



International Journal of Mining and Mineral Engineering

ISSN online: 1754-8918 - ISSN print: 1754-890X

<https://www.inderscience.com/ijmme>

Hybrid machine learning techniques to predict fuel consumption of dump trucks in an open-pit mine in Peru

Marco Cotrina, Jairo Marquina, Jose Mamani, Solio Arango, Joe Gonzalez, Eduardo Noriega, Eusebio Antonio

DOI: [10.1504/IJMME.2025.10066537](https://doi.org/10.1504/IJMME.2025.10066537)

Article History:

Received:	21 December 2023
Last revised:	05 August 2024
Accepted:	16 August 2024
Published online:	08 April 2025

Hybrid machine learning techniques to predict fuel consumption of dump trucks in an open-pit mine in Peru

Marco Cotrina* and Jairo Marquina

Department of Mining Engineering,
Universidad Nacional de Trujillo,
Trujillo, 13001, Perú
Email: mcotrinat@unitru.edu.pe
Email: h2223400318@unitru.edu.pe
*Corresponding author

Jose Mamani

Faculty of Engineering,
Universidad Nacional del Altiplano de Puno,
Puno, 21001, Peru
Email: jomamani@unap.edu.pe

Solio Arango

Department of Mining Engineering,
Universidad Nacional de Trujillo,
Trujillo, 13001, Perú
Email: sarango@unitru.edu.pe

Joe Gonzalez

Department of Industrial Engineering,
Universidad Nacional de Trujillo,
Trujillo, 13001, Perú
Email: jgonzalezv@unitru.edu.pe

Eduardo Noriega and Eusebio Antonio

Department of Mining Engineering,
Universidad Nacional de Trujillo,
Trujillo, 13001, Perú
Email: emnoriegav@unitru.edu.pe
Email: eantonio@unitru.edu.pe

Abstract: Mining dump trucks account for 31% of energy consumption and up to 60% of operating costs in open pit mines. This study aims to predict the fuel consumption of mining dumpers in an open pit mine in Peru by applying innovative hybrid machine learning models, including artificial neural networks (ANN), decision trees (DT) and random forests (RF), optimised with genetic algorithms (GA). The models evaluated the efficiency and cost savings of replacing 24 m³ dump trucks with 26 m³ dump trucks. The data were divided into 70% for training and 30% for testing. Input variables included utilisation, operating hours and total trips. The RF + GA model achieved the highest accuracy ($R^2 = 0.89$, RMSE = 250.05, MAE = 185.09), followed by ANN + GA with an R^2 of 0.87 and DT + GA with R^2 of 0.86. This novel approach significantly improves the prediction accuracy, showing that replacing dump trucks can save an average of US\$1.76 per hour.

Keywords: mining dump trucks; fuel consumption; hybrid techniques; machine learning; genetic algorithms.

Reference to this paper should be made as follows: Cotrina, M., Marquina, J., Mamani, J., Arango, S., Gonzalez, J., Noriega, E. and Antonio, E. (2025) 'Hybrid machine learning techniques to predict fuel consumption of dump trucks in an open-pit mine in Peru', *Int. J. Mining and Mineral Engineering*, Vol. 16, No. 1, pp.1–20.

Biographical notes: Marco Cotrina is a Professor and Researcher in Mining Engineering at the National University of Trujillo. With a PhD in science and engineering, his research interests revolve in the field of mining, testing algorithms and seeking to provide solutions in optimization problems of geostatistics, planning, costs, design, and automation with deep and machine learning.

Jairo Marquina is a Mining Engineer from Universidad Nacional de Trujillo. His current research focuses on areas such as mine planning, geostatistics and the implementation of artificial intelligence in mining stages.

Jose Mamani is a Chemical Engineer from Universidad Nacional del Altiplano de Puno. He has a Master's degree in Engineering Sciences, mention in extractive metallurgy (Mining Engineering) from the University of Chile and is a member of the Peruvian College of Engineers. His current research focuses on deep learning and machine learning models applied to mining and metallurgical stages.

Solio Arango is a Mining Engineer from the Universidad Nacional de San Cristobal de Huamanga, PhD in Environmental Sciences from the Universidad Nacional de Trujillo, with more than 25 years of experience in mining in areas such as safety and occupational health, his research interests are to apply machine learning algorithms in mining operations.

Joe Gonzalez is a Full-Time Professor and researcher with a PhD in Science and Engineering from the Universidad Nacional de Trujillo. He has management experience in operations management and public administration. His research interests are based on new topics in artificial intelligence applied to engineering sciences.

Eduardo Noriega is a Mining Engineer with a second specialty in Civil Engineering, professional with more than ten years of experience in mining operations and in recent years dedicated to the research of numerical methods, as well as predictive models, finite elements, and discrete elements. He

completed Master's studies in Concrete Engineering at the Polytechnic University of Valencia and Master's in Engineering with a mention in Project Management and Direction.

Eusebio Antonio is a Mining Engineer from Universidad Nacional San Cristóbal de Huamanga. He holds a Master's degree in Environmental Management and a PhD in Environmental Sciences, both from Universidad Nacional de Trujillo.

1 Introduction

In mining activities, the extraction of valuable resources involves a series of unit operations (Isnafitri et al., 2021). One of the most significant is mineral transportation, which generates the majority of operating costs, representing 60% of the total mining cost (Dadhich et al., 2016). This is due to excessive fuel consumption related to the capacity of haul trucks, maintenance costs, tire costs, equipment performance, and haul cycle time. Improving the performance and efficiency of both equipment and operators is crucial since prolonged times increase fuel costs and reduce operational efficiency. Therefore, it is necessary to reduce operating times (Nday and Thomas, 2019). Additionally, in all mining operations, it is essential to allocate an optimal amount of financial resources to carry out one of the processes involved in mineral extraction, with the aim of generating a profit (Botín and Vergara, 2015).

Dump trucks are the main transportation equipment in open-pit mining and represent a considerable amount of capital and operational costs. Fuel consumption constitutes approximately 30% of energy usage in open-pit mines. Additionally, a fleet of dump trucks is the primary source of greenhouse gas (GHG) emissions (Siami and Abbaspour, 2011). Gibson (2015) reported that eliminating the idle time of mining trucks in a typical surface mining operation with 20 trucks results in a savings of 145,000 gallons per year, which equates to US\$577,100 (at a fuel price of US\$3.98 per gallon). This also implies a reduction of 64.5 metric tons of carbon and 772 engine hours.

Researchers have studied fuel consumption in various earth-moving vehicles. He et al. (2005) examined fuel consumption trends in road transport in China. Kyoungho et al. (2002) investigated fuel consumption in various light vehicles using multiple regression methods. Delgado et al. (2012) attempted to predict fuel consumption in heavy trucks based on driving cycle operations. Thuy and Bruce (2010) predicted fuel consumption during seven different daily activities of garbage collection trucks using global positioning systems (GPS). Rahimi and Abbaspour (2011) applied artificial neural networks (ANN) and multiple regression techniques to predict the fuel consumption of tractors. Kara and Baysec (2010) proposed an ANN to predict the specific fuel consumption of a gasoline engine.

