Journal of Mining and Environment (JME) Vol. 17, No. 2, 2026, 501-518

Shahrood University of Technology

Journal of Mining and Environment (JME)

Journal homepage: www.jme.shahroodut.ac.ir

Iranian Society of Mining Engineering (IRSME)

Explainable Ensemble Machine Learning Model for Blast-Induced Noise Prediction

Clement Kweku Arthur[1]*, Yao Yevenyo Ziggah[2], and Victor Amoako Temeng[1]

1. Department of Mining Engineering, Faculty of Mining and Minerals Technology, University of Mines and Technology, Tarkwa, Western Region, Ghana
2. Department of Geomatic Engineering, Faculty of Geosciences and Environmental Studies, University of Mines and Technology, University of Mines and Technology, Tarkwa, Western Region, Ghana

Article Info

Received 11 June 2025
Received in Revised form 21 July 2025
Accepted 6 August 2025
Published online 6 August 2025

DOI: 10.22044/jme.2025.16351.3195

Keywords

Shapley additive explanations
Sustainable mining
Random search
Interpretability
Environmental impact assessment

Abstract

Blast-induced noise is one of the most persistent environmental challenges in surface mining, posing significant health risks to workers and nearby communities. Accurate prediction of noise levels prior to blasting is essential for mitigating its adverse impacts. This study proposes an explainable ensemble machine learning framework for predicting blast-induced noise using data from an open-pit gold mine in Ghana. Four ensemble models namely: Extreme Gradient Boosting (XGBoost), Gradient Boosting, Adaptive Boosting (AdaBoost), and Categorical Boosting (CatBoost), were developed and evaluated using a comprehensive dataset of 324 blasting events. Performances of the developed models were assessed using coefficient of determination (R^2), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and coefficient of the variation of the root mean squared error (CVRMSE), with XGBoost emerging as the best-performing model (R^2 = 1.0000, RMSE = 0.0005, MAE = 0.0004, MAPE = 0.0010, CVRMSE = 0.0013). To address the black-box nature of ensemble method, Shapley Additive exPlanations (SHAP) was employed, offering both global and local interpretability. SHAP analysis identified the distance from the blast site to the monitoring point as the most influential factor. This integrative approach not only enhances predictive accuracy but also improves model transparency, supporting sustainable mining practices aligned with United Nations Sustainable Development Goals (SDGs) 3 and 15.

1. Introduction

Blasting is one of the most economical and effective methods of fragmenting in-situ rock masses to access valuable mineral resources such as gold, manganese, iron, and copper. However, despite its efficiency in rock breakage, blasting is known to have significant adverse environmental impacts: many of which are unavoidable, even under optimal blasting conditions. Studies have shown that only 20–30% of the explosive energy contributes directly to rock fragmentation, while the remaining 70–80% is dissipated as ground vibrations, noise, air overpressure, and flyrock. These externalities not only pose operational safety risks but also threaten the environment and nearby communities. This study focuses specifically on blast-induced noise, which has been identified as a major concern due to its health implications. According to Paik et al. [1], blasting is an extreme form of noise exposure capable of causing cochlear damage, a condition that often results in permanent hearing loss—a severe and irreversible health hazard, particularly for mine workers and residents near mining sites.

Addressing blast-induced noise is essential not only from a health and safety perspective but also within the broader context of sustainable mining and the United Nations Sustainable Development Goals (SDGs). In particular, this research aligns with SDG 3 (Good Health and Well-being) [2] by aiming to reduce occupational health risks. It also

Corresponding author: ckarthur@umat.edu.gh (C.K. Arthur)

supports SDG 15 (Life on Land) [3] by mitigating the environmental footprint of mining activities. By developing predictive models to estimate the intensity of blast-induced noise before each blast, this study contributes to minimising the negative impacts of mining operations, fostering a more responsible and environmentally conscious approach to mineral extraction. Such advancements are crucial in transitioning the mining industry toward sustainable practices that balance economic benefits with environmental stewardship and social responsibility.

Despite the several extensive research works that have been done to study the underlying phenomenon of ground vibration, air overpressure, and fly rocks so as to effectively model and predict their occurrences, less work has been done in the area of blast induced noise modelling and prediction. From literature, it has been found that only Temeng et al. [4] and Ziggah et al. [5] have done works in the area of modelling and prediction of blast-induced noise levels. Temeng et al. [4] applied brain inspired emotional neural network (BENN), backpropagation neural network (BPNN), radial basis function neural network (RBFNN), generalised regression neural network (GRNN), group method of data handling (GMDH), least squares support vector machine (LSSVM) and support vector machine (SVM) in predicting blast-induced noise levels. Ziggah et al. [5] on the other hand developed a synergistic model based on neighbourhood component analysis (NCI) and artificial intelligence method (AI) for the prediction of blast-induced noise levels.

This study however, applies four ensemble machine learning techniques namely: extreme gradient boosting, gradient boosting, adaptive boosting and categorical boosting to predict blast-induced noise levels. Ensemble Learning algorithm denotes techniques in machine learning that amalgamates the predictions of various weak learners to enhance performance efficacy in both regression and classification problems [6]. These methodologies, despite their complexity and reduced interpretability compared to individual models, have demonstrated improved outcomes in numerous engineering applications notably in the area of hydrology [7-9], blast-induced ground vibration [10-13], slope stability [14-16] among others. The analysis of these engineering examples shows that the ensemble machine learning algorithm can deal well with the relationship between input and output parameters and give accurate and reliable prediction results, which has good applicability for blast-induced noise prediction.

Despite ensemble learning models being able to make accurate predictions based on input data, the decision-making process and the reasons behind those predictions are not readily accessible as they are “black box” (opaque) in nature [17, 18]. This poses a significant limitation as transparency and explainability of machine learning models are being recognised increasingly as critically important in recent years [17]. To address this limitation, Explainable Artificial Intelligence (XAI) [19] is adopted to help explain the behaviour of the ensemble models, offering insights into their decision-making processes. Thus, in this study, Shapley Additive exPlanations (SHAP) [20] which is one of the widely used Explainable Artificial intelligence (XAI) methods is employed to provide an explanation of the optimum ensemble machine learning model in predicting blast-induced noise. Here, SHAP provides a global and local interpretation of the prediction results. For the global interpretation, a SHAP summary plot is employed based on relevant contributions, to provide insight into the importance of each feature in the decision process of the developed model. Whereas for the local interpretation, the contribution of each feature for a particular instance is determined. This is visually done using the SHAP force plot.

The rest of the paper is structured as follows: Section 2 elaborates on the description of the study area and the dataset used in the development of the various models. Section 3 presents the methodology. Here, the mathematical descriptions of the ensemble machine learning techniques are presented. Furthermore, the mathematical framework of XAI is presented. Finally in this section, the model development procedure adapted as well as the criteria used in evaluating the developed models are presented. Section 4 details the results and discussion. Here the developed models as well as their prediction results are presented. Furthermore, a justification of the developed models in relation to work done by other researchers is presented. The section finally delves into use of XAI to improve the interpretability of the best ensemble machine learning model. This study ends with Conclusion on Section 5.

2. Study area and dataset description

2.1. Study area

The dataset used for this study was collected from Mine X which is an open pit gold mine

located in the Tarkwa, in the Western Region of Ghana. The mine is geographically positioned at latitude 5° 15' N and longitude 2°00' W. Figure 1 shows the location of the study area.

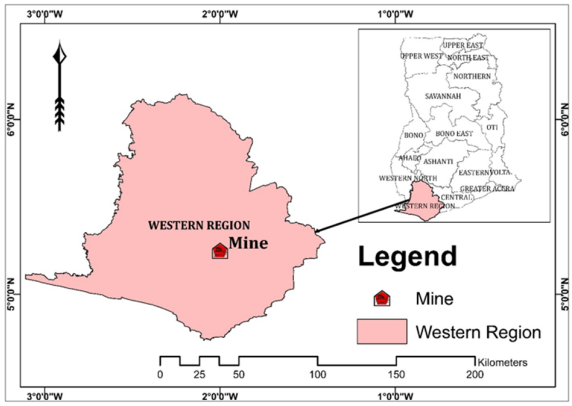


Figure 1. Location of Mine X on the map of Ghana

Mine X employs drill and blast operations to fragment its in-situ rock mass which is loaded by excavators into dump trucks and transported to either waste dump or crusher based on whether material is waste or ore. For the drilling operation, the study mine employs Sandvik's DP 1600i, Di 550i and DP 1500i drill rigs. For blasting, the mine uses emulsion as the main explosive with either an electronic initiation system or non-electric initiation system depending on the proximity of the blasting area to the neighbouring community. For the loading operation, the mine uses both Caterpillar and Liebherr excavators whereas for the hauling operation, the mine uses Caterpillar rear dump trucks.

