Journal of Advanced Research Design 144, Issue 1 (2026) 74-89

Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

Comparative analysis of Machine Learning Models for Predicting Particle Size Parameters from Drilling Data

Md. Aumio Tajrian[1,*], Shaolin Jahan Eidee[2], Md Ashfak Hossain Bhuiyan[1], MD Rakibul Islam Shanto[1], Faria Tasnia[1], A S Daiyan Ahmed[1]

[1] Department of Petroleum and Mining Engineering, Military Institute of Science and Technology, Dhaka, Bangladesh
[2] Department of Chemical Engineering, Oklahoma State University, Stillwater, Oklahoma, USA

ARTICLE INFO

Keywords:

Machine learning in drilling; particle size prediction; ROP model; drilling operational parameters; MSE evaluation

ABSTRACT

The efficiency of drilling operations is determined by numerous aspects, including the particle size of the material being drilled. To achieve efficiency, drilling engineers must take into consideration the size, shape, and density of the cuttings generated during the drilling process. Ineffective drilling can result in increasing expenses and delays for projects involving the extraction of natural resources. The objective of this study is to enhance drilling efficiency by investigating the correlation between drilling parameters such as weight on bit, revolutions per minute, torque, and rate of penetration and features of particle size distribution such as mean particle size and coarseness index as well as mechanical specific energy (MSE). The influence of particle size on drilling has been evaluated through the application of machine learning techniques and comprehensive datasets. The study highlighted relationships between particle size characteristics and the effectiveness of drilling, offering valuable insights into the optimal particle size for tonalite formations that are bordered by mica gneiss. Three machine learning techniques were employed to determine the closest relationship between drilling characteristics and particle size, with the Random Forest approach exhibiting the strongest correlation. This technique may be employed to forecast the size attributes of particles for data points that are not available within the usual range of drilling parameters. This work successfully emphasizes the significance of particle size in drilling operations and showcases the practical use of machine learning in enhanced drilling efficiency.

1. Introduction

Shale formations and other unconventional reserves are explored and produced via large diameter drilling. Copper and gold are extracted from underground mines using large diameter drilling. Drilling parameters are the numerous aspects of the drilling process that are managed and monitored for the best possible results. Drilling operational variables include the bit size and type, revolutions per minute (RPM), weight on bit (WOB), and properties of the drilling fluid. Optimizing drilling parameters requires a balance between achieving the desired drilling objectives and minimizing the wear and tear on equipment. Drilling efficiency and efficacy are both affected by

[*] *Corresponding author*
E-mail address: aumitajrian3@gmail.com

particle size distribution, such as mean particle size and making it a crucial metric. Particle size distribution in drilling is the variation in particle size caused by the cutting action of the drill bit [1]. The effectiveness of drilling can be influenced by the size and shape of the particles produced. Increased particle size might result in abrasion of the drilling bit and limit drilling velocity. Ensuring the size of particles is carefully observed and managed is essential for optimizing efficiency and reducing problems such as the wearing down of bits and the blockage of wellbores [2].

1.1 Background of Study

Drilling is a prevalent technique employed to extract various resources, including oil, gas, and minerals such as coal, copper, gold, iron, silver, and zinc. The petroleum and mining sectors extensively employ large diameter drilling for purposes such as exploration, production, and mineral extraction. With the rising need for oil, gas, and minerals, there is an expectation for the ongoing use of large-diameter drilling. Monitoring and managing drilling parameters, such as the rate of penetration (ROP), is crucial during the drilling process. ROP quantifies the velocity at which the drill bit enters the geological formation and can be determined using several methods. The publications of Warren *et al.,* [3]; Detournay and Defourny *et al.,* [4]; Hareland and Rampersad *et al.,* [5]; Graham and Muench *et al.,* [6]; Maurer *et al.,* [7]; Bingham *et al.,* [8]; Young *et al.,* [9]; Bourgoyne and Young [10] have been included for comparison with data-driven models or because of their significance in the industry. An outline of the development of ROP models and drilling optimization up to 2010 can be obtained from the Doctorate thesis of Eren and Ozbayoglu [11]. Because ROP is expressed primarily as a function of both WoB and rotational speed. Some early ROP models are also represented as R-W-N (ROP, WoB, Rotary Speed) [12]. The Maurer model serves as a model, assuming complete bit tooth penetration and flawless bottom-hole cleaning. For rolling cutter bits, Maurer *et al.,* [7] established the following Eq. (1):

$$ROP = \frac{k}{s^2}\left[\frac{W}{db} - \frac{W0}{db}\right]^2 N \tag{1}$$

Where *K* is the constant of proportionality, *S* is compressive rock strength, *W* is the weight of bit, *db* is the drill bit parameter, N is the rotary speed.

Bingham suggested another R-W-N:

$$ROP = K\left(\frac{W}{db}\right)^{a5} N \tag{2}$$

Where a^5 is the weight on bit exponent and *K* is the constant of proportionality, taking into account the influence of rock strength [13]. One of the most significant ROP models was created by Bourgoyne and Young [12] in 1974 and is used by the sector. The Bourogyne and Young model (BYM), according to Soares and Gray [14] has eight parameters and is expressed as follows:

$$\frac{dD}{dt} = \exp(a1 + \sum_{j=2}^{8} ajxj) \tag{3}$$

Where *D* is the well depth, *t* is the time, the coefficient *a1* is related to the formation strength parameter, *a2* is the formation compaction, *a3* to the pore pressure, *a4* to differential pressure, *a5* to the WOB exponent, *a6* to rotary drilling (N), *a7* to drill-bit tooth wear, *a8* to the bit hydraulic jet impact. Subsequently, Bourogyne *et al.,* [14] suggested the following modification to their initial ROP model:

$$ROP = (f1) * (f2) * \dots * (f8) \tag{4}$$

where the empirical coefficients *a1* through *a8* are included in *f1* through *f8*. Soares and Gray [14] state that the final function is the primary distinction between the two formulations. The revised version of the BYM uses a power law function of the hydraulic jet impact force [10,13]. All of the significant components of drilling are represented by the BYM equations; however, some model parameters, such as drill bit wear and differential pressure, are not computed in real time [14]. A general drag bit model was proposed by Hareland and Rampersad [5] based on cutter rock interaction:

$$ROP = \frac{14.14 * Nc * N * Av}{db} \tag{5}$$

The area of compressed rock ahead of the cutter, or Av, is determined by the type of drill bit used and varies depending on the number of cutters (Nc) [5]. As it is already indicated, there is a great deal of complexity and ignorance regarding the true link between the drilling factors [15]. Consequently, an attempt has been made to gain a better knowledge of the drilling variables and how they impact the ROP [16-18]. The reason that specifically improved the accuracy of the theoretical model. Al-Abduljabbar *et al.,* [18] presented a new model for ROP that was created by regression analysis.

$$ROP = 16.96\ \frac{W^{a} * N * T * SSP * Q}{db^{2} * \rho * PV * UCS^{b}} \tag{6}$$

