International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V74I1P120

Volume 74 Issue 1, 259-274, January 2026

Original Article

Stepwise Multiple Regression Analysis for Development of a Model to Predict the Performance of Surface Miners

Om Prakash Singh[1*], P Y Dhekne[1], Manoj Pradhan[1], Romil Mishra[1]

[1]*Department of Mining Engineering, National Institute of Technology, Raipur, Chhattisgarh, India.*

**Corresponding Author : opsinghnitrr@gmail.com*

Received: 02 August 2025 Revised: 31 December 2025 Accepted: 06 January 2026 Published: 14 January 2026

Abstract - *The present study puts forward the model formulations of Stepwise Multiple Regression Analysis for obtaining the value of Normalized Production Rate and Pick consumption per 1000t of surface miners operating in opencast coal mines of Mahanadi Coalfields Limited. A total of 143 data entries have been compiled to develop models. The entries contain Uniaxial Compressive Strength Index, Cerchar Abrasivity Index, In-situ P-Wave Velocity, and Normalized Production Rate and Pick Consumption per 1000t. Two models have been formed independently to determine the Normalized Production Rate and Pick Consumption per 1000t. The two models have been developed and generated with the help of Minitab. The models have been formed with the forward selection and backward elimination method of stepwise regression techniques. The Student's T-tests have been carried out on models to determine which of the predictors are most significant. The results also reveal that the accuracy of models formed using statistical models is high and provide easy accessibility to predisposed engineers of surface miners to obtain estimations of Normalized Production Rate and Pick Consumption per 1000t. The models formed with statistical techniques provided appropriate results and can be effectively employed in opencast coal mines with similar geotechnical conditions.*

Keywords *- Cutting Performance, PickConsumption, Opencast Coal Mines, Stepwise Multiple Regression Analysis, Surface Miner.*

1. Introduction

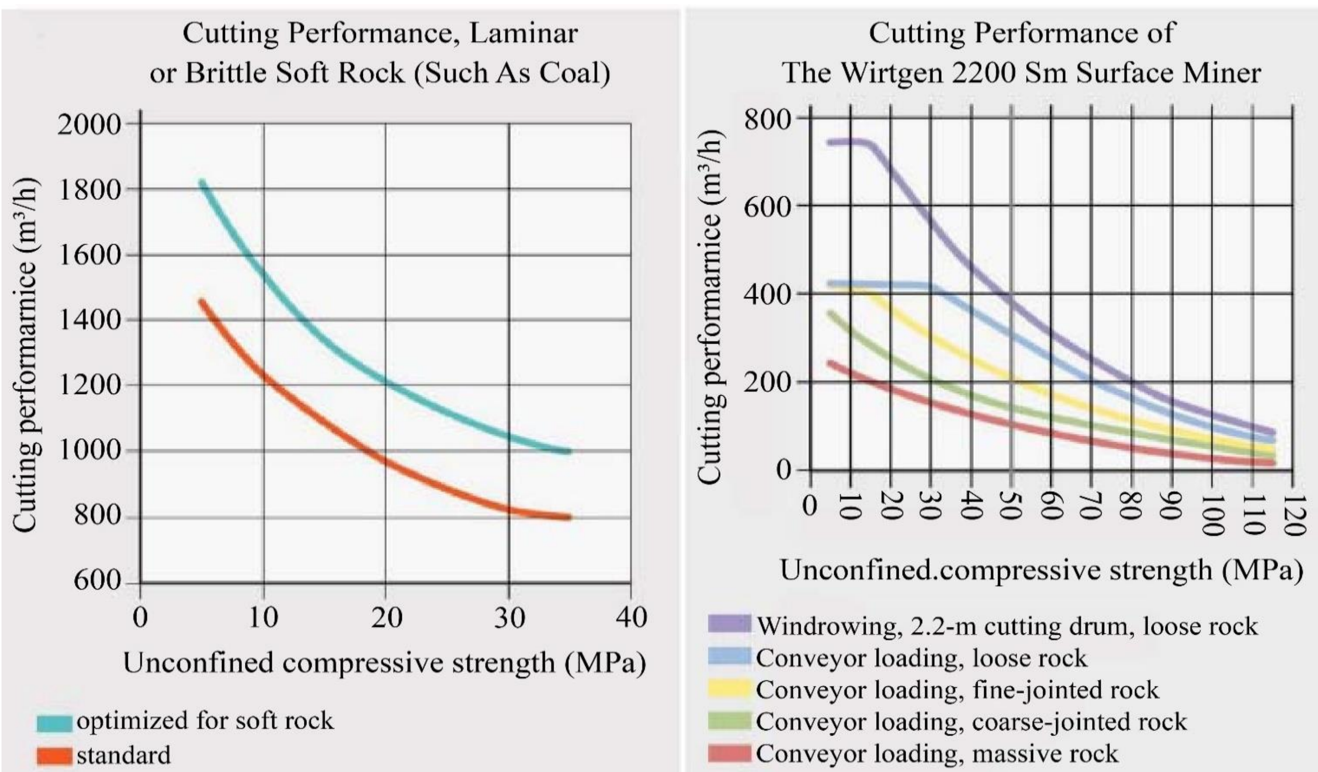
To satisfy the growing demand for faster production, the opencast coal mines have gradually adopted the use of mass production methods, mainly through the incorporation of surface miners. A surface miner is also known as a continuous surface miner. It combines extraction, crushing, and loading in a single process operation [1, 2]. Coal extraction through the usage of surface miners has proved more efficient, attracting a significant amount of research work for improving their efficiency. There has been a substantial amount of work carried out related to cutting performance predictions in surface miners. A significant part of existing models has been generalized predictive models that have been derived based on data drawn from different types of rocks [3-7]. Although models based upon machine parameters alone have underemphasized the importance of rock mass parameters in a substantial way, models based upon rock mass parameters often neglect a few important machine-related parameters in surface miner models. Recent models attempted based upon a broad spectrum of machine-related parameters that encompass intact rock properties, rock mass properties, machine properties, and geo-mining parameters have been generalized and are strongly predictive in nature. For instance, models based on multiple regressions were developed by Prakash et. al. (2024) [8] in evaluating cutting speeds in surface miners in coal and limestone settings. Although they are considered valuable, such models may also pose limitations when applied to specific pairs of geo-mining factors and machine parameters. Additionally, the impact caused by the dirt bands on the cutting performance of surface miners has received little or inadequate attention and research in past research efforts. As such, it has become necessary to study further the simultaneous influence exerted by intact rock, rock mass, and machine parameters on the geo-mining process concerning coal seams with intercalated dirt bands.

2. Literature Review

Various parameters tend to affect the performance level of surface miners. Four major groups of these parameters have been identified for classification and understanding purposes. They include rock parameters for intact rock and rock masses, parameters for machines and equipment, and operational parameters. The intact rock parameters that tend to affect the cuttability process include several rock properties that tend to be either mechanical, physical, or strength-related. On the other hand, rock parameters take into consideration the structural and geological discontinuities that exist within the rock environment. Moreover, parameters that relate to the machines tend to have an intensive influence on performance [9]. Understanding these variables is essential for selecting appropriate excavation systems and optimizing surface mining operations. Consequently, several empirical models have been developed in previous studies to predict Key Performance Indicators (KPIs) such as productivity, pick Consumption, and diesel consumption, using various combinations of these influencing parameters [10].

Table 1. Types of variables influencing the cutting performance of surface miners

Type	Variables
Intact Rock Parameters	Brazilian tensile strength, Brittleness index, Cerchar abrasivity index, Density, Firmness index, Moisture content, Point load strength index, P-wave velocity, Specific energy consumption, Uniaxial Compressive Strength (UCS), Young's modulus
Rock Mass Parameters	Ash/impurities/silica content, Dirt bands/intrusions, In-situ P-wave velocity, Joints/discontinuities, Rock Quality Designation (RQD), Schmidt rebound hardness number, Stickiness, Volumetric joint count.
Machine Parameters	Breakout angle, Cutter power, Drum diameter, Drum width, Ratio of Energy transfer to the cutting drum, Engine power, Machine weight, Number of picks, Pick lacing pattern, Pick material, Pick orientation
Operational Parameters	Available face length, Available face width, continuous mining method) Cutting speed, Depth of cut, Direct loading, empty travel back method, Mining technique (e.g., Operator efficiency, Side-casting, turn back method, Wet or dry cutting, Windrowing



(a) (b)

Fig. 1 Cutting performance of surface miner: (a) 2200 SM 3.8, and (b) 2200SM.

