



Integrating Bayesian, ensemble, and hybrid machine learning models for predicting overbreak and underbreak in underground stope mining

by S.Taheri¹, E.Topal¹, R.S. Faradonbeh¹, H.Jang²

Affiliation:

- ¹ Western Australian School of Mines:
Minerals, Energy and Chemical
Engineering, Curtin University, WA, Australia
² Hanwha Mining Services Australia,
Leeming, WA, 6149, Australia

Correspondence to:

S. Taheri

Email:

sina.taheri@me.com

Dates:

Received: 05 Jun. 2025
Revised: 23 Oct. 2025
Accepted: 21 Jan. 2026
Published: Sep. 2026

How to cite:

Taheri S., Topal E., Faradonbeh R.S.,
Jang, H. 2026. Integrating Bayesian,
ensemble, and hybrid machine learning
models for predicting overbreak and
underbreak in underground stope mining.
*Journal of the Southern African Institute of
Mining and Metallurgy*, vol. 126, no. 9,
pp. 525 – 540

DOI ID:

<https://doi.org/10.17159/2411-9717/3746/2026>

ORCID:

S.Taheri
<https://orcid.org/0009-0009-7229-6414>
E.Topal
<https://orcid.org/0000-0002-9564-3484>
R.S. Faradonbeh
<https://orcid.org/0000-0002-1518-3597>
H.Jang
<https://orcid.org/0000-0002-3978-5840>

Abstract

Overbreak and underbreak in underground stoping operations cause significant challenges, impacting safety and operational efficiency, leading to dilution and ore loss. This study employs machine learning methods to predict overbreak and underbreak at individual drillhole locations, enabling dynamic prediction of stope shapes before blasting in underground mining. Based on a comprehensive database collected from a longhole sublevel stoping mine in Western Australia, different machine learning-based models, including multiple linear regression, random forest regression, gradient boosting machine, and Gaussian process regression, were developed systematically. The best-performing model was the initial Gaussian process regression with the Matérn 3/2 kernel, achieving R^2 value of 0.820 and a root mean squared error value of 0.452. A series of hybrid models were also developed to improve prediction accuracy by combining Gaussian process regression, gradient boosting machine[A5.1], random forest regression, and artificial neural network with linear meta-learners. The best hybrid configuration achieved with R^2 value of 0.721 and RMSE value of 0.425, but further complexity led to overfitting with no gains in generalisation. SHapley Additive exPlanations analysis across all models highlighted the Mathews stability number, ratio of burden to spacing, and void ratio as the most influential variables, varying in impact between the random forest regression, gradient boosting machine, and Gaussian process regression models.

Keywords

Overbreak, underbreak, random forest, Gaussian process, gradient boosting machine, hybrid models

Introduction

Underground mining operations face significant challenges in managing overbreak (OB) and underbreak (UB), which can substantially affect operational costs, safety, and overall efficiency, from extraction through to processing in mine-to-mill workflows. Accurate prediction of these events is essential for optimising mine design, execution processes, and costs. Over the years, the majority of researchers have analysed machine learning applications in surface mining, as reviewed by McCoy (2019), however, there has been limited analysis into more complex machine learning techniques and their application in underground mining. Additionally, other underground-related studies focus on general underground issues, such as underground rock bursts and OB and UB studies in tunnelling and development, rather than specifically addressing OB and UB in underground stoping operations. Some approaches, such as those using point cloud data, oversimplify the models, potentially overlooking the complexities of underground drilling and blasting dynamics. Jung and Choi (2021) reviewed 109 papers that revealed active research in machine learning (ML) within the mining industry, particularly since 2018, with a focus on mineral exploration. While empirical methods and regression models have traditionally been employed, the inherent complexities of geological conditions and operational variables demand more sophisticated analytical approaches. Sellers and Salmi (2020) analysed blasting in various underground operations, highlighting the associated challenges. The study illustrates how the lack of continuous, accurate 3D measurements of crucial properties like blastability, in situ structures, drillhole deviation, and fragmentation, along with a limited understanding of mining-induced stresses, frequently lead to issues such as overbreak, dilution, production delays, reduced excavation stability, and poor plant performance. Their study highlights the necessity for a data-driven approach to effectively monitor and predict important factors in the dynamic mining environment. In a recent study, Abd Elwahab et al. (2023) reviewed the application of ML methods in mine blasting, noting that hybrid and ensemble models outperformed other approaches in 61% of case studies. The study highlighted trends in ML adoption, particularly for surface blast issues, where hybrid models were applied in 72% of cases,

Integrating Bayesian, ensemble, and hybrid machine learning models

compared to 45% for underground blasting. The review emphasised the growing role of ML in optimising blasting processes and decision-making while highlighting the need for further research, especially in underground blasting operations. Additionally, studies involving artificial neural network (ANN)-based classifiers suffer from limited validation on larger datasets, making it difficult to assess their broader applicability. While methods like histogram gradient boosting trees (HGBT) and Bayesian model averaging show potential, they often lack comparison with other models and fail to incorporate more complex geological and geotechnical parameters. Previous researchers have faced other challenges in their studies, which involved limited access to the broad drill and blast datasets; as Suglo (2012) discussed, the challenges of assessing dilution and ore loss in sublevel caving due to poor data quality.

A considerable body of research has focused on OB, UB, dilution, and ore loss in underground mining. These issues significantly impact mine economics, safety, and the downstream processing efficiency. Various predictive models and analytical frameworks have been developed, ranging from statistical and empirical methods to advanced machine learning and simulation techniques. For instance, studies by Liu et al. (2023) applied ensemble learning for rockburst prediction in relation to geological conditions, while Zhao and Jia'an (2020) employed neural networks to model unplanned dilution using overbreak-based indicators. Others, such as Papaioanou and Suorineni (2016), developed stability graphs tailored to stope dilution prediction using extensive case study data. Despite these advances, limitations remain, ranging from reliance on simplified models or site-specific data to lack of validation and limited generalisability. The growing complexity of underground mining environments and data availability continues to drive the need for robust, adaptable predictive tools. Table 1 summarises key studies focused on OB and UB in underground mines, highlighting the parameters considered, methods applied, results obtained, and limitations of each approach.

The predictors used in this study (see Table 2) represent both geotechnical and drill-and-blast design factors that collectively influence OB and UB. Parameters such as Mathews stability number (N') and hydraulic radius (HR) describe rock-mass stability and hanging-wall geometry, while the burden-to-spacing ratio (BS), powder factor (PF), and primer location (PRIM) drive explosive energy distribution and confinement. The void ratio (VR), defined as the available void space ahead of the blast directly affects the extent of break (Papaioanou, Suorineni, 2016). These predictors capture the dominant physical and design mechanisms responsible for deviations from the planned excavation boundaries.

The approach in this study marks a shift from the traditional reliance on 2D OB and UB prediction, the limitations of applying models in longhole stoping with small datasets, and the restricted use of features in previous studies, most of which focused only on geotechnical parameters and a few drill and blast features, without considering the full range of factors influencing OB and UB. Some of the earlier studies also relied on simpler methods, which, while foundational in mining predictive modelling, struggle with capturing the nonlinear and complex nature of geological data and the variables affecting OB and UB. This study builds on these limitations by adopting a unique hole-by-hole prediction approach, where all 15 features were collected and analysed at the level of individual drillholes. This allows for the first time a more detailed and localised prediction of OB and UB at each drillhole location, rather than generalised stope-level estimations. All

drillholes referenced in this study are production blastholes used in longhole stoping. The dataset comprises operational drilling and blasting, geotechnical, and geological parameters recorded at the drillhole level. Despite the growing interest in advanced predictive techniques, their practical application in predicting OB and UB in underground production mining remains limited. This study addresses these gaps by systematically assessing the performance of various predictive models to tackle dilution and ore loss challenges in underground mining. It advances existing work by applying RFR, GBM, GPR, and hybrid models with hyperparameter optimisation and sensitivity analysis to better understand the influence of different predictors on OB and UB outcomes. Importantly, it further extends the analysis through SHAP-based interpretation, providing a detailed view of how geological, drilling, blasting, and design variables influence the output, marking a novel shift toward hole-by-hole OB and UB prediction in underground stoping operations. The findings and methodologies presented in this paper highlight the potential of integrating ensemble, Bayesian, and hybrid techniques to improve prediction accuracy in underground stoping, leading to safer and more cost-effective mining practices.

Site description and data collection process

The study was conducted at an underground gold mining operation located to the north of Kalgoorlie, Western Australia, and has been operational for many years. This site utilises various extraction methods, notably the longhole sublevel stoping supplemented by paste-filling. In this research, an individual analysis of each drillhole was carried out to collect essential data for predicting OB and UB at specific drillhole locations, a critical component setting the foundation for developing a dynamic 3D predictive model. In this study, the term “3D prediction” refers to generating OB/UB predictions at every measured point along each drillhole. Because the drillholes define the perimeter of the stope envelope, the point-wise predictions of rings collectively form the three-dimensional representation of the final excavation boundary. The model, therefore, predicts the 3D deviation surface by first predicting each hole individually, which is the fundamental unit of spatial reconstruction in underground stoping.

Additionally, the research included an examination of the spatial relationships between adjacent drillholes and their collective setup within the stope's ring and panels. Data were precisely collected through analysis of individual drillholes, the examination of reconciliation reports, a thorough review of existing drill and blast designs, and operational planning parameters at the mine site. In total, 193 distinct production drillholes from different zones of the mine were analysed, each containing 15 predictor variables and one output variable (OB/UB), resulting in 3,088 data entries used for model development and evaluation. Initially, this information was used to construct a linear regression model as a foundational benchmark. Subsequently, the study focused on evaluating the performance of RFR, GPR, and GBM methodologies. This study aimed to determine the effectiveness of ensemble and Bayesian techniques in accurately predicting dynamic, hole-by-hole OB and UB conditions in underground stoping environments, where OB and UB vary spatially and form part of a three-dimensional predictive framework rather than a single averaged outcome. Leveraging recent advancements in predictive modelling, this research explores specific variables critical for predicting OB and UB in underground mining environments. Table 2 lists a detailed set of factors employed in developing the models for this analysis.