2.2 Data description

For this study, a total of 324 datasets were collected from the mining and environmental department of the study area. The dataset comprised of the following parameters: number of holes, cooperating charge (kg), hole depth (m), stemming height (m), total charge (kg), powder factor (kg/m3), distance from the blasting area to the monitoring point and noise levels (dBL). Parameters of number of holes, cooperating charge (kg), hole depth (m), stemming height (m), total charge (kg), powder factor (kg/m3) and distance from the blasting area to the monitoring point served as the inputs to the development of the various ensemble machine learning models while the noise levels served as the output parameter. Table 1 outlines the statistical descriptions of the input and output parameters.

Table 1. Statistical descriptions of input and output parameters

Parameters	Abbrev.	Type	Mean	Std Dev	Std Err	Min	Max	Skewness
Number of Holes	NH	Input	106.75	65.31	3.63	13	423	2.07
Cooperating Charge (kg)	CC		72.27	31.40	1.74	10.76	175	0.52
Hole Depth (m)	HD		8.26	1.95	0.11	3.4	10.2	-0.42
Stemming (m)	S		3.20	0.38	0.02	1.8	3.5	-1.96
Total Charge (kg)	TC		8091.99	6782.20	376.79	495	49100	2.43
Distance (m)	D		1312.40	516.37	28.69	518	3862	1.55
Powder Factor (kg/m^3)	PF		0.67	0.18	0.01	0.24	2.07	1.86
Noise (dBL)	N	Output	42.88	4.16	0.23	33.1	53.4	-0.25

The visual representation of the correlation between the input and output parameters was done using pair plot as shown in Figure 2.

With reference to Table 1 and Figure 2, it can be observed that number of holes and total charge are strongly skewed to the right. This indicates that most of the blast data had lower values of number of holes and total charge with a few high values. Similarly, distance and powder factor are moderately to strongly positively skewed, showing that shorter distances and lower powder factors are more common in the dataset, with a few larger values. The cooperating charge shows mild positive skewness, suggesting a near-normal distribution with a slight lean toward lower values. Stemming is strongly negatively skewed. This shows that higher stemming values are common in the data with few low values. Hole depth and noise display mild negative skewness, indicating slightly more high-end values in their distributions.

Furthermore, it can be observed that, number of holes is strongly positively correlated to total charge and moderate correlation with distance, but shows weak correlation with cooperating charge, hole depth, powder factor and noise levels. It also shows negligible correlation with stemming. Cooperating charge on the other hand exhibit strong correlation with hole depth, moderate correlation with total charge and powder factor but exhibits weak correlation with stemming, distance and noise. Hole depth has a moderate correlation with total charge and weak correlation with

distance and powder factor. It has negligible correlation with stemming, noise level. Stemming shows weak correlation with total charge and power factor but exhibited negligible correlation with distance and noise. Total charge also exhibits a moderate correlation with distance and weak correlation with powder factor. It also shows negligible correlation with noise. It can be seen that distance exhibits strong negative correlation with noise but has a weak correlation with powder factor. Power factor on the on the other had exhibits weak correlation with noise. It is evident that with the exception of distance all other input features exhibited a negligible to weak correlation to noise. Nonetheless, according to Kuhn and Johnson [21], machine learning models can accurately predict outcomes even when individual features exhibit weak correlation with the target variable, as they are capable of capturing complex nonlinear relationships and interactions that traditional correlation metrics may overlook. Thus, in this study all seven input features were used in the development of the ensemble machine learning models.

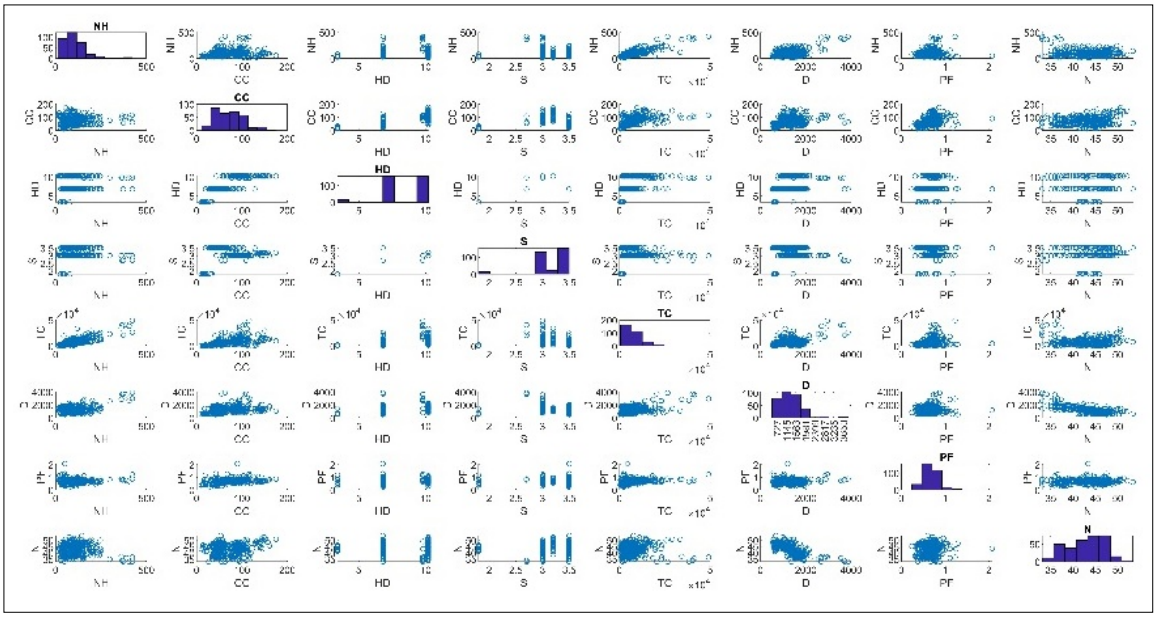


Figure 2. Pair plot of the input and output parameters

Values of the noise levels were obtained from readings from blasting seismograph (precisely an Instantel Minimate Plus), an instrument used in measuring noise levels. Values of other parameters such as number of holes, cooperating charge (kg), hole depth (m), stemming height (m), total charge (kg), powder factor (kg/m3) were obtained from the daily blasting plan. Finally, the distance *d* values were obtained using Equation 1.

$$d = \sqrt{\left(x_b - x_m\right)^2 + \left(y_b - y_m\right)^2} \quad (1)$$

Where are:

x_b – x coordinates of the blasting point,

y_b –y coordinates of the blasting point,

x_m – x coordinates of the monitoring station,

y_m – y coordinates of the monitoring station.

3. Methodology

In this section concise descriptions of the various ensemble machine learning models are presented. Furthermore, the framework of XAI is explained. The procedure adopted in the development of the various ensemble models as well as the criteria used in validating the various models are presented.

3.1. Ensemble learning models

3.1.1. Extreme gradient boosting (XGBoost)

Extreme gradient boosting algorithm developed by Chen and Guestrin [22] is a scalable end-to-end tree boosting system used for both classification and regression problems. It works in the following manner [7, 8]: Given a dataset, *S* having *t* features and *r* observations, $S = \left\{\left(X_k, Z_k\right): k = 1,\ldots,r, X_k \in \square^t\right\}$, an output

$\hat{Z}_k$ is predicted by an ensemble tree model using m additive functions as shown in Equation 2.

$$\hat{Z}_k = \sum_{j=1}^{M} f_j\left(X_k\right), f_j \in F \tag{2}$$

Where are:

F – space of all possible regression trees,

M – number of regression functions,

X_k – input vector,

f_j – an individual regression tree.

To find the best set of functions, the loss and regularisation objective is minimised as shown in Eq. (3).

$$L = \sum_{k=1}^{r} \ell\left(Z_k, \hat{Z}_k\right) + \sum_{j=1}^{M} \Phi\left(f_j\right) \tag{3}$$

Where are:

ℓ – loss function which is the difference between the actual output and predicted output,

Z_k – actual output,

$\hat{Z}_k$ – predicted output,

Φ – regularisation term (as shown in Equation 4) which measures the model complexity and helps in avoiding overfitting.

$$\Phi\left(f_j\right) = \nu G + \frac{1}{2}\sigma\left\|\omega\right\|^2 \tag{4}$$

Where are:

ν – degree of complexity of each leaf,

G – number of decision tree leaves,

σ –trade-off parameter to scale the penalty;

ω –leaf weight.

