Rock Mechanics Bulletin 4 (2025) 100173

Contents lists available at ScienceDirect

Rock Mechanics Bulletin

journal homepage: www.keaipublishing.com/en/journals/rock-mechanics-bulletin

Full Length Article

Prediction of rock fracture pressure in hydraulic fracturing with interpretable machine learning and mechanical specific energy theory

Xiaoying Zhuang [a,b,*], Yuhang Liu [a], Yuwen Hu [a], Hongwei Guo [b], Binh Huy Nguyen [c]

[a] *Department of Geotechnical Engineering, Tongji University, Shanghai, 200092, China*
[b] *Computational Science and Simulation Technology, Department of Mathematics and Physics Leibniz Universita t Hannover, Hannover, Germany*
[c] *Sensor and Actuator Technologies, imec, Leuven, Belgium*

ARTICLE INFO

Keywords:
Hydraulic fracturing
Mechanical specific energy
Nonlinear regression
Interpretable machine learning
Fracture pressure

ABSTRACT

Hydraulic fracturing stimulation technology is essential in the oil and gas industry. However, current techniques for predicting rock fracture pressure in hydraulic fracturing face significant challenges in precision and reliability. Traditional approaches often result in inadequate accuracy due to the complex and diverse nature of underground formations. However, recent advances in computational power and optimization techniques have enabled the application of machine learning in mining operations, resulting in improved prediction and feedback. In this study, various machine learning techniques are employed to predict hydraulic fracturing pressure based on the concept of mechanical specific energy. Additionally, the study interprets the models through feature importance analysis. The findings suggest that most machine learning models deliver highly accurate predictions. Feature importance analysis indicates that for an approximate assessment of fracture pressure, the characteristics of well depth and torque are sufficient. For more precise predictions, incorporating additional characteristics from the mechanical specific energy framework into the machine learning model is essential. The study emphasizes the feasibility of employing machine learning methods to predict fracture pressure and their usefulness in determining optimal engineering sites.

1. Introduction

The effectiveness of hydraulic fracturing stimulation is crucial in oil and gas extraction. However, the variability of underground conditions often leads to inconsistent outcomes in sweet spot cluster modifications during real-world operations. The competition between fractures in clusters results in both ‘over-reformed’ and ‘under-reformed’ perforation clusters, which hinder oil and gas production and significantly increase extraction costs. The complex and varied flow of oil and gas, coupled with dynamic changes in physical properties during hydraulic fracturing, complicates the understanding of the underlying physical and chemical mechanisms. Consequently, precise mathematical models and numerical simulation tools for fracturing design are lacking. Additionally, the significant variability in the distribution of subsurface oil and gas resources necessitates accurate prediction and assessment of compressibility at specific locations.

The advent of computer science and enhanced data accessibility has catalyzed the application of machine learning for data mining and nonlinear physical pattern recognition. The continuous evolution of machine learning algorithms has led to the development of various regression algorithms tailored to diverse scenarios and sample sizes. Artificial neural networks have emerged as a pivotal branch of machine learning for pattern recognition and data regression. Compared to traditional experimental and theoretical methods, ML techniques can predict fracture pressure in hydraulically fractured reservoirs at lower costs and in shorter timeframes through their prediction mechanisms. Models such as XGB, RF, GPR, and AdaBoost are being more widely adopted due to their robustness, prediction accuracy, and capability to handle non-linearity and flexibility in evaluating properties and datasets. The performance and accuracy of machine learning methods can be enhanced through the utilization of datasets derived from actual engineering projects. This becomes particularly crucial when considering that data obtained from laboratory settings may not precisely reflect field conditions (Salami et al., 2022; Ababneh et al., 2020; Inqiad et al., 2023).

* Corresponding author. Computational Science and Simulation Technology, Department of Mathematics and Physics Leibniz Universita t Hannover, Hannover, Germany.
E-mail address: zhuang@iop.uni-hannover.de (X. Zhuang).