Recent studies, such as that by Wang et al. (2021), used a regression model and the XGBoost algorithm to predict fuel consumption in mining trucks, identifying factors like lift height, load, and haul distance. Nejad et al. (2021) proposed a model for optimal fuel allocation in mining equipment using partial least squares structural equation modelling (PLS-SEM). Alamdari et al. (2022) modelled and evaluated diesel fuel consumption of mining trucks using various machine learning techniques. Arhinful and Agyei (2023)

applied and tested three training algorithms to build an ANN for predicting truck fuel consumption. Zargarneshad et al. (2019) studied fuel consumption of maintenance and supervisory vehicles for disruption elimination. Soofastaei et al. (2016) developed a comprehensive artificial intelligence model to reduce fuel consumption in the mining industry.

This study aims to predict the monthly fuel consumption of dump trucks in open-pit mining based on monthly transport activities. The results will be used to estimate costs and evaluate the replacement of 24 m³ dump trucks with 26 m³ ones. The contribution of this study is to fill the knowledge gap in fuel consumption in open-pit mining, addressing two main gaps:

- 1 the lack of development of hybrid machine learning models to predict fuel consumption based on monthly cyclic parameters
- 2 the absence of comprehensive statistical analyses on the parameters affecting fuel consumption and fuel cost estimation to evaluate the replacement of dump trucks with different capacities.

In this study, important parameters were identified based on material transport activities, which can be easily collected using data logging instruments. Additionally, this is the first study that allows for analysis and decision-making regarding the replacement of lower-capacity equipment with higher-capacity ones, benefiting the mining company in terms of production and cost savings. Kumar et al. (2014) proposed a deterministic approach to find a functional relationship between fuel consumption and various explanatory variables but did not consider the stochastic nature of transport cycle activities. This study proposes a hybrid neural network model with a genetic algorithm based on historical data, considering uncertainties in input variables and addressing the limitations of multiple regressions. The novelty of this study lies in predicting fuel consumption based on transport cycle activities, such as equipment utilisation, operating hours, and total material trips (ore and waste). Furthermore, a fleet of 24 m³ and 26 m³ mining dump trucks was analysed over a period of 7 months of continuous operation.

2 Methodology

2.1 Database description

The database used in this study comes from mining dump trucks at an open-pit mine located in northern Peru. The selected explanatory variables were: utilisation (U), operating hours (HO), and total trips of waste and ore material (TT). The output variable was the total monthly fuel consumption (F). To ensure the quality and reliability of the data, the interquartile range (IQR) method was applied to eliminate outliers. The IQR is a statistical technique used to identify and remove extreme values in a dataset by calculating the interquartile range, which is the difference between the third quartile (Q3, the top 75%) and the first quartile (Q1, the bottom 25%). This methodology helps ensure that subsequent analyses are conducted on representative and unbiased data (Sanjeev et al., 2023). Table 1 presents the descriptive statistics of all variables after outlier removal. These statistics include the count, mean, standard deviation, minimum and

maximum values, variance, and kurtosis, providing a detailed view of the data distribution and variability.

Table 1 Descriptive statistics of study variables

Statistic	U (%)	HO (h)	TT (quantity)	F (gal/month)
Count	89.00	89.00	89.00	89.00
Mean	0.82	494.31	1,164.65	2,492.56
Std dev	0.07	77.37	224.97	497.08
Minimum	0.61	304.06	609.00	1,417.00
Maximum	0.94	592.00	1,051.00	3,280.00
Variance	0.05	5,985.75	1,612.00	247,085.76
Kurtosis	0.97	−0.40	−0.02	−1.16

In Figure 1, the boxplots of the variables used in the study after the removal of outliers are shown. For utilisation (U), most of the data range between 0.80 and 0.86, with a median of 0.82. Operating hours (HO) are concentrated between 450 and 525 hours, with a median of 500 hours. Total trips (TT) mainly vary between 1,000 and 1,400, with a median of 1,200 trips. Fuel consumption (F) ranges between 2,200 and 2,800 gallons per month, with a median of 2,500 gallons per month. These visualisations help understand the distribution and central tendency of each variable, which are essential for subsequent analysis.

Figure 1 Boxplot of study variables without outliers (see online version for colours)

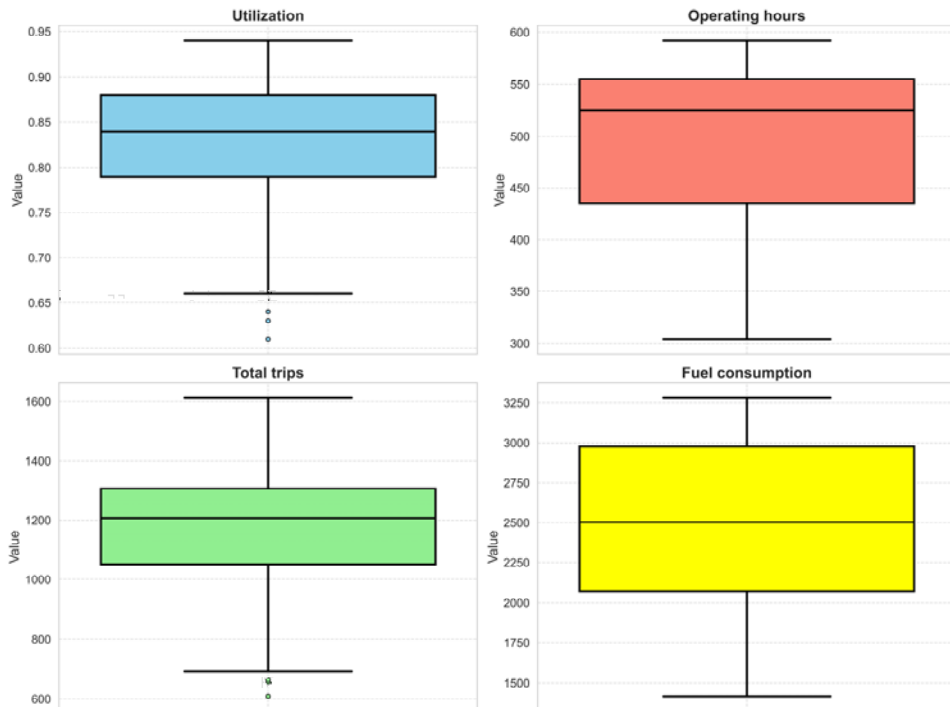


Figure 2 Histograms of the study variables without outliers (see online version for colours)