As the model continues to train, a new function f_u is added. Thus, the new function is added at the *u-th* iteration as shown in Equation 5 to minimise the objective.

$$L^{(u)} = \sum_{k=1}^{r} \ell\left(Z_k, \hat{Z}_k^{(u-1)} + f_u\left(X_k\right)\right) + \Phi\left(f_u\right) \tag{5}$$

Optimisation of Equation 4 is made simpler by expanding the loss function in accordance with the second-order Taylor series (Equation (6))

$$f\left(y\right) = f\left(b\right) + \frac{f'\left(b\right)}{1!}\left(y-b\right) + \frac{f''\left(b\right)}{2!}\left(y-b\right)^2 \tag{6}$$

When y is $L^{(u)} = \sum_{k=1}^{r} \ell\left(Z_k, \hat{Z}_k^{(u-1)} + f_u\left(X_k\right)\right) + \Phi\left(f_u\right)$ and b is $\hat{Z}_k^{(u-1)}$, the objective function be rewritten as Equation 7.

$$L^{(u)} = \sum_{k=1}^{r}\left[\ell\left(Z_k, \hat{Z}_k^{(u-1)}\right) + c_k f_u\left(X_k\right) + \frac{1}{2} d_k f_u^2\left(X_k\right)\right] + \Phi\left(f_u\right) \tag{7}$$

where are:

$c_k = \partial_{\hat{Z}^{(u-1)}} \ell\left(Z_k, \hat{Z}_k^{(u-1)}\right)$ – first order gradient statistics on the loss function,

$d_k = \partial^2_{\hat{Z}^{(u-1)}} \ell\left(Z_k, \hat{Z}_k^{(u-1)}\right)$ – second order gradient statistics on the loss function.

Expanding Φ and removing the constant terms in Equation 7 results into Equation 8.

$$\begin{aligned} \tilde{L}^{(u)} &= \sum_{k=1}^{r}\left[c_k f_u\left(X_k\right) + \frac{1}{2} d_k f_u^2\left(X_k\right)\right] + \nu G + \frac{1}{2}\sigma\sum_{m=1}^{G}\omega^2 \\ &= \sum_{m=1}^{G}\left[\left(\sum_{k\in I_m} c_k\right)\omega_m + \frac{1}{2}\left(\sum_{k\in I_m} d_k + \sigma\right)\omega_m^2\right] + \nu G \end{aligned} \tag{8}$$

where is:

$I_m = \left\{k \middle| q\left(X_k\right) = m\right\}$ – leaf m instant set.

Taking the derivative of Equation 8 with respect to ω_m and setting it equal to zero gives an optimal weight ω_m^* of leaf m as shown in Equation 9 and a

corresponding optimal loss value shown in Equation 10.

$$\omega_m^* = -\frac{\sum_{k\in I_m} c_k}{\sum_{k\in I_m} d_k + \sigma} \quad (9)$$

$$\omega_m^* = -\frac{1}{2}\sum_{m}^{G} \frac{\left(\sum_{k\in I_m} c_k\right)^2}{\sum_{k\in I_m} d_k + \sigma} + vG \quad (10)$$

Given I_A and I_B are the left and right nodes' instant set after splitting respectively and I is their union, then the loss reduction after splitting is presented in Equation 11 as:

$$L_{split} = \frac{1}{2}\left[\frac{\left(\sum_{k\in I_A} c_k\right)^2}{\sum_{k\in I_A} d_k + \sigma} + \frac{\left(\sum_{k\in I_B} c_k\right)^2}{\sum_{k\in I_B} d_k + \sigma} - \frac{\left(\sum_{k\in I} c_k\right)^2}{\sum_{k\in I} d_k + \sigma}\right] - v \quad (11)$$

3.1.2. Gradient boosting

One kind of tree ensemble model that makes use of the boosting ensemble learning technique is gradient boosting (GBRT) developed by Friedman [23]. The performance of the model is improved by the integration of several decision trees using boosting techniques because a single decision tree is a weak learner with low robustness and estimation accuracy [24]. Modification of weights of the training data are done iteratively, with samples difficult to predict having higher weight values than those well predicted. The boosting algorithm is known to be made up of three components, namely: a loss function, weak learners and an additive model [25]. Given s is a set of input variables and $\hat{y}$ is the predicted output variable, a tree ensemble model using T additive functions f_t is expressed in Equation 12 as:

$$\hat{y} = \sum_{t=1}^{T} f_t(s) = \sum_{t=1}^{T} \sigma_t d(s; b_t) \quad (12)$$

Where are:

$d(\cdot)$ – additive expansion of the basis function,

σ_t – weights of the end node of the t^{th} tree,

b_t – mean of the end nodes in the t^{th} tree.

The forward stagewise optimisation of parameters σ_t and b_t is carried out in the iteration process. To reduce the overall model's objective function $L(.)$, a new tree is added for every iteration. The estimated function after the m^{th} iterations is shown in Equation 13. Furthermore, the optimal σ_t can be determined using Equation 14.

$$f_m(s) = f_{m-1}(s) + \sigma_m d(s; b_m) \quad (13)$$

$$\sigma_m = arg\,min \sum_{r=1}^{R} L\left(y_r, f_m(s_r)\right) = arg\,min \sum_{r=1}^{R} \left(y_r - f_m(s_r)\right)^2 \quad (14)$$

Where are:

R – number of observations in the training sample,

s_r – set of input variables of the r^{th} training sample,

y_r – output variable of the r^{th} training sample.

To avoid overfitting and produce accurate results, a lower learning rate ψ is also added to Equation 13 to regulate the impact of new regression trees as shown in Equation 15.

$$f_m(s) = f_{m-1}(s) + \psi\sigma_m d(s; b_m), \psi \in (0,1) \quad (15)$$

3.1.3. Adaptive boosting

Adaptive Boosting (AdaBoost) is an ensemble strategy for regression that builds a strong predictive model by iteratively combining several weak predictive models. Given actual output y and predicted output $f(t)$, where t is the input vector, AdaBoost starts with a base model $f_0(t)$ (Equation 16) that minimises the loss $L = \left(y - f(t)\right)^2$.

$$f_0(t) = arg\,min_{\mu} \sum_{k=1}^{M} L(y_k, \mu) \quad (16)$$

Where is:

μ – base model hyperparameter.

Boosting process is done iteratively. In each iteration, the residual (Equation 17) is computed and a weak model $g_m(t)$ is trained to predict these residuals.

$$r_k^{(m)} = y_k - f_{m-1}(t) \quad (17)$$

The weight σ_n of the weak model is computed using Equation 18.

$$\sigma_m = arg\,min_\sigma \sum_{k=1}^{M} L\left(y_k, f_{m-1}(t_k) + \sigma g_m(t_k)\right) \quad (18)$$

The overall model is updated as shown in Equation 19 and this process is repeated for a predefined number of iterations.

$$f_m(t) = f_{m-1}(t) + \sigma_m g_m(t) \quad (19)$$

The final model is the weighted sum of all weak models as shown in Equation 20

$$f(t) = \sum_{m=1}^{M} \sigma_m g_m(t) \quad (20)$$

AdaBoost develops a powerful regressor that successfully lowers prediction error by concentrating on data points with bigger errors at each stage.

3.1.4. Categorical boosting

Categorical Boosting (CatBoost) proposed by Prokhorenkova et al. [26] is a gradient boosting algorithm optimised for handling categorical data. It minimises a loss function *L* between a predicted output *f(t)* and actual output y, where *t* is the input vector for *K* number of observations. CatBoost starts with an initial model designed to minimise the overall loss. It then iteratively incorporates additional models to further reduce the loss function. During the iteration process, pseudo-residuals $p_k^{(s)}$ (Equation 21) which is the negative gradient of the loss with respect to the current model's predictions are computed.

$$p_k^{(s)} = -\frac{\partial L\left(y_k, f_{s-1}(t_k)\right)}{\partial f_{s-1}(t_k)}, k = 1,..., K \text{ and } s = 1,2,...,S \quad (21)$$

Where are:

s – iteration number,

S – maximum number of iterations.

A decision tree model $q_s(t)$ is trained on these residuals, and the model is updated as shown in Equation 22.

$$f_s(t) = f_{s-1}(t) + \xi \cdot q_s(t) \quad (22)$$

Where is:

ξ – learning rate.