Integrating Bayesian, ensemble, and hybrid machine learning models

Table 1

Summary of recent studies on dilution and ore loss prediction methods applied in underground mining

Ref.	Parameters	Methods	Results	Critique
(Liu et al., 2023)	Focus on geological conditions.	Histogram gradient boosting tree (HGBT).	Addressed rock burst prediction, improved dataset handling.	Lack of comparison with other methods.
(Zhang et al., 2024)	Ground deformations, model and parameter uncertainties.	Bayesian model averaging.	Enhanced estimation of tunnelling-induced deformations.	Focus on tunnelling, not specific to stope mining or UB/OB.
(Jang et al., 2015)	Geological variability, OB/UB	Linear/nonlinear regressions, ANN.	Introduced uneven break and predicted unplanned dilution and ore loss.	Limited to simpler methods, lacking advanced ML applications or 3D models.
(Hong et al., 2023)	Point cloud data, 3D model for blasting.	3D Modelling and simulation using point cloud.	Simulated distribution of OB/UB.	The simplified model may not fully capture complex blasting dynamics.
(Adoko et al., 2022)	Stope design, stability graph method.	ANN-based classifiers.	Critiqued traditional methods, suggested ANN for stability design.	Limited testing of ANN models, lacks validation on large datasets.
(Cordova et al., 2022)	Stope/drift design, dilution assessment.	ELOS method, stope design assessment.	Assessed actual dilution, the impact of stope design on stability.	ELOS method limited by its reliance on traditional design metrics.
(Han et al., 2020)	Drillhole diameter and depth, powder factor, burden/spacing.	Random forest and Bayesian network.	Eliminated less impactful parameters for predicting fly rock.	Limited to surface mining and lack of geotechnical parameters.
(Chimunhu et al., 2024)	Stope geometry, geotechnical factors, geological factors.	Classification and regression tree (CART), hybrid model.	Achieved F1 score of 70% and 72% for the SSE-CART and PCA-CART with R2 of 72% and 93%, respectively.	The models utilised basic geological, geotechnical, and stope design attributes available during the early stages of mine planning, making them practical only for preliminary assessment.
(Yu et al., 2021)	Blast design, boundary measurement.	Support vector regression and Gaussian process regression.	Improved boundary control, reduced ore loss and dilution.	Lacks detailed validation metrics, requires high-quality boundary data.
(Ibarra-Gutierrez, Laflamme, 2021)	Stope limits, geotechnical and geological factors, unrecovered blasted ore.	Analysis of reconciliation practices.	Identified inconsistencies in accounting for blasted ore left unrecovered in stopes. Highlighted the impact on mineral reserve and stope performance assessment.	Lacks quantitative data or case studies, primarily a discussion on reporting practices.
(Forster et al., 2007)	Stope geometry, rock mass quality, stress conditions, cable support, blasting method, sequencing.	Detailed analysis of 3D stope survey data combined with geological assessment, focused on hanging wall stability.	Identified that multiple factors simultaneously influence stability and dilution in open stopes, making it challenging to assess the effect of individual changes.	The study provides valuable insights into the factors affecting hanging wall dilution, it does not offer specific quantitative results or predictive models. The findings are based on observations from a single mine site, which may limit the generalisability of the conclusions.

Integrating Bayesian, ensemble, and hybrid machine learning models

(Mottahedi et al., 2018)	Tunnel section area, perimeter holes powder factor (PPF), specific drilling (SD), spacing-to-burden ratio (S/B), and rock mass rating (RMR).	Developed predictive models using soft computing techniques, specifically adaptive neuro-fuzzy inference system (ANFIS) and a hybrid ANFIS optimised with particle swarm optimisation (ANFIS-PSO).	The ANFIS-PSO model outperformed the standard ANFIS model, achieving a higher coefficient of determination ($R^2 = 0.961$) and lower root mean square error (RMSE = 0.064), indicating improved predictive accuracy.	Although the models demonstrated high accuracy, they were developed based on specific case studies, which may limit their generalisability and only applicable to development areas.
(Himanshu et al., 2023)	Blasting design parameters, rock mass characteristics, stope geometry, and operational constraints.	Discusses innovative blasting techniques tailored for underground hard rock mining, focusing on optimising blast designs to enhance fragmentation, reduce overbreak, and improve overall operational efficiency.	Implementation of the discussed innovative blasting practices has led to improved fragmentation, reduced overbreak, enhanced safety, and increased productivity in underground hard rock mining operations.	The study provides insights into advanced blasting techniques, but it primarily offers qualitative assessments without extensive quantitative data or case studies to substantiate the claimed improvements. Further empirical research and detailed case analyses would strengthen the applicability of these practices across diverse mining scenarios.
(Papaioanou, Suorineni, 2016)	Hydraulic radius (HR), modified stability number (N'), orebody width.	Statistical analysis of 226 open stope case studies from six Australian underground metalliferous mines using logistic regression and Bayesian likelihood discrimination to develop a quantitative dilution-based stability graph independent of orebody width.	The developed graph provides quantitative predictions of dilution levels, enabling mining engineers to design open stope sizes based on acceptable dilution thresholds, applicable to both narrow and wide orebodies.	While the model offers a more generalised approach than previous qualitative graphs, its reliance on data from Australian mines may limit its applicability in different geological settings without further validation. In addition, it only classifies stopes based on stability criteria (i.e., not predicting overbreak or underbreak).
(Zhao, Jia'an, 2020)	Modified stability number (N'), hydraulic radius (HR), average borehole deviation, and powder factor.	Developed a three-layer back-propagation (BP) neural network with six neurons in the hidden layer to predict unplanned ore dilution, using equivalent linear overbreak slough (ELOS) as the output variable. The model was trained and tested on 120 datasets.	Achieved a high prediction accuracy with an R^2 of 0.9761 and a mean square error of 0.0001. When applied to a test stope in the Sandaoqiao lead-zinc mine, the model predicted an unplanned dilution of 0.717 m, closely matching the measured value of 0.70 m, resulting in a relative error of 2.4%.	While the model demonstrates high accuracy, its validation was limited to a single case study with only 120 data points. In addition, this study only focuses on dilution and did not consider ore loss.
(Hefni, Abdellah Wael, 2020)	Ore dip angle, stope width, stope height, In situ stress ratio	2D elasto-plastic numerical modelling to assess the impact of ore inclination and stope geometry on hanging wall (HW) stability and ore dilution.	Steeper ore dips (85°) improve hanging wall stability and reduce dilution, while wider and taller stopes increase instability and dilution. Ore dip angle has little effect on wall deformation.	The study provides valuable insights into the interplay between stope geometry and HW stability. However, it is based on 2D models, which may not capture all 3D complexities of actual mining scenarios.

Integrating Bayesian, ensemble, and hybrid machine learning models

These predictors, instrumental in understanding the dynamics of OB and UB within the context of a specific mine site, include a range of variables from basic drill characteristics such as positioning and drillhole length to more complex drill and blast design and planning parameters like proximity from drillhole end to the planned stope shape, and burden and spacing. Additionally, factors like the location of primers in the blast plan, the void ratio relevant to drill and blast design stages, and explosive density are considered alongside parameters influencing stope stability including the hydraulic radius, Mathews stability number (Mathews et al., 1981), and the Q-value (Barton et al., 1974).

RQD is rock quality designation, Jn is joint set number, Jr is the joint roughness number, Ja is the joint alteration number

Figure 1 shows some key predictors impacting OB and UB, discussed in detail in Table 1, within an example stope from this study. This figure shows the OB and UB areas, the spatial arrangement of drillholes, the distance between adjacent drillholes, the distance between each drillhole end, and the planned stope shape boundary, the position of the final primer within the drillholes, and drillholes that do not break through to other underground areas or levels.

OB and UB were quantified as the linear deviation between the designed stope boundary and the actual excavated surface obtained from CMS data. Positive values indicate material removed beyond

the design limit (OB), while negative values indicate insufficient breakage within the design (UB). These deviations were measured

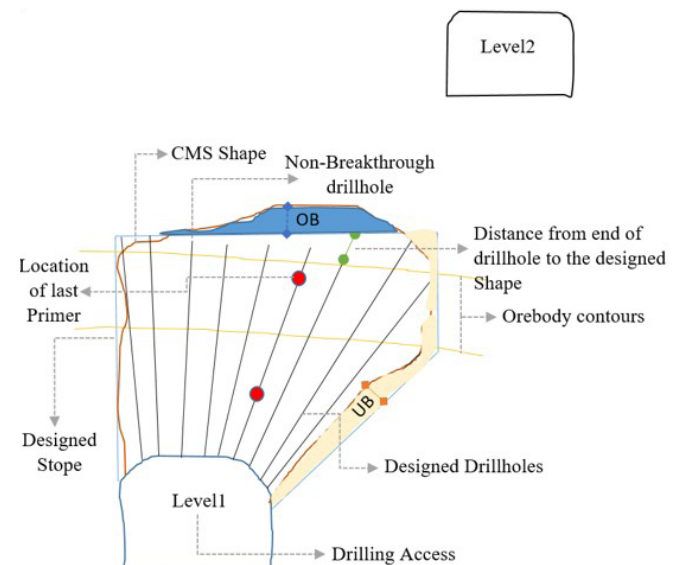


Figure 1—Some of the key predictors affecting OB and UB in an example underground mining stope