Peer review under the responsibility of Chinese Society for Rock Mechanics & Engineering.

https://doi.org/10.1016/j.rockmb.2024.100173
Received 25 August 2024; Received in revised form 10 November 2024; Accepted 14 December 2024
Available online 30 December 2024

Furthermore, this study employs a feature selection strategy for machine learning models (Cheng et al., 2024; Chu and Wang, 2024; Huang et al., 2006; Ibrahim et al., 2024; Ions et al., 2024) that is fundamentally based on parameters derived from Mechanical Specific Energy (MSE) theory. MSE theory is commonly employed to describe drill bit performance. By collecting and processing bit data during hydraulic fracturing, real-time drilling performance evaluation is facilitated. Utilizing drill bit data to reflect fracturing effects mitigates challenges arising from uneven bottom layer distribution and energy source variability in fracturing effect analysis. Currently, MSE theory is widely applied in monitoring, prediction, optimization design, drilling method evaluation, and bottom rock mechanical properties assessment during drilling (Teale, 1965; Rafatian et al., 2010; Minghui et al., 2016; Chen et al., 2018; Wiśniowski et al., 2015; Rashidi et al., 2010). Additionally, machine learning models are essentially black boxes, making it difficult for observers to understand the underlying mechanisms and processes behind their predictions. To enhance transparency and increase the credibility of prediction results, interpretable frameworks have been introduced. In recent years, machine learning has been applied in hydraulic fracturing projects. (Fasola, 2023; Yu et al., 2024; Chen et al., 2024; Bolton et al., 2024; Zheng et al., 2024; Lin et al., 2024).

This paper investigates and compares the effectiveness of twenty machine learning models driven by drilling parameters base on MSE theory to predict fracture pressure. Additionally, sensitivity and correlation analysis methods are employed to provide local explanations for the best-performing models, making the machine learning models more transparent and the results more convincing. The models analyzed include Gaussian Process Regression, Decision Trees, and Support Vector Machines, among others. These models aim to quantify the response of different drill bits to fracture pressure. The predictive results and their explanations will guide the selection of sweet spots in fracture construction.

2. Feature selection and model building

2.1. Feature selection based on mechanical specific energy theory

MSE theory is typically defined as the mechanical work done to excavate a unit volume of formation. It is widely used to quantify drilling efficiency and maximize the rate of penetration. Drilling parameters can characterize the mechanical properties relevant to hydraulic fracturing. The most commonly used MSE model, proposed by Teale (1965), is based on extensive experiments with various drill bits and rock types:

$$MSE = \frac{WOB}{A_b} + \frac{120\pi \cdot RPM \cdot TOR}{A_b \cdot ROP} \tag{1}$$

where MSE is the mechanical specific energy (J/m^3), WOB is the weight on bit (N), A_b is the drill bit area (m^2), RPM is the turntable speed (rad·s^{-1}), TOR is the bit torque (N·m) and ROP is the mechanical drilling speed (m/s).

The concept of MSE provides a framework for feature selection. Due to the variability in reservoir geology and the difficulty in obtaining continuous formation parameters, drilling parameters are chosen to characterize formation parameters and establish the relationship between drilling parameters and fracture pressure. Based on the principle of MSE and the actual working conditions of the project, well depth, drill bit diameter, torque, ROP, weight on bit, and rotation speed are selected as input features for predicting fracture pressure. The MSE value itself is not used as a feature variable because MSE theory is derived from empirical formulas with artificially assigned coefficients, which vary across different MSE theories and can lead to deviations from actual conditions. To avoid errors caused by human factors, drill bit response is directly used as the characteristic value, and the calculated MSE is not used as a feature.

Due to the large variation in each parameter and the significant difference in magnitude between features, regression processes that calculate Euclidean distances can produce significant errors in predicted results. Therefore, it is essential to use feature scaling methods, such as the minimum-maximum normalization as:

$$x_i^* = (b-a)\frac{x_i - x_{\min}}{x_{\max} - x_{\min}} + a \tag{2}$$

to standardize the range of independent features. Where x_i is the initial feature variable, x_i^* is the standardized feature variable, $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of the initial feature values, respectively. The parameters a and b represent the lower and upper limits of the normalized feature values.