CatBoost leverages unique methods for handling categorical data, including encoding categorical variables with target statistics while preventing label leakage through ordered boosting [26]. Additionally, it employs regularisation techniques to minimise the risk of overfitting. The final model (Equation 23) is an integration of decision trees, forming a powerful predictive ensemble.

$$f(t) = \sum_{s=1}^{S} \xi \cdot q_s(t) \quad (23)$$

3.2. Explainable artificial intelligence

This study employed the use of Shapley Additive exPlanations (SHAP) a widely used Explainable Artificial Intelligence model to interpret predictions made by the optimum machine learning model. SHAP developed by Lundberg and Lee [20] is primarily designed to analyse the contribution of each feature on the prediction, providing valuable insights into model behaviour. As a game theory-based approach, SHAP provides a methodical way to divide the overall prediction value among features according to each one's unique contribution. Here each feature acts as a 'player' and the model prediction serves as the value. The contribution of each feature is determined by computing the Shapley value [27]. To determine the Shapley value for a feature *k* in a prediction *f(t)*, every possible subset of features is taken into account. The Shapley value represents the variation in the model's prediction when feature *k* is included in each subset. The Shapley value is computed using Equation 24.

$$\psi_k = \sum_{R \subseteq Q \setminus \{k\}} \frac{|R|!\left(|Q| - |R| - 1\right)}{|Q|} \left[f\left(R \cup \{k\}\right) - f(Q)\right] \quad (24)$$

where are

ψ_k – Shapley value for feature *k*,

Q – all the possible subset of features,

R – subset of R that does not include feature k,

$|R|$ – number of features in subset *R*,

$f(Q)$ – model's prediction using only the features in subset *R*,

$f(R \cup \{k\})$ – model's prediction using features in subset *R* including feature *k*.

The explanation model *h* is then derived by aggregating the SHAP values of all features as shown in Equation 25.

$$h(t') = \psi_0 + \sum_{k=1}^{S} \psi_k t' \quad (25)$$

where are:

t' – feature vector,

S – number of features,

ψ_k – computed Shapley values and

ψ_0 – model output when all features are unavailable.

3.3. Model development procedure

3.3.1. Data partitioning

Prior to the development of the ensemble machine learning models, the total 324 blasting events were partitioned into two separate datasets. Two hundred and fifty-nine (259) blasting events which is 80% of the total blasting events was used as the training dataset to develop the various models. The remaining 20% (65 blasting events) was used as the test dataset to independently test the performances of the developed models. The partitioning ratio of 80:20 based on holdout cross validation was adopted because several researchers have proven that this ratio is not liable to overfitting and/or underfitting as they help produce optimal models with accurate prediction results [28, 29].

3.3.2. Data normalisation

In this study, the input parameters were normalised into the range [-1,1] [30] to prevent the inputs having higher values from influencing the prediction results using Equation 26.

$$T_k = T_{min} + \frac{(S_k - S_{min}) \times (T_{max} - T_{min})}{(S_{max} - S_{min})} \quad (26)$$

where are:

T_k – normalised input data,

S_k – actual input variables,

S_{max} – maximum values of the actual input variables,

S_{min} – minimum values of the actual input variables,

T_{min} – -1,

T_{max} – 1.

3.3.3. Hyperparameter tuning

Machine learning models depend on hyperparameters that define the configurable part of their learning process. Thus, hyperparameters tuning is a critical step in enhancing the performance of the machine learning models. In this study, the random search technique [31] was adopted to determine the set of optimal hyperparameters that define the ensemble machine learning models applied in this study. Random Search involves defining a range of possible values for each ensemble machine learning model's hyperparameter and then randomly sampling combinations to find the best-performing one usually based on test set performance. The combination that performs the best is then chosen as the optimal set of hyperparameter values in the training process of the machine learning model. Table 2 provide the list of hyperparameters of the ensemble machine learning models [32] evaluated in this study.

Table 2 List of hyperparameters evaluated in development of models

Models	Hyperparameters
XGBoost	Number of trees Learning rate Maximum depth Number of needed leaves Regularisation
AdaBoost	Base Estimator Number of estimators Learning rate Loss function
CATBoost	Number of trees Learning rate Maximum depth Number of needed leaves Regularisation
Gradient Boosting	Number of trees Learning rate Maximum depth

3.3.4. Flowchart of model development procedure

Figure 3 shows the flowchart of the model development procedure. It also showcases the flow of logic adopted in this study.

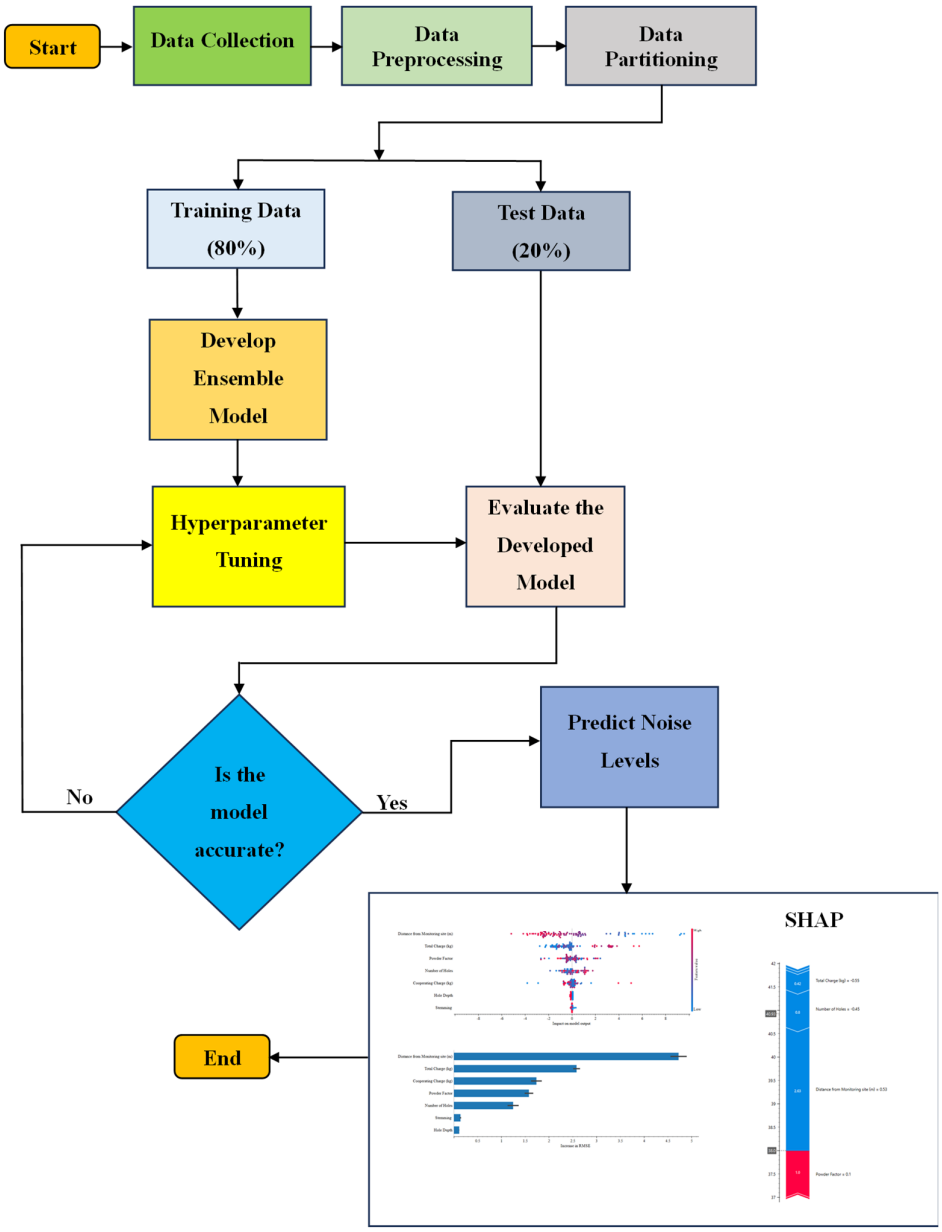


Figure 3. Framework of explainable ensemble machine learning model for predicting noise levels

3.4. Model evaluation criteria

In evaluating the various models developed, statistical performance metrics of root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), coefficient of determination (R^2) and coefficient of the variation of the root mean squared error (CV_{RMSE}) were used. These metrics are mathematically presented in Equation 27 to Equation 31.

$$RMSE = \sqrt{\frac{\sum_{w=1}^{P}\left(a_w - t_w\right)^2}{P}} \quad (27)$$

$$MAE = \frac{\sum_{w=1}^{P}\left|a_w - t_w\right|}{P} \quad (28)$$

$$MAPE = \frac{\sum_{w=1}^{P} \left| \frac{a_w - t_w}{a_w} \right|}{P} \times 100\% \quad (29)$$

$$R^2 = \left(\frac{\sum_{w=1}^{P} \left(a_w - \bar{a}_w\right)\left(t_w - \bar{t}_w\right)}{\sqrt{\sum_{w=1}^{P} \left(a_w - \bar{a}_w\right)^2 \times \sum_{w=1}^{P} \left(t_w - \bar{t}_w\right)^2}} \right)^2 \quad (30)$$

$$CV_{RMSE} = \frac{\sqrt{\frac{1}{P} \times \sum_{w=1}^{P} \left(a_w - t_w\right)^2}}{\frac{1}{P} \times \sum_{w=1}^{P} a_w} \quad (31)$$

where are:

P – number of samples in the dataset,

a_w – actual noise level values,

$\bar{a}_w$ – mean of the actual noise level values,

t_w – predicted noise level values,

$\bar{t}_w$ – mean of the values of the predicted noise level values.

