



## REAL-TIME LITHOLOGY PREDICTION FROM DRILLING DATA USING MACHINE LEARNING: A CASE STUDY CA TAM OIL FIELD, OFFSHORE VIETNAM

H. D. Vu<sup>1</sup>, T. H. Nguyen<sup>\*1</sup>, M. H. Nguyen<sup>1</sup>, X. D. Nguyen<sup>2</sup>

<sup>1</sup>Hanoi University of Mining and Geology, Hanoi, Vietnam

<sup>2</sup>Vietnam Petroleum Institute, Hanoi, Vietnam

### ABSTRACT

Accurate prediction of lithological layers during drilling operations remains a fundamental challenge that directly impacts drilling efficiency, wellbore stability, and overall hydrocarbon exploration success. Traditional formation evaluation methods often rely on periodic sampling or wireline logging, which can introduce delays and limit real-time decision-making capabilities during critical drilling phases. This study presents a machine learning approach for real-time lithology prediction using continuously measured drilling parameters from the Ca Tam oil field, offshore Vietnam. The study utilized seven key drilling parameters: weight on bit (WOB), rotary speed (RPM), torque, standpipe pressure (SPP), rate of penetration (ROP), mud weight (MW IN), and flow rate (FLW). These measurements were collected from multiple wells across the Ca Tam field, representing diverse geological conditions and operational scenarios. Data preprocessing involved removing outliers to focus on measurements representing actual formation properties. Subsequently, feature selection analysis identified the most discriminative parameters for lithological classification, focusing on variables that behave differently in each rock type. Four supervised machine learning algorithms were implemented and compared: Random Forest, Gradient Boosting (XGBoost), Support Vector Machines, and Artificial Neural Networks. Each model was trained to classify formations into three primary lithological categories: Sand, Claystone, and Shale. The Random Forest algorithm achieved the best performance, demonstrating the highest cross-validation accuracy and showing good generalization capabilities when tested on independent well data. Model validation confirmed consistent performance across varying geological conditions, successfully handling the complex stratigraphic variations characteristic of the Cuu Long basin. This approach provides real-time formation identification, improving drilling decisions and reducing operational costs.

**Keywords:** lithology; drilling data; machine learning; data-driven.

**Date submitted:** 03.06.2025

**Date accepted:** 01.09.2025

© 2025 «OilGasScientificResearchProject» Institute. All rights reserved.

### 1. Introduction

In oil and gas exploration and development projects, drilling operations consistently account for a significant portion of total costs. Statistical data from operators and contractors indicate that during the exploration phase, drilling design and execution costs comprise 65–75 % of the budget, while in the field development phase, these costs range from 25–35 % [1]. Consequently, enhancing the rate of penetration (ROP) to minimize time, costs, and risks of complications is a primary objective for drilling engineers and operators. ROP is influenced by multiple factors, including geological characteristics of the drilled formations, drilling parameters (e.g., weight on bit, rotary speed, drilling fluid flow rate, pressure losses during circulation, drilling fluid density...), bit type, and wellbore diameter [2-5]. Among

technical solutions to improve drilling efficiency, real-time identification of lithology layers optimizes various drilling aspects, such as selecting appropriate drill bits, estimating circulation losses, mitigating drilling issues and ensuring wellbore stability [6-10]. This is particularly critical in complex formations, where lithology prediction is essential for optimizing technical parameters, reducing risks, and enhancing drilling performance.