As seen in Figures 1(a) and 1(b), the manufacturers of surface miners only used one rock parameter-the undefined compressive strength-to characterise the cutting performance. In this case, cutting performance (m3/hr) is determined by the volume of material cut relative to cutting time, or effective hours of cutting. Manoeuvring and servicing time are not accounted for. Rock's UCS is thought to be the most accurate measure of cuttability, and as compressive strength rises, cutting rate falls [11]. The cutting performance of surface miners, as illustrated in Figures 1(a) and 1(b), demonstrates a strong inverse relationship with the Unconfined Compressive Strength (UCS) of the rock. Figure 1(a) presents the cutting performance of a surface miner in laminar or brittle soft rock (e.g., coal), comparing two operational configurations-standard and optimized for soft rock. The optimized configuration achieves significantly higher productivity, particularly at lower UCS values (up to ~1800 m³/h at UCS ≈ 0 MPa), with performance gradually decreasing as UCS increases to 35 MPa. In contrast, the standard setup yields lower performance throughout, with the productivity gap narrowing as UCS increases. Whereas Figure 1(b) displays the cutting performance of the Wirtgen 2200 SM surface miner under different operating conditions, including windrowing and conveyor loading in various rock mass structures (loose, fine-jointed, coarse-jointed, and massive) [9]. The highest cutting performance (~750 m³/h) is achieved during windrowing in loose rock at low UCS values.

In contrast, performance declines markedly in more competent and massive rocks under conveyor loading, reaching below 100 m³/h for UCS values near 100 MPa. The performance hierarchy clearly shows that rock mass structure and the method of material handling substantially influence cutting efficiency [4].

Figures 1(a) and 1(b) collectively emphasize that surface miner cutting performance decreases with increasing UCS [12]. Optimized machine configurations (Figure 1(a)) and favourable operating conditions, such as windrowing in loose rock (Figure 1(b)), significantly enhance productivity. Machine efficiency is highest in low-strength, well-fragmented rock masses and lowest in high-strength, massive formations, underlining the importance of selecting suitable equipment and operational strategies based on geotechnical site conditions [13].

Abrasiveness is a key characteristic of rock or coal that affects cutting pick wear and pick maintenance costs, and consequently affects output rate. Because coal quality varies, the abrasivity differs from site to site. According to reports, the average pick life at SECL mines ranged from 275 to 681 hours, and the rate of abrasion of cutting picks is strongly impacted by the coal quality (silica content) [22].

The Cerchar abrasivity test can be used to evaluate the abrasivity of rock. Pick wear is an ongoing process, and in order to precisely measure it, the weight loss of all the picks in relation to the amount of material cut must be correlated. This is not a practical method because it requires stopping the machine for a considerable amount of time. Picks are changed only after the tungsten carbide insert is totally worn out, and it is determined that it is no longer able to cut the material. If no pick is replaced on a given day, it indicates that none of the picks have deteriorated to the point where they need to be replaced, not that there is no pick wear. Therefore, picks that are changed every day in terms of material cut will vary greatly and cannot be utilised to indicate pick Consumption or wear rate. According to a report, the coefficient of correlation between pick consumption and the daily and monthly production of coal was 0.6643 and 0.9092, respectively [15]. In order to account for changes and provide a more accurate picture of pick consumption, the authors of this research averaged pick replacements over a period of one month or more.

Origiliasso C et al. (2014) have created an empirical relation for calculating the production rate of surface miners by taking into consideration UCS, Cerchar's Abrasivity Index (CAI), and engine power (P_w) in kW as the main characteristics influencing productivity. This relationship is founded on data from equipment makers and experimental data. In addition to cutting time, the output rate accounts for time spent on ancillary tasks like manoeuvring and servicing. Usually, the materials having a CAI value of 0.5 are considered easy to dig and non-abrasive [3]. Furthermore, the production rate will be influenced by the machine's power. Rocks with a higher UCS can be cut by a machine that is heavier and more powerful. This equation's primary flaw, though, is that it totally ignores the properties of the rock mass. Dey and Ghose (2008) created a cuttability index for the first time. It is a composite of the following factors: machine power, rock abrasivity, volumetric joint count, direction of cutting relative to the major joint orientation, and point load strength index. This index provides an initial assessment of the surface miner's applicability. If the value of the cuttability index is found to be greater than 80, then the deployment of surface miners is not recommended. Based on the cuttability index, a model was developed to predict the production of surface mining in m³/hr using the rated capacity of the machine in m³/hr as another variable. The relationship considers machine and operating factors, intact rock, and rock mass. However, there is ample room for investigation because the value of the proportionality constant falls between 0.5 and 1. All the above researchers have developed relationships for estimating the productivity of surface miners only.

On the other hand, in their outstanding research, Prakash et al. (2015) created the Rock Cuttability Index for Surface Miners which is used to estimate Key Performance Indicators (KPIs) of surface miners viz., TPH (production in tonnes per hour), DCT (diesel consumption/1000 t) and PCT (pick consumption/1000 t) [16]. The models are summarized in the table below.

Table 2. Models developed by different researchers

Author	Year	Model
Jones and Kramadibrata	1995	$P_R = 1005 - 559 \text{ Log (UCS)}$
Origiliasso et al.	2014	$P_R = (2 \times \text{EP} - 600) \times e^{-0.024\{10\times(\text{CAI}-0.5)+UCS\}}$
Dey and Ghose	2008	$L^* = \left(1 - \frac{\text{CI}}{100}\right) \text{K} \times \text{Mc}$
Prakash et. al.	2015	$\text{TPH} = 181.5\, \text{I}_{SM}{}^{0.245}$ $\text{DCT} = 338 \times \text{I}_{SM}{}^{-0.19}$ $\text{PCT} = 2 \times \text{I}_{SM}{}^{-0.18}$
Acronyms - P_R = production rate of surface miner; UCS = uniaxial compressive strength; EP = engine power; CAI = Cerchar's Abrasivity Index; L* = production of surface miner; CI = cuttability index which is a composite of the		

following factors: Machine Power (M), Rock Abrasivity (Aw), Direction Of Cutting With Regard To Major Joint Orientation (Js), Volumetric Joint Count (Jv), and point load strength Index (Is) calculated using the equation $\mathrm{CI} = \mathrm{Is} + \mathrm{Jv} + \mathrm{Aw} + \mathrm{Js} + \mathrm{M}$; Mc = rated capacity of the machine; k = proportionality constant whose value ranges from 0.5 to 1; TPH = production in tonnes per hour; DCT = diesel consumed/1000 t; PCT = pick consumed/1000 t; I_{SM} = Rock Cuttability Index for Surface Miners calculated using equation $I_{SM} = \frac{1000\ \mathrm{MF}}{\mathrm{IRF\ x\ RMF}}$ where, MF = Machine factor calculated using $\mathrm{MF} = \frac{\mathrm{EP\ x\ CA\ x\ CS}}{1000}$; IRF is Intact Rock Factor calculated using $\mathrm{IRF} = \mathrm{E\ x\ CAI\ x\ LVp}$; RMF = Rock Mass Factor calculated using $\mathrm{RMF} = \frac{\mathrm{IVp}}{\mathrm{RN}}$; CA = contact area of the cutting drum calculated as $\mathrm{CA} = \frac{2\pi \mathrm{R} \cos^{-1}[(\mathrm{R}-\mathrm{D})/\mathrm{R}]}{360}\mathrm{DW}$; CS is the speed of cutting (m/min); E = Young's modulus (GPa); LV_p = laboratory p-wave velocity (km/s); IV_p = in-situ p-wave velocity (m/s); RN = rebound harness number

Cutting speed and Depth of cut are two input factors that depend on rock strength in these relationships. Additionally, the developed relationships encompass ten distinct factors, making KPIs time-consuming and intricate. The machine's operational weight is a crucial factor that will affect the surface miner's KPIs. Because the cutting drum is towards the bottom, the machine's weight makes cutting easier and enhances pick penetration [17]. Furthermore, to achieve enough reaction force and vibration-free cutting motion, the engine power-to-operating weight ratio must be proportionate. The fuel efficiency of the machine improves as the ratio increases. This becomes noticeable while cutting through the hard or dirt bands that are interwoven throughout the coal seam. When cutting dirt bands, there is a potential that the machine will vibrate if its weight is lower.