Table 2

Parameters used in the overbreak and underbreak prediction models

Variable Number	Predictors	Symbol	Explanation
1	Drillhole sequence	HN	The drillhole number serves to establish the sequence in a ring blast, specifying the location of drillholes (numeric).
2	Explosive density	EXP	Density of explosive used in the specific drillhole (g/cm ³).
3	Distance from end of drillhole to the stope shape	DIS	The distance from the end of the drillhole to the designed/planned stope shape (m).
4	Ratio of burden to spacing	BS	Burden: distance from point of blast to the free face (m). Spacing: distance between adjacent blastholes (m).
5	Primer location	PRIM	Distance from the last primer to the end of the drillhole (m).
6	Void ratio	VR	Ratio of void in relation to the blasted rock (%).
7	Powder factor	PF	Amount of explosive used per volume of rock (kg/tonne).
8	Paste-filled or not	PASTE	Whether the previous stope has been paste-filled or not (i.e., blasting against paste-fill or not).
9	Blasthole diameter	HD	Blasthole diameters ranging from 89 mm to 102 mm
10	Drillhole length	HL	Length of the designed drillhole in metres.
11	Breakthrough or non-breakthrough drillhole	BT	Whether the drillhole breakthrough into the drift or level is above or not (value of 1 for breakthrough and 0 for non-breakthrough).
12	Ring dump	RD	Ring designed dump, where 0° is vertical, and 90° is horizontal.
13	Hydraulic radius	HR	Hanging wall hydraulic radius
14	Mathews stability number	N'	Mathews stability number, specifically assessing the stability of hanging wall in the dataset.
15	Q-system1	Q'	$Q' = \left(\frac{RQD}{J_n} \times J_r / J_a \right)$

Integrating Bayesian, ensemble, and hybrid machine learning models

along each blasthole to enable hole-by-hole analysis and three-dimensional prediction of boundary accuracy.

Table 3 presents descriptive statistics for both predictor variables and the output parameter. It includes important measures such as the mean, median, standard deviation (StdDev), and the 25th and 75th percentiles, providing an overview of the data distribution.

To illustrate the data composition and the structure of predictor-output relationships, Table 4 presents a small representative subset of the dataset used for model development, including sample drillholes with their corresponding predictor values and measured OB/UB outcomes.

Data filtering

In the age of technology, where data is extensively utilised across various sectors, failing to detect anomalies can lead to unreliable results and misinterpretations (Falahi et al., 2023). In this study, outliers were detected using a distance-based method and the chi-squared distribution by measuring how far each point is from the data centre, based on the number of variables in the dataset (Filzmoser et al., 2008). Conventionally, the 95th or 99th percentiles serve as the cut-off benchmarks for identifying outliers. For this particular analysis and due to the size of the dataset, the 99th percentile was selected as the threshold criterion. The Mahalanobis distance test identified five data points with values exceeding the critical threshold of 30.577. After careful review, these points were confirmed as extreme but valid observations representative of genuine variability in OB and UB conditions rather than data-entry errors. Given the study's focus on modelling the full range of operational variability, all observations were retained to preserve dataset integrity and ensure realistic predictive modelling.

The correlation matrix in Figure 2 shows several key relationships among the variables and the output. The output variable shows the highest positive correlations with HR (0.4568) and PF (0.3924) and the highest negative correlations with VR (-0.3968) and PASTE (-0.3113), indicating these variables are potentially significant predictors for OB and UB. In addition, PF and VR, as well as PASTE and HL, show strong negative correlations of -0.75 and -0.71, respectively. In contrast, Q' and HR, as well as HR and PF, show positive correlations of 0.65 and 0.66, respectively. The highly correlated variables and multicollinearity will be investigated and discussed in the next section. Although multicollinearity was observed among some variables, principal component analysis (PCA) was not applied. PCA can effectively reduce dimensionality but transforms correlated variables into composite components, which could obscure the physical meaning

of key geotechnical and blasting parameters (Jolliffe, Cadima, 2016). Because the purpose of this study is to interpret variable-specific effects on OB and UB rather than solely maximise variance, all variables were retained in their original form. Model interpretability was maintained using SHAP analysis, which quantifies each variable's influence while preserving its engineering significance.

Methodology

Multiple linear regression analysis

This study used MLR as the foundational approach to define relationships between independent and target variables and evaluated the model's performance using statistical metrics, such as RMSE and R^2 . Initially, the linear regression model produced an RMSE of 0.580, and R^2 of 0.629. Collinearity analysis was conducted for linear regression as it can produce parameter estimates with incorrect signs or doubtful magnitudes, causing parameter estimates to vary widely with small data changes (Belsley et al., 1980). The VIF and tolerance are both widely used measures to assess the degree of multicollinearity among the independent variables in a regression model (O'Brien, 2007). The VIF for each predictor was calculated as:

$$VIF_i = \frac{1}{1 - R_i^2}$$

where R_i^2 represents the coefficient of determination when predictor X_i is regressed on all other predictors (Montgomery et al., 2012).

VIF was calculated for all predictor variables to detect the potential multicollinearity problem. The findings are presented in Table 5.

VIF values exceeding 10 usually indicate that associated regression coefficients are poorly estimated because of multicollinearity, impacting the reliability of regression coefficients and model interpretability (Montgomery et al., 2012). This issue was noticed for the predictors VR, PF, PASTE, and HR, [A8.1] which showed VIF values of 16.2248, 12.5457, 16.5485, and 32.3352, respectively, indicating moderate to strong linear relationships with other predictors and potentially leading to high standard errors. Initially, to maintain key drill and blast predictors in the dataset, a decision was made to remove VR and PASTE from the four variables identified with VIF values higher than 10. This choice was based on the practical relationships between void ratio and hydraulic radius, as well as between paste and powder factor, ensuring that HR and PF remained in the analysis. The resulting linear regression model produced an R -squared of 0.439, and an RMSE of 0.742, indicating that while multicollinearity was partially mitigated, the model still explained only a moderate portion of the variance. Additionally, based on strong correlations observed in the correlation matrix, specifically between HR and PF, and HR and Q', a revised approach was tested by instead removing PASTE and HR while retaining VR and PF. The performance of this adjusted model declined slightly, with an R -squared of 0.399 and RMSE of 0.739. These results suggest that, although different variable exclusion strategies can address multicollinearity to varying degrees, the overall explanatory power of the linear model remains limited and additional factors or more advanced modelling techniques may be required to improve predictive accuracy. Furthermore, as explained by O'Brien, (2007), values of the VIF of 10, 20, 40, or even higher do not, by themselves, invalidate the results of regression analysis,

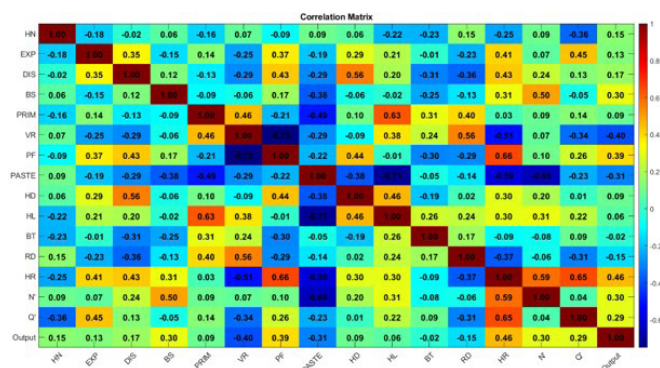


Figure 2—Correlation matrix representing also the relationships between variables

Integrating Bayesian, ensemble, and hybrid machine learning models

Table 3

Basic statistical summary of predictors and output

Predictors	Mean	Median	StdDev	25th percentile	75th percentile
HN	3.461	3	2.245	2	5
EXP	1.031	1	0.111	1	1
DIS	0.115	0	0.455	-0.24	0.455
BS	0.913	0.875	0.391	0.682	1.094
PRIM	1.259	1	1.041	0.425	2
VR	53.527	52	17.107	30	70
PF	1.758	1.96	0.447	1.35	2.24
PASTE	0.471	0	0.500	0	1
HD	90.818	89	4.521	89	89
HL	12.763	10.873	9.085	3.813	23.18
BT	0.186	0	0.390	0	0
RD	22.393	26.5	10.35	18.4	28.5
HR	4.574	4.88	0.873	4.6	5.15
N'	7.459	6.23	3.609	5.4	11.15
Q'	21.832	26.67	11.477	11.8	28.33
OB/UB	-0.241	0	0.956	-0.49	0.275

Table 4

Representative sample of drillhole-level data used in this study

HN	EXP	DIS	BS	PRIM	VR	PF	PASTE	HD	HL	BT	RD	HR	N'	Q'	OB UB
1	1.2	0.5	0.86	0.00	30.00	1.96	1.00	89.00	1.40	0.00	0.00	4.88	6.23	28.33	0.18
8	1	0.13	2.38	0.50	48.80	2.24	0.00	89.00	1.80	0.00	28.50	5.15	11.15	11.8	-0.12
1	1	0	0.80	2.00	52.00	2.26	0.00	89.00	7.14	0.00	29.30	5.34	5.4	42.5	0.24
1	1	0	0.81	1.00	74.00	1.35	1.00	89.00	7.85	0.00	35.00	3.30	2.50	8.89	-2.10
3	1.2	0	0.30	7.00	64.00	1.39	0.00	89.00	20.67	0.00	29.50	4.60	5.40	26.67	0.92
1	1	0	0.92	1.00	76.00	1.12	0.00	89.00	16.85	0.00	28.30	5.00	14.10	28.33	-0.80
1	1	-0.41	1.33	1.00	70.00	1.20	1.00	89.00	4.90	0.00	32.80	2.64	4.10	5.80	-0.60

require the removal of independent variables from the analysis, suggest the need for ridge regression, or require the combination of independent variables into a single index.