The RMSE, MAE, MAPE and CV_{RMSE} show how well a regression model can predict the value of an output parameter in absolute terms. Thus, lower values imply higher accuracy of a regression model. The R^2 tells how well the predictor variables can explain the variation in the response variable. Thus, a higher value implies an accurate model [33].

4. Results and discussion

4.1. Developed models

After the hyperparameters tuning, the determined optimal hyperparameter values for the various ensemble learning machine models are presented in Table 3.

Table 3. Optimal hyperparameters of developed models

Models	Hyperparameters	Range of Values
XGBoost	Number of trees	65
	Learning rate	1
	Maximum depth	3
	Regularisation	0.001
AdaBoost	Base Estimator	Tree
	Number of estimators	50
	Learning rate	1
	Loss function	Exponential
CATBoost	Number of trees	100
	Learning rate	0.7
	Maximum depth	6
	Regularisation	3
Gradient Boosting	Number of trees	50
	Learning rate	0.9
	Maximum depth	3

The results based on RMSE and R, after training and testing the developed ensemble machine models are presented in Table 4.

4.2. Model evaluation

Using the test dataset, the performance of the various models based on the performance metrics of RMSE, MAE, MAPE, R^2 and CV_{RMSE} are outlined in Table 5.

To analyse the results presented in Table 5 so as to ascertain the best performing model, the model scoring and ranking method [20] was adopted. Here, for each metric, the model with the best metric value was given a high score value. Since there are four models, the highest score was set to 4, followed by 3, then 2, then finally 1, for the lowest performing method. For models having the same metric values, the average of the total scores for the models, were used. After obtaining the individual score values for each metric for the various models, the total score is obtained by summing up the individual score for the various models. The model with the highest score is then ranked as the best model. Table 6 shows the results of the applied model scoring and ranking method.

Table 4. Training and testing results of the various ensemble machine learning models

Model	Training		Testing	
	RMSE	R^2	RMSE	R^2
XGBoost	0.031	0.9999	0.0005	1.0000
AdaBoost	0.072	0.9997	0.0746	0.9996
CATBoost	0.081	0.9996	0.0016	1.0000
Gradient Boosting	0.075	0.9996	0.0017	1.0000

Table 5. Performance results of the various models

Performance Indicators	XGBoost	AdaBoost	CATBoost	Gradient Boosting
RMSE	0.0005	0.0746	0.0016	0.0017
MAE	0.0004	0.0126	0.0010	0.0012
MAPE	0.0010	0.0307	0.0030	0.0031
R^2	1.0000	0.9996	1.0000	1.0000
CV_{RMSE}	0.0013	0.1824	0.0039	0.0041

Table 6. Ranking results of the various models

Performance Indicators	XGBoost	AdaBoost	CATBoost	Gradient Boosting
RMSE	4	1	3	2
MAE	4	1	3	2
MAPE	4	1	2	3
R^2	3	1	3	3
CV_{RMSE}	4	1	3	2
Total Score	19	5	14	12
Rank	**1**	**4**	**2**	**3**

From Table 6 it can be seen that XGBoost had the lowest RMSE, MAE, MAPE and CV_{RMSE} and thus had the highest score of 4. This was followed by CATBoost which had a score value of 3 for RMSE, MAE and CV_{RMSE}. This was then followed by Gradient Boosting which had a score value of 2 for RMSE, MAE and CV_{RMSE}. For the R^2, XGBoost, CATBoost and Gradient Boosting had the same value, thus a score value of 3 (ie. $\frac{4+3+2}{3}=\frac{9}{3}=3$) was assigned. The Adaboost had the lowest metric values and thus had a score of 1 for the metric values. Summing up the scores, XGBoost had a total score of 19 and thus was ranked as the best model in predicting noise levels. This was followed by CATBoost with a total score of 14. Then by Gradient Boosting with a total score of 12. Adaboost was the least performing model with a total score value of 5.

4.3. Validation of models

In order to justify the XGBoost technique as an accurate estimator of blast-induced noise level, the prediction results obtained for XGBoost was compared to other research works done by other researchers in estimating blast-induced noise level. This is as shown in Table 7.

Table 7. Research works conducted

Reference	Methods	R	RMSE
Ziggah et al. [5]	Optimised backpropagation neural network with neighbourhood component analysis (NCA)	0.912	1.558
Temeng et al. [4]	Brain Inspired Emotional Neural Network (BENN)	0.911	1.619
Proposed model	Extreme Gradient Boosting (XGBoost)	1.000	0.0005

With reference to Table 7, it can be gleaned that XGBoost is a great tool to estimate blast-induced noise level as it had better R value and lower error indicators relative to methods proposed by Temeng et al. [4] and Ziggah et al. [5], as the same dataset was used in their work.

4.4. XAI based on best ensemble learning model

4.4.1. Global explanations

Using the feature importance plot as shown in Figure 4, the global importance of each input parameter with respect to the predicted noise level based on the training dataset can be evaluated.

It can be seen from Figure 4 that distance from the monitoring site to the blasting point, having the highest mean SHAP value is the most influencing input parameter in the prediction of blast-induced noise. This is followed by total charge, powder factor, number of holes, cooperating charge, hole depth and stemming, in that order of importance.

Furthermore, the SHAP beeswarm summary plot as shown in Figure 5 uses an information dense summary to show the impact of each observation of the input data on the output of the predictive model. Each row of input observation is represented by colour coded dot in the summary plot. The red colours show high input values whereas the blue colour shows low values. The *y*-axis shows the names of each input variable whereas the x-axis shows the SHAP values for each observation. The input variables are ranked from the top to the bottom in order of importance. Positive SHAP value means the input contributes positively to the output of the predictive model whereas negative value means the input contributes negatively to the output of the predictive model.

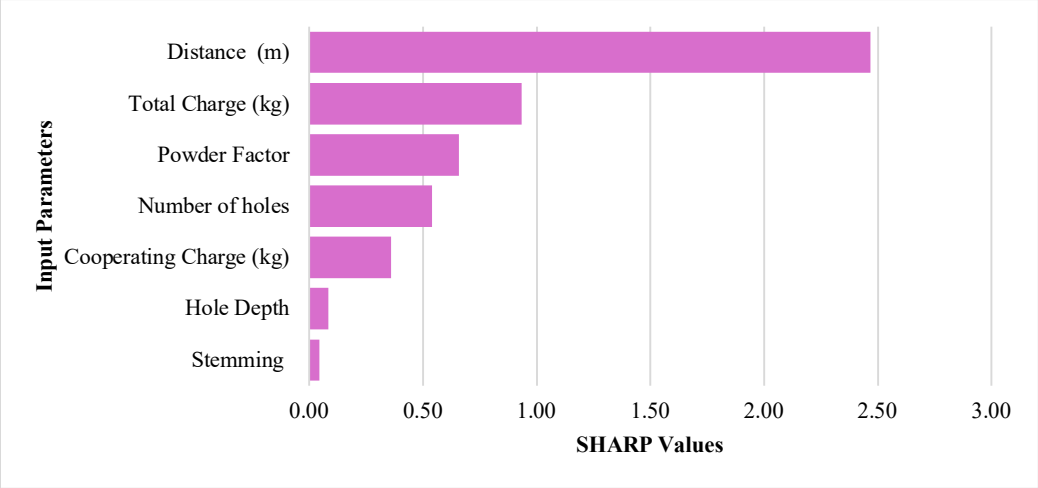


Figure 4. SHAP values of input parameters in estimating blast-induced noise levels

