capturing important feature splits particularly in large-dimensional industrial data, due to its speed and leaf-wise strategy of tree growth. Despite slightly more mistakes than XGBoost, LightGBM manages the nonlinearities in silica concentration and it has good generalization on the data that has not been seen. The findings suggest that LightGBM provides a good trade-off between accuracy and computing efficiency, and thus it is a viable competitor to industrial prediction tasks.

6.2 Deep Learning Model Results

The MLP model gave an MAE of 0.5001, MSE of 0.4963, RMSE of 0.7045 and an R2 value of 0.6022, which shows that the model has low predictive ability compared to the machine learning models that were tested. Even though the neural network is able to capture general nonlinear trends in the flotation data, it is not as effective in modeling complex interactions between variables as tree-based boosting approaches. The larger error values and the smaller R2 indicate constraints to learning finer relationships of features with the provided architecture meaning more sophisticated deep learning methods or attention-based processes might be required to get higher accuracy.

6.3 AutoInt Model Results

AutoInt model gave an MAE of 0.3601, MSE of 0.2774, RMSE of 0.5267, and R 2 value of 0.7777 which is the best in all models considered in the present study. These measures show that AutoInt does not only generate much lower prediction errors, but also characterizes a much greater share of variance in silica concentration. The fact that it is more accurate points to successfulness of automatic feature interaction learning in which the model relies on self-attention learning to determine how variables (amine dosage, pulp density and air flow conditions) relate. AutoInt learns such interactions dynamically and not based on fixed or manually designed combinations and is thus able to better adapt to nonlinear and changing behaviour of industrial flotation processes.

In addition to enhanced accuracy, the good performance of the model can also be attributed to its capability of representing high-order dependencies, which are generally ignored by other traditional ML regressors and normal deep learning architectures. The layers of self attention allow AutoInt to concentrate on the most influential variables at every step of the prediction resulting in more stable and context-sensitive output. This enhanced enables engineers to visualise the weight of attention and comprehend the reason why silica levels vary in various conditions of operation. In general, the AutoInt model proves to be extremely suitable in real-time industrial prediction tasks with a combination of high predictive performance, strong generalization, and clear understanding of feature interactions.

6.4 Findings of the Study and Superiority of the AutoInt Model

The AutoInt model showed an overall and consistent high level of superiority compared to all the machine learning and deep learning models that were tested during this work. Although Random Forest, XGBoost, LightGBM, and the MLP network could model some nonlinear relationships in the flotation data, they were limited to fixed feature interaction, shallow learning features or lack of dynamic contextual knowledge. AutoInt in contrast recorded the lowest values of MAE, MSE and RMSE as well as the highest value of R2, which proves that it was able to explain a much larger percentage of the variability of the process. These findings indicate the multivariate and complex nature of silica behaviour in industrial processes that the simple models are not able to capture.

The greatness of AutoInt is largely motivated by the fact that it has a multi-head self-attention mechanism, which provides the automatic identification of low-order and high-order feature interactions. As opposed to other conventional models that rely on manual feature engineering or tree-based splits, AutoInt is able to learn the interactions between reagent flows, pulp properties, air rates, and column indicators under varying operating conditions. This adaptive weighting enables the model to concentrate on the most impactful variables at each given time, and thus, subtle dependencies and dynamic trends can be observed which is not the case with other models. Learnable embeddings, stacked attention layers, and capability to generalize effectively make AutoInt more precise, more resilient, and make it the most efficient

framework in the context of real-time silica concentration prediction in mineral processing plants.

7 Conclusion and Future Work

7.1 Conclusion

The proposed study presented a novel state-of-the-art attention-based AutoInt framework to predict silica concentration in industrial flotation processes to overcome significant drawbacks of other machine learning and deep learning models. The model achieved this by automatically learning both low- and high-order feature interactions with multi-head self-attention and removing the limitations of pre-defined feature engineering, as well as giving a clearer insight into which operational variables affect impurity behaviour. Extensive experiments with real plants data revealed that AutoInt performed much better than traditional models such as Random Forest, XGBoost, LightGBM and MLP that produced the best accuracy and strongest generalization in changing industrial conditions. All in all, this work provides a scalable and trusted solution that enhances the real-time monitoring of impurities, increases stability in the processes and helps making more informed decisions in the process of mineral processing.