Where ρ is the mud density, *T* is the torque, *SSP* is the standpipe pressure, *Q* is the flow rate, *PV* is the plastic viscosity, and *UCS* is the uniaxial compressive strength. The unit conversion factor employed by the authors is 16.96. Using non-linear regression, the coefficients (*a* and *b*) were determined. The quantity of energy needed to drill through a unit volume of rock is known as mechanical specific energy (MSE). Usually, it is stated as energy per unit volume, such as foot-pounds per cubic inch (ft-lb/in3) or joules per cubic centimeter (J/cm3). Different definitions of MSE exist, depending on the context in which it is intended to be applied [19-22]. MSE is displayed by:

$$MSE = \frac{WOB}{Ab} + \frac{120 * \pi * RPM * TOB}{Ab * ROP} \tag{7}$$

Where *WOB* is Weight on Bit, *RPM* is the rotary Speed, *TOB* is the torque on bit, A_b is the bit area, (in^2). Additionally, Teale observed that the crushing strength of the drilled medium is correlated with the minimum amount of specific energy (SE) [22]. To describe torque as a function of WOB on the mean squared error (MSE) correlation, Pessier and Fear [23] developed a bit specific coefficient of sliding friction to express torque as a function of WOB on the MSE correlation, as follows:

$$MSE = \frac{WOB}{Ab} + \frac{13.33 * \mu * RPM * WOB}{Db * ROP} \tag{8}$$

For field applications, μ is usually assumed to be equal to 0.25 for tricone bits, and 0.5 for PDC bits. Dupriest and Koederitz [20] assumed that drilling efficiency remains at 35%, independent of bit type or WOB, based on field data.

$$MSE = 0.35 * (\frac{WOB}{Ab} + \frac{120* \pi*RPM*TOB}{Ab*ROP}) \quad (9)$$

Rabia *et al.,* [24] provided the following simplified explanation of the bit selection-SE correlation:

$$SE = \frac{20*WOB*RPM}{Db*ROP} \quad (10)$$

Among of these models, the Maurer *et al.,* [7] model has been employed in this work to determine the ROP. The term "mean particle size" denotes the average dimension of particles or cuttings generated during the process of drilling. The drilling rate is dependent upon the specific rock being drilled and the drilling settings. The bit generates cuttings by crushing or fracturing rock particles. The dimensions and configuration of these cuttings give crucial insights into the formation being drilled. Techniques such as laser diffraction and visual inspection can be employed to ascertain the average particle size. Microscopy and imaging methods may be employed to examine the morphology and surface characteristics of drill cuttings, providing more comprehensive insights.

$$d = 5 * 10^{-8} * (CI)^{2.755} \quad (11)$$

Where, *d*= Mean Particle Size(mm), *CI* = Coarseness Index. Coarseness Index (CI) is a parameter used in drilling to quantify the distribution of particle sizes in the cuttings generated by the drilling process [25]. CI parameter is calculated by using the following formula:

$$CI = \left(\frac{K}{SE}\right)^{\frac{1}{n}} \quad (12)$$

For *CCS* disk cutter, $K = 2*10^{15}$, $n= 5.5$. Increased hole cleaning efficiency can be achieved by decreasing and micronizing the cutting size [26]. The generated cuttings increase with the penetration rate, and a greater amount can be moved toward the output due to the annulus space limitation. The likelihood of various mechanical pipes sticking rises in such situations [27]. Many academics have sought to predict and optimize the ROP in order to increase drilling efficiency. The project typically involves drilling input parameters such bit characteristics, drilling fluid qualities, WOB, and RPM. Recent advancements in machine learning algorithms have created new opportunities for ROP optimization and general drilling efficiency gains. According to Mitchell and Miska [15], the ROP is now an open-ended inquiry used in drilling engineering to comprehend the impact of drilling factors. The ROP was optimized using several supervised model types [28]. These investigations were limited to examining the drilling variables. Recent research on drilling optimization using factors, such as MSE, has demonstrated that proper optimization of the drilling operation as a whole cannot be achieved by depending just on drilling parameters [29]. Therefore, in order to improve the overall drilling efficiency, consideration must be given to the particle size distribution and MSE. Mud loggers have utilized cuttings to construct lithology columns during drilling operations. Reservoir cuttings aid in comprehending the characteristics of porosity and permeability [42]. The efficiency and prevention of difficulties with real-time drilling depend on the analysis of cutting. Accurate sampling, measurement, and analysis of cuttings helps prevent problems and improve the efficiency of drilling operations [2]. Despite the fact that researchers have emphasized the significance of particle size analysis on different occasions, they have not addressed the impact of particle size characteristics on other drilling parameters. The purpose of this work is to examine how the ROP relates to particle characteristics like mean particle size and CI and parameters such as MSE.

Three distinct supervised machine learning models have been employed to forecast and ascertain the closely related features of drilling and particle parameters. Out of all the methods, random forest demonstrates the most consistent connection among these features. The investigation required excavating a substantial diameter hole to extract a narrow gold vein from under the surface. The data were gathered from the Research Tunnel, situated at a depth of 60 meters in the VLJ repository. The Olkiluoto nuclear power plant utilizes this subterranean repository for the containment of low- and intermediate-level waste. The tunnel was constructed using the conventional drill-and-blast technique and consists of gneissic tonalite and pegmatite rock formations. The tonalite is composed of quartz, plagioclase, biotite, and hornblende minerals. The investigation also included the excavation of three complete deposition holes utilizing an innovative full face boring technique. The holes had a diameter of 1.527 meters and a depth of 7.5 meters. The tedious procedure entailed the utilization of rotating rock crushing and subsequent suction of the pulverized rock. The Subterranean-OSL-137 raise drilling machine was employed for the purpose of rotational crushing [30].

2. Application of Supervised Machine Learning Models

Supervised learning is a subfield of machine learning and artificial intelligence. It involves the utilization of labeled datasets to instruct algorithms in making predictions or classifying data. Weights of the model are modified during the cross-validation procedure, with the aim of progressively enhancing the model's performance using the training set [41]. Supervised learning is applicable to classification and regression issues, since it enables the examination of the correlation between dependent and independent variables [31].

Linear regression is a statistical technique used for predicting the relationship between a dependent variable and one or more independent variables. The line of best fit is determined using the method of least squares. Simple linear regression entails the use of a single independent variable, whereas multiple linear regression involves the utilization of numerous independent variables.

$$y = a0 + a1x + \varepsilon \tag{13}$$

Where, *Y*= Dependent Variable (Target Variable), *X*= Independent Variable (predictor Variable), *a0*= intercept of the line (Gives an additional degree of freedom), *a1* = Linear regression coefficient (scale factor to each input value), *ε* = random error. The values for the x and y variables are training datasets for Linear Regression model representation [32].