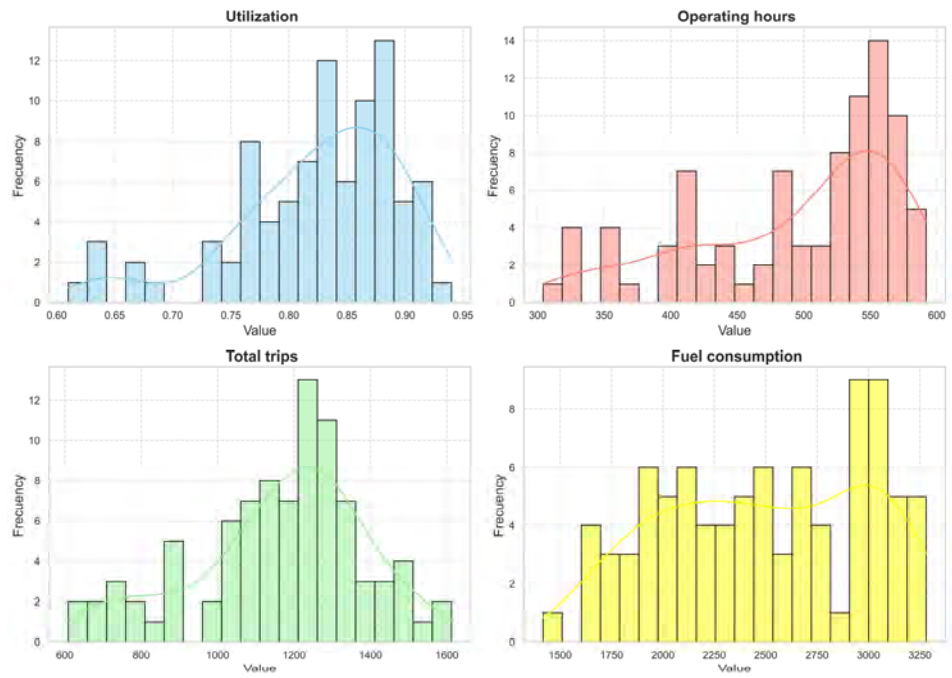


Figure 3 Correlation matrix of the variables without outliers (see online version for colours)

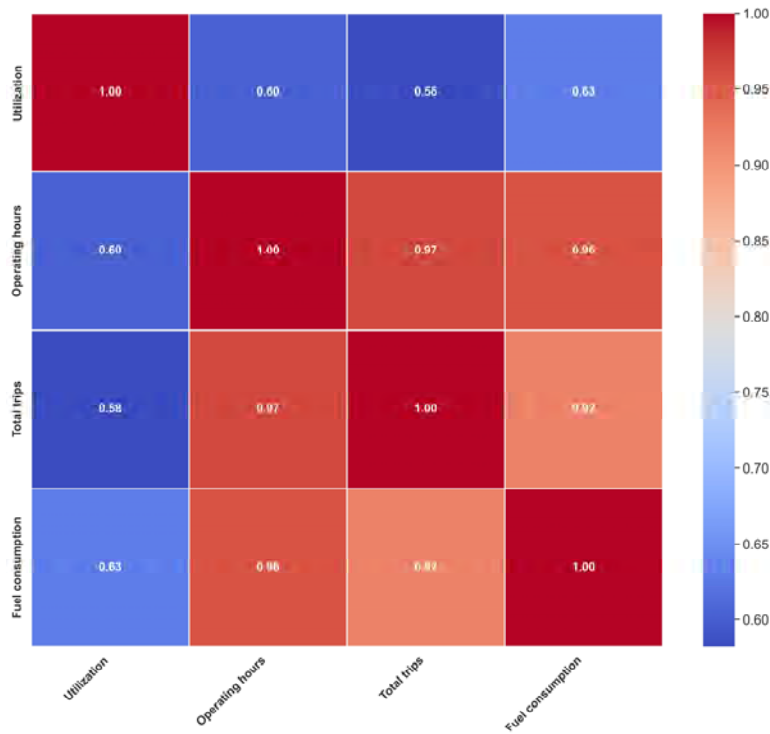


Figure 2 shows the histograms of the study variables without outliers. Utilisation (U) is concentrated between 0.80 and 0.85, operating hours (HO) present two peaks around 450 and 525 hours, total trips (TT) are mostly between 1,000 and 1,400, and fuel consumption (F) varies uniformly between 2,200 and 2,800 gallons per month. These distributions are key for predictive analysis.

Figure 3 shows the correlation matrix of the variables presented in Table 1. Operating hours (HO) and total trips (TT) show the highest positive correlation with average monthly fuel consumption (F), with values of 0.96 and 0.92, respectively. Additionally, utilisation (U) has a moderate positive correlation of 0.63 with average monthly fuel consumption (F). This indicates that as operating hours and total trips increase, fuel consumption also increases.

2.2 Predictive machine learning models

2.2.1 Artificial neural network

Artificial neural networks (ANN) employ a massive interconnection of simple processing elements that learn incrementally from their environment, capturing essential linear and nonlinear trends in complex data. This allows ANNs to provide reliable predictions even in situations with noisy and partial information (Shalkoff, 1997). They are extremely effective for fitting nonlinear models to multidimensional data, offering desired accuracy (Smith, 1993). Functionally, they can be considered extensions of multivariate techniques such as multiple linear regression and nonlinear regression (Samarasinghe, 2006). The appeal of ANNs lies in their learning and generalisation capabilities. A neural network gains its intelligence from the collective behaviour of simple computational mechanisms in individual neurons (Kweku et al., 2023). Each neuron receives inputs associated with weights that correspond to electrochemical impulses and synaptic connections in the brain. These synaptic weights are determined as the neural network learns (Turban et al., 2001).

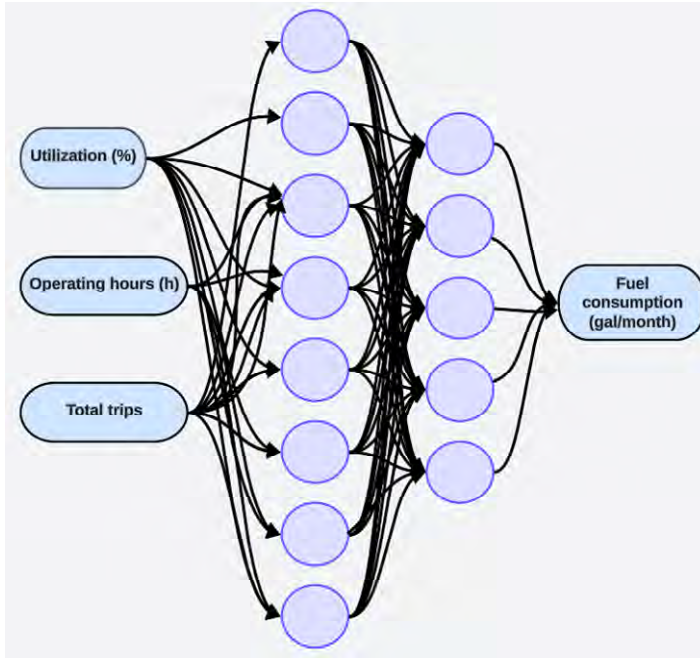
Table 2 Artificial neural network hyperparameters

<i>Artificial neural network</i>	<i>Initial range</i>	<i>Optimised by GA</i>
Learning rate	0.001–0.1	Yes
Batch size	10–100	Yes
Epochs	10–100	Yes
Input layers	1	No
Hidden layers	2	No
Output layer	1	No
Activation function	ReLU	No
Optimiser	Adam	No
Loss function	MSE	No

In this research, the following initial hyperparameters were applied: a learning rate between 0.001 and 0.1, batch size between 10 and 100, and epochs between 10 and 100, which were optimised using the genetic algorithm (GA). The structural parameters such as input layers, hidden layers, and output layers were set with 1, 2, and 3, respectively. The input layer had 3 neurons, the hidden layers had 128 and 64 neurons, and the output

layer had 1 neuron. Other hyperparameters are shown in Table 2. The structure of the artificial neural network is depicted in Figure 4.