Contemporary developments in rock excavation mechanics and energetics involve improving the understanding of the basic process controls that define rock excavation efficiency, force requirement, or cutter wear mechanisms. Using fracture mechanics theories, Wang & Su (2019b) studied the cutting process using a conical pick in rocks, showing that rock fracture toughness, elastic modulus, or Depth of cut are primary controls in specific energy consumption or force coefficient, whereby specific energy consumption increases remarkably with rock elastic modulus while only a small fraction of the mechanical work input contributes to the generation of new fracture surfaces during rock breakage [18]. On the other hand, applying full-scale tests to jet-assisted rotary drilling, J. Yang et al. (2019) studied full-scale tests to investigate rock responses to different rock-drilling conditions using jet-assisted rock drilling technology, providing a comparative energy analysis functional in designing optimized drilling mechanisms [19].

Thermal phenomena at the tool-rock contact have also been of concern. Kumar et al. (2020) carried out an analysis involving experiments and modeling with conical pick cutting to understand related thermal phenomena and utilize the design of experiments for parameter optimization. In this analysis, results indicated optimized solutions for parameter sets that can minimize tool tip temperature and prevent wear factors that would otherwise affect operational reliability [20]. On a more general scale, Zhang and Ouchterlony (2022) have carried out a study combining research findings for rock breakage specific energy and established related perspectives for specific energy in surface miners that relate to minimum specific energy models applied for studies related to analysis and optimization [21].

More recent works include the use of data-intensive models for cutting force prediction and tool performance analysis. Morshedlou et al. (2024) introduced an ensemble learning and regression model for cutting force estimation in conical cutters based on rock mechanical properties and cutter and rock contact area. Their best ensemble models using Explainable AI (XAI) delivered the best results for cutting force estimation and have potential applications for tool selection and cutting force estimation [22]. Zhao et al. (2024) further extended these studies by combining experimental results and numerical analysis to assess the dynamic reliability and wear behavior of picks mounted on cutting drums operating under harsh working conditions. They identified the typical wear behavior of picks and stress concentrations for designing cutting drums and assessing pick wear life [23].

Considered collectively, these works represent a progression from fundamental fracture mechanics analysis to system-level testing and analysis through to modern machine learning models for prediction. This work establishes a strong scientific foundation for advancing cutting efficiency, minimizing tool wear, and providing optimal mechanical excavation system designs and usage practices.

3. Research Gap

Although there is an existing literature scope on surface miner performance, the research covering these aspects has remained deeply oriented to the parameters of intact rock, rock masses, machine setup, and operating conditions. This paved the way for the creation of performance indicators, namely the Cuttability Index (CI) and Rock Cuttability Index for Surface Miners (RCISM), and Empirical Relations Involving Uniaxial Compressive Strength (UCS), Rock Abrasity, and Engine Power. Nonetheless, most of this research has remained concerned with relatively homogenous rock, namely coal and limestone, having relatively low variations within the seams. Conversely, most of the Indian opencast mines, where coals are excavated, tend to display pronounced stratigraphic heterogeneity, where dirt bands, shales, and variations of ash content manifest with prompt changes in surface miner cuttability, cutting force, picks, and machine utilization. Such heterogeneities are currently underrepresented in current modeling approaches despite the fact that parameters such as the level of silica content, ash percentage, joint density, and moisture content are proven to significantly impact pick life, cutting energy demand, and cutting rates. This aforementioned fact makes the above-mentioned performance dynamics of surface miners, when operating in coal mines where the seam appears together with bands of intercalated dirt layers, inadequately forecasted, thereby emphasizing an existing gap in the used methods. This thereby forms the justification that an immense possibility exists within the design of an easy-to-use predictive version able to forecast significant performance indicators of surface miners concerning normalized cutting rates and pick wear consumption.

4. Site Description and Methodology

In an effort to achieve a better insight into the variables that determine the performance of Surface Miners, especially while coal cutting in the presence of intercalated dirt bands, an in-depth survey has been done in the ten opencast coal mines of Mahanadi Coalfields Limited (MCL) in the state of Odisha, wherein the coal mines have high ash content and the regular occurrence of interlaced dirt bands of varying thickness from 10 cm to 150 cm. The research emphasized the development of predictive models on the basis of Key Performance Indices (KPIs) of surface miners, viz., production rate and pick consumption. The diesel consumption component of the KPI has not been taken into account, being influenced by various non-geological and non-operational parameters alongside machine age, maintenance level of engines, and operator expertise, apart from the tonnage cut. As a measure to ensure equal comparison of production rates with varying makes and models of surface miners, the chosen parameter replaced the absolute value of the production rate called the Normalized Production Rate (NPR) with a definition that stated the tonnage of material removed per unit area of the drum per hour and expressed in t/h/m². The Pick Consumption Per 1000 tonnes of material removed (PCM) parameter has been considered, along with generally accepted norms of earlier research works [24]. The prime aim of the predictive models has been to reasonably predict the removal rates within geo-mining conditions, with a special focus on coal seams with intercalated dirt bands.

5. Data Collection

To achieve this objective, a dataset comprising 143 observations has been compiled from the designated study locations. Out of the total number of observations, 127 were used to train the model, while 16 others were set aside to test the model. The observations were gathered through machine usage hours, material excavated quantities and types, and picks consumed. Machine usage hours did not factor in standby and maintenance time. The material quantities variable considered the total coal and dirt bands excavated by the surface miners. The pick consumption data were obtained through daily inspections, whereby the picks were inspected at the start of the working day, and the worn-out picks were replaced with new ones.

The movement of the surface miners was traced through the high-accuracy navigation system. A non-destructive technique was used to measure the in-situ sonic velocity of the coal seam. The technique is called In-situ seismic refraction tomography, and the procedure followed was according to the guidelines set by the ASTM E494-20 standard. Moreover, the NX-size cores, 3 m long, were extracted through the core drilling equipment for a better insight into the characteristics of the coal seam and the dirt band. The standards used for laboratory analysis are presented in Table 3. The sample preparation for performing the tests was conducted according to IS:9179-1979.

Table 3. Standards adopted for the laboratory determination of intact rock properties

Properties	Standard adopted
Bulk density	IS:13030-1991
CAI	ASTM D7625-10
E	IS:9221-1979
UCS	IS:9143-1979

A statistical summary of the input and output variables used for development of the model (excluding the data kept for validation) is presented in the table.

Table 4. Statistical summary of the variables for model development

Type	Variable	Count	Minimum	Maximum	Mean	StDev	Median
Input	UCS	127	13.5	25.01	21.429	3.404	22.45
Input	CAI	127	0.125	0.216	0.17727	0.02105	0.18
Input	E	127	1.11	3.781	2.1751	0.6291	2.23
Input	IVp	127	511	1369	966.1	193.6	1011
Output	NPR	127	14720	32352	20132	4484	18646
Output	PCM	127	0.67	3.42	2.0461	0.5699	2.1

The dataset has four input variables: Uniaxial Compressive Strength (UCS), Cerchar Abrasivity Index (CAI), Young's Modulus (E), and P-wave Velocity (IVp), and two output variables: the Normalized Production Rate (NPR) and the Pick Consumption per 1000 tonnes (PCM). The minimum and maximum values of the UCS are observed to be

13.5 and 25.01 MPa, respectively, with the average and standard deviation at 21.429 MPa and 3.404 MPa, respectively, thereby implying moderate variation. The CAI ranges between 0.125 and 0.216, with the average and standard deviation at 0.17727 and 0.02105, respectively, thereby implying low variation in the abrasivity aspects. Young's modulus varies between 1.11 GPa and 3.781 GPa, with the average and standard deviation at 2.1751 GPa and 0.6291 GPa, respectively, thereby implying moderate variation. The P-wave Velocity contains the highest level of variation among the variations presented, ranging between 511 m/s and 1369 m/s, with the average and standard deviation at 966.1 m/s and 193.6 m/s, respectively. For the output parameters, the range for NPR is between 14.720 and 32.352 thousand tons/day, with an average of 20.132 thousand tons/day and a standard deviation of 4.484 thousand tons/day, which is highly variable in the production process. The range for PCM is between 0.67 and 3.42, with an average of 2.0461 and a standard deviation of 0.5699, indicating moderate variability.