A popular machine learning technique, random forest is a combination of the results of numerous decision trees. Its adaptability and simplicity have led to its widespread usage, particularly since it can solve both classification and regression issues. The predictor space is divided in the set of possible values for X1,X2, . . . ,Xp — into J distinct and non-overlapping regions, R1,R2, . . . ,RJ . For every observation that falls into the region Rj , to make the same prediction. To find boxes R1, . . ., Rj that minimize the RSS, given by

$$\sum_{j=1}^{j} \sum (yi - \hat{y}Rj)^2 \tag{14}$$

Where $\hat{y}Rj$ is the mean response for the training observations within the j^{th} box [32]. In greater detail, for any *j* and *s*, the pair of half-planes can be defined by: R1(j, s) = {X|Xj < s} and R2(j, s) = {X|Xj ≥ s} and the value of *j* and *s* that minimize the equation can be found by the following equation:

$$\sum_{i:xi\in R1(j,s)}(yi-\hat{y}\,R1)^2+\sum_{i:xi\in R2(j,s)}(yi-\hat{y}\,R2)^2 \tag{15}$$

where $\hat{y}R1$ is the mean response for the training observations in $R1(j, s)$, and $\hat{y}R2$ is the mean response for the training observations in $R2(j, s)$. Pruning for each value of a there corresponds a subtree $T \subset T0$ [32].

$$\sum_{m=1}^{|T|}\sum_{i:x\in Rm}(yi-\hat{y}Rm)^2+\ \alpha|T| \tag{16}$$

Here $|T|$ indicates the number of terminal nodes of the tree T, Rm is the rectangle (i.e., the subset of predictor space) corresponding to the m^{th} terminal node, and Rm is the predicted response associated with Rm—that is, the mean of the training observations in $\hat{y}Rm$ [33].

XGBoost is a machine learning algorithm that is used for supervised learning tasks such as regression, classification, and ranking. It is based on a gradient-boosting framework that uses decision trees as base models. Unlike fitting a single large decision tree to the data, which amounts to fitting the data hard and potentially overfitting, the boosting approach instead learns slowly.[33] The XGBoost regression can be built by [33]:

i. Setting $\widehat{f(x)}=0$ and $r_i = y_i$ for all i in the training set.

ii. For b = 1, 2, . . ., B,
repeating:
(a) A tree $\hat{f}^b$ will be fit with d splits (d+1 terminal nodes) to the training data (X, r).
(b) $\hat{f}$ will be updated by adding in a shrunken version of the new tree:

$$\widehat{f(x)}\leftarrow\widehat{f(x)}+\ \lambda\hat{f}^b(x) \tag{17}$$

(c) The residuals will be updated by,

$$ri\leftarrow ri-\lambda\hat{f}^b(xi) \tag{18}$$

iii. Output of the boosted model,

$$\widehat{f(x)}=\sum_{b=1}^{B}\lambda\,\widehat{f(x)}^b \tag{19}$$

A regression model can only forecast values that are more than or less than the actual value. Therefore, residuals are the sole method to assess the model's correctness. Residuals represent the discrepancy between observed and expected values. The residuals may be conceptualized as a distance. Therefore, the closer the residual gets to zero, the more accurately the model predicts.

$$ei=yi-y^i \tag{20}$$

In the equations above, e_i =represents the residual value, Y_i =represents the true value, y^i= represents the expected value [33]. The most important assessment measures for regression issues include R^2 Score, Mean Absolute Error (MAE), Mean squared error (MSE), Mean Square Root Error (RMSE). The R^2 score is utilized to assess the model's distance or residual accuracy. R2 score may be computed using the following formula:

$$R^2=1-\frac{RSS}{TSS} \tag{21}$$

$$RSS=\sum(yi-\hat{yi})^2 \tag{22}$$

$$TSS = \sum(yi - \bar{y})^2 \tag{23}$$

Where y_i is the actual value, $\hat{yi}$ is the predicted value, y_i is the actual value and $\bar{y}$ is the mean value of the variable feature [34]

$$Mean\ Absolute\ Error = \left(\frac{1}{N}\right) * \sum_{i=1}^{N} |yi - \hat{y}| \tag{24}$$

Where, ∑=Greek sign for summing, y_i=Observation of I^{th} actual value, $\hat{y}$=Calculated value for observation number I^{th} Total number of occurrences [35].

MSE can be determined using the following equation [36]:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (yi - \hat{y})^2 \tag{25}$$

RMSE can be determined by the following equation [35]:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (yi - \hat{y})^2} \tag{26}$$

3. Methodology

The data for this investigation is collected from the Research Tunnel, located at a depth of 60 meters in the VLJ repository. Prior to use in machine learning models, the data must undergo preprocessing to enhance its effectiveness. Data pre-processing is an essential step in preparing data for machine learning algorithms. The objective is to transform unprocessed data into a well-organized and standardized format that can be efficiently utilized by machine learning algorithms. Data preparation encompasses the subsequent stages: Data cleaning, data transformation, feature selection, data scaling, and data splitting. Furthermore, data visualization serves the purpose of enhancing understanding of data, promoting the identification of patterns, and facilitating the transmission of discoveries. Subsequently, a quantitative metric is employed to evaluate the degree to which a model can effectively forecast results for new data. The process is referred to as "Model Fitting". For the purpose of this investigation, three separate supervised machine learning models were employed to make predictions. Prediction in machine learning refers to the outcome produced by an algorithm that has been trained using a dataset. It provides estimated values for unknown variables in new data inputs. The models' accuracy has been evaluated using many criteria, including R2 score, MSE, MAE, and RMSE. The random forest model demonstrated superior performance, as seen in Figure 1 depending on its level of accuracy. Subsequently, the data variable is incremented by employing the "Range Function". The augmented data variables are subsequently utilized for further prediction.

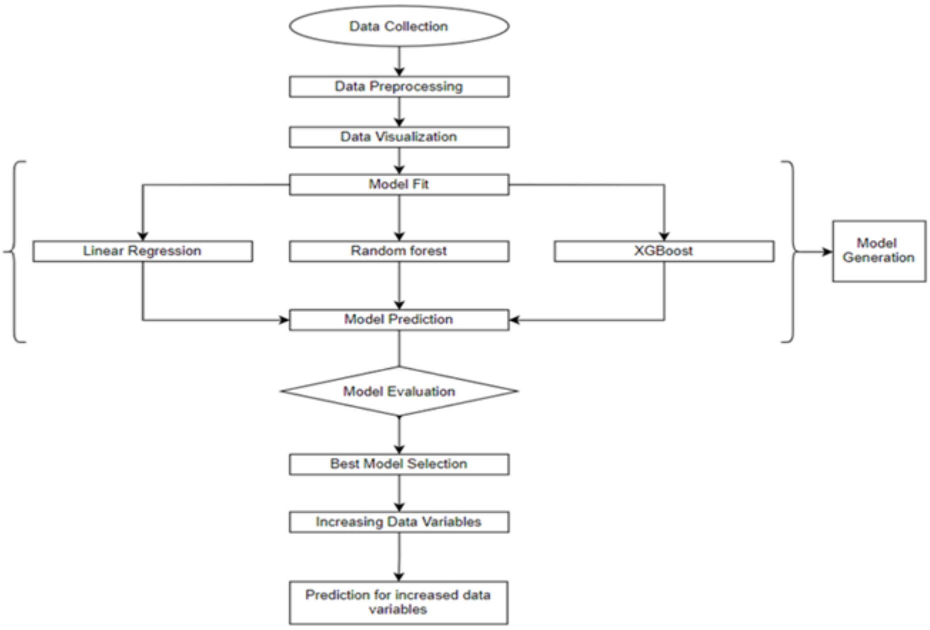

Fig. 1. Flowchart showing methodology for predicting particle parameters from drilling parameters

4. Results

The following Figure 2 demonstrates the correlation between various drilling parameters and particle parameters. The correlation between ROP and Mean Particle size is 0.95, indicating a strong positive relationship. This correlation is visually represented by the colour yellow.