Figure 4 Structure of the artificial neural network (see online version for colours)



2.2.2 Decision tree

The decision tree (DT) is a widely used machine learning algorithm for both classification and regression tasks. This model has a tree-like structure, where each internal node represents a test on an attribute, each branch indicates the outcome of the test, and each leaf node represents a class label or a numerical value. The algorithm works by recursively splitting the data based on the most important attribute at each level of the tree, thus forming a decision path from the root to the leaf node (Owusu et al., 2023).

The initial hyperparameters of the decision tree include a maximum depth ranging from 1 to 20 and a minimum of samples to split a node between 2 and 20. These hyperparameters were optimised using GA to enhance the model's performance.

2.2.3 Random forests

Random forests (RF) have the advantage of handling imbalanced, multiclass, and small sample data without the need for prior data preparation. RF, developed by Breiman (2001), is a regression approach that uses a wide range of decision tree techniques to predict or characterise the value of a variable. Essentially, RF builds multiple regression trees and averages their results after receiving an input vector composed of the characteristic values of a specific training set (Hastie et al., 2009).

The RF model reduces variance through the bagging method, developing trees from various subsets of training data and eliminating the correlation between different decision trees. To generate the training data, RF uses the bagging approach through the resampling with replacement technique from the original dataset (Cotrina et al., 2024). The initial hyperparameters for the random forest model include the number of estimators (10–200), maximum depth (1–20), and minimum samples per split (2–20), which were optimised using G) to enhance the model's performance.

2.2.4 Genetic algorithm

The GA is a population-based heuristic search algorithm. Its fundamental idea is to simulate the principle of 'survival of the fittest' in nature, replicating biological processes such as selection, reproduction, crossover, and mutation through computer programs. The search for solutions is conducted at a group level, ultimately seeking the optimal solution in the solution space (Liu, 2021). During the GA optimisation process, possible solutions are first considered as independent individuals. These individuals are encoded to form an initial population, and through evolution generation after generation, an approximate optimal solution is generated (Gu et al., 2022).

The genetic algorithm works as follows: consider a random forest (RF) model. Initially, a population of possible hyperparameter combinations (number of estimators, maximum depth, minimum samples per split) is created. Each combination is evaluated using a fitness function, such as model accuracy. The best individuals are selected to reproduce, combining their characteristics and adding small mutations to explore new solutions. This process of selection, crossover, and mutation is repeated over several generations. Finally, the hyperparameter combination that maximises model accuracy is obtained, demonstrating the effectiveness of the GA in finding optimal solutions.

2.2.5 Hybrid machine learning model

To address the problem of predicting fuel consumption in mining dump trucks, hybrid machine learning models were implemented, combining ANN, DT, and RF, optimised using GA. These hybrid models, referred to as ANN-GA, DT-GA, and RF-GA, are used to predict the average monthly fuel consumption in mining dump trucks.

The process begins by splitting the dataset into 70% for training and 30% for testing. The training set is used for the model to learn the nonlinear relationships between features and labels, while the test set evaluates the model's performance. The hyperparameter vectors of ANN, DT, and RF are randomly initialised as encoded individuals, and the prediction error is used as the objective function. Through the genetic algorithm, evolutionary iteration is performed to search for the best hyperparameter combination in the solution space.

When executing the combined models, the basic GA hyperparameters must be initialised, such as population size, number of generations, and crossover and mutation probabilities. Table 3 shows the genetic algorithm hyperparameters used in the optimisation process. In the case of ANN, the learning rate, batch size, and epochs were optimised. For DT, the maximum depth and minimum samples per split were adjusted. Finally, in RF, the number of estimators, maximum depth, and minimum samples per split were optimised.

The GA hyperparameters are essential to guide the evolutionary process. The population (10) determines the number of solutions in each generation, and the number of generations (10) defines the repetitions of the process. The crossover probability (0.5) and mutation probability (0.2) influence the creation and alteration of solutions. The tournament size (3) is the number of solutions competing to be selected. The Gaussian mutation parameters mu (0) and sigma (1) control the distribution of mutations, and the individual mutation rate (indpb) of 0.2 determines the probability of each parameter undergoing mutation. These hyperparameters optimise the ANN, DT, and RF models.

Table 3 GA hyperparameters used in machine learning models

<i>Genetic algorithm</i>	<i>Range</i>
Population	10
Number of generations	10
Crossover probability	0.5
Mutation probability	0.2
Tournament size	3
Gaussian mutation mu	0
Gaussian mutation sigma	1
Individual mutation rate (indpb)	0.2

2.3 Model evaluations metrics

Several evaluation metrics were applied to measure and compare the performance of the prediction model. In this work, the metrics considered were the coefficient of determination (R^2), mean squared error (MSE), and mean absolute error (MAE). These metrics were used to evaluate and compare the success of the performance of the machine learning models (Sebbeh-Newton and Binti, 2021). The mathematical formulas for these metrics are:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y - y')^2}{\sum_{i=1}^N (y - \bar{y})^2} \quad (1)$$

$$RMSE = \frac{1}{N} \sum_{i=1}^N (y - y')^2 \quad (2)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y - y'| \quad (3)$$

In the above equations, y represents the actual value, $\bar{y}$ indicates the mean of y , and y' is the predicted value of y . The total number of observations is N .

3 Results

3.1 Artificial neural network with genetic algorithm

Table 4 shows the hyperparameters of the ANN that were optimised using a GA. The initial values for the learning rate, batch size, and epochs were in the ranges of 0.001 to 0.1, 10 to 100, and 10 to 100, respectively. GA determined the optimal values as a learning rate of 0.1, a batch size of 28, and 67 epochs. These optimised settings significantly improved the performance of the ANN model in predicting the fuel consumption of mining dump trucks.

Table 4 Artificial neural network hyperparameters optimised with genetic algorithm

Artificial neural network	Initial range	Optimal value determined by GA
Learning rate	0.001–0.1	0.1
Batch size	10–100	28
Epochs	10–100	67

Figure 5 compares the actual and predicted fuel consumption values using the ANN model optimised with GA. The graph on the left shows the prediction on the training set, with a coefficient of determination (R^2) of 0.95, indicating a high accuracy in model fitting. The graph on the right shows the prediction on the test set, with an R^2 of 0.87, demonstrating a robust and generalisable predictive capability of the model. These results highlight the effectiveness of the ANN + GA model in capturing the nonlinear relationship between explanatory variables and fuel consumption.