2.2. Evaluation metrics and cross-validation

In this paper, the accuracy of the regression prediction model is evaluated using the coefficient of determination (R^2) and the root mean square error ($RMSE$). A coefficient of determination closer to 1 and a smaller $RMSE$ value indicate a more accurate prediction.

$$R^2 = 1 - \frac{\sum_i (y_i' - y_i)^2}{\sum_i (\bar{y} - y_i)^2} \tag{4}$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i' - y_i)^2} \tag{5}$$

where, y_i is the true value output in the sample data, y_i' is the predicted value of the model, and $\bar{y}$ is the mean value of the true values.

Fig. 1. Schematic diagram of 10-fold-cross-validation principle (remake from scikit-learn).

Table 1
Comparison of comprehensive indicators of twenty regressors for fracture pressure prediction.

Model	*T*(s)	R^2	*RMSE*
Extra Trees	0.011	0.946	0.102
Voting	0.002	0.944	0.103
HGB	0.002	0.944	0.103
Random Forest	0.052	0.941	0.106
KRR	0.001	0.941	0.106
GPR	0.001	0.941	0.106
poly svm	≤0.001	0.940	0.107
GB	0.001	0.937	0.110
rbf svm	0.001	0.935	0.112
ELM	0.001	0.935	0.112
Ada Boost	0.013	0.932	0.114
Stacking	0.002	0.930	0.115
Gen-ELM	0.001	0.927	0.118
Kernel SGD	0.001	0.925	0.120
Bagging	0.004	0.916	0.127
SGD	≤0.001	0.914	0.128
KNN distance	0.001	0.911	0.130
NN	0.002	0.911	0.130
KR	0.002	0.908	0.133
Decision tree	≤0.001	0.876	0.154

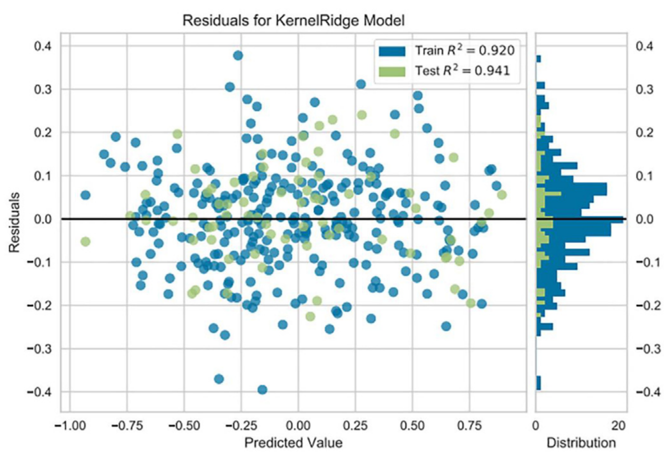

Fig. 2. Residuals for KRR

Cross-validation, also known as rotation estimation, is commonly used to evaluate machine learning models and prevent overfitting, particularly with limited datasets. The most widely used type is K-fold cross-validation, as illustrated in Fig. 1. In this paper, 10-fold cross-validation is employed to calculate accuracy.

3. Machine learning performance analysis

The performance of the machine learning models listed in Section 2.2, supported by the Scikit-learn library, was evaluated using 10-fold cross-validation. This fracture pressure prediction utilized 537 sets of data from the same oil well, including well depth, drill bit pressure, drill bit torque, drill bit size, drilling speed, rotational speed, and fracture pressure. By fixing the random seed and dividing the dataset, 80% of the data was randomly assigned as the training set and 20% as the test set.

To conduct a comparative analysis and select the most suitable machine learning regression algorithm for this problem, the calculation metrics of different models have been listed in the following Table 1, *T*(s) represents the algorithm's running time, R^2 represents the coefficient of determination, and *RMSE* represents the root mean square error.