Lithology identification during drilling typically relies on anticipated stratigraphic columns and the expertise of drilling engineers. For exploration wells, stratigraphic columns are constructed by interpreting wireline logging data and correlating geological information from nearby wells with similar inclinations, azimuths, and formation boundaries [11–13]. This approach is impractical in frontier areas with limited prior drilling data and is prone to significant errors, especially in geologically complex regions. Moreover, traditional lithology prediction methods, such

\*E-mail: [nguyenminhhoa@humg.edu.vn](mailto:nguyenminhhoa@humg.edu.vn)  
<http://dx.doi.org/10.5510/OGP20250301096>

as analyzing core or cutting samples or interpreting wireline logs, have a major limitation: these data are collected post-drilling, precluding real-time identification of lithology and immediate adjustments to drilling methods or parameters. Thus, a method utilizing real-time drilling data to directly identify lithological layers during drilling is needed to optimize techniques, minimize complications and enhance efficiency. Specifically, optimizing techniques refers to adjusting drilling parameters such as weight on bit (WOB), rotary speed (RPM), and mud flow rate in real-time based on the identified lithology to achieve optimal penetration rates and hole quality. Minimizing complications involves preventing drilling problems such as stuck pipe, lost circulation, wellbore instability, and formation damage by anticipating challenging formations and adjusting drilling practices accordingly. Enhanced efficiency is achieved through improved rate of penetration (ROP), reduced non-productive time (NPT), and better wellbore trajectory control by adapting drilling strategies to specific rock types as they are encountered. Reduced drilling costs result from decreased drilling time, fewer drilling incidents requiring remedial operations, optimized mud programs tailored to specific formations, and reduced need for additional logging runs or coring operations. Improved drilling outcomes result from reduced drilling time, fewer drilling incidents requiring remedial operations, optimized mud programs tailored to specific formations, and decreased need for additional logging runs or coring operations.

Advances in drilling technology and downhole data monitoring systems have generated increasingly rich and diverse datasets. This has spurred a new approach: leveraging modern computational technologies, such as artificial intelligence, to analyze and predict geological characteristics of drilled formations. The use of big data and machine learning for formation identification and mechanical property prediction is a rapidly growing research area. Various models employing machine learning techniques, such as Support Vector Machines (SVMs), Random Forests (RF), and deep learning, have been developed to identify lithology from diverse input data. While some rely on traditional wireline logging data [14–18], recent studies have increasingly utilized real-time drilling parameters, including ROP, rotary speed (RPM), standpipe pressure (SPP), torque, weight on bit (WOB), and drilling fluid density (MW) [19–22].

Notable studies have demonstrated the efficacy of machine learning algorithms in lithology classification from drilling data. Proposed models enable real-time prediction of lithology and formation boundaries, optimizing drilling processes and mitigating technical risks. For instance, Mahmoud et al. (2021) [23] employed Artificial Neural Networks (ANN), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), and Functional Neural Networks (FNN) to predict lithology and boundaries, achieving over 98% accuracy on test datasets. Khalifa et al. (2023) [24] developed a web-based application integrating machine learning to predict lithology in real-time from drilling data, attaining 95% accuracy in classifying rocks such as claystone, marl, and sandstone. Elkatatny et al. (2024) [25] applied SVM, Random Forest, KNN, and MLP algorithms to predict formation boundaries on the Norwegian continental shelf, achieving up to 99% accuracy. Hansen et al. (2024) [26] compared machine learning methods for classifying rock mass quality from drilling

data in tunneling environments, achieving a balanced accuracy of 0.86. Collectively, these studies underscore the growing importance of advanced data analytics in drilling operations, providing valuable insights for real-time decision-making and optimization. However, challenges remain in handling noisy or incomplete drilling data and generalizing models across diverse geological contexts.

This study aims to develop a robust machine learning model for real-time lithology prediction using drilling data from the Cá Tầm oil field, Vietnam. By integrating key drilling parameters and applying advanced feature extraction techniques, the model will be evaluated for accuracy, reliability, and practical applicability in real-time drilling operations.