Figure 5 Comparison of actual and predicted values with the ANN + GA model, (a) training set prediction (b) prediction in test set (see online version for colours)

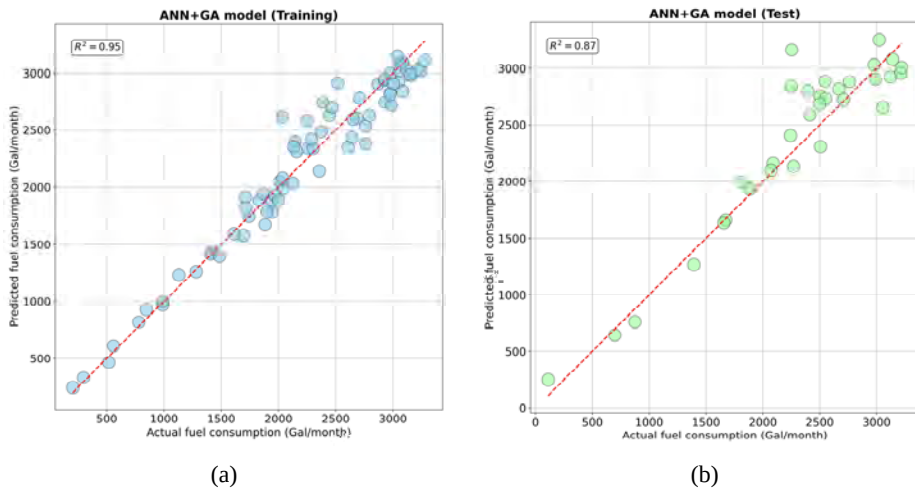
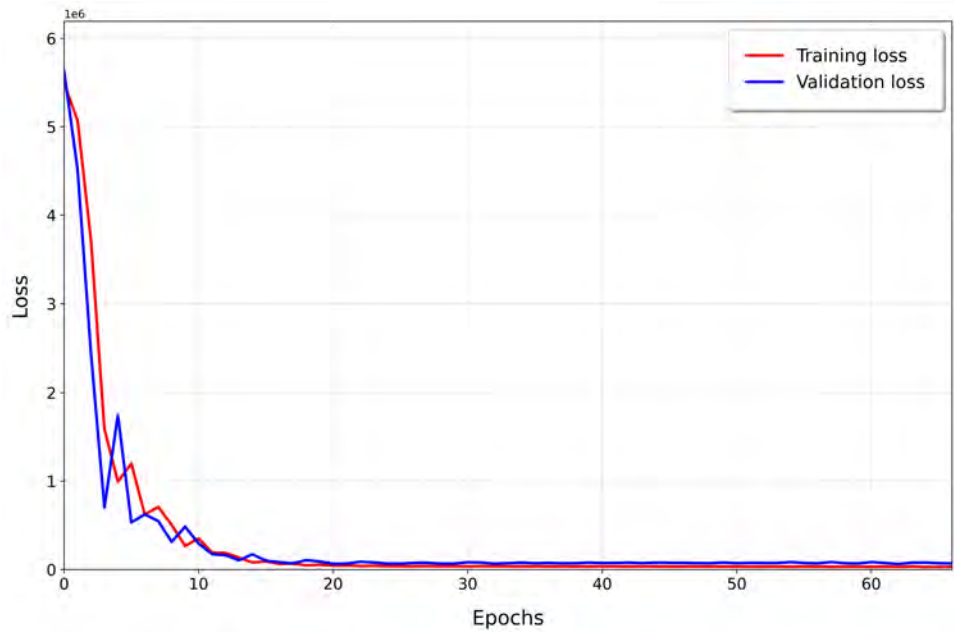


Figure 6 presents the loss curve during the training and validation of the ANN + GA model. Both curves show a rapid and steady decrease in loss during the first 10 to 20 epochs, stabilising around 30 epochs. The final loss in the training and validation sets is close to zero, indicating that the model has effectively learned the features of the data.

without overfitting. The proximity between the training and validation curves suggests that the model generalises well to unseen data, confirming the robustness of the optimisation process with GA.

Figure 6 Loss curve in training and validation (see online version for colours)



3.2 Decision tree with genetic algorithm

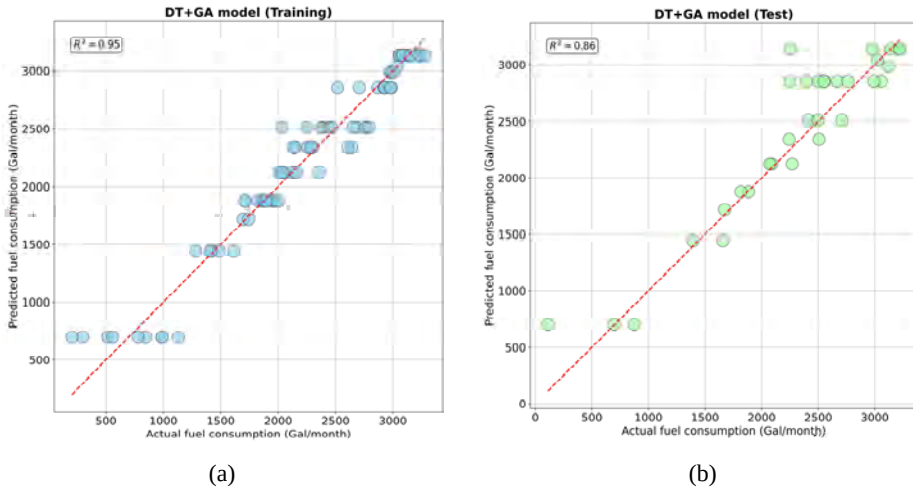
Table 5 shows the hyperparameters of the DT model that were optimised using a GA. The initial ranges for the maximum depth and minimum samples per split were between 1–20 and 2–20, respectively. The GA determined the optimal values as a maximum depth of 15 and a minimum samples per split of 12. These optimised settings significantly improved the performance of the DT model.

Table 5 Hyperparameters used in the decision tree model

Decision tree	Initial range	Optimal value determined by GA
Max depth	1–20	15.00
Min sample split	2–20	12.00

Figure 7 compares the actual and predicted fuel consumption values using the DT model optimised with GA. The graph on the left shows the prediction on the training set, with a coefficient of determination (R^2) of 0.95, indicating a high accuracy in model fitting. The graph on the right shows the prediction on the test set, with an R^2 of 0.86, demonstrating a robust and generalisable predictive capability of the model.

Figure 7 Comparison of actual and predicted values with the DT + GA model, (a) training set prediction (b) prediction in test set (see online version for colours)



3.3 Random forests with genetic algorithm

Table 6 presents the hyperparameters of the RF model that were optimised using a GA. The initial ranges for the number of estimators, maximum depth, and minimum samples per split were 10–200, 1–20, and 2–20, respectively. The GA determined the optimal values as 110.86 estimators, a maximum depth of 6.42, and a minimum samples per split of 2.07. These optimised values significantly improved the predictive capability of the RF model in estimating the fuel consumption of mining dump trucks.

Table 6 Hyperparameters used in the random forest model

Random forest	Initial range	Optimal value determined by GA
N estimators	10–200	110.86
Max depth	1–20	6.42
Min samples split	2–20	2.07

Figure 8 shows the comparison between the actual and predicted fuel consumption values using the RF model optimised with GA. In the training set [Figure 8(a)], the model showed a coefficient of determination (R^2) of 0.99, indicating excellent model fitting accuracy. In the test set [Figure 8(b)], the model achieved an R^2 of 0.89, demonstrating its robust predictive and generalisation capability. These results highlight the effectiveness of the RF + GA model in capturing the complex relationships between explanatory variables and fuel consumption, surpassing other approaches in terms of accuracy and reliability.