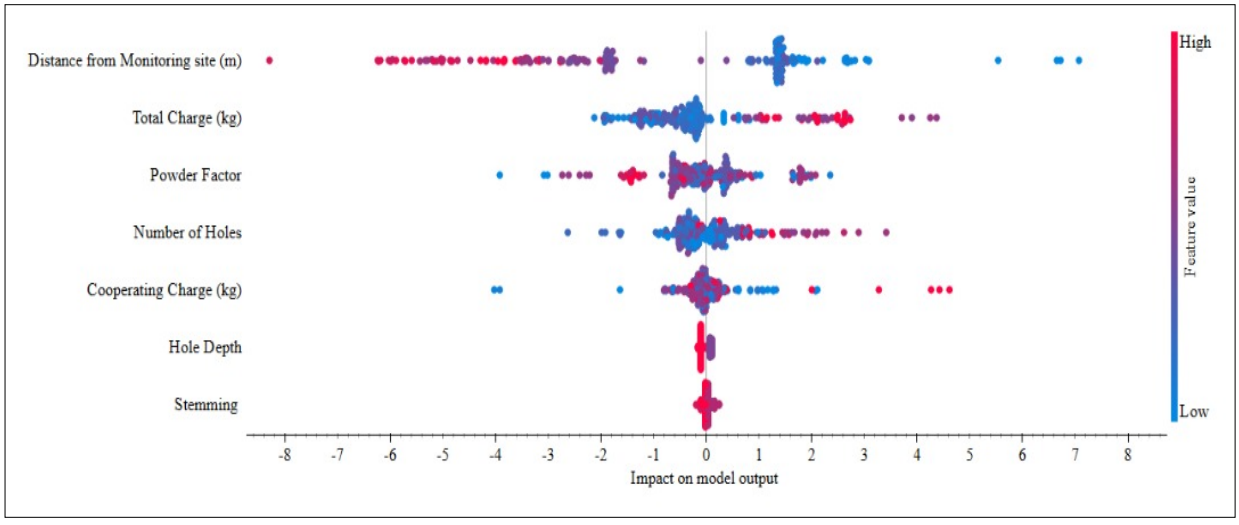


Figure 5. SHAP beeswarm summary plots showing impact of each input feature in predicting noise level

A study of the Figure 5 reveals that the higher values of distance had negative SHAP values whereas the lower values had positive SHAP value. This indicates a negative corelation between distance and noise levels meaning as distance values increase, the noise level decreases. The next

influencing feature is the total charge. For total charge, the higher values had positive SHAP values whereas the lower values had negative SHAP values. This also indicates a positive correlation between total charge and noise level, meaning noise level increases as total charge increases. The next influencing feature is powder factor. It is noteworthy that a scattered distribution of SHAP values for powder factor can be observed across both negative and positive impacts. Nonetheless, a critical analysis reveals that the red dots (high powder factor) are mostly located on the region of the positive SHAP values while the blue dots (low powder factor) are more concentrated toward the region of the negative SHAP values. This suggests a positive correlation between powder factor and noise level. It can also be observed that number of holes which is the next influencing feature is positively correlated to noise. This is because majority of the blue dots (low number of hole values) were in the region of the negative SHAP values and the red dots (high number of holes values) were in the region of the positive SHAP values. This is intuitive: a greater number of holes usually implies a larger-scale blast, resulting in higher cumulative explosive energy release, which correlates with higher noise emissions. Furthermore, it can be observed that majority of the values of the cooperating charge clustered around the zero SHAP value. Nonetheless, some observations had positive SHAP value indicating a moderate positive correlation between cooperating charge and noise level. Hole depth and stemming appears to have a relatively centred and narrow distribution of SHAP values, implying they have a more neutral or low impact on noise level.

4.4.2. Local explanations

Apart from a global explanation of the input variables in all instances, SHAP provides a local explanation of the input variables on specific instances based on the test dataset. In that regard, two force plots as shown in Figure 6(a) and (b) for two prediction instances were adopted because it visually explains the contribution of each input parameter to the prediction made. Additionally, a force plot illustrates how individual features impact a model's prediction for a specific instance by comparing the base value (the model's average prediction) to the final output. The base value serves as a starting point, typically based on a background dataset. Features then act as forces pushing the prediction up or down. Positive SHAP values (in red) raise the prediction, while negative SHAP values (in blue) lower it. The length of each bar reflects the strength of that feature's effect. This visual "tug-of-war" helps explain how each feature contributes to the deviation from the base value in a prediction.

In Figure 6(a), the base value is approximately 40.93 dBL, and the model predicts a lower final output of 38.0 dBL. The decrease is primarily driven by strong negative contributions from Distance from Monitoring Site, Number of Holes, and Total Charge with values of –2.63, –0.8, and –0.42, respectively. These three features are shown in blue, indicating that they significantly reduce the predicted value. The only positive influence comes from the Powder Factor with a value of +1.0 and is shown in red, pulling the prediction upward slightly. However, this effect is not enough to counteract the stronger negative forces, resulting in an overall lower prediction than the base value.

In Figure 6(b), the base value starts slightly higher at around 41.97 dBL and the final predicted value is 42.6 dBL. In this instance, there is an increase in the final predicted value relative to the base value. Compared to Figure 6(a), only Distance from Monitoring Site had a substantial blue colour and bar length, indicating a strong negative contribution with a value of –1.58. This pushed the prediction downwards. Nonetheless, features namely: Total Charge, powder factor, number of holes and cooperating charge which are shown in red pulled the prediction upward by +1.61, +0.38, +0.16 and +0.09 respectively. The base value plus the positive contributions minus the negative contribution resulted in a higher final prediction relative to the base value. These examples in Figure 6(a) and 6(b) show how SHAP force plots not only identify the direction and strength of each feature's influence but also help compare how different input combinations affect predictions across multiple instances.

It is worth noting that features such as hole depth and stemming had little to no effect on the prediction, as they had SHAP values close to zero for this instance. Thus, they were omitted from the force plot (Figures 6(a) and 6(b)) because they didn't meaningfully influence the output.

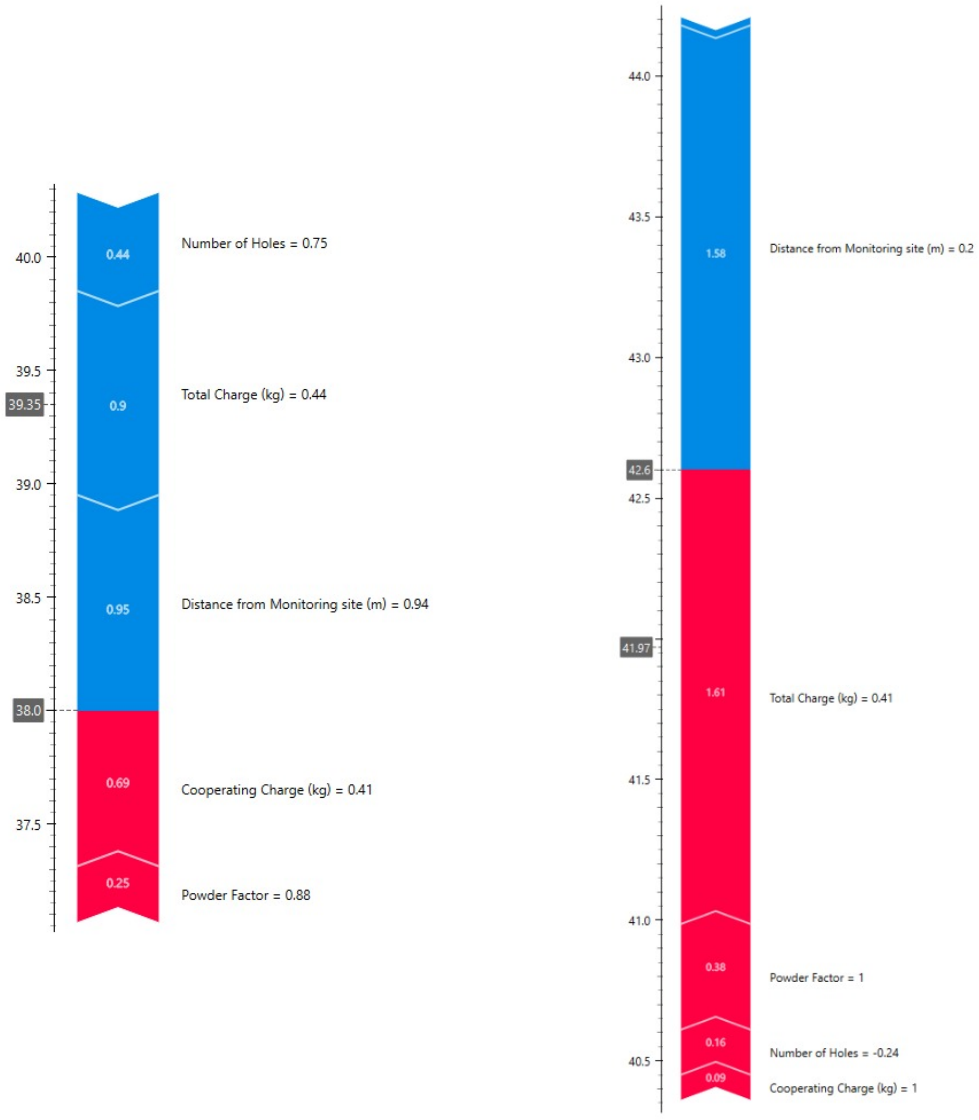
4.5. Practical implication

The integration of the XGBoost model into daily blasting operations presents a significant opportunity for proactive noise control and regulatory compliance in surface mining. One of the most direct applications is in the design of noise mitigation strategies. By forecasting noise levels

prior to each blast, the blast engineer can implement dynamic adjustments such as reducing charge mass, increasing stemming height, or altering blast geometry when predicted values approach or exceed acceptable thresholds. This allows for targeted interventions without compromising the overall productivity of the blasting operation.