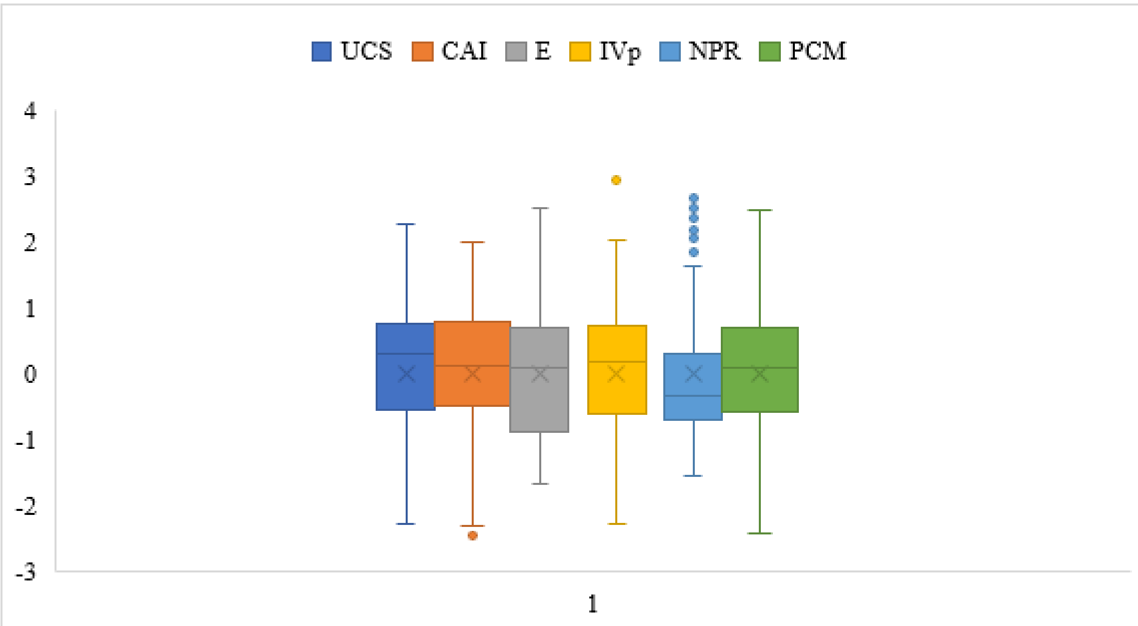


Fig. 2 Box-plot showing distribution of the standardised values of input and output variables

It can be observed that this dataset ensures a good variability in input and output variables that can be used in the development of predictive models that are meaningful from a statistical perspective. Variability in the data provides a reasonable basis for determining relationships between variables related to rock mass properties and surface miner performance indicators.

6. Stepwise Multiple Regression Analysis

Stepwise Multiple Regression Analysis was performed using the Minitab software. Multiple Linear Regression (MLR) is an analytical method used for modeling the relationship between the dependent variable and a set of independent variables. This represents the mathematical relationship of the variables, which are commonly employed for predicting the response and understanding the relative impact of measures of the variables within the environment.

MLR assumes a linear relationship exists between the dependent and independent variables. Even in situations where the proper functions are less linear, linear approximations can still be applicable for modelling. Indeed, in relation to other evolved methods, for instance, Artificial Neural Network (ANN) models, Adaptive Neuro-Fuzzy Inference System, or Tree-Based models, Multiple Linear Regression appears to be less complex in terms of computations while still being efficient. The general equation of the proper function in the form of the regression function appears in the following equation.

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \varepsilon$$

Where Y denotes the estimated response or dependent variable, β_0 refers to the intercept of the regression line; β_i represents the slope parameters or coefficients associated with the predictors; x_i denotes the independent variables; and ε signifies the random error component that captures the portion of variability unexplained by the actual regression function and cannot be entirely removed or minimized.

The presence of a large number of potential predictors in a single model may lead to problems such as overfitting, multicollinearity, and reduced interpretability. It is therefore essential to identify the optimal subset of predictors to ensure model strength. The stepwise multiple linear regression procedure is one common technique for accomplishing variable selection.

Stepwise regression is a semi-automated process whereby the addition or deletion of variables is done based on predetermined criteria set by the researcher using appropriate mathematical or computational models and processes. The process aims to strike a balance between achieving simplicity within the model and maximizing predictability, with the understanding that not all variables may influence the outcome equally or independently. These criteria include, but are not limited to, the p-value and the adjusted R^2, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC), discussed later on. Out of the various methods used for stepwise regression, the two most commonly employed methods are forward selection and backward elimination.

The forward selection method involves beginning the process by considering a model that includes the constant term alone. The next step involves selecting the predictor that results in the most significant improvement in the model at each stage and adding it to the existing model. The process continues this way until no additional variable enters the model that meets the specified criterion for entry. The forward selection method is practical when many variables are available for selection, though few of them are expected to enter the equation as significant variables. Additionally, this method does not introduce multicollinearity at the initial stages of the process.

In contrast, backward elimination uses a reverse approach. It begins with a complete model including all candidate predictor variables. At each step, the variable having the highest p-value-that is, the variable which shows least statistical significance-is eliminated from the model. The model is then re-estimated and the process repeated in order to ensure that the remaining variables have p-values below some pre-specified value. This method has many advantages when the initial set of predictors is somewhat limited in number, and many of these predictors are presumed to be statistically insignificant. Nonetheless, the backward elimination process may be computationally demanding when the number of predictors is significant, and its performance may be adversely affected by the inherent multicollinearity of the whole model.

Although it is powerful for model simplification and increasing interpretability, they do suffer from some limitations, i.e., neither of the methods guarantees identification of the globally optimal model, as both exclude a comprehensive evaluation of all possible combinations of predictors. Furthermore, stepwise methods are susceptible to the sequence effects of variable entry or removal, such that slightly different data and/or criteria may yield divergent results. Neither method considers variable interactions if they are not explicitly included. Sometimes, stepwise regression could be considered better suited for a given scenario than machine learning models. By using stepwise regression, parameter estimates are readily determinable, and statistical inference is possible in a simple and comprehensible manner. This is specifically important in a particular scenario where explaining variable effects is of prime importance rather than predicting them. Additionally, in stepwise regression, smaller amounts of data are sufficient for the proper functioning of regression analysis, and the process is less computer-intensive compared to most machine learning models, which require larger amounts of data and quite often extensive parameter tuning to counteract overfitting.

7. Evaluation Criteria for the Developed Models

Evaluation of multiple linear regression models involves statistical soundness and predictive performance. For the purpose at hand, several quantitative criteria have been developed for investigating model adequacy, parsimony, and reliability in making predictions for unseen data. The commonly used measures are p-values, the coefficient of determination R2, adjusted R2, AIC, BIC, and Cp of Mallows. Each metric performs a different evaluative function to contribute to informed variable selection and model comparison.

7.1. p-Value

The p-value is an assessment of the significance of individual predictors in the regression model, assuming that a specific regression coefficient equals zero. The dependent variable remains unaffected. A lower p-value, typically below 0.05, suggests that the associated predictor contributes significantly to the model. In stepwise regression analysis, p-values provide a basis for the addition of a variable or for shrinkage. However, relying exclusively on the p-value is not always appropriate, especially when there are problems with multicollinearity and small sample size.

7.2. Coefficient of Determination (R^2)

The coefficient of determination, denoted as R^2, indicates the percentage of the dependent variable's variance that can be accounted for by the independent variables. Its range is from 0 to 1. It reflects the better performance of a model when it is high.

For instance, 0.75 of the R^2 would suggest that the predictors account for 75% of the variability in the response variable. Although intuitively appealing, R^2 has one natural Achilles heel: its value never decreases with the addition of more predictors, even those that bear no explanatory relevance.

7.3. Adjusted R^2

Adjusted R^2 overcomes the limitation of R^2 by penalizing the addition of non-informative predictors. It adjusts the coefficient of determination based on the number of predictors and the sample size, hence guarding against overfitting. In contrast to R^2, adjusted R^2 may go down when an added variable does not increase the performance of the model, and it is, therefore, more reliable in comparing models with different numbers of explanatory variables.