Figure 8 Comparison of actual and predicted values with the RF + GA model, (a) training set prediction (b) prediction in test set (see online version for colours)

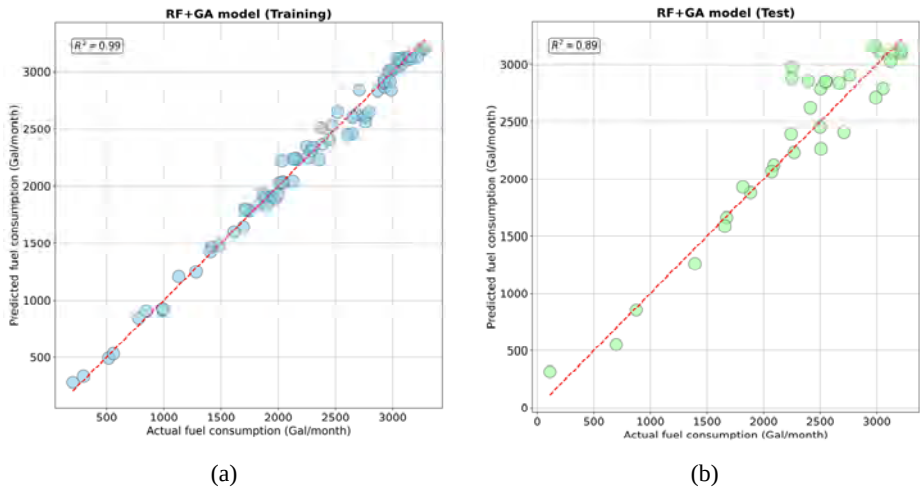


Figure 9 Importance of the variables used unprediction (see online version for colours)

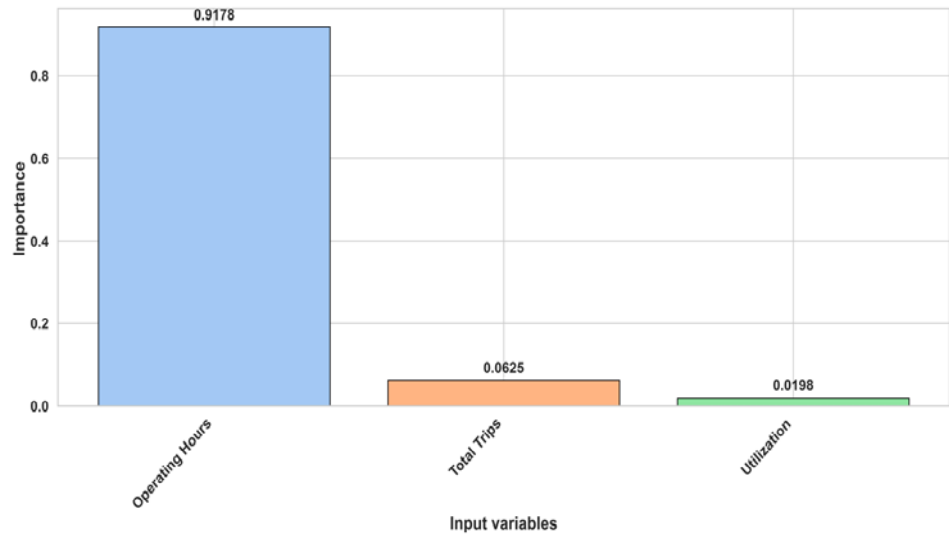


Figure 9 highlights the relative importance of the variables used in predicting fuel consumption. Operating hours emerged as the most significant variable, with an importance of 0.9178, followed by total trips with 0.0625 and utilisation with 0.0198. These results emphasise that operating hours are the predominant factor in determining fuel consumption, while total trips and utilisation have a much smaller contribution. This variable importance analysis provides valuable insights for optimising mining operations by focusing on the factors that most influence fuel consumption.

3.4 Comparison of hybrid predictive models

Figure 10 presents the comparison of hybrid models in predicting fuel consumption. On the left, the prediction on the training set is shown, and on the right, the prediction on the test set. The models compared include ANN optimised with a genetic algorithm (ANN + GA), DT optimised with a genetic algorithm (DT + GA), and RF optimised with a genetic algorithm (RF + GA). The results indicate that all models accurately predict fuel consumption in the training set, while in the test set, the RF + GA model demonstrates a slight advantage in terms of accuracy.

Figure 10 Model comparison in forecasting, (a) training set prediction (b) prediction of the test set (see online version for colours)

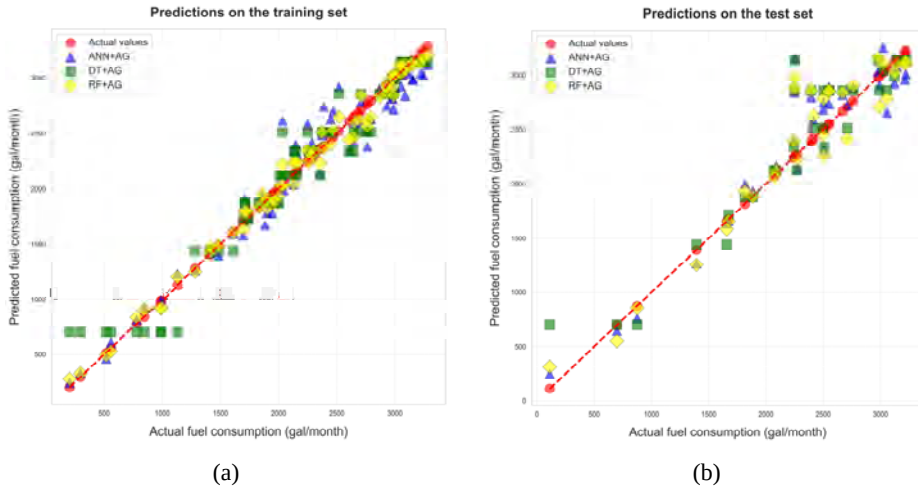


Table 7 summarises the performance of the models in terms of the coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE) in both the training and test sets. In the training set, the RF + GA model achieves an R^2 of 0.99, an RMSE of 78.22, and an MAE of 62.25, outperforming the ANN + GA and DT + GA models, both with an R^2 of 0.95. In the test set, RF + GA maintains the best performance with an R^2 of 0.89, an RMSE of 250.05, and an MAE of 185.09, compared to ANN + GA (R^2 of 0.87, RMSE of 262.30, and MAE of 190.71) and DT + GA (R^2 of 0.86, RMSE of 272.55, and MAE of 185.89).

Table 7 Comparison of model performance in predicting fuel consumption

Model	Training			Test		
	R^2	RMSE	MAE	R^2	RMSE	MAE
ANN + AG	0.95	167.74	127.87	0.87	262.30	190.71
DT + AG	0.95	172.22	127.80	0.86	272.55	185.89
RF + AG	0.99	78.22	62.25	0.89	250.05	185.09

3.5 Cost evaluation and replacement of mining dump trucks

The RF + GA model predicted the fuel consumption of mining dump trucks with different capacities, providing a basis for evaluating efficiency and replacement. Table 8 shows the monthly fuel consumption cost for 24 m³ and 26 m³ capacity dump trucks. The 24 m³ dump trucks consume an average of 2,331.65 gallons per month, resulting in a total cost of US\$10,492.43 per month, with a fuel price of US\$4.5 per gallon. On the other hand, the 26 m³ dump trucks consume an average of 2,250.46 gallons per month, which translates to a total cost of US\$10,127.07 per month. The difference in fuel consumption is –81.19 gallons, representing a monthly saving of US\$365.36 by using higher capacity dump trucks. This indicates that replacing 24 m³ dump trucks with 26 m³ dump trucks can significantly reduce fuel costs. This compares with the research by Anticona (2023), where 24 m³ dump trucks had a total cost of US\$11,524.62 per month, and 26 m³ dump trucks had a cost of US\$7,771.08 per month, achieving a saving of US\$3,753.55 per month.