Furthermore, the explainable nature of the model, powered by SHAP analysis, provides practical insight into the influence of specific blast parameters. For instance, the model identifies distance to the monitoring point as the most critical factor affecting noise propagation. This allows planners to make adaptive decisions on blast site placement or to install temporary buffer barriers to reduce noise impact in directions where sensitive receptors are located.

Lastly, the predictive outputs of the model can be integrated into Environmental Management Plans (EMPs) as part of a broader environmental monitoring framework, supporting compliance documentation and auditing processes. In this way, the model becomes not only a predictive tool but also a valuable asset for risk-based planning, regulatory alignment, and sustainable blasting practices in the mining industry. Overall, the model serves as a decision-support tool that enhances both operational efficiency and environmental accountability in blasting practices.



(a) **(b)**

Figure 6. Force Plots of two instances of predicted noise levels

4.6. Limitations and future works

Despite the strong performance of the proposed XGBoost model in predicting blast-induced noise levels, several limitations should be considered when interpreting the results and deploying the model in broader operational contexts. First, the dataset used for model training and evaluation was derived from a single open-pit gold mine in Ghana. While this enables focused model development and tuning, it may limit the generalisability of the results to other sites with different geological, topographical, or operational characteristics. Future research should consider validating the model across multiple mining environments to enhance its transferability.

The current study does not incorporate time-series or seasonal features such as weather variations, humidity, or changes in atmospheric pressure, which could influence noise propagation patterns. Integrating temporal components into future models may improve robustness and allow for real-time predictive adjustments under varying environmental conditions. Furthermore, the current study does not incorporate the effect of topography on noise propagation. The topography around a blasting site plays a critical role in noise propagation by influencing reflection, diffraction, and atmospheric effects. Hard, reflective surfaces such as rocky terrain or steep slopes can amplify noise through constructive interference, while valleys may focus sound waves, increasing noise levels in specific areas [34]. Conversely, hills and ridges act as natural barriers, diffracting sound and reducing noise in shadow zones behind them. According to Hannah [35], acoustically "soft" surfaces such as grass, forest floors, or cultivated land absorb more sound energy compared to "hard" surfaces like concrete, rock, or water, which tend to reflect sound. This ground absorption can lead to reductions of up to 10 dB over 100 meters at higher frequencies (e.g., 2000 Hz), with greater attenuation observed as distance increases. Wind and temperature gradients, shaped by topography, further modify noise propagation. Temperature inversions can bend sound downward, increasing ground-level noise, while uphill winds may carry sound upward, reducing its impact at lower elevations [36]. These factors underscore the importance of incorporating site-specific topography characteristics into predictive noise models to improve accuracy and inform blast design and mitigation strategies.

While the SHAP-based interpretability provides valuable insights into feature contributions, it remains somewhat abstract in the absence of domain-specific noise thresholds or regulatory benchmarks. Future efforts could focus on integrating regulatory cutoffs or health-based exposure limits into the interpretability framework to enhance decision-making and policy alignment.

Lastly, while the XGBoost model demonstrated exceptional predictive performance with an R^2 of 1.0 and extremely low error metrics on both training and test datasets, such near-perfect results may raise concerns regarding potential overfitting, despite the use of an 80:20 train-test split based on the holdout cross validation. Although the model generalised well to the test set, its performance may be overly optimistic without robust validation. As part of future work, applying k-fold cross-validation, bootstrapping techniques, or evaluating performance on external independent datasets is recommended to assess the model's generalisation capacity more comprehensively. Addressing these limitations will help enhance the model's credibility, scalability, and practical utility in real-world mining operations and environmental monitoring systems.

5. Conclusions

In this study, four ensemble learning models namely: XGBoost, AdaBoost, CATBoost and Gradient Boosting were implemented to predict blast-induced noise levels making up the first phase of the study. A total of 324 blasting events was used in that regard to ascertain the viability of the developed models. The prediction results showed that XGBoost achieved the best accuracy with R^2 = 1.0000, RMSE = 0.0005, MAE = 0.0004, MAPE = 0.0010, CV_{RMSE} = 0.0013, compared to the CATBOOST (R^2 = 1.0000, RMSE = 0.0016, MAE = 0.0010, MAPE = 0.003, CV_{RMSE} = 0.0039), Gradient Boosting (R^2 = 1.0000, RMSE = 0.0017, MAE = 0.0012, MAPE = 0.0031, CV_{RMSE} = 0.0041), AdaBoost (R^2 = 0.9996, RMSE = 0.0746, MAE = 0.0126, MAPE = 0.0307, CV_{RMSE} = 0.1824). In the second phase of the study, the prediction results of the XGBoost model (best ensemble learning model) were combined with SHAP to provide a comprehensive explainable model in predicting blast-induced noise level. Here, the global interpretation as well as the local interpretation of the explainable ensemble model revealed distance between the blasting point and monitoring station to be the dominant feature in the prediction of blast-induced noise level. This study contributes to minimising the negative impacts of mining operations, fostering a more responsible,

sustainable and environmentally conscious approach to mineral extraction.

Conflict of interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article

Ethical approval

There are no human participants in this article and informed consent is not required.

Data availability statement

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Funding statement

The authors thank the Ghana Chamber of Mines (GCM) for providing funding to support this work through the Ghana Chamber of Mines' Tertiary Education Fund (GCM-TEF) Faculty Members' Research Grant with number UMaT-Gh_Chamber_of_Mines-FR/023/24 at the University of Mines and Technology (UMaT), Ghana.

Author contributions

Conceptualisation: Yao Yevenyo Ziggah; Methodology: Clement Kweku Arthur, Yao Yevenyo Ziggah; Formal analysis and investigation: Clement Kweku Arthur, Yao Yevenyo Ziggah; Writing - original draft preparation: Clement Kweku Arthur; Writing - review and editing: Yao Yevenyo Ziggah, Victor Amoako Temeng; Funding acquisition: Clement Kweku Arthur, Yao Yevenyo Ziggah; Resources: Clement Kweku Arthur; Supervision: Victor Amoako Temeng

References

[1]. Paik, C. B., Pei, M., & Oghalai, J. S. (2022). Review of blast noise and the auditory system. *Hearing Research*, 425, 108459.

[2]. Das, T., Holland, P., Ahmed, M., Husain, L., Ahmed, M., & Husain, L. (2021). Sustainable development goal 3: Good health and well-being. In *South-East Asia Eye Health: Systems, Practices, and Challenges* (pp. 61-78). Springer.

[3]. Klein, M. (2020). SDG 15: life on land. *Jean Monnet Sustainable Development Goals Network Policy Brief Series*, *1*, 1-6.

[4]. Temeng, V. A., Ziggah, Y. Y., & Arthur, C. K. (2021). Blast-induced noise level prediction model based on brain inspired emotional neural network. *Journal of Sustainable Mining*, *20*(1), 28-38.

[5]. Ziggah, Y. Y., Temeng, V. A., & Arthur, C. K. (2023). A new synergetic model of neighbourhood component analysis and artificial intelligence method for blast-induced noise prediction. *Modeling Earth Systems and Environment*, *9*(3), 3483-3502.

[6]. Rincy, T. N., & Gupta, R. (2020, February). Ensemble learning techniques and its efficiency in machine learning: A survey. In *2nd international conference on data, engineering and applications* (IDEA) (pp. 1-6). IEEE.

[7]. Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J., & Liu, J. (2020). Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. *Journal of Hydrology*, *586*, 124901.

[8]. Osman, A. I. A., Ahmed, A. N., Chow, M. F., Huang, Y. F., & El-Shafie, A. (2021). Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal, 12*(2), 1545-1556.