The results from the three linear regression equations highlight several important variables. In the first equation, significant predictors include PASTE, HR, and EXP, where higher negative and positive values are associated with increased UB and OB, respectively. In the second equation, after removing PASTE and VR due to multicollinearity, PF, EXP and HR remain crucial predictors, with PF showing a strong positive impact and EXP and HR showing strong negative impacts on OB and UB and in the third equation, EXP, BS and PRIM emerge as strong predictors.

Table 6 displays the equation and results of the first and second linear regression models.

OLS¹: Ordinary least squares (OLS) is a standard method for estimating the coefficients in linear regression, aiming to minimise the sum of squared deviations (residuals) between data points and the regression line

Ensemble methods

Tree-based methods such as RFR and GBM are generally less sensitive to multicollinearity compared to linear regressions. Each tree in these ensemble models splits the data based on a subset of features and individual trees are not affected by multicollinearity in the same way that linear models are (Cha et al., 2021). Similarly, Gaussian process regressions, being a non-parametric Bayesian approach, are primarily concerned with the covariance structure of the data rather than the individual impact of each predictor. Gaussian process uses kernel functions to define the similarity between data points, which helps in prediction (Rasmussen, Williams, 2006). In summary, while multicollinearity is a crucial consideration in linear models due to its impact on coefficient estimates and statistical tests, it is less of a concern in tree-based methods and GPR, which are the focus of this study. These models can handle multicollinearity without significant issues to their predictive capabilities. However, it may affect the interpretation of feature importance and, in the case of GPR, the model's

Integrating Bayesian, ensemble, and hybrid machine learning models

Table 5
Collinearity statistics results

Predictor number	Predictors	VIF
1	HN	1.5641
2	EXP	1.7733
3	DIS	2.2493
4	BS	2.0038
5	PRIM	2.535
6	VR	16.2248
7	PF	12.5457
8	PASTE	16.5485
9	HD	2.7901
10	HL	4.0637
11	BT	1.5483
12	RD	2.0897
13	HR	32.3352
14	N'	8.000
15	Q'	6.6617

configuration. In this study, RFR and GBM approaches were utilised due to their ability to handle complex, non-linear relationships and interactions between variables, leveraging ensemble learning techniques to improve the accuracy of OB and UB predictions. Both models combine outputs from multiple decision trees, effectively reducing variance and bias (Ghosal, Hooker, 2020). The model development process is discussed in more detail in the following sections.

Random forest regression

Random forests are ensembles of decision trees built from bootstrap samples, where each tree is trained on a random subset of predictors (Cutler et al., 2011). The generalisation error converges as the number of trees increases and depends on both individual tree strength and inter-tree correlation. RFRs have become widely used in high-dimensional regression tasks because they require limited hyperparameter tuning and have a low tendency to overfit (Malhotra, Karanickolas, 2020). The model in this study was trained on 193 samples with 15 predictor variables. Hyperparameters were optimised using a grid search (number of trees = 50 – 200; minimum leaf size = 1 – 20; maximum splits = 10 – 100) and five-fold cross-validation. The optimal configuration, 100 trees, 1 minimum leaf size, and 100 maximum splits, achieved the highest cross-validated $R^2 = 0.682$. The final model yielded $R^2 = 0.812$ (RMSE = 0.395) on the training set and $R^2 = 0.749$ (RMSE = 0.521) on the testing set (Table 7). Hyperparameter optimisation was

Table 6
Linear regression models

1st Model	Method	Equation	RMSE	R ²
Linear regression2	OLS ¹	$OB, UB = 14.937 + .047443HN - .50342EXP + .013756DIS + .019678BS + .11296PRIM - .098005VR - .23934PF - 2.88PASTE - .0099673HD - .020184HL + .32883BT + .0066685RD - 2.0473HR + 0.155571N' + 0.054729Q'$	Linear regression1	OLS ¹
2nd Model	Method	Equation	RMSE	R ²
Linear regression2	OLS	$OB, UB = 1.19 + .11934HN - 1.13387EXP + .25597DIS + .3718BS + .28364PRIM + 1.7076PF - .029085HD - .011805HL + .51258BT - .0088097RD - .74715HR + .13736N' + .050129Q'$	0.742	0.439
3rd Model	Method	Equation	RMSE	R ²
Linear regression3	OLS	$OB, UB = -.0106 - .6458EXP + .0857DIS + .51BS + .2855PRIM - .0308VR + .0378PF + .0033HD - .0076HL + .2218BT + .011RD + .0612N' + .0108Q'$	0.739	0.399

Integrating Bayesian, ensemble, and hybrid machine learning models

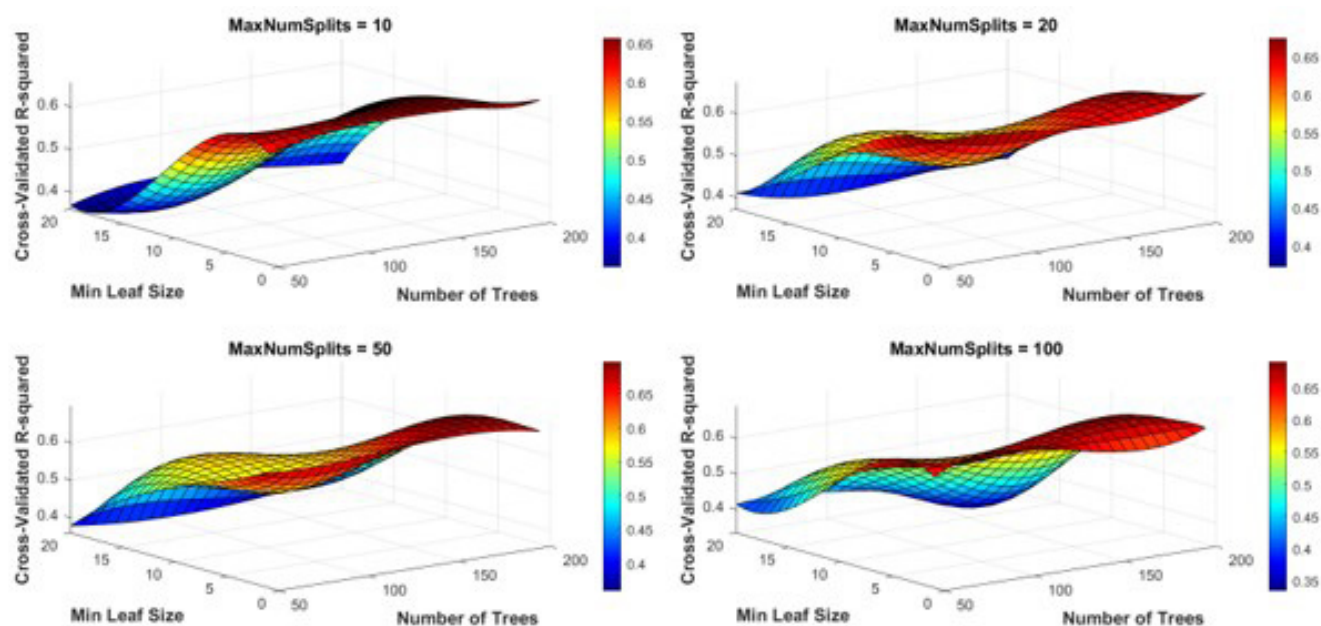


Figure 3—3D plots of cross-validated R² values for different combinations of hyperparameters in the random forest model

Table 7

Training and testing dataset performance metrics for the tuned RFR model

Dataset	R ²	RMSE
Training	0.812	0.395
Testing	0.749	0.521

conducted exclusively using the training data through a five-fold cross-validation grid search. The test set was withheld throughout this process and used only for final model evaluation to avoid information leakage.

As illustrated in Figure 3, the cross-validated R² surfaces from the grid search confirm that increasing the number of trees and maximum splits improves model performance and stability. The figure also shows that very small or very large leaf sizes reduce accuracy. These patterns guided the selection of the optimal configuration (100 trees, minimum leaf size = 1, maximum splits = 100).

Increasing the number of trees and maximum splits improved accuracy by reducing variance, whereas excessively large or small leaf sizes degraded performance (Hastie et al., 2009; Cutler et al., 2007). The selected configuration balanced stability and computational efficiency. Predictor importance analysis (Figure 4) identified N', BS, VR, and PF as the dominant predictors, linking geotechnical stability (e.g., Mathews stability number) with blast and design factors (burden-to-spacing ratio, powder factor, and void ratio).

Various methods have recently been proposed to help users interpret the predictions of complex models such as SHAP analysis (Lundberg, Lee, 2017) that confirmed some variables as the most influential, with N' and HR increasing predicted OB and features such as PF and PRIM showing context-dependent effects (Figures

5a – 5b). In addition, negative SHAP values (in Figure 5b) mainly associated with lower VR, shorter HL, and lower PF, which drive UB. These effects are smaller in mean magnitude because UB cases are fewer and more localised.

The SHAP baseline value represents the mean predicted OB/UB across the training data, serving as the reference point from which each predictor's contribution shifts the prediction toward OB (positive values) or UB (negative values).