It can be observed from the table above that the R^2 of all models exceeds 0.85 and *RMSE* is below 0.15 demonstrating the high feasibility of using machine learning methods to predict fracture pressure.

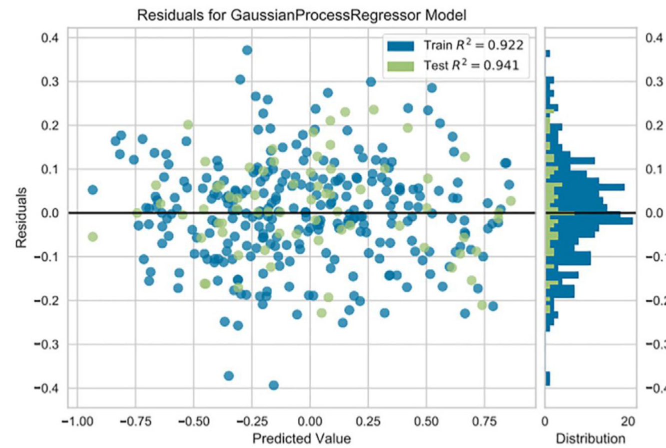

Fig. 3. Residuals for GPR

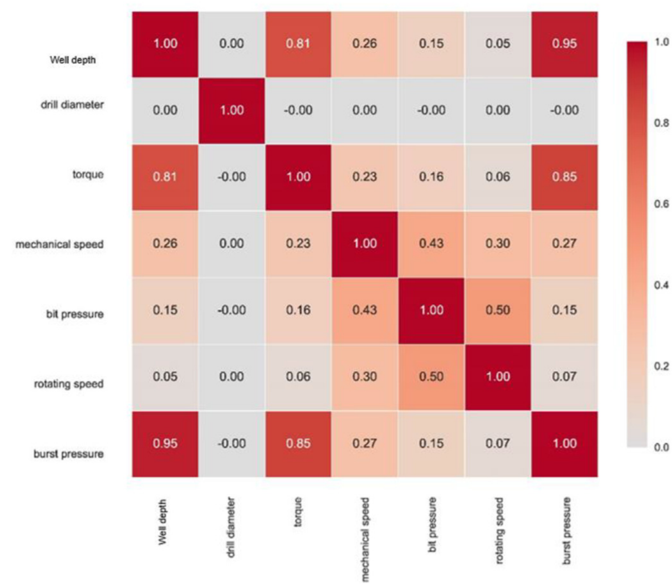

Fig. 4. Heat map of correlation between coefficients.

Additionally, the drill bit response data directly reflects hydraulic fracturing conditions. These machine features are easy to read and have high data quality, contributing to the successful implementation of this regression prediction method. This indicates that using machine learning to predict fracture pressure is both feasible and efficient.

In the column for running time, the 20 regression algorithms with the shortest running times are marked. The six algorithms with the highest accuracy are highlighted, and the six algorithms with the smallest error are marked based on *RMSE*. A comprehensive analysis of these results indicates that, apart from decision trees, most machine learning models demonstrate good regression performance. Kernel Ridge Regression (Nguyen et al., 2020) and Gaussian Process Regression (Swiler et al., 2020) exhibit the lowest model complexity and computational cost while maintaining high accuracy. Additionally, an analysis of the residual plots shown in Figs. 2 and 3 reveals that the prediction results of the KRR and GPR methods are accurate and stable, with few instances of large errors. Therefore, these two methods are considered the most suitable machine learning regression algorithms for predicting fracture pressure using drill bit response data.