[9]. Jarajapu, D. C., Rathinasamy, M., Agarwal, A., & Bronstert, A. (2022). Design flood estimation using extreme Gradient Boosting-based on Bayesian optimization. *Journal of Hydrology*, *613*, 128341.

[10]. Ragam, P., Komalla, A. R., & Kanne, N. (2022). Estimation of blast-induced peak particle velocity using ensemble machine learning algorithms: A case study. *Noise & Vibration Worldwide*, *53*(7-8), 404-413.

[11]. Hosseini, S., Pourmirzaee, R., Armaghani, D. J., & Sabri Sabri, M. M. (2023). Prediction of ground vibration due to mine blasting in a surface lead–zinc mine using machine learning ensemble techniques. *Scientific Reports*, *13*(1), 6591- 6610.

[12]. Chandrahas, N. S., Choudhary, B. S., Teja, M. V., Venkataramayya, M. S., & Prasad, N. K. (2022). XG boost algorithm to simultaneous prediction of rock fragmentation and induced ground vibration using unique blast data. *Applied Sciences*, *12*(10), 5269.

[13]. Gu, Z., Xiong, X., Yang, C., Cao, M., & Xu, C. (2024). Research on prediction of PPV in open pit mine used on intelligent hybrid model of extreme gradient boosting. *Journal of Environmental Management*, *371*, 123248.

[14]. Ahmad, F., Samui, P., & Mishra, S. S. (2024). Probabilistic slope stability analysis using subset simulation enhanced by ensemble machine learning techniques. *Modeling Earth Systems and Environment*, *10*(2), 2133-2158.

[15]. Yadav, D. K., Chattopadhyay, S., Tripathy, D. P., Mishra, P., & Singh, P. (2025). Enhanced slope stability prediction using ensemble machine learning techniques. *Scientific Reports*, *15*(1), 7302.

[16]. Soltanalinejad, S., & Moomivand, H. (2024). Development of a novel empirical approach to control overbreak, surface quality, and slope angle of benches following blasting. *Canadian Geotechnical Journal*, *62*, 1-22.

[17]. Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: an analytical review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *11*(5), 1-13. https://doi.org/10.1002/widm.1424

[18]. Gadde, N., Mohapatra, A., Tallapragada, D., Mody, K., Vijay, N., & Gottumukhala, A. (2024). Explainable AI for dynamic ensemble models in high-stakes decision-making. *International Journal of Science and Research Archive*, *13*, 1170-1176.

[19]. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138-52160.

[20]. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems,* 30, 4765–4774.

[21]. Kuhn, M., & Johnson, K. (2013). Applied predictive modelling. Springer.

[22]. Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794).

[23]. Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, *29*(5), 1189-1232.

[24]. Wang, T., Hu, S., & Jiang, Y. (2021). Predicting shared-car use and examining nonlinear effects using gradient boosting regression trees. *International Journal of Sustainable Transportation*, *15*(12), 893-907.

[25]. Singh, U., Rizwan, M., Alaraj, M., & Alsaidan, I. (2021). A machine learning-based gradient boosting regression approach for wind power production forecasting: A step towards smart grid environments. *Energies*, *14*(16), 5196.

[26]. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, *31*, 1-23.

[27]. Antonini, A. S., Tanzola, J., Asiain, L., Ferracutti, G. R., Castro, S. M., Bjerg, E. A., & Ganuza, M. L. (2024). Machine Learning model interpretability using SHAP values: Application to Igneous Rock Classification task. *Applied Computing and Geosciences*, *23*, 100178.

[28]. Rácz, A., Bajusz, D., & Héberger, K. (2021). Effect of dataset size and train/test split ratios in QSAR/QSPR multiclass classification. *Molecules*, *26*(4), 1111.

[29]. Kulkarni, S. (2023, November). Impact of various data splitting ratios. In *Proceedings of the Second International Conference on Emerging Trends in Engineering (ICETE 2023)* (pp. 96–104). Springer Nature.

[30]. Brantson, E. T., Ju, B., Omisore, B. O., Wu, D., Selase, A. E., & Liu, N. (2018). Development of machine learning predictive models for history matching tight gas carbonate reservoir production profiles. *Journal of Geophysics and Engineering*, *15*(5), 2235-2251.

[31]. Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *The journal of Machine Learning Research*, 13(1), 281-305.

[32]. Alshboul, O., Shehadeh, A., Almasabha, G., & Almuflih, A. S. (2022). Extreme gradient boosting-based machine learning approach for green building cost prediction. *Sustainability*, *14*(11), 6651-6670. https://doi.org/10.3390/su14116651

[33]. Chugh, A. (2020). MAE, MSE, RMSE, coefficient of determination, adjusted R squared—which metric is better?. Medium.

[34]. Long, M. (2014). Environmental noise. In: M. Long (Ed), Architectural Acoustics (pp. 175-219), Academic Press.

[35]. Hannah, L. (2006). Ground, terrain and structure effects on sound propagation. *New Zealand Acoustics*, *20*(3): 22-29.

[36]. Hannah, L. (2006). Wind and temperature effects on sound propagation. *New Zealand Acoustics, 20*(2): 22-29.

انجمن مهندسی معدن ایران
نشریه مهندسی معدن و محیط زیست
نشانی نشریه: www.jme.shahroodut.ac.ir
دانشگاه صنعتی شاهرود

مدل یادگیری ماشین گروهی قابل توضیح برای پیش‌بینی نویز ناشی از انفجار

کلمنت کوئکو آرتور[1*]، یائو یونیو زیگا[2] و ویکتور آمواکو تمنگ[1]

۱ .گروه مهندسی معدن، دانشکده فناوری معدن و مواد معدنی، دانشگاه معادن و فناوری، تارکوا، منطقه غربی، غنا
۲. گروه مهندسی ژئوماتیک، دانشکده علوم زمین و مطالعات محیطی، دانشگاه معادن و فناوری، دانشگاه معادن و فناوری، تارکوا، منطقه غربی، غنا

اطلاعات مقاله

تاریخ ارسال: ۲۰۲۵/۰۶/۱۱
تاریخ داوری: ۲۰۲۵/۰۷/۲۱
تاریخ پذیرش: ۲۰۲۵/۰۸/۰۶

DOI: 10.22044/jme.2025.16351.3195

کلمات کلیدی

توضیحات افزودنی‌های شپلی
استخراج پایدار
جستجوی تصادفی
قابلیت تفسیر
ارزیابی اثرات زیست محیطی

چکیده

نویز ناشی از انفجار یکی از پایدارترین چالش‌های زیست‌محیطی در معادن سطحی است که خطرات سلامتی قابل توجهی را برای کارگران و جوامع اطراف ایجاد می‌کند. پیش‌بینی دقیق سطح نویز قبل از انفجار برای کاهش اثرات نامطلوب آن ضروری است. این مطالعه یک چارچوب یادگیری ماشین گروهی قابل توضیح برای پیش‌بینی نویز ناشی از انفجار با استفاده از داده‌های یک معدن طلای روباز در غنا ارائه می‌دهد. چهار مدل گروهی به نام‌های: تقویت گرادیان شدید (XGBoost)، تقویت گرادیان، تقویت تطبیقی (AdaBoost) و تقویت طبقه‌بندی (CatBoost) با استفاده از یک مجموعه داده جامع از ۳۲۴ رویداد انفجار توسعه داده شده و ارزیابی شدند. عملکرد مدل‌های توسعه‌یافته با استفاده از ضریب تعیین (R^2)، جذر میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE)، میانگین درصد خطای مطلق (MAPE) و ضریب تغییرات جذر میانگین مربعات خطا (CVRMSE) ارزیابی شد و XGBoost به عنوان بهترین مدل ($R^2 = 1.0000$، RMSE = 0.0005، MAE = 0.0004، MAPE = 0.0010، CVRMSE = 0.0013) ظاهر شد. برای پرداختن به ماهیت جعبه سیاه روش گروهی، از توضیحات افزایشی شپلی (SHAP) استفاده شد که قابلیت تفسیر جهانی و محلی را ارائه می‌دهد. تجزیه و تحلیل SHAP فاصله از محل انفجار تا نقطه نظارت را به عنوان تأثیرگذارترین عامل شناسایی کرد. این رویکرد یکپارچه نه تنها دقت پیش‌بینی را افزایش می‌دهد، بلکه شفافیت مدل را نیز بهبود می‌بخشد و از شیوه‌های پایدار استخراج معادن همسو با اهداف توسعه پایدار سازمان ملل متحد (SDGs)3 و ۱۵ پشتیبانی می‌کند.

نویسنده مسئول مکاتبات: ckarthur@umat.edu.gh (ک.ک. آرتور)