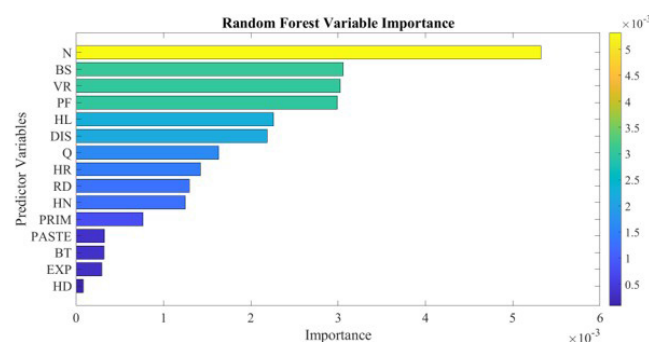


Figure 4—Random forest variable importance plot

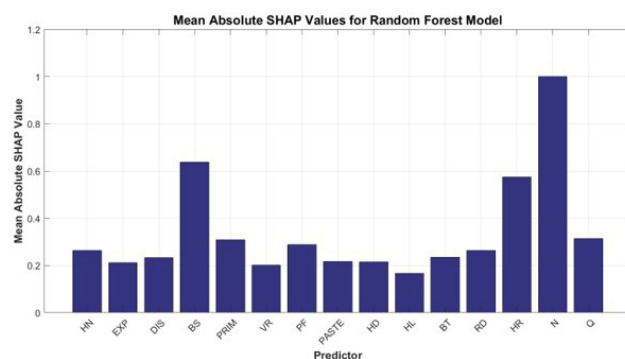


Figure 5a—Mean absolute SHAP values for the optimised RFR model

Integrating Bayesian, ensemble, and hybrid machine learning models

Gradient boosting machine

The GBM model is an ensemble learning method that iteratively builds decision trees to correct the residuals of previous weak learners, thereby reducing bias and variance (Friedman, 2001). Similar to the random forest approach, the dataset was divided into 70% training and 30% testing to identify the best-performing model for predicting OB and UB. Hyperparameters were tuned using a grid search across number of trees = 50 – 200, learning rate = 0.01 –

0.2, and maximum splits = 10 – 100, with five-fold cross-validation used to assess each configuration. The best model achieved a cross-validated $R^2 = 0.731$ with 200 trees, learning rate = 0.01, and 50 splits. Retraining produced $R^2 = 0.897$ (RMSE = 0.269) for the training set and $R^2 = 0.686$ (RMSE = 0.656) for the test set, while the full-dataset RMSE = 0.422. As illustrated in Figure 6, the 3D surface plots show how the GBM model's performance (cross-validated R^2) varies with tree number, learning rate, and split depth. Increasing the maximum number of splits generally improved accuracy, whereas lower learning rates produced more stable results. Optimal performance occurred with moderate-to-high tree counts, lower learning rates, and deeper splits.

To address the small performance gap between training and test sets, a refined search (GBM2) was conducted using narrower hyperparameter ranges: up to 250 trees, learning rate = 0.015, and maximum splits = 60. The resulting model achieved cross-validated $R^2 = 0.726$, training $R^2 = 0.946$ (RMSE = 0.206), and test $R^2 = 0.780$ (RMSE = 0.503), with full-dataset RMSE = 0.323, indicating improved generalisation. The 3D surface plots for the refined model (Figure 7) confirm that increasing the number of trees and the learning rate enhances predictive strength, particularly when combined with higher split counts (40 – 60). The summary of both models is provided in Table 8, highlighting the improved robustness

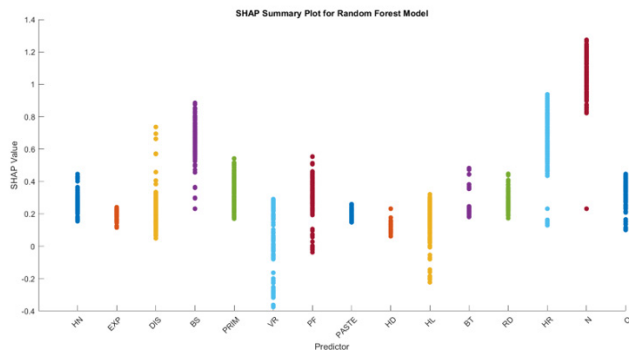


Figure 5b—SHAP summary plot for the RFR model demonstrating variability

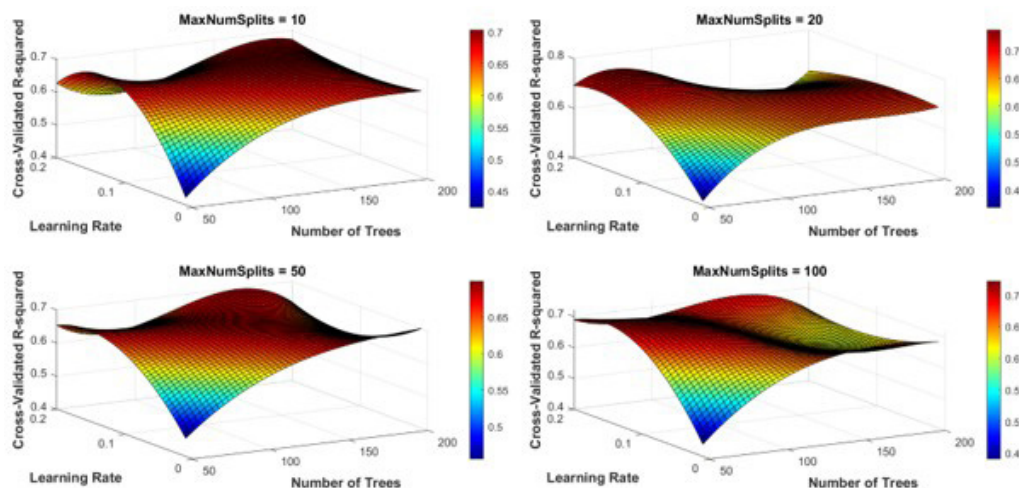


Figure 6—3D surface plots visualising the performance of the GBM model based on different hyperparameter configurations

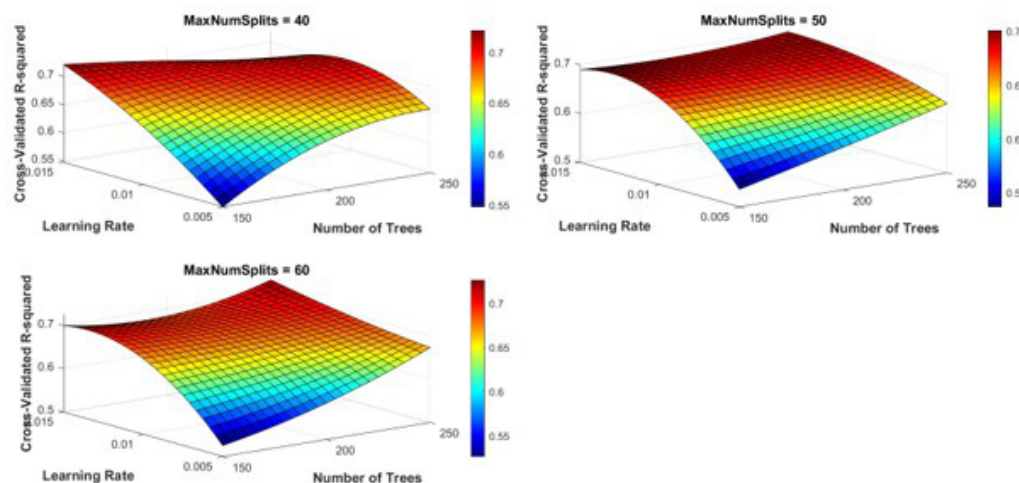


Figure 7—3D surface plots visualising the performance of the GBM2 model based on updated hyperparameter configurations

Integrating Bayesian, ensemble, and hybrid machine learning models

and lower error achieved by GBM2.

Feature-importance analysis based on the SHAP framework is presented in Figures 8a and 8b. Figure 8a presents the mean absolute SHAP values for each predictor variable, quantifying their relative contribution to the GBM model's predictions. The results show that N' is by far the dominant predictor (~ 2.28), applying a much stronger influence on model outputs than any other variable. BS follows as the second-most influential factor (~ 0.52), while predictors such as HD, PASTE, and Q' contribute only marginally. This hierarchy indicates that N' primarily governs the model's predictive capability, with BS providing secondary influence related to burden-spacing effects. The SHAP summary plot (Figure 8b) further shows the direction and variability of these influences: higher N' values tend to increase predicted OB, while lower HL (shorter drillholes) and VR (smaller available void) are associated with UB. The spread of SHAP values reflects sample-specific variation, emphasising that the same predictor can have different effects under varying geological and blasting conditions.

Bayesian methods

Gaussian process regression

GPR is a non-parametric, probabilistic regression method that infers a distribution over possible functions rather than a single deterministic mapping (Rasmussen, Williams, 2006). This allows the model to provide both mean predictions and uncertainty bounds, making it valuable for applications requiring confidence estimation. By incorporating prior information through kernel functions, GPR captures complex relationships and quantifies prediction uncertainty (Snoek et al., 2012). The kernel function, defining the covariance between data points, is central to model performance. Four kernel types were evaluated in this study: squared exponential, Matérn 3/2, Matérn 5/2, and rational quadratic (Murphy, 2012). A grid search was applied to determine the optimal kernel based on cross-validated R^2 (Bergstra, Bengio, 2012). The initial model identified the Matérn 3/2 kernel as best, achieving cross-validated $R^2 = 0.752$. Using a 70/30 train-test split, the model produced $R^2 = 0.933$ (RMSE = 0.232) on training data and $R^2 = 0.820$ (RMSE = 0.452) on the test set, with an overall RMSE = 0.314. Figure 9 presents heatmaps of cross-validated R^2 values across kernel functions and hyperparameter combinations (length scale, sigma), where warmer colours denote higher performance. The best configurations included: squared exponential (length = 20, sigma = 1, $R^2 = 0.752$), Matérn 3/2 (1, 0.1, $R^2 = 0.766$), Matérn 5/2 (20, 0.1, $R^2 = 0.761$), and rational quadratic (20, 10, $R^2 = 0.767$). These results highlight the impact of kernel selection and parameter tuning on GPR performance.