7.2 Future Work

There are also a number of ways that one can expand this study to make the attention-based impurity prediction systems even more reliable and applicable in the industry. To begin with, it can be noted that incorporation of other real-time sensor streams like acoustic signals, advanced imaging data and reagent chemical profiles could be used to capture more complex process behaviours and further enhance the accuracy of the model. Second, the ability to track long-term trends in the operation of the system and abrupt process disruptions might be enhanced by the integration of the temporal attention mechanisms or sequence models based on transformers. Third, online or incremental learning would enable the model to constantly update itself as new data about the plant is received such that it operates reliably with varying ore properties and operating conditions. Future studies might also consider creating a hybrid prediction-control scheme, in which the AutoInt model directly drives automated control policies to optimize reagent dosage, air flow and other process variables. Finally, the cross-validation of the system on several industrial sites and various mineral processing circuits would also serve to evaluate its generalizability and make it a potential scalable solution in real-time quality monitoring of the mining and metallurgical industries.

References

Anjum, M., Khan, K., Ahmad, W., Ahmad, A., Amin, M. N., & Nafees, A. (2022). Application of ensemble machine learning methods to estimate the compressive strength of fiber-reinforced nano-silica modified concrete. Polymers, 14(18), 3906.

Chen, J., Yang, Q., Li, D., & Song, P. (2025, July). Time-series prediction of silicon content in froth flotation based on MI-PCA and improved LSTM. In 2025 44th Chinese Control Conference (CCC) (pp. 1279–1284). IEEE.

Cherkaev, A. V., Erwee, M., Reynolds, Q. G., & Swanepoel, S. (2024). Prediction of silicon content of alloy in ferrochrome smelting using data-driven models. Journal of the Southern African Institute of Mining and Metallurgy, 124(2), 67–74.

Deng, P., Li, Y. G., & Li, J. X. (2022, July). Prediction of matte grade in copper flash smelting process based on LSTM and mechanism model. In 2022 41st Chinese Control Conference (CCC) (pp. 2613–2620). IEEE.

Fan, Y., Zhang, G., Li, S., Zhang, L., Guo, J., & Feng, C. (2024). Mineral liberation and concentration characteristics of apatite comminated by high-pressure GRU. Minerals, 14(11), 1148.

Hao, Q., Liu, W., Gao, W., & Wang, X. (2025). A multi-scale fusion convolutional network for time-series silicon prediction in blast furnaces. Mathematics, 13(8), 1347.

Harsha, S. S., Prasad, K. V., & Raghuveer, K. (2021). Prediction of silica impurity using deep learning techniques for mining environment. REVISTA GEINTEC – Gestão Inovação e Tecnologias, 11(3), 506–517.

Ismael, A., Embaby, A. K., Ali, F., Farag, H., Gomaa, S., Elwageeh, M., & Mousa, B. (2024). Prediction of iron ore grade using artificial neural network, computational method, and geo-statistical technique at El-Gezera area, Western Desert, Egypt. Journal of Mining and Environment, 15(3), 889–905.

Kołodziej, D., Bałazy, P., Knap, P., Lalik, K., & Krawczykowski, D. (2025). Energy optimisation of industrial limestone grinding using ANN. Applied Sciences, 15(14), 7702.

Kraiem, Z., Maalaoui, F., Bouhlel, S., Sadik, C., & Sdiri, A. (2024). Predicting iron contents in the Tamra-Douahria mining site using a deep neural network approach. Journal of Taibah University for Science, 18(1), 2422646.

Kukartsev, V., & Degtyareva, K. (2024). Application of neural networks to predict the quality of iron ore concentrate based on flotation data. In E3S Web of Conferences (Vol. 583, p. 01014). EDP Sciences.

Myshchenko, I., Pawlaczyk-Luszczynska, M., Dudarewicz, A., & Bortkiewicz, A. (2024). Health risks due to co-exposure to noise and respirable crystalline silica among workers in the open-pit mining industry—Results of a preliminary study. Toxics, 12(11), 781.

Nakhaei, F., Rahimi, S., & Fathi, M. (2022). Prediction of sulfur removal from iron concentrate using column flotation froth features: Comparison of k-means clustering, regression, backpropagation neural network, and convolutional neural network. Minerals, 12(11), 1434.

Nagarjuna, P., & Ananthan, T. (2022, October). Machine learning-based estimation of silicon concentration in froth flotation process. In 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT) (pp. 1–6). IEEE.

Nistala, S. H., Biswas, J., Kumar, R., Pandya, R., Rathore, P., Mynam, M., Runkana, V., Raj, S., & Ganguly, A. (2024). Machine learning-based knowledge discovery and modeling of silicon content of molten iron from a blast furnace.

Saavedra, M., Risso, N., Momayez, M., Nunes, R., Tenorio, V., & Zhang, J. (2025). Blending characterization for effective management in mining operations. Minerals, 15(9), 891.

Wang, X., Hu, T., & Tang, L. (2021). A multiobjective evolutionary nonlinear ensemble learning with evolutionary feature selection for silicon prediction in blast furnace. IEEE Transactions on Neural Networks and Learning Systems, 33(5), 2080–2093.