## 2. Database

### 2.1. Overview data

In the oil and gas industry, drilling operations are complex processes requiring meticulous control of drilling parameters to ensure efficiency, safety, and cost optimization. Key parameters derived from drilling data – including Rate of Penetration (ROP), Rotary Speed (RPM), Torque (TORQUE), Flow Rate (FLOWIN), Mud Weight (MWz), Weight on Bit (WOB) and Standpipe Pressure (SPP) not only facilitate operational control but also reflect the lithological characteristics of the formations penetrated [9, 27]. A comprehensive understanding of the interplay between these parameters and the geological properties of rock is critical for enhancing drilling performance and mitigating operational risks.

Each lithological layer encountered during drilling exhibits distinct petrophysical properties, requiring operators to adjust drilling parameters to achieve optimal performance. The relationship between drilling parameters and lithology is not unidirectional but involves complex, instantaneous interactions. For instance, a sudden ROP decrease may indicate a transition from sandstone to harder limestone, necessitating increased WOB or a different bit type. Similarly, elevated TORQUE may signal abrasive rock, requiring reduced RPM to protect the bit. In summary, parameters such as ROP, RPM, TORQUE, FLOWIN, MW, and WOB serve not only as operational controls but also as critical indicators of subsurface lithology. This intricate relationship underscores the importance of integrating drilling mechanics with geological knowledge and highlights the potential of data-driven techniques to address the challenge of real-time lithology identification, thereby enhancing drilling efficiency.

Figure 1 illustrates the drilling parameters directly measured during operations alongside detailed stratigraphic columns from three wells in the Cá Tầm oil field, Vietnam, which serve as the research dataset for this study. Based on the analyzed data, three primary lithological categories were identified within the wells: sand, shale, and claystone, with layer thicknesses ranging from a few centimeters to several meters. The configuration of rock layers across the three wells is notably complex, characterized by interbedded strata and predominantly thin layer thicknesses. Consequently, it can be inferred that, despite their proximity within the same field, the depth intervals of lithological layers vary significantly among the wells. This variability poses substantial challenges for predicting the lithological sequence in a new well without reference to sample data.