Fig. 2. Data visualization of the relationship between different drilling parameters by heatmap

The CI also has a value of 0.94. Conversely, the correlation between ROP and SE is shown by a purple value of -0.89. A negative value signifies a negative connection or association between the variables. In this scenario, there is an inverse relationship between the two variables, where an increase in one variable corresponds to a reduction in the other variable.

4.1 Relationship between ROP and Mean Particle Size

The accompanying diagram in Figure 3 illustrates the connection between the mean particle size and the ROP, where the Y-axis represents the dependent parameter and the X-axis represents the independent variable. ROP and Mean Particle size are shown to have a proportional relationship in this illustration. This supports the statement by the research work from Altindag *et al.,* [27].

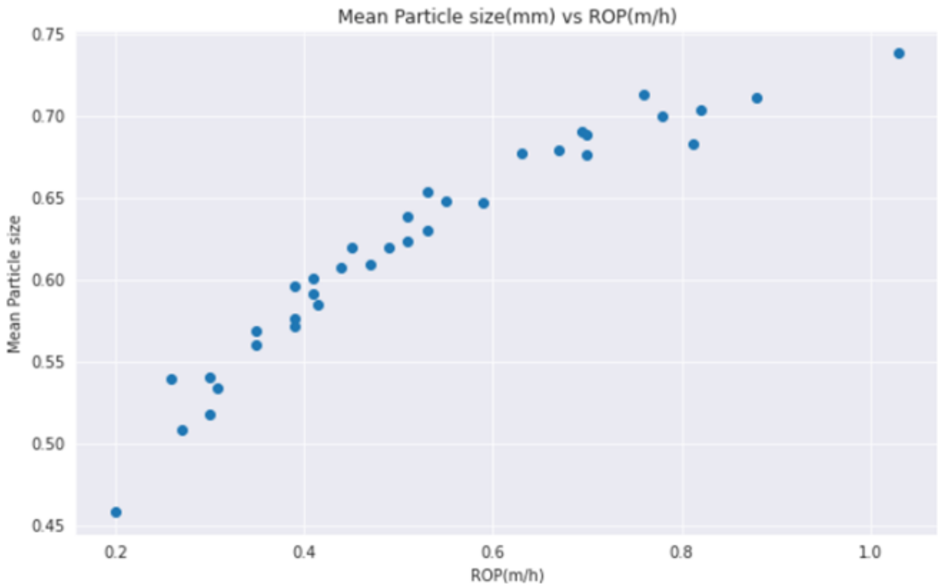

Fig. 3. Visualization of relationship between ROP and Mean particle size

4.2 Relationship between ROP and Coarseness Index (CI)

The graph shows how the CI and ROP are related, with the dependent parameter (Y-axis) and the independent variable (X-axis) being shown. This result replicates the proportional link between ROP and the CI from the research work of Kumar *et al.,* [37].

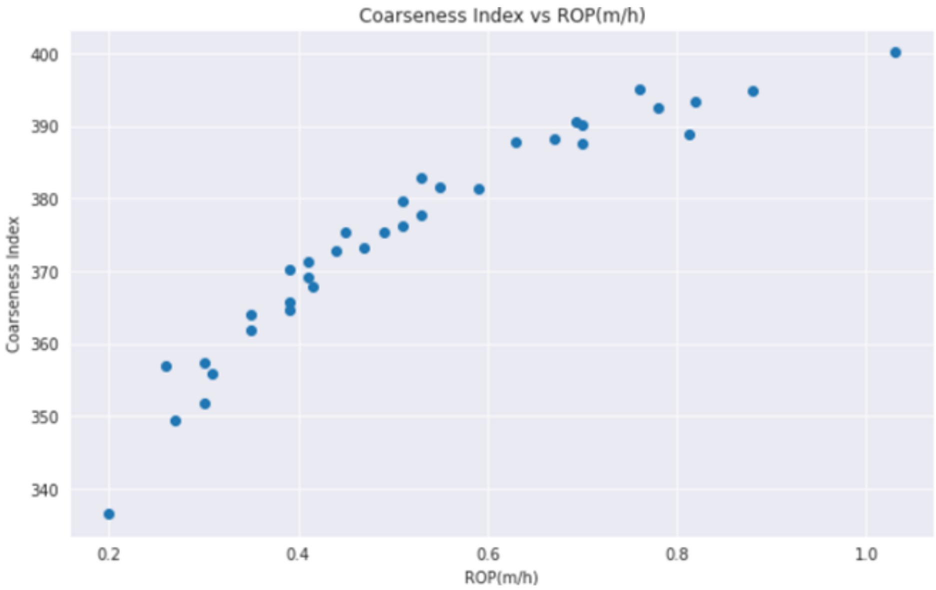

Fig. 4. Visualization of relationship between ROP and coarseness index

4.3 Relationship between Mean Particle Size and Specific Energy (SE)

The graph depicts the relationship involving mean particle size and SE, with the dependent parameter (Y-axis) and independent variable (X-axis) shown. This figure illustrates the inverse proportional relationship between the mean particle size and SE. This supports the work from Mohammadi *et al.,* [38] and Kim *et al.,* [39].