7.4. Akaike Information Criterion (AIC)

The Akaike Information Criterion is one of the most commonly used information-theoretic measures for comparing models.

By evaluating the goodness-of-fit against the model complexity, it provides a relative quality of the model and is defined as:

$$AIC = 2k - 2\ln(L)$$

Where k represents the number of parameters estimated, and L is the likelihood function. The model with the smaller AIC is regarded as superior. Since AIC penalizes models with too many parameters, it is considered advantageous for parsimonious model selection and is especially useful when models are non-nested.

7.5. Bayesian Information Criterion (BIC)

The Bayesian Information Criterion, like AIC, also considers both model explanatory power and parsimony; however, BIC has a more substantial penalty for complexity, particularly in large samples. It is defined as:

$$BIC = k\ln n - 2\ln(L)$$

Where (n) is the sample size. As in AIC, smaller values of BIC indicate better models. Because its penalty term is more stringent than that of AIC, BIC tends to favour simpler models. Both AIC and BIC have wide applications in model comparisons, and the model that results in the smallest value of the criterion is usually considered the best.

7.6. Mallows' Cp

Mallows' Cp is one of the significant criteria for regression subset selection, which assesses the trade-off between bias and variance. Mallows' Cp is calculated as:

$$C_p = \frac{SSE_p}{\hat{\sigma}^2} - (n - 2p)$$

Where n is the sample size, $\hat{\sigma}^2$ is the estimate of the error variance, and SSE_p Is the sum of squared errors for the model with p predictors. A model with a Cp value close to is considered desirable. Larger values of Cp suggest overfitting, whereas minimal values may indicate underfitting.

Thus, Cp offers a model selection diagnostic that strikes a balance between explanatory adequacy and model complexity.

8. Model Development for KPIs

8.1. NPR Model by Backward Elimination

The stepwise regression equation for predicting the NPR was developed using the backward elimination approach available in the Minitab software environment.

The backward elimination technique involves the elimination of variables that do not show an appreciable level of significance, and the process continues until an equation with a suitable level of simplicity and explanatory power has been derived. The variables considered for the multiple linear equation in the present study were UCS, CAI, E, and IVp.

In the first cycle, the regression model comprised all four variables. The p-value for E (Young's Modulus) was 0.158, exceeding the selection criterion for retaining the variable, which was set at α = 0.10. E was thus eliminated, and the model comprised of the remaining three variables: UCS, CAI, and IVp.

Table 5. Metrics of the stepwise backward elimination method for the NPR model

	Step 1		Step 2	
	Coefficient	**P value**	**Coefficient**	**P value**
Constant	52125		51413	
UCS	-592.5	0	-591.7	0
CAI	-87233	0	-82046	0
E	433	0.158		
IVp	-4.94	0	-4.2	0
Metrics				
S	858.085		861.621	
R^2	96.45%		96.40%	
R^2(adj)	96.34%		96.31%	
Mallows' Cp	5		5.02	
AICc	2083.7		2083.58	
BIC	2100.07		2097.31	
α to remove = 0.1				

The regression equation is:

$$NPR = 51413 - 591.7(UCS) - 82046(CAI) - 4.20\ (IV_p)$$

This formula represents a negative relationship between NPR and each of these three variables, suggesting that higher values of UCS, CAI, and IVp are associated with lower values of normalized production rate. From the coefficient table of the final model, UCS, CAI, and IVp are found to be equally significant predictors because they all share the exact value of 0.000. The absolute values of the t-statistics associated with UCS (33.44), CAI (25.94), and IVp (27.86) are tremendous, and thus, there is robust support for the inclusion of predictors.

The Variance Inflation Factor indices of UCS (5.90), CAI (7.22), and IVp (7.20) are less than 10. Thus, the multicollinearity between the predictors CAI and IVp. The model Summary reports an R^2 of 96.40, which means that the chosen predictors are collectively explaining approximately 96.40% of the variance in NPR. The Adj R^2 of 96.31 is very close to this and further supports the robustness of the model to the number of variables being used. The Predicted R^2 is 96.16, which denotes that the regression equation has maintained its goodness of fit and is free from overfitting.

Table 6. Coefficients table for the backward elimination NPR model

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	51413	822	62.54	0	
UCS	-591.7	54.8	-10.8	0	5.9
CAI	-82046	9798	-8.37	0	7.22
IVp	-4.2	1.06	-3.95	0	7.2

Table 7. Model summary for the backward elimination NPR model

S	R-sq	R-sq(adj)	R-sq(pred)
861.621	96.40%	96.31%	96.16%

Table 8. ANOVA metrics for the backward elimination NPR model

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	2.44E+09	8.14E+08	1096.56	0
UCS	1	86561721	86561721	116.6	0
CAI	1	52056833	52056833	70.12	0
IVp	1	11564823	11564823	15.58	0
Error	123	91314083	742391		
Total	126	2.53E+09			

The results of the ANOVA test establish the overall significance of the Regression Model. The F-statistic for the Regression Model is 1096.56, and its significance is given by its p-value of 0.000. The F-values for UCS, CAI, and IVp are 116.60, 70.12, and 15.58, respectively, which are significant at their respective levels. The Standard error of the Estimate (S = 861.621) is the standard error of the regression. Information-theoretic values also support the model.

The Corrected Akaike Information Criterion (AICc) of 2083.58 and Bayesian Information Criterion (BIC) of 2097.31 suggest a good trade-off between model simplicity and fit. Mallows' Cp statistic is also at 5.02, almost hitting the target of p+1=4, implying that this model is unbiased and efficient. Therefore, the backward elimination method provides a statistically sound and significant predictive model for NPR based on UCS, CAI, and IVp. The deletion of E does not weaken the model since R-squared and Mallows' Cp remain stable with each elimination. Hence, the final result provides a good representation of the rock and material variables' effect on the normalized production rate.

8.2. NPR Model by Forward Selection

On the contrary, in the forward selection method, the process starts with a null model, and predictors are gradually added according to their relevance in enhancing model performance, which might be assessed using significance tests and information criteria.

In this analysis, UCS was added to the first step due to its significant value (p = 0.000) and considerable contribution to reducing the standard deviation to 1349.14. In this first step, UCS accounted for approximately 91.02% of the variation in NPR. In the second step, CAI was added, increasing the value of R^2 to 95.94% while decreasing the standard error to 910.861.

Adding IVp in the third step further improved the model fit. The value of R^2 climbed to 96.40%, while the standard error decreased to 861.621. Analysis of E in the fourth step revealed a probability value of 0.158, which exceeds the level of significance. Therefore, E is not included in the model. The process is illustrated in a table.

The forward selection method produced the following final regression equation:

$$NPR = 51413 - 591.7(UCS) - 82046(CAI) - 4.20(IVp)$$

The model constructed using forward selection is identical to the one constructed using backward elimination. The coefficients table offers detailed information about the regression parameters. The constant term is 51413, while the coefficients for UCS, CAI, and IVp are all negative: -591.7, -82046, and -4.20, respectively. Each of the three factors is significant (p = 0.000) with high t-statistics, establishing the importance of the factors for explanation. The VIFs are precisely the same as the values shown in the backward elimination test, establishing the extent of multicollinearity.

Table 9. Metrics of the stepwise forward selection method for the NPR model

	Step 1		Step 2		Step 3		Step 4	
	Coef	**P**	**Coef**	**P**	**Coef**	**P**	**Coef**	**P**
Constant	47064		53223		51413		52125	
UCS	-1256.8	0	-682.6	0	-591.7	0	-592.5	0
CAI			-104156	0	-82046	0	-87233	0
IVp					-4.2	0	-4.94	0
E							433	0.158
Metrics								
S	1349.14		910.861		861.621		858.085	
R-sq	91.02%		95.94%		96.40%		96.45%	
R-sq(adj)	90.95%		95.87%		96.31%		96.34%	
Mallows' Cp	186		18.72		5.02		5	
AICc	2195.22		2096.56		2083.58		2083.7	
BIC	2203.56		2107.61		2097.31		2100.07	
Achieved minimum AICc = 2083.58								

Table 10. Coefficients table for the forward selection NPR model

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	51413	822	62.54	0	
UCS	-591.7	54.8	-10.8	0	5.9
CAI	-82046	9798	-8.37	0	7.22
IVp	-4.2	1.06	-3.95	0	7.2

Table 11. Model summary for the forward selection NPR model

S	R-sq	R-sq(adj)	R-sq(pred)
861.621	96.40%	96.31%	96.16%

On comparing the performance metrics of the forward selection method, it verifies the reliability of the resultant model. R^2, adj R^2 (96.31%), and predicted R^2 (96.16%) match exactly as in the backward elimination, which verifies consistency in the result. The value of the standard error, 861.621, matches precisely as calculated earlier. Information criteria showed a constant decrement upon the addition of the significant variables, which attained minimum points (AICc = 2083.58, BIC = 2097.31) at the addition of variables UCS, CAI, and IVp. Mallows' Cp values for this situation are also the same at 5.02, revealing no bias within the predictions. The ANOVA table for the forward model identifies the significance the model assigns to the retained factors in the explained variance.