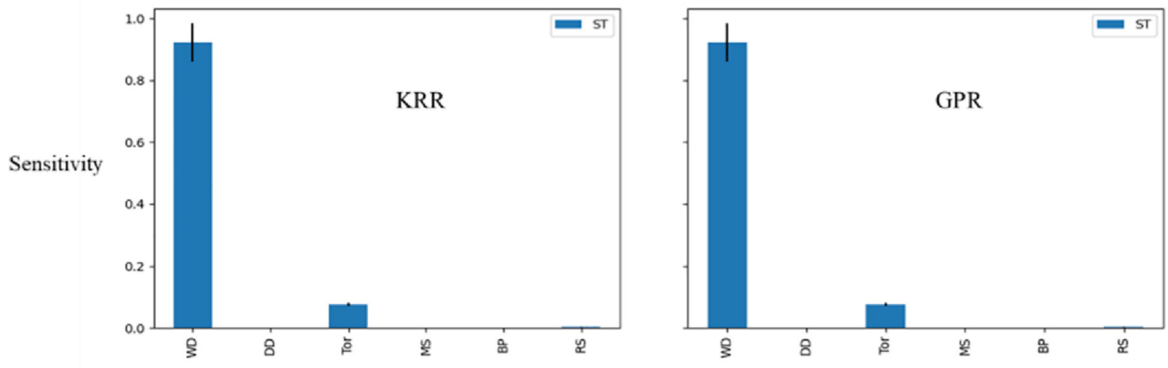

Fig. 5. The sensitivity analysis of the KRR and GPR.

4. Interpretable machine learning

The machine learning algorithms discussed have demonstrated high accuracy in predicting fracture pressure. However, assessing whether these complex predictive processes align with actual conditions can be challenging. To interpret the machine learning models and analyze whether their predictive processes adhere to inherent physical laws and empirical observations, the impact and significance of various features were investigated on the predictive outcomes. Therefore, based on the framework of mechanical specific energy and in conjunction with the two optimal models selected in the previous section, this paper conducts sensitivity analysis and correlation analysis to explore the importance of features, thereby explaining the machine learning models.

A correlation analysis was conducted on the dataset by calculating the Pearson correlation coefficient, as shown in Eq. (6). Although the Pearson correlation coefficient cannot fully capture nonlinear relationships between variables, it numerically characterizes the correlation between two datasets and provides a statistical indicator of their closeness. Figure 4 displays the Pearson correlation coefficients between features. The selected drill bit characteristics are used to predict fracture pressure, and the resulting variable heat map is shown below.

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{6}$$

In the above, x_i and y_i contain n sets of corresponding variables, $\overline{x}$ and $\overline{y}$ are the average values of the two sets of variables. r represents the correlation coefficient of the two sets of variables.

Sensitivity analysis is a method used to evaluate how changes in various factors affect the outcomes of a research project (Saltelli, 2004). In this paper, sensitivity analysis is employed to verify and quantify the influence of changes in input features on fracture pressure. The results of the sensitivity analysis for the KRR and GPR models are presented below in Fig. 5. The Hilbert-Schmidt Independence Criterion (HSIC)-based global sensitivity analysis was employed, where ST represents the total sensitivity index of variables, the sampling range spans from the minimum to maximum values in the dataset, with a sampling size of 512*(6 * 2 + 2). The independent variables include well depth, borehole diameter, torque, feed rate, rotation speed, and drilling pressure, while the dependent variable is fracture pressure. Furthermore, the sensitivity analysis results are similar for both models. As shown in the figure, WD exhibits the highest sensitivity, followed by TOR, while the sensitivity of other features is negligibly small.

However, using only these two variables for prediction yields an R-squared (R^2) value of around 85%, which does not exceed the R^2 value of approximately 95% shown in Table 1. Additionally, the correlation analysis in Fig. 4 reveals that while well depth and torque are most correlated with fracture pressure, other variables also exhibit a linear relationship with fracture pressure. Therefore, it is crucial to consider the combined contribution of all relevant features, despite their smaller individual impacts, to achieve an R-squared (R^2) value close to 1. According to the physical significance of the MSE theoretical formula, it can roughly describe the physical properties of rocks, while burial depth mainly affects the stress state of rocks. For deep-buried rocks' fracture pressure prediction, the stress state plays a dominant role, whereas the physical properties of rocks have relatively less influence. However, to achieve more accurate predictions, the physical properties of rocks cannot be neglected. This underscores the importance and validity of selecting characteristic variables based on the principle of mechanical specific energy and employing machine learning for fracture pressure prediction.