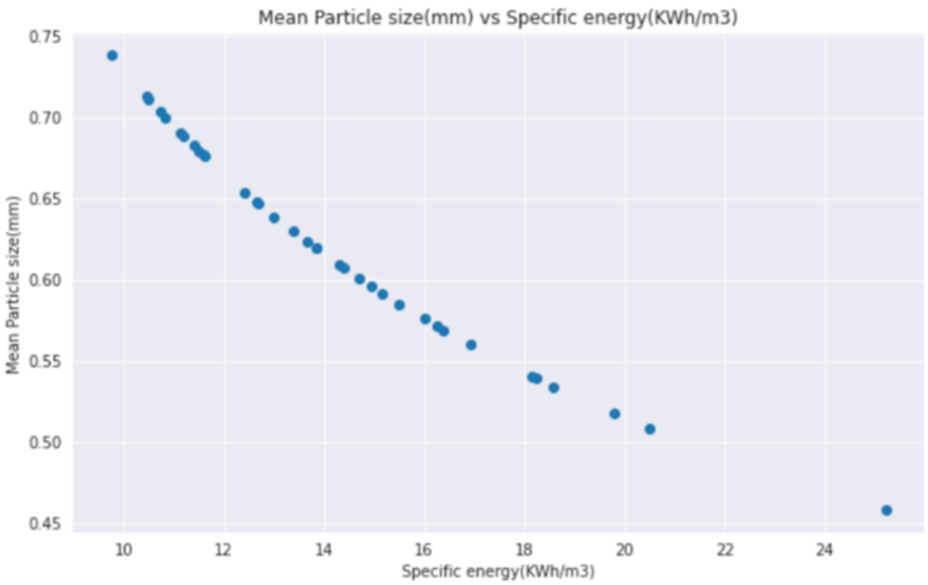

Fig. 5. Visualization of relationship between mean particle size and specific energy

4.4 Relationship between Coarseness Index (CI) and Specific Energy (SE)

The graph illustrates the relationship between the CI and the SE, with the Y-axis representing the dependent variable and the X-axis representing the independent variable. The graph demonstrates an inverse relationship between the CI and the SE. This conclusion is supported by the correlation seen between the specific energy and CI in a study conducted by Tuncdemir *et al.,* [28] on Kartal

Limestone using constant cross-section disc cutter tests in unrelieved cutting mode. Additionally, this supports the work from Mohammadi *et al.,* [38], Kim *et al.,* [39] and Abu Bakar and Gertsch [40].

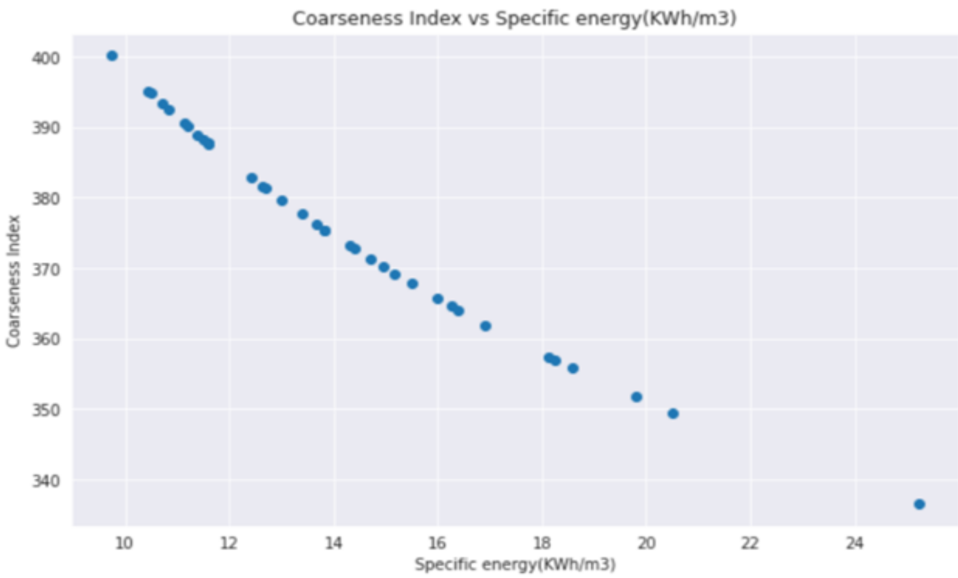

Fig. 6. Visualization of relationship between coarseness index and specific energy

4.5 Prediction of mean particle size from ROP

Actual and expected mean particle size are shown on the Y axis, while ROP is plotted on the X axis in the graph (Figure 7). The blue dots in the graph represent the estimated particle size. In contrast, the orange dots represent the actual mean particle size values. The predicted values are values absent from the training dataset.

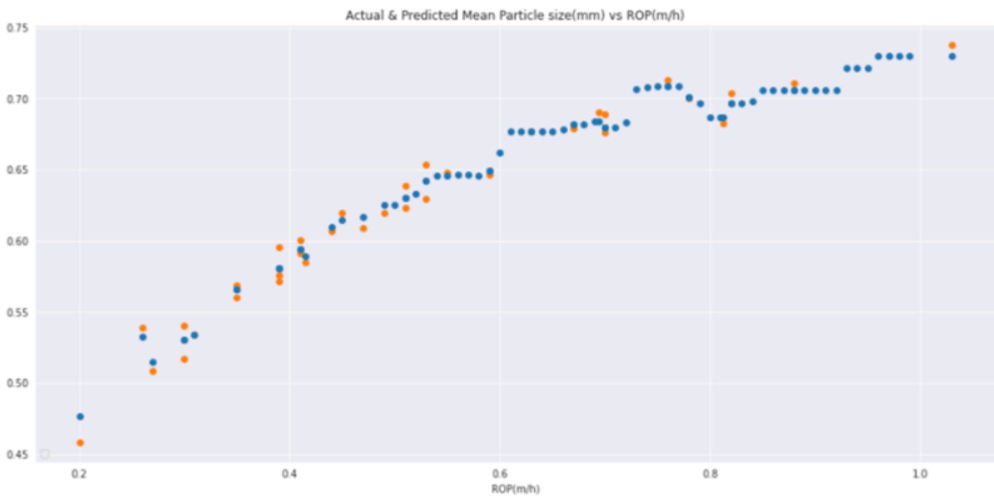

Fig. 7. Visualization of comparison between actual and predicted mean particle size from ROP

4.6 Prediction of Coarseness Index (CI) from ROP

Actual and predicted CI are displayed on the Y axis, while ROP is presented on the X axis in the Figure 8. The graph's blue points indicate the estimated CI. The orange dots, in comparison, indicate the real CI values. The anticipated values do not exist in the training dataset.

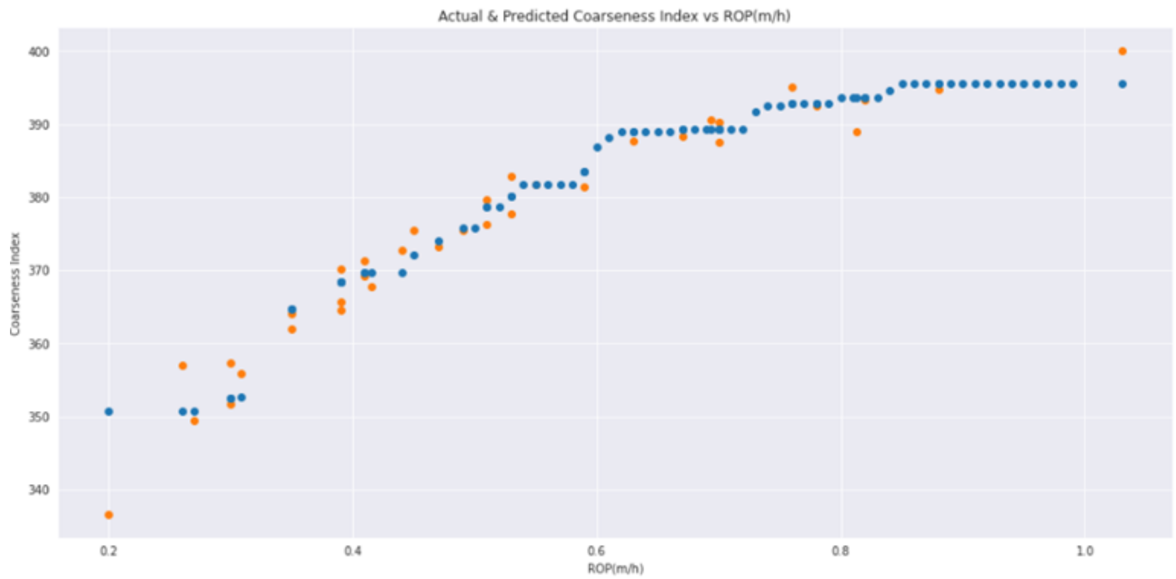

Fig. 8. Visualization of comparison between actual and predicted coarseness index from ROP