Table 12. ANOVA metrics for the forward selection NPR model

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	2.44E+09	8.14E+08	1096.56	0
UCS	1	86561721	86561721	116.6	0
CAI	1	52056833	52056833	70.12	0
IVp	1	11564823	11564823	15.58	0
Error	123	91314083	742391		
Total	126	2.53E+09			

The F-statistic for E, however, is 2.02 with a p-value of 0.158, again suggesting that E is not significant in predicting NPR in the current model context.

Notwithstanding the absence of statistical significance in the current model context, Young's Modulus is physically relevant because it is a measure that affects the amount of energy needed for fracture.

8.3. PCM Model by Backward Elimination

The construction of a regression equation for predicting PCM was carried out using the methods of backward elimination and forward selection using the Minitab software. The backward elimination method began with the initial complete model, which included the predictor variables UCS, CAI, E, and IVp.

The first step eliminated the variable UCS on the basis of its high probability value (0.635), which surpassed the alpha to remove value set for the test criterion of 0.10.

The process continued by eliminating the variables E (probability value = 0.283) and IVp (probability value = 0.437) for the same reason, eventually terminating with the final step, Step 4, where only the significant predictor variable CAI with a probability value of 0.000, well within the criterion, was retained as the sole predictor for interpretation, with the regression equation being:

$$PCM = -2.554 + 25.949(CAI)$$

The process has been summarised in a table.

Table 13. Metrics of the stepwise backward elimination method for the PCM model

	Step 1		Step 2		Step 3		Step 4	
	Coef	P	Coef	P	Coef	P	Coef	P
Constant	-2.524		-2.525		-2.628		-2.554	
UCS	-0.0049	0.635						
CAI	26.76	0	26.39	0	27.14	0	25.949	0
E	0.0627	0.283	0.0624	0.283				
IVp	-0.00021	0.351	-0.00025	0.231	-0.00014	0.437		
Metrics								
S	0.163404		0.162889		0.162995		0.162739	
R-sq	92.04%		92.02%		91.95%		91.91%	
R-sq(adj)	91.78%		91.83%		91.82%		91.85%	
Mallows' Cp	5		3.23		2.38		0.99	
AICc	-92.12		-94.09		-95.06		-96.57	
BIC	-75.75		-80.36		-84.01		-88.24	
α to remove = 0.1								

The coefficient values not only support the significance of the intercept variable (–2.554) and CAI coefficient (25.949) with a value of 0.000, but also support the validity of the VIF value of 1.00, which rejects the multicollinearity assumption.

The model signified excellent explanatory and predictive capabilities with an R^2 value of 91.91%, adjusted R^2 value of 91.85%, and predicted R^2 value of 91.66%, which indicates that CAI has been able to explain a significant variation of the variability of PCM.

Furthermore, the ANOVA table verified the validity of the robust model with an F-statistic value of 1420.14 and a regression p-value of 0.000, which not only indicated that CAI has a high degree of significance to predict PCM, but also supported that the model is free from multicollinearity.

Table 14. Coefficients table for the backward elimination PCM model

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-2.554	0.123	-20.78	0	
CAI	25.949	0.689	37.68	0	1

Table 1. ANOVA metrics for the backward elimination PCM model

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	37.611	37.6111	1420.14	0
CAI	1	37.611	37.6111	1420.14	0
Error	125	3.311	0.0265		
Total	126	40.922			

8.4. PCM Model by Forward Selection

A parallel modelling was performed employing Minitab's forward selection technique. In contrast to backwards elimination, this process starts from a null or empty model and stepwise adds predictors depending upon their role in enhancing model fitness.

Assessment of each of the first candidate variables (UCS, CAI, E, and IVp) indicated that CAI had the least p-value (0.000) and thus was introduced into the model first. After inclusion in the model, the remaining candidates were tested in turn; none were found to be macro-statistically significant (p= 0.635 in UCS, p= 0.283 in E, p= 0.231 in IVp). Therefore, no further candidates were added to the model, and a quadratic optimality criterion function was reached that was equivalent to that from backward elimination:

$$PCM = -2.554 + 25.949(CAI)$$

This equation precisely defines the linear relationship, in which PCM is estimated using the value of CAI. The steps have been summarised in a table.

Table 16. Metrics of the stepwise forward selection method for the PCM model

	Step 1		Step 2		Step 3		Step 4	
	Coef	P	Coef	P	Coef	P	Coef	P
Constant	-2.554		-2.628		-2.525		-2.524	
CAI	25.949	0	27.14	0	26.39	0	26.76	0
IVp			-0.00014	0.437	-0.00025	0.231	-0.00021	0.351
E					0.0624	0.283	0.0627	0.283
UCS							-0.0049	0.635
Metrics								
S	0.162739		0.162995		0.162889		0.163404	
R-sq	91.91%		91.95%		92.02%		92.04%	
R-sq(adj)	91.85%		91.82%		91.83%		91.78%	
Mallows' Cp	0.99		2.38		3.23		5	
AICc	-96.57		-95.06		-94.09		-92.12	
BIC	-88.24		-84.01		-80.36		-75.75	
Achieved minimum AICc = -96.57								

Table 17. Coefficients table for the forward selection PCM model

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-2.554	0.123	-20.78	0	
CAI	25.949	0.689	37.68	0	1

Table 18. Model summary for the forward selection PCM model

S	R-sq	R-sq(adj)	R-sq(pred)
0.162739	91.91%	91.85%	91.66%

The coefficients table once again verifies the strong statistical significance of the two terms (p = 0.000) along with the absence of multicollinearity in the model (VIF = 1.00). The summary table mentions the Standard Error of Regression (SER) as 0.162739, which informs about the scale of the deviations.

The value of R^2, Adjusted R^2, and Predicted R^2 is the same as that of the backward elimination model (91.91%, 91.85%, and 91.66%, respectively), signifying perfect explanatory fitness and generality. Finally, the Analysis of Variance (ANOVA) table confirms the statistical significance of the created model. The output further strengthens the significance of the model, with an F-statistic value of 1420.14 (p = 0.000). The degree of freedom values: 1 for regression and 125 for error – specify that the test has been performed on a dataset of 127 observations. Furthermore, the forward selection model has given a minimum Akaike Information Criterion (AIC) of -96.57. In conclusion, both the backward elimination method and the forward selection technique yield the same single-predictor regression equation, in which CAI emerges as the lone, significant predictor of PCM. The equation possesses high explanatory power, predictability, and simplicity.

Table 19. ANOVA Metric for the forward selection PCM model

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	37.611	37.6111	1420.14	0
CAI	1	37.611	37.6111	1420.14	0
Error	125	3.311	0.0265		
Total	126	40.922			

9. Model Validation

The proposed models were verified using the unseen data subset, which was obtained from the original dataset. The data subset for verification was designed to incorporate every type of surface miner used at the chosen study sites. Furthermore, the data subset for verification was prepared to retain significant variations for the most material influencing factors, such as the UCS, IVp, CAI, and power-to-weight ratio of surface miners.