5. Conclusion

In this study, bit characteristics (selected based on MSE theory) and well depth were employed as indicators to reflect both rock physical properties and rock stress state to predict the fracture pressure. And the performance of twenty machine learning models was calculated and compared for fracture pressure prediction. The fracturing pressure prediction method provides guidance and support for fracturing sweet spot selection and fracturing feedback design.

Based on the framework of mechanical specific energy theory and actual working conditions, drill bit parameters related to fracture pressure, such as well depth, drill bit diameter, torque, ROP, WOB, and rotational speed, were selected as input features. Additionally, twenty machine learning algorithms, including Kernel Ridge Regression, Gaussian Processes, Decision Trees, Random Forest, and others, were employed to construct a robust fracture pressure prediction model. With the exception of the Decision Tree, the predictive performance of the other machine learning models was strong, demonstrating the high feasibility of using machine learning methods to predict fracture pressure. Among these methods, different algorithms exhibited distinct characteristics: Kernel Ridge Regression showed excellent performance in handling nonlinear relationships but was sensitive to outliers; ensemble learning demonstrated good generalization ability and noise resistance but had relatively weak model interpretability; decision trees, while offering good interpretability, showed relatively poor predictive performance. Based on the final results of this study, Gaussian Process Regression proved to be the optimal choice, featuring low computational cost, high computational accuracy, and the ability to demonstrate uncertainty. Through sensitivity and correlation analyses, the importance of features was determined, revealing that well depth and torque contribute most significantly to fracture pressure prediction, while the combination of other features provides a smaller contribution. For a rough estimate of fracture pressure, a model can be established using only well depth and torque. However, for more accurate predictions, it is necessary to combine other features within the mechanical specific energy framework. This approach validates the correctness and necessity of selecting feature variables based on mechanical specific energy principles and using machine learning to predict fracture pressure. This method can guide the selection of fracturing sweet spots, real-time fracturing process monitoring, and fracturing feedback design.

CRediT authorship contribution statement

Xiaoying Zhuang: Software, Resources, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Yuhang Liu:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yuwen Hu:** Project administration, Methodology, Investigation, Funding acquisition, Formal analysis. **Hongwei Guo:** Visualization, Supervision, Resources, Methodology, Formal analysis. **Binh Huy Nguyen:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ababneh, A., Alhassan, M., Abu-Haifa, M., 2020. Predicting the contribution of recycled aggregate concrete to the shear capacity of beams without transverse reinforcement using artiffcial neural networks. Case Stud. Constr. Mater. 13. https://doi.org/10.1016/j.cscm.2020.e00414.

Bolton, W.J., Wilson, R., Gilchrist, M., Georgiou, P., Holmes, A., Rawson, T.M., 2024. Personalising intravenous to oral antibiotic switch decision making through fair interpretable machine learning. Nat. Commun. 15 (1), 506.

Chen, X., Yang, J., Gao, D., 2018. Drilling performance optimization based on mechanical specific energy technologies. Drilling 1 (1), 133–162.

Chen, Y., Calabrese, R., Martin-Barragan, B., 2024. Interpretable machine learning for imbalanced credit scoring datasets. Eur. J. Oper. Res. 312 (1), 357–372.

Cheng, T.S., Lucchi, A., Kratsios, A., Dokmanić, I., Belius, D., 2024. A theoretical analysis of the test error of finite-rank kernel ridge regression. Adv. Neural Inf. Process. Syst. 36.

Chu, X., Wang, J., 2024. Distribution system state estimation based on enhanced Kernel Ridge Regression and ensemble empirical mode decomposition. Processes 12 (4), 823.