4.7 Prediction of Mean Particle Size from Specific Energy (SE)

The Y-axis depicts the distinction between measured and expected mean particle size, while the X-axis shows the SE in this Figure 9. The estimated actual and expected mean particle size is shown in green on the graph. For comparison, the orange dots represent the actual SE levels. There are no instances of the expected values in the dataset used for training.

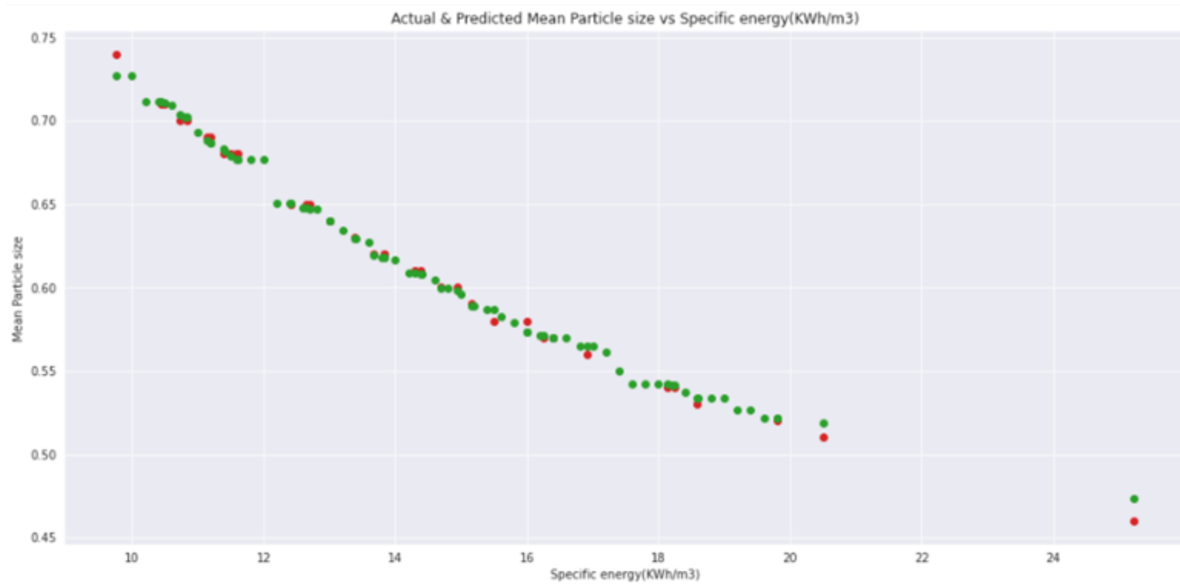

Fig. 9. Visualization of comparison between actual and predicted mean particle size from specific energy

4.8 Prediction of Mean Coarseness Index (CI) from Specific Energy (SE)

Variance in CI between actual and predicted values is shown against SE on the y-axis. In green, the graph depicts the assessed actual and predicted CI. The CI shown by the orange dots is for purposes of comparison. Training data lacks occurrences of the target values.

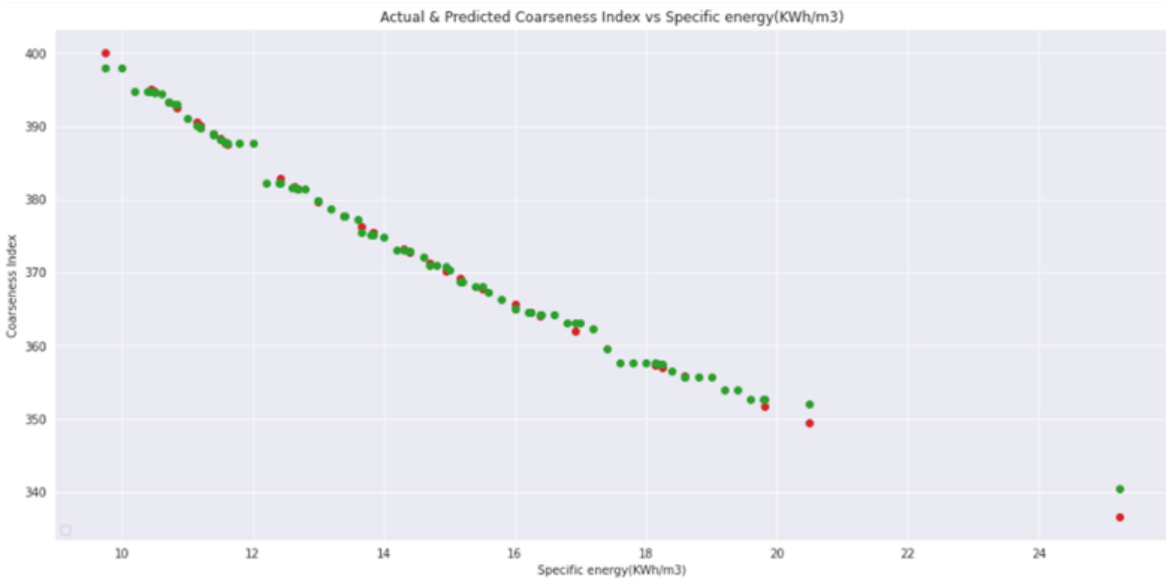

Fig. 10. Visualization of comparison between actual and predicted coarseness index from specific energy

The best model is selected based on the above evaluation parameters. The evaluation shows as follows:

Here, it is highlighted that, among the three models, the random forest has the best accuracy. The previously described parameter *R* Squared demonstrates the strongest link between ROP and Mean Particle Size, with a score of 0.94, and between ROP and CI, with a score of 0.97. The MSE, MAE, and RMSE indicate the mistakes of the prediction model. Among these models, random forest has the lowest score, indicating that it is capable of preventing "Over Fitting" and "Under Fitting." This demonstrates that random forest is the most appropriate model for forecasting particle size parameters based on drilling parameters (Table 1).