Table 20. Statistical summary of the validation dataset

Variable	Count	Mean	StDev	Minimum	Median	Maximum	IQR
UCS	16	21.45	4.15	13.61	22.44	29.31	5.25
CAI	16	0.17763	0.02424	0.129	0.179	0.22	0.03575
E	16	2.122	0.735	1.15	2.06	3.6	1.257
IVp	16	946.8	257.4	529	913	1555	323.5
NPR	16	20621	5131	13220	19233	32003	6791
PCM	16	2.053	0.607	0.88	2.065	3.47	0.615

For validation purposes, the values of the predicted Normalized Production Rate (NPR) and pick Consumption per 1000 tonnes (PCM) were compared with the actual values obtained from field measurements. The models are restated below for ready reference.

$$NPR = 51413 - 591.7(UCS) - 82046(CAI) - 4.20(IVp)$$

$$PCM = -2.554 + 25.949(CAI)$$

The percentage error of prediction was calculated between actual and predicted values using the standard expression:

$$\%\,Error = \frac{(Predicted\ Value - Actual\ Value)}{Actual\ Value} \times 100$$

Table 21. Actual vs Predicted values of NPR and PCM

Actual NPR	Predicted NPR	Percentage error	Actual PCM	Predicted PCM	Percentage error
13219.82	9489.153	-28.22	3.47	3.15478	-9.08
17591.61	16051.61	-8.75	2.29	2.557953	11.70
16303.44	16186.31	-0.72	2.65	2.428208	-8.37
17546.29	16784.16	-4.34	2.27	2.402259	5.83
18957.56	18218.56	-3.90	2.07	2.246565	8.53
16788.13	16093.85	-4.14	2.58	2.661749	3.17
16163.78	16590.14	2.64	2.36	2.6358	11.69
18324.61	18260.13	-0.35	2.23	2.428208	8.89
19631.69	20251.58	3.16	2.06	1.935177	-6.06
19507.85	20622.42	5.71	1.81	1.909228	5.48
21355.36	21391.72	0.17	1.99	1.831381	-7.97
24360.3	23706.38	-2.68	1.7	1.519993	-10.59
21993.2	22776.05	3.56	1.83	1.909228	4.33
28553.89	27508	-3.66	1.3	1.260503	-3.04
27636.85	28214.53	2.09	1.36	1.208605	-11.13
32002.55	30554.23	-4.53	0.88	0.793421	-9.84

To better assess the predictability, actual versus predicted scatter plots were employed. Scatter plots allow the direct rendering of the observed values on the x-axis, the values predicted by the model on the y-axis, with the 1:1 line connecting the values in a perfect correlation manner. The spatial arrangement of the data in relation to the line can convey the level of prediction accuracy, the nature of the systematic error, and the error patterns of the data. If the spatial arborization of the data aligns closely with the line, the data has made accurate predictions, while a larger spacing from the line implies less precise prediction accuracy. Examination of the scatter plots reveals that the NPR model has a pretty good predictive ability across a majority of the data points, except for one strong outlier where the model significantly underpredicts the value. For the remaining points, the absolute percentage errors in the predictions made by the NPR model are relatively small, with most points having a percentage error of less than 5%. For the PCM, the error pattern is relatively evenly spread, although the percentage errors look relatively larger because the actual magnitude of the measured quantities is much smaller. For the PCM, the error spread is roughly between −11% and +12%, implying stable, although not uniformly accurate, predictive performance. There are positive as well as negative errors for the NPR and the PCM, implying that the predictive errors contain instances where the prediction either overestimated or underestimated the actual quantity. Although the −28.22% error in the NPR, classed as an outlier, contributes significantly to the overall error distribution, the remaining predictive results for the NPR imply that the model is functioning correctly for most operating conditions. The PCM predictions, although necessarily more affected by slight variations in numbers, are generally around a stable error level, implying stable performance at the various data points. From the above results, it is verified that the models generated are giving reliable predictions for both the NPR and PCM, and the differences in predictions are more accurate at determined conditions.

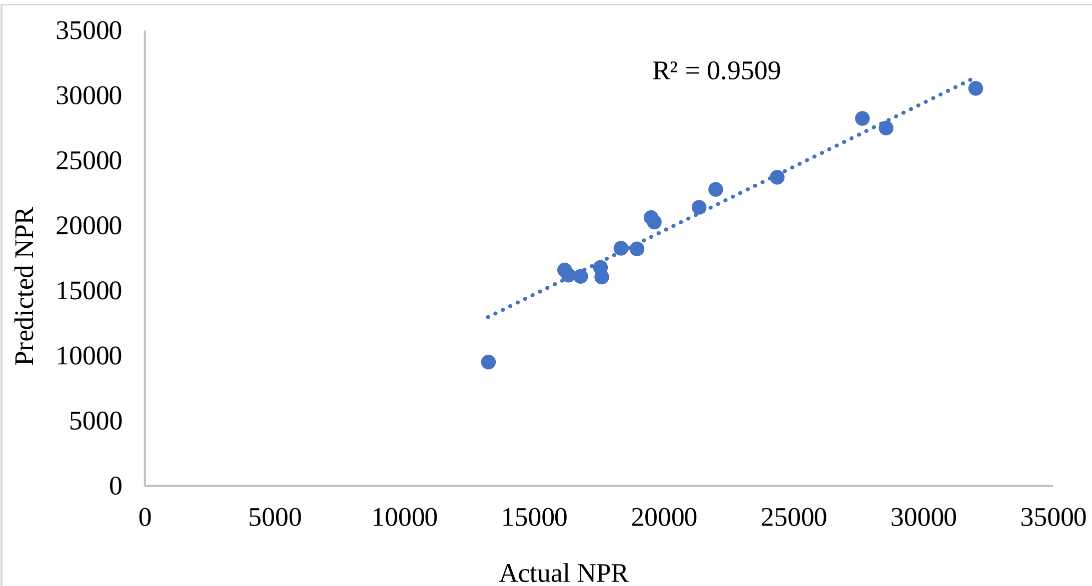


Fig. 3 Actual vs Predicted scatter Plot for NPR

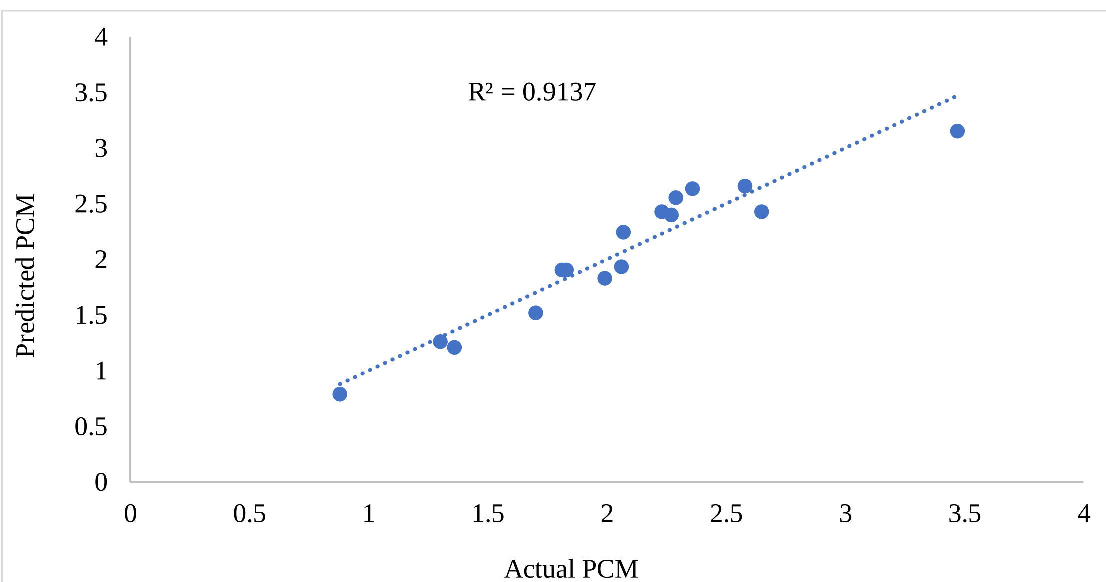


Fig. 4 Actual vs Predicted scatter Plot for PCM

10. Limitations of the Model

Models developed are based on data gathered in coal mines where the dirt bands are present in the coal seams. Dirt bands are known to have a significant influence on cutting and total machine productivity, and the presence of these bands in certain coal seams exerts a significant effect on the cutting process and the overall efficiency and productivity at which the machines work. Hence, these models may not apply to coal seams having fewer dirt bands, as the conditions are pretty different. Additionally, the models do not include operational aspects such as cutting direction, operation mode, loading practices, and production planning, among others. These operational factors exercise considerable impact on the productivity rates for surface miners, whereas they were handled indirectly instead of being handled directly in the current modelling approach.