Table 8 Monthly fuel consumption cost

<i>Dump truck capacity (m³)</i>	<i>Fuel consumption (gal/month)</i>	<i>Cost per fuel (US\$/gal)</i>	<i>Total cost (US\$/month)</i>
24	2,331.65	4.5	10,492.43
26	2,250.46	4.5	10,127.07
Difference	–81.19	-	–365.36

Table 9 Comparison of operational variables of dump trucks

<i>Dump truck capacity (m³)</i>	<i>Number of trips</i>	<i>Utilisation (%)</i>	<i>Operating hours (h)</i>	<i>Fuel cost (US\$/h)</i>
24	1,035.53	0.79	442.11	23.73
26	1,116.50	0.87	460.83	21.97
Difference	80.97	0.08	18.72	–1.76

Table 9 compares the operational variables of 24 m³ and 26 m³ capacity dump trucks. The 26 m³ dump trucks make 1,116.50 trips per month, with a utilisation rate of 0.87% and 460.83 operating hours, while the 24 m³ dump trucks make 1,035.53 trips per month, with a utilisation rate of 0.79% and 442.11 operating hours. The fuel cost per hour is US\$23.73 for 24 m³ dump trucks and US\$21.97 for 26 m³ dump trucks. The difference indicates that the higher capacity dump trucks have superior utilisation (0.08), more operating hours (18.72), and a lower fuel cost per hour (–US\$1.76), highlighting their greater operational efficiency and lower operating cost. These predictions and comparisons, provided by the RF + GA model, emphasise the economic and operational benefits of replacing 24 m³ dump trucks with 26 m³ dump trucks in terms of fuel consumption, costs, and operational efficiency in mining.

4 Discussions

The results obtained in our research using hybrid models ANN + GA, DT + GA and RF + GA demonstrate a high level of accuracy in predicting the fuel consumption of

mining dump trucks. In particular, the RF + GA model achieved a coefficient of determination (R^2) of 0.89 and an RMSE of 250.05 in the test set, outperforming the ANN + GA ($R^2 = 0.87$, RMSE = 262.30) and DT + GA ($R^2 = 0.86$, RMSE = 272.55) models. These results represent a significant improvement compared to the research by Alamdari et al. (2022), where the best performance was obtained by the RF model with an R^2 of 0.90 and an RMSE of 511.34. Although their R^2 is slightly higher, our RF + GA model managed to reduce the RMSE considerably, indicating greater accuracy in predicting fuel consumption.

In comparison with the study by Wang et al. (2021), which employed the XGBoost algorithm and obtained an R^2 of 0.93 and an RMSE of 8.78, our hybrid models, particularly RF + GA, show competitive performance. However, the XGBoost of Wang et al. offers higher accuracy, suggesting that this algorithm can be extremely effective for this type of prediction. This comparison underscores the importance of considering multiple machine learning and optimisation approaches to identify the most suitable model for specific contexts.

Table 10 Comparison of prediction models with hybrid models in fuel consumption prediction

<i>Investigation</i>	<i>Models</i>	<i>R²</i>	<i>RMSE</i>
Actual investigation	ANN + GA	0.87	262.30
	DT + GA	0.86	272.55
	RF + GA	0.89	250.05
Siami and Abbaspour (2011)	ANN	-	1.5819
Wang et al. (2021)	XGBoost	0.93	8.78
Alamdari et al. (2022)	MLR	0.81	957.63
	RF	0.90	511.34
	ANN	0.90	485.28
	SVM	0.89	568.30
	K-NN	0.88	607.29
Zargarneshad et al. (2019)	ANN	0.95	0.517

On the other hand, the research by Zargarneshad et al. (2019) with ANN achieved an R^2 of 0.95 and an RMSE of 0.517, significantly outperforming the results of our ANN + GA model. This indicates that although our hybrid models optimised with genetic algorithms have proven effective, there is still considerable room for improvement, particularly in the optimisation of artificial neural networks. These results highlight the need to continue exploring and refining hyperparameter optimisation techniques to further enhance the accuracy and robustness of fuel consumption prediction models in the mining sector.

5 Conclusions

In this study, hybrid machine learning models were implemented and optimised using artificial neural networks (ANN), decision trees (DT), and random forests (RF) with genetic algorithms (GA) to predict the fuel consumption of mining dump trucks. The results indicate that the RF + GA model achieved the best performance, with a coefficient of determination (R^2) of 0.99 in the training set and 0.89 in the test set, demonstrating its

robustness and accuracy in predicting fuel consumption. Compared to other models, the RF + GA not only outperformed ANN + GA and DT + GA but also stood out against previous studies, showing a significant reduction in RMSE and higher predictive capability. The ANN + GA model achieved an R^2 of 0.95 in the training set and 0.87 in the test set, with an RMSE of 262.30 and an MAE of 190.71 in the test set. The DT + GA model obtained an R^2 of 0.95 in the training set and 0.86 in the test set, with an RMSE of 272.55 and an MAE of 185.89 in the test set.

Using the best predictive model, RF + GA, fuel consumption predictions and operational cost calculations were performed. The 26 m³ dump trucks showed lower fuel consumption and fuel operating costs compared to the 24 m³ dump trucks. With an average monthly fuel cost of US\$10,127.07 for the 26 m³ dump trucks compared to US\$10,492.43 for the 24 m³ dump trucks, and a cost per hour of US\$21.97 for the 26 m³ dump trucks compared to US\$23.73 for the 24 m³ dump trucks, the replacement of dump trucks results in a significant saving of US\$365.36 per month and US\$1.76 per hour.

Despite the success of the hybrid models, there are some limitations. Hyperparameter optimisation can be computationally intensive, which may limit its applicability in resource-constrained scenarios. Additionally, reliance on historical data can affect adaptability to abrupt changes in operating conditions. For future work, it is recommended to explore the integration of advanced techniques like XGBoost and the combination of GA with other optimisation methods such as particle swarm optimisation (PSO). It would also be beneficial to expand the dataset to include external factors such as weather conditions and terrain characteristics. Implementing these models in a real-time environment could provide a valuable tool for the management and optimisation of mining operations, allowing dynamic adjustments based on accurate fuel consumption predictions.

References

- Alamdari, S., Basiri, M., Mousavi, A. and Soofastaei, A. (2022) 'Application of machine learning techniques to predict haul truck fuel consumption in open-pit mines', *Journal of Mining and Environmental*, Vol. 13, No. 1, pp.69–85, <https://doi.org/10.22044/jme.2022.11577.2145>.
- Anticona, S. (2023) *Reemplazo de volquetes de 24 por 26 m³ para reducir el costo horario en ciclo de acarreo contrata La Mamita*, Universidad Nacional de Trujillo, Trujillo.
- Arhinful, K. and Agyei, G. (2023) 'Fuel consumption prediction in shovel-truck system of surface mine using artificial neural network', *Journal of Science and Technology Research*, Vol. 5, No. 2, <https://doi.org/10.5281/zenodo.8075149>.
- Botín, J. and Vergara, M. (2015) 'A cost management model for economic sustainability and continuous improvement of mining operations', *Resources Policy*, Vol. 46, pp.212–218, <https://doi.org/10.1016/j.resourpol.2015.10.004>.
- Breiman, L. (2001) 'Random forests', *Mach. Learn.*, Vol. 45, No. 1, pp.5–32.
- Cotrina, M., Marquina, J., Mamani, J., Arango, S., Gonzalez, J., Ccatamayo, J. and Noriega, E. (2024) 'Predictive model using machine learning to determine fuel consumption in CAT-777F mining equipment', *International Journal of Mining and Mineral Engineering*, Vol. 15, No. 2, pp.147–160, <https://dx.doi.org/10.1504/IJMME.2024.140073>.
- Dadhich, S., Bodin, U. and Andersson, U. (2016) 'Key challenges in automation of earth-moving machines', *Automation in Construction*, Vol. 68, pp.212–222, <https://doi.org/10.1016/j.autcon.2016.05.009>.