A refined model (GPR2) expanded the grid search to jointly optimise the kernel type, length scale, and sigma. The rational quadratic kernel achieved the best overall accuracy (cross-validated

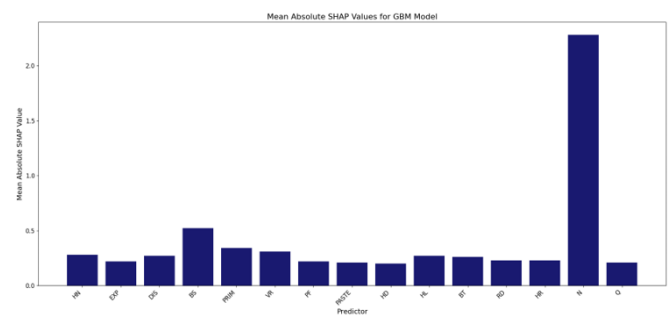


Figure 8a—Mean absolute SHAP values for the optimised GBM model

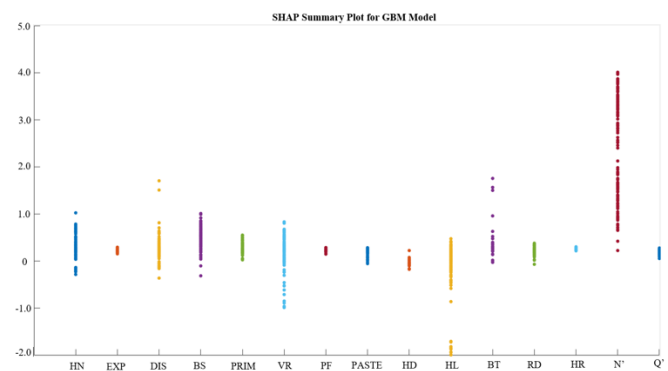


Figure 8b—SHAP summary plot for the optimised GBM model demonstrating variability

Kernel: squareexponential				
Length Scale \ Sigma	0.01	0.1	1	10
0.1	0.7158	0.7299	0.7182	0.7185
1	0.6655	0.7191	0.7053	0.7003
5	0.5344	0.7096	0.7036	0.7130
20	0.7133	0.7260	0.7524	0.7210

Kernel: matern32				
Length Scale \ Sigma	0.01	0.1	1	10
0.1	0.6975	0.7276	0.7373	0.7465
1	0.7378	0.7665	0.7157	0.7049
5	0.7420	0.7532	0.7068	0.6973
20	0.7373	0.7417	0.7425	0.7429

Kernel: matern52				
Length Scale \ Sigma	0.01	0.1	1	10
0.1	0.7262	0.7337	0.7241	0.7003
1	0.7004	0.7165	0.6980	0.7068
5	0.6866	0.7323	0.7083	0.7455
20	0.7218	0.7612	0.7083	0.7037

Kernel: rationalquadratic				
Length Scale \ Sigma	0.01	0.1	1	10
0.1	0.7359	0.7536	0.7250	0.7159
1	0.7335	0.7017	0.7039	0.7359
5	0.7549	0.7009	0.7333	0.7500
20	0.7474	0.7360	0.7434	0.7670

Figure 9—Heatmap of cross-validated R2 values for different kernel functions

Table 8 Training and testing dataset performance metrics						
Models	Best R ² (5-fold CV)	Training set R ²	Training set RMSE	Test set R ²	Test set RMSE	Full dataset RMSE
Initial GBM	0.731	0.897	0.269	0.686	0.656	0.422
GBM2	0.726	0.946	0.206	0.780	0.503	0.323

Integrating Bayesian, ensemble, and hybrid machine learning models

$R^2 = 0.767$). The final configuration (length = 20, sigma = 10) produced $R^2 = 0.920$ (RMSE = 0.237) for training and $R^2 = 0.737$ (RMSE = 0.601) for testing, with a full-dataset RMSE = 0.323. Comparatively, the initial Matérn 3/2 model generalised slightly better (test $R^2 = 0.820$, RMSE = 0.452), suggesting that while the refined model achieved higher training accuracy, it exhibited mild overfitting. As a result, the first model was selected as the preferred configuration for its balanced performance and stability. Figures 10a and 10b present the SHAP-based feature-importance results. The mean absolute SHAP values (Figure 10a) show VR, Q', and HL as the most influential predictors (~ 0.45 – 0.50), followed by RD, HN, and N' (~ 0.30). Less influential features such as EXP, DIS, BS, PRIM, PF, PASTE, HD, and BT exhibit moderate-to-low impact (~ 0.25). The SHAP summary plot (Figure 10b) further illustrates directionality: higher HL (longer holes) generally increases predicted OB, while smaller VR (limited available void) and lower HL correspond to UB conditions. The variability in SHAP distributions indicates that feature influence is context-dependent, reflecting local geological and blasting variability.

The SHAP analysis not only quantifies the statistical influence of each predictor but also supports their physical relevance in underground stoping. For instance, High SHAP values for N' indicate that greater wall stability reduces the likelihood of excessive OB, consistent with established stability graph principles. Similarly, higher BS values correspond to more balanced explosive energy distribution, reducing both OB and UB. The VR parameter reflects the available void space for explosive expansion; insufficient void can lead to confined energy release and UB, while excessive void increases the risk of OB. These insights confirm that the model captures both the statistical and physical relationships driving stope performance, reinforcing its interpretability and field applicability.

Model evaluation and comparison – Bayesian and ensemble

This study initially used multiple linear regression to define the relationships between independent variables and the target variable. The first linear regression model produced an RMSE of 0.580, an R^2 of 0.629, and an adjusted R^2 of 0.598. However, the second model's performance declined after addressing multicollinearity, resulting in an RMSE of 0.742 and R^2 of 0.439. The RFR model utilised 100 decision trees, with the best-performing model achieving a cross-validated R^2 of 0.682, a training set R^2 of 0.812, and an RMSE of 0.396. On the test set, it achieved an R^2 of 0.749 and an RMSE of 0.521, showing strong generalisability. Similarly, the first GBM model achieved a cross-validated R^2 of 0.731, with a training set R^2 of 0.898 and an RMSE of 0.270. However, its test set R^2 of 0.686 and RMSE of 0.656 suggested potential overfitting. To address this, a refined GBM model (GBM2) was developed, achieving a cross-validated R^2 of 0.726, a training set R^2 of 0.946 with an RMSE of 0.206, and a test set R^2 of 0.780 with an RMSE of 0.503, which indicated improvements in generalisability. The initial GPR model, utilising a basic grid search over kernel functions, identified the Matérn 3/2 kernel as the best-performing option, achieving an R^2 of 0.752, a training set R^2 of 0.933 with an RMSE of 0.233, and a test set R^2 of 0.820 with an RMSE of 0.452. A second GPR model with a refined hyperparameter tuning process using the rational quadratic kernel achieved a cross-validated R^2 of 0.767. Although it improved training set performance ($R^2 = 0.921$, RMSE = 0.237), its test set R^2 of 0.737 and RMSE of 0.600 indicated potential overfitting compared to the initial model. As shown in Figure 11, the initial

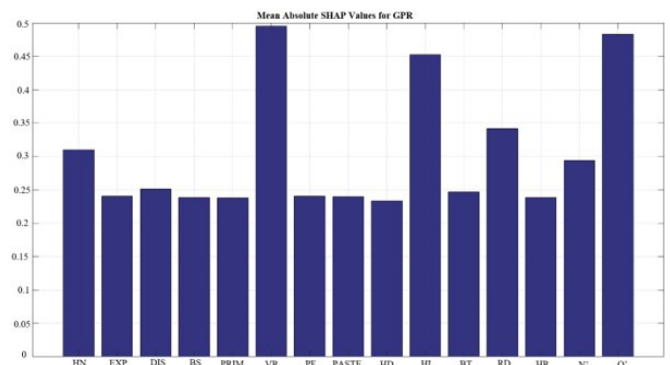


Figure 10a—Mean absolute SHAP values for GPR model

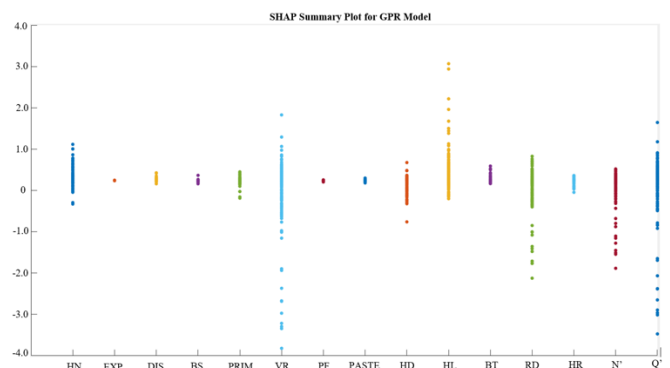


Figure 10b—SHAP summary plot for GPR model demonstrating variability

GPR model became the preferred model due to its better test set performance and balanced fit.

SHAP analysis was used to compare variable importance across all models. The bar chart in Figure 12 shows the mean absolute SHAP values for all variables across the three models (RFR, GBM, and GPR) and assesses their importance in predicting OB and UB. As highlighted in the chart, N', BS, and VR consistently appeared as the most important variables in all models.

Table 9 further highlights the findings, showing that N' is the most important variable in both RFR and GBM, BS is the second most important in RFR and GBM, and VR is the top variable in GPR, where it is also the third most important in GBM.