11. Results and Discussion

The models of regression designed for calculating the Estimated Normalized Production Rate (NPR) can predict well and faithfully reproduce the behavior of the values for all points except one, which acted as an obvious outlier. In the case of the Pick Consumption per 1000 tonnes (PCM), the deviations of the predictions are more equally distributed but within a slightly broader range in percentage because of the low value of the PCM measurements. Additionally, it is easy to implement these models because of their linear form. Moreover, these models can be applied by practitioners in the field. Hence, the predictive accuracy obtained can be of great benefit to the production planning department, where it can be used as an approximation of production time and operational costs.

However, there is a good deal of potential for improvement within models through the usage of machine learning methods, particularly when the datasets include geotechnical conditions from wider regions. It can efficiently decode complex non-linear relations, along with the processing of categorical variables, unlike linear models.

Further, the current models include various key influencing parameters like Depth of cut, skill level of the operator, properties of intact and rock mass, machine characteristics, mine-specific habits of replacing picks, climatic conditions like rainfall, and operating method as constraints. They are not considered as predictors. Adding them to the models will enable improved predictive accuracy.

12. Conclusion

The study provides a complete overview of an experiment conducted to develop predictive models for the critical performance indicators of surface miners in coal mines. A total of 143 data observations were obtained for the locations shortlisted for this study. Out of these, 127 observations were used to train the models, and 16 observations were used for testing purposes.

Table 22. Models developed for the prediction of KPIs of surface miners

S. No	KPI	Predictive Model	R^2
1.	Normalized Production Rate	$NPR = 51413 - 591.7(UCS) - 82046(CAI) - 4.20(IVp)$	0.9617
2.	Pick Consumption per 1000 tonnes	$PCM = -2.554 + 25.949(CAI)$	0.9166

Finally, the developed models have been validated using a separate test dataset drawn from the original sample in a manner that ensured representation of all surface miner types employed at the selected sites.

However, large data sets of other parameters, including machine life and human-machine interface from different geological settings, may be included in future studies to create more generalized models.

Acknowledgements

The authors acknowledge the support and facilities provided by the Department of Mining Engineering, National Institute of Technology Raipur, India.

References

[1] Prakash Amar, Murthy Vemavarapu Mallika Sita Ramachandra, and Singh Kalendra Bahadur, "Rock Excavation using Surface Miners: An Overview of Some Design and Operational Aspects," *International Journal of Mining Science and Technology*, vol. 23, no. 1, pp. 33-40, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[2] A.K. Ghose, "New Technology for Surface Mining in the 21st Century-Emerging Role for Surface Miners," *Journal of Mines, Metals and Fuels*, vol. 56, no. 3, pp. 41-43, 2008. [Google Scholar] [Publisher Link]

[3] Chiara Origliasso, Marilena Cardu, and Vladislav Kecojevic, "Surface Miners: Evaluation of the Production Rate and Cutting Performance based on Rock Properties and Specific Energy," *Rock Mechanics and Rock Engineering*, vol. 47, no. 2, pp. 757-770, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[4] Prakash Amar, Murthy Vemavarapu Mallika Sita Ramachandra, and Singh Kalendra Bahadur, "Performance Simulaton of Surface Miners with Varied Machine Parameters and Rock Conditions: Some Investigations," *Journal of Geology and Mining Research*, vol. 5, no. 1, pp. 12-22, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[5] Lvinash Prasad, Sandeep Prasad, and G.K. Pradhan, "Performance Appraisal of Surface Miner in Opencast Mines: A Case Study," *The Indiln Mining & Engineering Journal*. [Google Scholar]

[6] Amar Prakash, and V.M.S.R. Murthy, "Hierarchy of Parameters Influencing Cutting Performance of Surface Miner through Artificial Intelligence and Statistical Methods," *Current Science*, vol. 112, no. 6, pp. 1242-1249, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[7] A. Singh et al., "Failure Analysis and Performance Improvement using Surface Miner on Field Breakdown Data: A Case Study," *Industrial Engineering Journal*, vol. 12, no. 9, pp. 1-31, 2019. [Google Scholar]

[8] Amar Prakash, and Vemavarapu M.S.R. Murthy, "Prediction of Cutting Speed of Surface Miner for Coal and Limestone Production under Varied Rock Mass Conditions," *Journal of the Geological Society of India*, vol. 100, no. 8, pp. 1122-1128, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Kaushik Dey, and A.K. Ghose, "Review of Cuttability Indices and A New Rockmass Classification Approach for Selection of Surface Miners," *Rock Mechanics and Rock Engineering*, vol. 44, no. 5, pp. 601-611, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[10] Yong Sun, Xing Sheng Li, and Wen Shao, "Influence of Cutting Parameters and Interactions on the Depth of Cut in Continuous Mining Operation," *Advanced Materials Research*, vol. 538-541, pp. 1422-1428, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[11] V. Raghavan, and Ch.S.N. Murthy, "Prediction of Cuttability from Rock Cutting Resistance," *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 118, no. 3, pp. 321-329, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[12] Wirtgen, Wirtgen Surface Miner, 2022. [Online]. Available: https://www.wirtgen-group.com/ocs/en-in/wirtgen/surface-miners-and-cross-application-miners-74-c/

[13] K. Dey, and P. Sen, "Continuous Surface Miner in Mining Limestone and Coal in India - A Comparative Study," *Journal of Mines, Metals and Fuels*, vol. 56, no. 3-4, pp. 51-54, 2008. [Google Scholar] [Publisher Link]

[14] Nilesh Pratap Singh et al., "Impact of Coal Quality on Longevity of Picks of Surface Miner-A Case Study," *International Conference on Deep Excavation, Energy Resources and Production*, 2017. [Google Scholar]

[15] Nilesh Pratap Singh et al., "An Investigation into Statistical Correlations between Coal Production and Key Productivity Indicators of Surface Miners in Indian Opencast Mines," *Mining, Metallurgy & Exploration*, vol. 40, no. 1, pp. 389-402, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[16] Amar Prakash, Vemavarapu Mallika Sita Ramachandra Murthy, and Kalendra Bahadur Singh, "A New Rock Cuttability Index for Predicting Key Performance Indicators of Surface Miners," *International Journal of Rock Mechanics and Mining Sciences*, vol. 77, pp. 339-347, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[17] Kaushik Dey, and Jayanta Bhattacharya, "Operation of Surface Miner: Retrospect of a Decade Journey in India," *Procedia Engineering*, vol. 46, pp. 97-104, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[18] Xiang Wang, and Okan Su, "Specific Energy Analysis of Rock Cutting based on Fracture Mechanics: A Case Study using a Conical Pick on Sandstone," *Engineering Fracture Mechanics*, vol. 213, pp. 197-205, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[19] Jian Yang et al., "Mechanical Property and Thermal Damage Factor of Limestone at High Temperature," *International Journal of Rock Mechanics and Mining Sciences*, vol. 117, pp. 11-19, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[20] C. Kumar et al., "Experimental Investigations on Thermal Behavior during Pick-Rock Interaction and Optimization of Operating Parameters of Surface Miner," *International Journal of Rock Mechanics and Mining Sciences*, vol. 133, pp. 1-29, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[21] Zong-Xian Zhang, and Finn Ouchterlony, "Energy Requirement for Rock Breakage in Laboratory Experiments and Engineering Operations: A Review," *Rock Mechanics and Rock Engineering*, vol. 55, no. 2, pp. 629-667, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[22] Amid Morshedlou, Jamal Rostami, and Omid Moradian, "Estimation of Cutting Forces Acting on Conical Cutters Based on Rock Properties and the Contact Area Between Cutter Tip and Rock," *Rock Mechanics and Rock Engineering*, vol. 57, no. 12, pp. 10307-10328, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[23] Guochao Zhao et al., "Experimental and Numerical Evaluation for Drum Dynamic Reliability under Extremely Complex Working Conditions," *Scientific Reports*, vol. 14, no. 1, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] V.M.S.R. Murthy et al., "Development of a Cuttability Index of Surface Miner for Performance Prediction in Different Geomining Conditions," *International Symposium on Rock Mechanics and Geo-Environment in Mining and Allied Industries*, Banaras Hindu University, Varanasi, pp. 12-14, 2009. [Google Scholar]