- Delgado, O., Clark, N. and Thompson, G. (2012) 'Heavy duty truck fuel consumption prediction based on driving cycle properties', *International Journal of Sustainable Transportation*, Vol. 6, No. 6, pp.338–361. <https://doi.org/10.1080/15568318.2011.613978>.
- Gibson, J. (2015) 'Going green with mobile mining equipment', *Mining Engineering*, Vol. 67, No. 3, p.30.
- Gu, Z., Cao, M., Wang, C., Yu, N. and Qing, H. (2022) 'Research on mining maximum subsidence prediction based on genetic algorithm combined with XGBoost model', *Sustainability*, Vol. 14, No. 16, pp.10421, <https://doi.org/10.3390/su141610421>.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Vol. 2, Springer, USA.
- He, K., Hou, H., Zhang, Q., He, D., An, F., Wang, M. and Walsh, M. (2005) 'Oil consumption and CO₂ emissions in China's road transport: current status, future trends, and policy implications', *Energy Policy*, Vol. 33, No. 12, pp.1499–1507, <https://doi.org/10.1016/j.enpol.2004.01.007>.
- Isnafitri, M., Rosyidi, N. and Aisyati, A. (2021) 'A truck allocation optimization model in open pit mining to minimize investment and transportation costs', *IOP Conference Series: Materials Science and Engineering*, p.1096, <https://doi.org/10.1088/1757-899X/1096/1/012024>.
- Kara, N. and Baysec, S. (2010) 'Prediction of torque and specific fuel consumption of a gasoline engine by using artificial neural networks', *Applied Energy*, Vol. 87, No. 1, pp.349–355, <https://doi.org/10.1016/j.apenergy.2009.08.016>.
- Kumar, L., Bandyopadhyay, S. and Banerjee, R. (2014) 'Benchmarking energy consumption for dump trucks in mines', *Applied Energy*, Vol. 113, pp.1382–1396, <https://doi.org/10.1016/j.apenergy.2013.08.058>.
- Kweku, S., Amoako, V. and Yevenyo, Y. (2023) 'Appraisal of artificial intelligence techniques for the prediction of on-site dump truck tyre operating hours', *International Journal of Mining and Mineral Engineering*, Vol. 14, No. 2, pp.157–179, <https://dx.doi.org/10.1504/IJMME.2023.133650>.
- Kyoungho, A., Hesham, R., Antonio, T. and Michel, V. (2002) 'Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels', *Journal of Transportation Engineering*, Vol. 128, No. 2, pp.182–190, [https://doi.org/10.1061/\(ASCE\)0733-947X\(2002\)128:2\(182\)](https://doi.org/10.1061/(ASCE)0733-947X(2002)128:2(182)).
- Liu, W. (2021) *Research on Optimization of A Company's Workshop Layout Based on Genetic Algorithm*, Diploma Thesis, China University of Mining and Technology, Beijing, China.
- Nday, I. and Thomas, H. (2019) 'Optimization of the cycle time to increase productivity at Ruashi Mining', *Journal of the Southern African Institute of Mining and Metallurgy*, Vol. 119, No. 7, <http://dx.doi.org/10.17159/2411-9717/624/2019>.
- Nejad, M., Farrashzadeh, T., Ejlali, B. and Torkayesh, A. (2021) 'Fuel consumption in mining industry using partial least squares structural equation modeling approach', *International Journal of Energy Sector Management*, Vol. 15, No. 6, pp.1122–1143, <https://doi.org/10.1108/IJESM-07-2020-0020>.
- Owusu, D., Tinoco, J., Lohrasb, F., Martins, F. and Matos, J. (2023) 'A decision tree for rockburst conditions prediction', *Applied Sciences*, Vol. 13, No. 11, p.6655, <https://doi.org/10.3390/app13116655>.
- Rahimi, F. and Abbaspour, Y. (2011) 'Artificial Neural Network and stepwise multiple range regression methods for prediction of tractor fuel consumption', *Measurement*, Vol. 44, No. 10, pp.2104–2111, <https://doi.org/10.1016/j.measurement.2011.08.006>.
- Samarasinghe, S. (2006) *Neural Networks for Applied Sciences and Engineering*, 1st ed., Auerbach Publications, New York, <https://doi.org/10.1201/9780849333750>.
- Sanjeev, K., Kumar, A., Dehuri, S. and Ghosh, A. (2023) 'An outliers detection and elimination framework in classification task of data mining', *Decision Analytics Journal*, Vol. 6, <https://doi.org/10.1016/j.dajour.2023.100164>.

- Sebbeh-Newton, S. and Binti, H. (2021) 'Application of artificial intelligence techniques for identifying rock mass quality in an underground tunnel', *International Journal of Mining and Mineral Engineering*, Vol. 12, No. 2, pp.132–148, <https://dx.doi.org/10.1504/IJMME.2021.116885>.
- Shalkoff, R. (1997) *Artificial Neural Networks*, Vol. 448, McGraw-Hill Higher Education, Michigan.
- Siami, F. and Abbaspour, Y. (2011) 'Artificial Neural Network and stepwise multiple range regression methods for prediction of tractor fuel consumption', *Measurement*, Vol. 44, No. 10, pp.2104–2111, <https://doi.org/10.1016/j.measurement.2011.08.006>.
- Smith, M. (1993) *Neural Networks for Statistical Modeling*, Vol. 240, Thomson Learning, New York.
- Soofastaei, A., Mostafa, S., Siddik, M. and Knights, P. (2016) 'Reducing fuel consumption of haul trucks in surface mines using artificial intelligence models', *Resource Operators Conference*.
- Thuy, N. and Bruce, W. (2010) 'Fuel consumption estimation for kerbside municipal solid waste (MSW) collection activities', *Waste Management and Research*, Vol. 28, No. 4, pp.289–297, <https://doi.org/10.1177/0734242X09337656>.
- Turban, E., Aronson, J. and Peng, T. (2001) *Decision Support Systems and Intelligent Systems*, Prentice International Hall, New Jersey.
- Wang, Q., Zhang, R., Lv, S. and Yangting, R. (2021) 'Open-pit mine truck fuel consumption pattern and application based on multi-dimensional features and XGBoost', *Sustainable Energy Technologies and Assessments*, Vol.43, <https://doi.org/10.1016/j.seta.2020.100977>.
- Zargarneshad, S., Dashti, R. and Ahmadi, R. (2019) 'Predicting vehicle fuel consumption in energy distribution companies using ANNs', *Transportation Research Part D: Transport and Environment*, Vol.74, pp.174–188, <https://doi.org/10.1016/j.trd.2019.07.020>.