The prominence of N', BS, and VR across all three models is consistent with their physical significance and with field observations from the case study site. N' represents the modified Mathews stability number and is a composite indicator of rock mass quality, stress environment, and structural orientation. Higher N' values generally correspond to more favourable stability conditions, while lower values are commonly associated with increased perimeter damage and OB potential. The burden-to-spacing ratio directly influences energy distribution and confinement during blasting; deviations from the optimal BS range typically result in uneven breakout, either as excessive OB or UB. VR, defined as the void ratio available in front of the advancing blast, represents the amount of open space created by the previous extraction. It governs how effectively the next blast can expand into the available void. Insufficient void ratio restricts rock movement, increases confinement, and contributes to underbreak, while excessive void ratio reduces confinement and can contribute to uncontrolled OB (Lovitt, 2008). The alignment between these physical mechanisms, field observations, and the SHAP-derived importance rankings

Integrating Bayesian, ensemble, and hybrid machine learning models

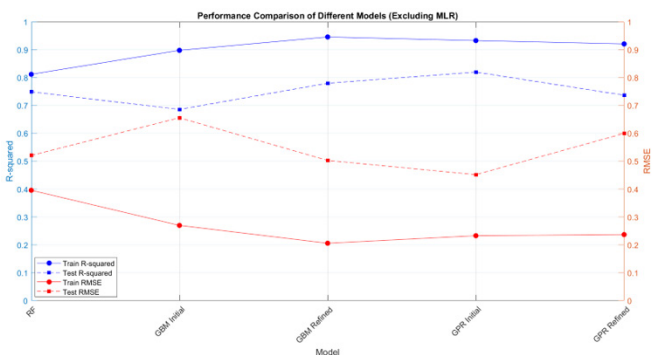


Figure 11—Performance comparison of different Bayesian and ensemble models

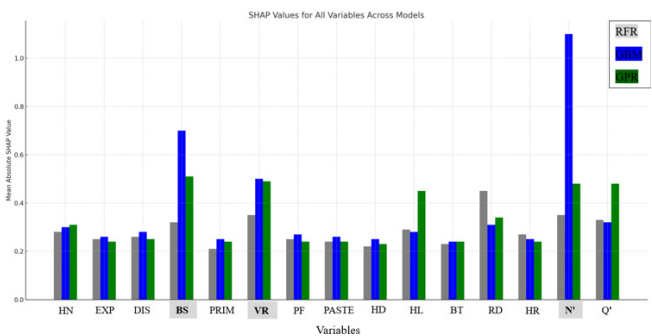


Figure 12—SHAP values for all variables across models

Table 9 SHAP analysis results for all variables			
Variables	RFR	GBM	GPR
HN	0.26	0.28	0.31
EXP	0.21	0.22	0.24
DIS	0.23	0.27	0.25
BS	0.64	0.52	0.24
PRIM	0.31	0.24	0.24
VR	0.2	0.31	0.5
PF	0.29	0.22	0.24
PASTE	0.22	0.21	0.24
HD	0.22	0.2	0.23
HL	0.17	0.27	0.45
BT	0.23	0.26	0.25
RD	0.26	0.23	0.34
HR	0.57	0.23	0.24
N'	1	2.28	0.29
Q'	0.31	0.21	0.48

provides a consistent explanation for why N', BS, and VR emerged as the leading predictors of OB and UB.

Hybrid models

While individual models such as RFR, GBM, and GPR demonstrated varying degrees of accuracy and generalisability, none offered consistently superior performance across all metrics.

For instance, the refined GBM model achieved strong training accuracy ($R^2 = 0.946$) but showed signs of overfitting on the test set ($R^2 = 0.780$, $RMSE = 0.503$). Similarly, the refined GPR model, despite improved hyperparameter tuning, resulted in a test R^2 of 0.737, lower than its initial version. To address these limitations and leverage the complementary strengths of different algorithms, a hybrid modelling approach was employed.

Table 10 summarises all seven hybrid models developed in this study. Each model is described in terms of its architecture, training and test performance (R^2 and RMSE), and a brief explanation of its structure and outcomes. The results show that while several hybrid configurations resulted in improved generalisation compared to basic models, the standalone GPR model remained the top performer. Nevertheless, the best hybrid model combining GPR, GBR, RFR, and ANN with a linear meta-learner achieved the highest test R^2 of 0.721 among all hybrid approaches.

Figure 13 demonstrates the architecture of the best-performing hybrid model, which integrates four base learners of GPR, GBR, RFR, and ANN, each trained on the full input feature set. The individual predictions ($\hat{y}_1$ to $\hat{y}_4$) from these models are then passed to a linear regression meta-learner, which combines them to produce the final OB/UB prediction amount at each drillhole location ($\hat{y}_{final}$).

Despite extensive experimentation with a range of hybrid models, including various combinations of GPR, GBR, RFR, and ANN, and different meta-learners such as linear regression, gradient boosting, ANN, and Bayesian ridge, further improvements in test set R^2 plateaued. As shown in Table 10, the strongest generalisation performance ($R^2 = 0.721$, $RMSE = 0.425$) was achieved by Model 6, which incorporated all tuned base learners and used a linear regression meta-model. More complex configurations provided better training fits but failed to improve test R^2 , indicating overfitting. Given this saturation point, additional attempts to improve R^2 were deemed unlikely to produce meaningful gains and were therefore not further pursued.

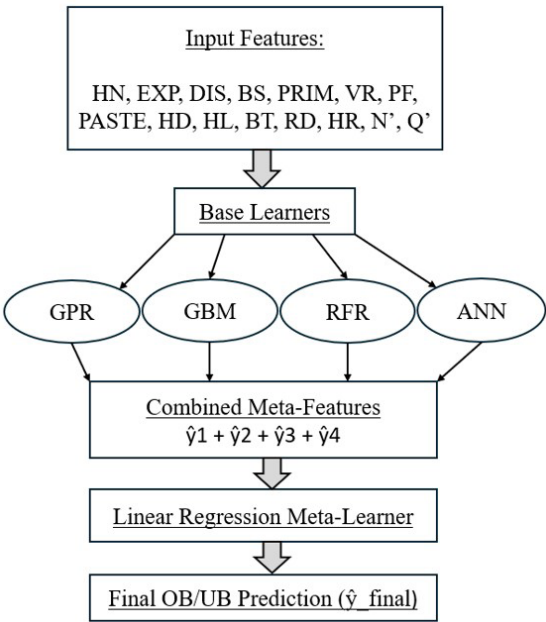


Figure 13—Architecture of the best hybrid model (Model #6), using GPR, GBR, RFR, and ANN for OB/UB prediction via a linear regression meta-learner

Table 10

Performance summary of developed hybrid models

Model #	Architecture	Train R ²	Train RMSE	Test R ²	Test RMSE	Description
1	GPR + GBR LR meta	0.915	0.295	0.644	0.479	Initial hybrid model with GPR and GBR as base learners, LR as meta-learner. Reasonable generalisation, but did not surpass standalone GPR.
2	GPR + GBR GBR meta	0.884	0.344	0.599	0.509	Same base learners, but nonlinear meta-learner (GBR). Performance dropped due to overfitting.
3	GPR + GBR + RFR LR meta	0.903	0.315	0.560	0.533	Added RFR to increase diversity, but generalisation weakened, suggesting redundancy in base learners.
4	Tuned GPR + GBR + RFR LR meta	0.900	0.319	0.679	0.455	All base learners tuned via GridSearchCV. Stronger generalisation, approaching standalone GPR.
5	Tuned GPR + GBR + RFR ANN meta	0.877	0.354	0.702	0.438	ANN used as meta-learner. Reduced overfitting and improved test performance modestly.
6	Tuned GPR + GBR + RFR + ANN LR meta	0.912	0.300	0.721	0.425	ANN added as a base learner. This was the strongest test performance among all hybrid models.
7	GPR + GBR Bayesian	0.957	0.21	0.677	0.457	Hybrid model combining Bayesian (GPR) and ensemble (GBR) base learners with a Bayesian ridge meta-model. Strong training fit and moderate generalisation.

Across all configurations, the standalone GPR model remained the highest-performing test set R² (0.820), making it the most accurate and generalisable model in this study. While several hybrid models approached this level of performance, particularly the combination of GPR, GBR, RFR, and ANN with a linear meta-learner (R² = 0.721), none fully surpassed GPR.

Conclusions

This study focused on the application of Bayesian and ensemble methods to predict OB and UB in underground stoping operations. Initially, multiple linear regression was used to establish foundational relationships between variables, revealing significant predictors and multicollinearity issues. Addressing these issues resulted in a decline in the model's performance, highlighting the limitations of linear methods to capture complex geotechnical and drill and blast relationships within the dataset. The RFR model significantly outperformed the MLR models, achieving a test R² of 0.749 and an RMSE of 0.521. This model's success can be

related to its robustness against multicollinearity and its ability to capture complex interactions between variables. The RFR model was optimised using a grid search over ranges for the number of trees, minimum leaf size, and maximum number of splits. The best model was determined based on the highest cross-validated R² value, leading to the selection of 100 trees, a minimum leaf size of 1, and 100 maximum splits as the optimal hyperparameters. SHAP value analysis for the RFR model highlighted N', BS, and HR as the most impactful variables, indicating their important role in predicting OB and UB. The first GBM model achieved a cross-validated R² of 0.731. A refined GBM model (GBM2) was developed to address potential overfitting. The GBM2 model achieved a cross-validated R² of 0.726, a training set R² of 0.946 with an RMSE of 0.206, and a test set R² of 0.780 with an RMSE of 0.503. SHAP analysis for this model showed N', BS, and VR to be the most impactful variables. The GPR model, particularly with the Matérn 3/2 kernel, demonstrated strong predictive performance with a test R² of 0.820 and an RMSE of 0.452. As the refined GPR model

Integrating Bayesian, ensemble, and hybrid machine learning models

showed signs of overfitting, the initial model's balanced fit and high predictive accuracy made it a preferred choice. SHAP analysis for the GPR model revealed that VR, Q', and HL were the most important variables to predict OB and UB. To further explore model performance, a series of hybrid stacking models were developed using various combinations of base learners, including GPR, GBR, RFR, and ANN, with meta-learners such as linear regression, ANN, and Bayesian ridge. Among the architectures trialled, the model that incorporated all four base learners with a linear regression meta-learner achieved the strongest generalisation, with a test R^2 of 0.721 and RMSE of 0.425. Deep stacking configurations and more complex meta-learners such as ANN and Bayesian ridge did not yield further improvements and showed signs of overfitting. Additional approaches, such as simple and performance-weighted averaging of tuned base model predictions, also failed to outperform the best hybrid configuration. These results demonstrate that while hybrid ensembles offer flexibility and robustness, there is a clear performance plateau when using the current feature set and data structure.