Table 1
Comparison of different machine learning techniques for different drilling parameters

Parameters	Linear Regression				Random Forest				XGBoost			
	R^2	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE
ROP vs Mean Particle Size	0.84	0.08	0.01	0.28	0.94	0.013	0.0003	0.11	0.92	0.017	0.0004	0.02
ROP vs CI	0.8	0.09	0.01	0.3	0.97	2.15	6.01	1.47	0.91	4.38	11.97	4.69

5. Conclusion

Simple linear regression can be used to establish the relationship between drilling parameters and particle size characteristics. The correlation is highly noteworthy due to its demonstration of a proportional and inverse-proportional relationship between these variables. In order to estimate different parameters in the field, these linkages can be utilized to establish links between the parameters. Additionally, it is beneficial for creating drilling configurations that optimize drilling productivity. For example, the intercept and co-efficient of the ROP vs Mean particle size linear regression is 0.452 and 0.321 respectively.

Both Random Forest and XGBoost demonstrate a strong correlation (*R* squared>=0.9), but the random forest regression findings suggest a more intimate relationship between the variables.

Predictive modelling techniques, such as random forest, can be used to accurately estimate particle size based on drilling parameters. An *R2* score close to 1 indicates that a significant percentage of the variability in the dependent variable can be accounted for by the independent variables in the model. Put simply, the model effectively corresponds to the data and successfully captures a substantial portion of the variability in the target variable. The metric quantifies the extent to which the independent variables accurately anticipate the fluctuations in the dependent variable. This illustrates the effective utilization of random forest for prediction in this dataset. Due to its enhanced precision and decreased margin of error, it is more probable to be employed in practical applications for forecasting.

This study illustrates the substantial significance of particle properties in optimizing ROP. Particle size has emerged as a crucial variable to examine due to its detrimental effect on drill bit performance and drilling efficiency. The heatmap demonstrates a robust link between the average particle size and the cutting depth. Increasing the depth of cut can result in a higher ROP, hence enhancing overall drilling efficiency. The conveyance of particles to the surface is significantly affected by the size of the particles, which can improve the efficiency of drilling operations.

The practical use of predicting particle size characteristics from drilling data is when its usefulness becomes evident. The practical implementation of such studies can demonstrate numerous possible benefits, including but not limited to the accuracy of the drilling can be assessed using the projected data. By employing a machine learning strategy, the field-based extraction process may be made more efficient, resulting in increased productivity and reduced costs. An identical approach will be available for comparable geological conditions.

There are specific constraints to the work. The work has been founded on the Maurer *et al.,* [7] ROP model. The current model is outdated, while the new models are more versatile and consider a wider range of variables to optimize ROP. Furthermore, there has been minimal study conducted, resulting in limited opportunities for adoption in specialized fields.

Acknowledgments

The authors acknowledge Novamera Inc. for providing the test data. The authors are also grateful to Mr. Daiyan Ahmed for his constant guidance and support throughout the whole research work. This research was funded by Research and Development Wing and Department of Petroleum and Mining Engineering, Military Institute of Science and Technology, Mirpur Cantonment, Dhaka-1216, Bangladesh

References

[1] Ahmed, D., Y. Xiao, J. de Moura, and S. D. Butt. "INVESTIGATION OF HOLE WIDENING DRILLING USING CUTTING ANALYSIS."

[2] Karimi, Moji. "Drill-cuttings analysis for real-time problem diagnosis and drilling performance optimization." In *SPE Asia Pacific oil and gas conference and exhibition*, pp. SPE-165919. SPE, 2013. https://doi.org/10.2118/165919MS

[3] Warren, T. M. "Penetration-rate performance of roller-cone bits." *SPE Drilling Engineering* 2, no. 01 (1987): 9-18. https://doi.org/10.2118/13259-pa

[4] Detournay, Emmanuel, and Paul Defourny. "A phenomenological model for the drilling action of drag bits." In *International journal of rock mechanics and mining sciences & geomechanics abstracts*, vol. 29, no. 1, pp. 13-23. Pergamon, 1992. https://doi.org/10.1016/0148-9062(92)91041-3

[5] Hareland, Geir, and Paul R. Rampersad. "Drag-bit model including wear." In *SPE Latin America and Caribbean Petroleum Engineering Conference*, pp. SPE-26957. SPE, 1994. https://doi.org/10.2118/26957-MS

[6] Graham, J. W., and N. L. Muench. "Analytical determination of optimum bit weight and rotary speed combinations." In *SPE Annual Technical Conference and Exhibition?*, pp. SPE-1349. SPE, 1959. https://doi.org/10.2118/1349-g

[7] Maurer, W. C. "The" perfect-cleaning" theory of rotary drilling." *Journal of Petroleum Technology* 14, no. 11 (1962): 1270-1274. https://doi.org/10.2118/408-pa