Fasola, S.L., Brudzinski, M.R., 2023. Machine learning reveals additional hydraulic fracture-induced seismicity in the eagle ford shale. J. Geophys. Res. Solid Earth 128 (2), e2022JB025436.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2006. Extreme learning machine: theory and applications. Neurocomputing 70 (1–3), 489–501.

Ibrahim, B., Tetteh-Asare, A., Ahenkorah, I., 2024. A transparent and valid framework for rockburst assessment: unifying interpretable machine learning and conformal prediction. Rock Mech. Rock Eng. 1–15.

Inqiad, W.B., Siddique, M.S., Alarifi, S.S., Butt, M.J., Najeh, T., Gamil, Y., 2023. Comparative analysis of various machine learning algorithms to predict 28-day compressive strength of Self-compacting concrete. Heliyon 9 (11).

Ions, K., Rahat, A., Reeve, D.E., Karunarathna, H., 2024. Gaussian process regression approach for predicting wave attenuation through rigid vegetation. Appl. Ocean Res. 145, 103935.

Lin, S., Dong, M., Liang, Z., Guo, H., Zheng, H., 2024. Image-based 3D reconstruction and permeability modelling of rock using enhanced interpretable deep residual learning. Eng. Anal. Bound. Elem. 160, 187–200.

Minghui, W., Gensheng, L., Huaizhong, S., Shuaishuai, S., Zhaokun, L., Yi, Z., 2016. Theories and applications of pulsed-jet drilling with mechanical specific energy. SPE Journal 21 (01), 303–310.

Nguyen, T., Chen, Z., Lee, J., 2020. Dataset meta-learning from kernel ridge-regression. arXiv preprint arXiv:2011.00050.

Rafatian, N., Miska, S., Ledgerwood, L.W., Ahmed, R., Yu, M., Takach, N., 2010. Experimental study of MSE of a single PDC cutter interacting with rock under simulated pressurized conditions. SPE Drill. Complet. 25 (1), 10–18.

Rashidi, B., Hareland, G., Fazaelizadeh, M., Svigir, M., 2010. Comparative study using rock energy and drilling strength models. In: ARMA US Rock Mechanics/Geomechanics Symposium, pp. ARMA–10. ARMA.

Salami, B.A., Iqbal, M., Abdulraheem, A., Jalal, F.E., Alimi, W., Jamal, A., Tafsirojjaman, T., Liu, Y., Bardhan, A., 2022. Estimating compressive strength of lightweight foamed concrete using neural, genetic and ensemble machine learning approaches. Cem. Concr. Compos. 133. https://doi.org/10.1016/jcemconcomp.2022.104721.

Saltelli, A., Tarantola, S., Campolongo, F., et al., 2004. Sensitivity Analysis in Practice: a Guide to Assessing Scientific Models [M]. John Wiley & Sons, Hoboken NJ.

Swiler, L.P., Gulian, M., Frankel, A.L., Safta, C., Jakeman, J.D., 2020. A survey of constrained Gaussian process regression: approaches and implementation challenges. Journal of Machine Learning for Modeling and Computing 1 (2).

Teale, R., 1965, March. The concept of specific energy in rock drilling. In: International journal of rock mechanics and mining sciences & geomechanics abstracts. Pergamon, pp. 57–73 (Vol. 2, No. 1.

Wiśniowski, R., Knez, D., Hytroś, Ł., 2015. Drillability and mechanical specific energy analysis on the example of drilling in the pomeranian basin. AGH drilling. Oil Gas 32 (1).

Yu, M., Yuan, Z., Li, R., Shi, B., Wan, D., Dong, X., 2024. Interpretable machine learning model to predict surgical difficulty in laparoscopic resection for rectal cancer. Front. Oncol. 14.

Zheng, D., Zhong, H., Camps-Valls, G., Cao, Z., Ma, X., Mills, B., et al., 2024. Explainable deep learning for automatic rock classification. Comput. Geosci. 184, 105511.