In conclusion, ensemble methods, particularly RFR and GBM, demonstrated reasonable performance in predicting OB and UB in underground stoping. However, The Bayesian approach, represented by the GPR model, showed significant promise. SHAP value analysis provided valuable insights into feature importance, consistently identifying geotechnical and drill and blast design variables such as N', BS, VR, and HR as key predictors across all models. These findings reinforce the critical role of design parameters like burden and spacing, the Mathews stability number (focusing on hanging wall stability), as well as blasting parameters such as void ratio and hydraulic radius, in effectively controlling OB and UB in open stoping operations.

While the models developed in this study provided strong predictive capabilities, several limitations remain. All models were trained and tested on historical data from different zones in a single mining operation, limiting the ability to generalise findings across different geological settings or operational conditions. Despite the extensive exploration of hybrid configurations, performance gains plateaued, suggesting that the current feature set may not fully capture all drivers of OB and UB. Future work should focus on expanding the dataset to include multiple operations with varying geological conditions, incorporating additional process and operational variables, and exploring real-time data integration to improve model adaptability.

References

Abd Elwahab, A., Topal, E., Jang, H. 2023. Review of machine learning application in mine blasting. *Arabian Journal of Geosciences*, 16, Article 133.

<https://doi.org/10.1007/s12517-023-13329-4>

Adoko, A. C., Saadaari, F., Mireku-Gyimah, D. 2022. A feasibility study on the implementation of neural network classifiers for open stope design. *Geotechnical and Geological Engineering*, vol 40, no. 2.

Barton, N., Lien, R., Lunde, J. 1974. Engineering classification of rock masses for the design of tunnel support. *Rock Mechanics*, 6, pp. 189–236.

Belsley, D. A., Kuh, E., Welsch, R.E. 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.

New York: Wiley.

Bergstra, J., Bengio, Y. 2012. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, vol. 13, (Feb), pp. 281–305.

Cha, G.W., Moon, H. J., Kim, Y.C. 2021. Comparison of Random Forest and Gradient Boosting Machine models for predicting demolition waste based on small datasets and categorical variables. *International Journal of Environmental Research and Public He.*

Chimunhu, P. et al. 2024. Development of Novel Hybrid Intelligent Predictive Models for Dilution Prediction in Underground Sub-level Mining. *Mining, Metallurgy & Exploration*.

Cordova, D.P., Zingano, A.C., Goncalves, I.G. 2022. Unplanned dilution back analysis in an underground mine using numerical models. *Mining • REM, Int. Eng. J.*, vol. 75, no. 4.

Cutler, A., Cutler, D.R., Stevens, J.R. 2011. Random forests. In C. Zhang & Y. Ma (Eds.), *Ensemble machine learning: Methods and applications*. pp. 157–176. Springer.

https://doi.org/10.1007/978-1-4419-9326-7_5

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A. 2007. Random forests for classification in ecology. *Ecology*, vol. 88, 11, pp. 2783–2792. <https://doi.org/10.1890/07-0539.1>

Falahi, T., Nasserddine, G., Younis, J. 2023. Detecting Data Outliers with Machine Learning. *Al-Salam Journal for Engineering and Technology*, vol. 2, no. 2. <https://doi.org/10.55145/ajest.2023.02.02.018>

Filzmoser, P., Maronna, R., Werner, M. 2008. Outlier identification in high dimensions. *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1694–1711. <https://doi.org/10.1016/j.csda.2007.05.018>

Forster, K., Milne, D., Pop, A. 2007. Mining and rock mass factors influencing hangingwall dilution. s.l., s.n.

Friedman, J.H. 2001. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232.

Ghosal, I., Hooker, G. 2020. Boosting Random Forests to Reduce Bias; One-Step Boosted Forest and its Variance Estimate. arXiv. Retrieved from <https://arxiv.org/abs/1803.08000>

Han, H. et al. 2020. Random Forest and Bayesian Network Techniques for Probabilistic Prediction of Flyrock Induced by Blasting in Quarry Sites. *Natural Resources Research*, vol. 29, vol. 7. <https://doi.org/10.1007>

Hastie, T., Tibshirani, R., Friedman, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed., pp. 587–588). Springer.

Hefni, M.A., Abdellah Wael, R.E. 2020. Factors influencing stope hanging wall stability and ore dilution in narrow-vein deposits. *Geotechnical and Geological Engineering*.

Himanshu, V., Mishra, A.K., Roy, M.P., Singh, P.K. 2023. Innovative blasting practices for underground hard rock mining, in *Blasting technology for underground hard rock mining*. Springer Nature Singapore.

Integrating Bayesian, ensemble, and hybrid machine learning models

Hong, Z. et al. 2023. Experimental and numerical studies of the blast-induced overbreak and underbreak in underground roadways. *Underground Space*, vol. 8, no.1.

Ibarra-Gutierrez, S., Laflamme, M. 2021. Blasted ore losses and mineral reserve: reconciliation approaches and impact on stope performance. *Mining, Metallurgy & Exploration*.

Jang, H., Topal, E., Kawamura, Y. 2015. Unplanned dilution and ore loss prediction in longhole stoping mines via multiple regression and artificial neural network analyses. *The Journal of The South African Institute of Mining and Metallurgy*, vol. 115, pp 115.

Jolliffe, I., Cadima, J. 2016. Principal Component Analysis: A Review and Recent Developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*.

Jung, D., Choi, Y. 2021. Systematic Review of Machine Learning Applications in Mining: Exploration, Exploitation, and Reclamation. *Minerals*, vol 11, no. 2, pp. 148.
<https://doi.org/10.3390/min11020148>

Liu, H., Zhao, G., Yin, Y. 2023. Ensemble Tree Model for Long-Term Rockburst Prediction in Incomplete Datasets. *Minerals*, vol. 13, no. 1, pp. 103.
<https://doi.org/10.3390/min13010103>

Lovitt, D. 2008. Low voidage blasting to improve underground stope productivity. s.l., Proceedings of the Tenth Underground Operators Conference, *The Australasian Institute of Mining and Metallurgy (AusIMM)*, pp. 59–68.

Lundberg, S.M., Lee, S.I. 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774. Retrieved from
<https://arxiv.org/abs/1705.07874>

Malhotra, S., Karanickolas, J. 2020. A numerical transform of random forest regressors corrects systematically-biased predictions. *Machine Learning*, LicenseCC BY 4.0. Retrieved from
<https://doi.org/10.1007/s11749-016-0481-7>

Mathews, K.E., Hoek, E., Wellie, D.C., Steward, S.B.V. 1981. Prediction of stable excavation spans for mining at depths below 1,000 metres in hard rock. *CANMET*, DSS Serial No. OSQ80-00081.

McCoy, J.T., Auret, L. 2019. Machine learning applications in minerals processing: A review. *Minerals Engineering*, vol. 132,

pp. 95–109.

Montgomery, D.C., Peck, E.A., Vining, G.G. 2012. *Introduction to Linear Regression Analysis* (5th ed.). John Wiley & Sons. s.l.:s.n.

Mottahedi, A., Sereshki, F., Ataei, M. 2018. Development of overbreak prediction models in drill and blast tunneling using soft computing methods. *Engineering with computers*.

Murphy, K.P. 2012. *Machine learning: A probabilistic perspective*. MIT Press.

O'Brien, R.M. 2007. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, vol. 41, no. 5, pp. 673–690. <https://doi.org/10.1007/s11135-006-9018-6>

Osborne, M.A.G.R., Osborne, R.S.J. 2009. *Gaussian Processes for Global Optimization*. Department of Engineering Science, University of Oxford, Oxford, UK.

Papaioanou, A., Suorineni, F. 2016. Development of a generalised dilution-based stability graph for open stope design. *Mining Technology*.

Rasmussen, C.E., Williams, C.K.I. 2006. *Gaussian Processes for Machine Learning*. 1st ed. s.l.:Cambridge, Mass.: MIT Press.

Sellers, E.J., Salmi, E.F. 2020. Breaking new ground: challenges and opportunities for maximising value from underground blasting. In J. Wesseloo (Ed.), *UMT 2020: Proceedings of the Second International Conference on Underground Mining Technology* (p.

Snoek, J., Larochelle, H., Adams, R.P. 2012. Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, 4. arXiv.

Suglo, R.S., Opoku, S. 2012. An assessment of dilution in sublevel caving at Kazansi Mine. *International Journal of Mining and Mineral Engineering*, vol. 4, no. 1, pp. 1–16.

Yu, Z. et al. 2021. Machine-Learning-Aided Determination of Post-blast Ore Boundary for Controlling Ore Loss and Dilution. *Natural Resources Research*.

Zhang, Z., Zhang, T., Li, X., Dias, D. 2024. Bayesian ensemble methods for predicting ground deformation due to tunnelling with sparse monitoring data. *Underground Space*, vol. 16, pp. 79–93.

Zhao, X., Jia'an, N. 2020. Method of predicting ore dilution based on a neural network and its application. *Sustainability*. ♦

**METALLURGY & MINING
CHEMICALS SOLUTIONS THAT
DRIVE RESULTS – GLOBALLY**