[8] Bingham, M. Grant. "A new approach to interpreting--rock drillability." (1965). https://cir.nii.ac.jp/crid/1130282272567413504
[9] Young Jr, F. S. "Computerized drilling control." *Journal of Petroleum Technology* 21, no. 04 (1969): 483-496. https://doi.org/10.2118/2241-pa
[10] Bourgoyne Jr, Adam T., and F. S. Young Jr. "A multiple regression approach to optimal drilling and abnormal pressure detection." *Society of Petroleum Engineers Journal* 14, no. 04 (1974): 371-384. https://doi.org/10.2118/4238-pa
[11] Eren, Tuna, and M. Evren Ozbayoglu. "Real time optimization of drilling parameters during drilling operations." In *SPE Oil and Gas India Conference and Exhibition?*, pp. SPE-129126. SPE, 2010. https://doi.org/10.2118/129126-MS
[12] Nascimento, Andreas, David Tamas Kutas, Asad Elmgerbi, Gerhard Thonhauser, and Mauro Hugo Mathias. "Mathematical modeling applied to drilling engineering: an application of Bourgoyne and Young ROP model to a presalt case study." *Mathematical Problems in Engineering* 2015 (2015). https://doi.org/10.1155/2015/631290
[13] Adam, T. Bourgoynejr, K. K. Millheim, Martin E. Chenevert, and F. S. Young. "Applied drilling engineering." *SPE EXTBOOK SERIES* 2 (1991).
[14] Soares, Cesar, and Kenneth Gray. "Real-time predictive capabilities of analytical and machine learning rate of penetration (ROP) models." *Journal of Petroleum Science and Engineering* 172 (2019): 934-959. https://doi.org/10.1016/j.petrol.2018.08.083
[15] Mitchell, Robert F., and Stefan Miska. "Fundamentals of drilling engineering." (2011). https://doi.org/10.2118/9781555632076
[16] Motahhari, Hamed Reza, G. Hareland, and J. A. James. "Improved drilling efficiency technique using integrated PDM and PDC bit parameters." *Journal of Canadian Petroleum Technology* 49, no. 10 (2010): 45-52. https://doi.org/10.2118/141651-pa
[17] Deng, Yong, Mian Chen, Yan Jin, Yakun Zhang, Daiwu Zou, and Yunhu Lu. "Theoretical and experimental study on the penetration rate for roller cone bits based on the rock dynamic strength and drilling parameters." *Journal of natural gas science and engineering* 36 (2016): 117-123. https://doi.org/10.1016/j.jngse.2016.10.019
[18] Al-AbdulJabbar, Ahmad, Salaheldin Elkatatny, Mohamed Mahmoud, Khaled Abdelgawad, and Abdulaziz Al-Majed. "A robust rate of penetration model for carbonate formation." *Journal of Energy Resources Technology* 141, no. 4 (2019): 042903. https://doi.org/10.1115/1.4041840
[19] Armenta, Miguel. "Identifying inefficient drilling conditions using drilling-specific energy." In *SPE Annual Technical Conference and Exhibition?*, pp. SPE-116667. SPE, 2008. https://doi.org/10.2118/116667-ms
[20] Dupriest, Fred E., and William L. Koederitz. "Maximizing drill rates with real-time surveillance of mechanical specific energy." In *SPE/IADC Drilling Conference and Exhibition*, pp. SPE-92194. SPE, 2005. https://doi.org/10.2118/92194-ms
[21] Chevreau, Marc-Aurèle. "New formulation of mechanical specific energy (MSE) taking into account the hydraulic effects for PDC bits." *PhD diss*., 2017.
[22] Teale, Robert. "The concept of specific energy in rock drilling." In *International journal of rock mechanics and mining sciences & geomechanics abstracts*, vol. 2, no. 1, pp. 57-73. Pergamon, 1965. https://doi.org/10.1016/0148-9062(65)90022-7
[23] Pessier, R. C., and M. J. Fear. "Quantifying common drilling problems with mechanical specific energy and a bit-specific coefficient of sliding friction." In *SPE Annual Technical Conference and Exhibition?*, pp. SPE-24584. Spe, 1992. https://doi.org/10.2118/24584-ms
[24] Rabia, Hussain. "Specific energy as a criterion for bit selection." *Journal of petroleum technology* 37, no. 07 (1985): 1225-1229. https://doi.org/10.2118/12355-pa
[25] Altindag, Raşİt. "Evaluation of drill cuttings in prediction of penetration rate by using coarseness index and mean particle size in percussive drilling." *Geotechnical & Geological Engineering* 22 (2004): 417-425. https://doi.org/10.1023/b:gege.0000025043.92979.48
[26] Tuncdemir, H., N. Bilgin, H. Copur, and C. Balci. "Control of rock cutting efficiency by muck size." *International Journal of Rock Mechanics and Mining Sciences* 45, no. 2 (2008): 278-288. https://doi.org/10.1016/j.ijrmms.2007.04.010
[27] Dehvedar, M., P. Moarefvand, A. R. Kiyani, and A. R. Mansouri. "Using an experimental drilling simulator to study operational parameters in drilled-cutting transport efficiency." *Journal of Mining and Environment* 10, no. 2 (2019): 417-428. https://doi.org/10.22044/jme.2019.7665.1624
[28] Hegde, Chiranth, Scott Wallace, and Ken Gray. "Using trees, bagging, and random forests to predict rate of penetration during drilling." In *SPE Middle East Intelligent Oil and Gas Symposium*, p. D011S001R003. SPE, 2015. https://doi.org/10.2118/176792-MS
[29] Hegde, Chiranth, and Ken Gray. "Evaluation of coupled machine learning models for drilling optimization." *Journal of Natural Gas Science and Engineering* 56 (2018): 397-407. https://doi.org/10.1016/j.jngse.2018.06.006

[30] Autio, Jorma, and Timo Kirkkomäki. *Boring of full scale deposition holes using a novel dry blind boring method*. No. SKB-TR--96-21. Swedish Nuclear Fuel and Waste Management Co., Stockholm (Sweden), 1996.
[31] Nasteski, Vladimir. "An overview of the supervised machine learning methods." *Horizons*. b 4, no. 51-62 (2017): 56. https://doi.org/10.20544/horizons.b.04.1.17.p05
[32] Hope, Thomas MH. "Linear regression." In *Machine Learning*, pp. 67-81. Academic Press, 2020. https://doi.org/10.1016/B978-0-12-815739-8.00004-3
[33] James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*. Vol. 112, no. 1. New York: springer, 2013. https://doi.org/10.1007/978-1-4614-7138-7_1
[34] Colton, James A., and Keith M. Bower. "Some misconceptions about R2." *International Society of Six Sigma Professionals, EXTRAOrdinary Sense* 3, no. 2 (2002): 20-22.
[35] Willmott, Cort J., and Kenji Matsuura. "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance." *Climate research* 30, no. 1 (2005): 79-82. https://doi.org/10.3354/cr030079
[36] Schluchter, Mark D. "Mean square error." *Encyclopedia of Biostatistics* 5 (2005). https://doi.org/10.1002/0470011815.b2a15087
[37] Kumar, Rakesh, V. M. S. R. Murthy, and L. A. Kumaraswamidhas. "Performance analysis of rotary blast-hole drills through machine vibration and coarseness index mapping–a novel approach." *Measurement* 165 (2020): 108148. https://doi.org/10.1016/j.measurement.2020.108148
[38] Mohammadi, Mehdi, Jafar Khademi Hamidi, Jamal Rostami, and Kamran Goshtasbi. "A closer look into chip shape/size and efficiency of rock cutting with a simple chisel pick: a laboratory scale investigation." *Rock Mechanics and Rock Engineering* 53 (2020): 1375-1392. https://doi.org/10.1007/s00603-019-01984-5
[39] Kim, Han-eol, Kyoung-min Nam, Hafeezur Rehman, Tae-su Kyeon, and Han-kyu Yoo. "Comparison Study on Coarseness Index and Maximum Diameter of Rock Fragments by Linear Cutting Tests." *Applied Sciences* 12, no. 17 (2022): 8793. https://doi.org/10.3390/app12178793
[40] Bakar, Muhammad Zubair Abu, and Leslie Sour Gertsch. "Evaluation of saturation effects on drag pick cutting of a brittle sandstone from full scale linear cutting tests." *Tunnelling and Underground Space Technology* 34 (2013): 124-134. https://doi.org/10.1016/j.tust.2012.11.009
[41] Eidee, Shaolin Jahan, Dionne Mayibeki, Geir Hareland, and Mohammed F. Al Dushaishi. "Hard Rock Drilling Efficiency Mapping Using Machine Learning." In *GRC Transactions*, vol. 48, 2024. Paper presented at the Geothermal Rising Conference, Hawaii, October 2024.
[42] Mannafi, A.S.M., Md Tauhidur Rahman, K. M. Haidarul Alam, Md Niamul Quader, Khairul Habib, and Tamanna Zafrin Orin. 2024. "Porosity and Density Determination from Well Log Data: Machine Learning and Simulation Approaches". *Journal of Advanced Research in Applied Sciences and Engineering Technology* 51 (1):116-28. https://doi.org/10.37934/araset.51.1.116128