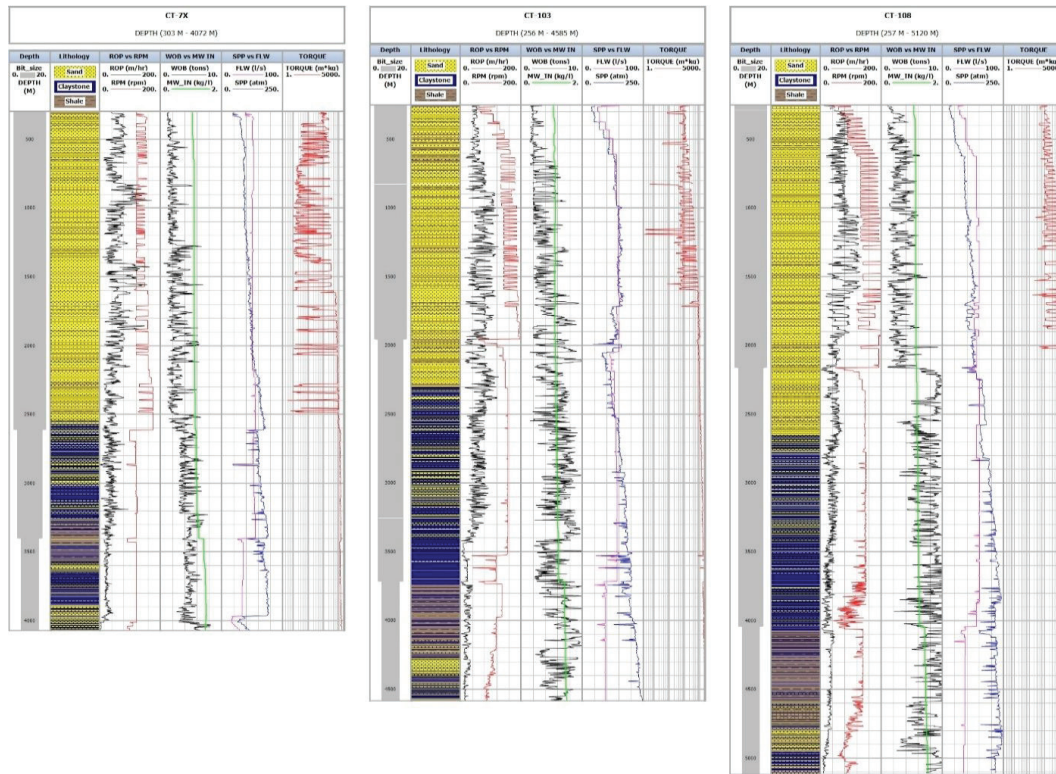


Fig. 1. Drilling data of three wells in study area

## 2.2. Detect and remove the outliers in database

In the realm of data analysis, real time drilling data often present challenges due to their susceptibility to noise, measurement errors, and skewed distributions. Drilling parameters such as ROP, WOB and TORQUE frequently exhibit non-normal behavior, influenced by geological heterogeneities and borehole condition (fig. 2).

Traditional statistical methods, such as the Z-score approach, which rely on the mean and standard deviation, are less effective in these scenarios as they are heavily distorted by extreme values and assume a Gaussian distribution.

Robust Statistics provides a more resilient alternative by employing estimators like the median and Median Absolute Deviation (MAD) to detect and remove outliers, ensuring data integrity for subsequent geological interpretations [25]. This method calculates a Modified Z-score, defined as

$$Mi = 0.6745 \times (xi - \text{median}(x)) / \text{MAD}$$

where  $xi$  denotes individual data points, and the constant 0.6745 aligns MAD with the standard deviation of a normal distribution. Data points with an absolute Modified Z-score exceeding a predefined threshold – typically 3 or 3.5 – are

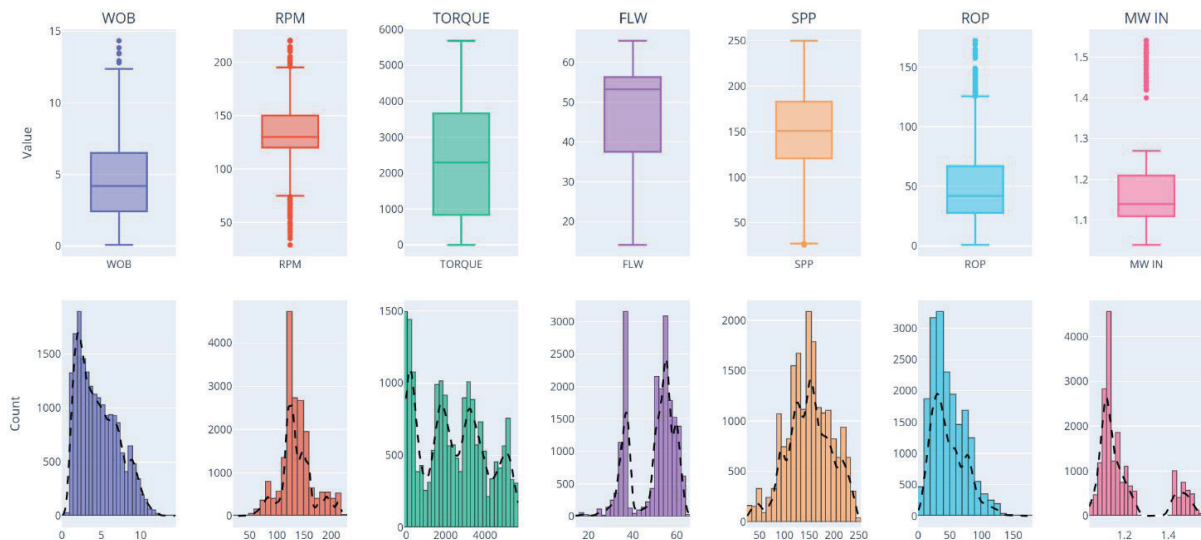
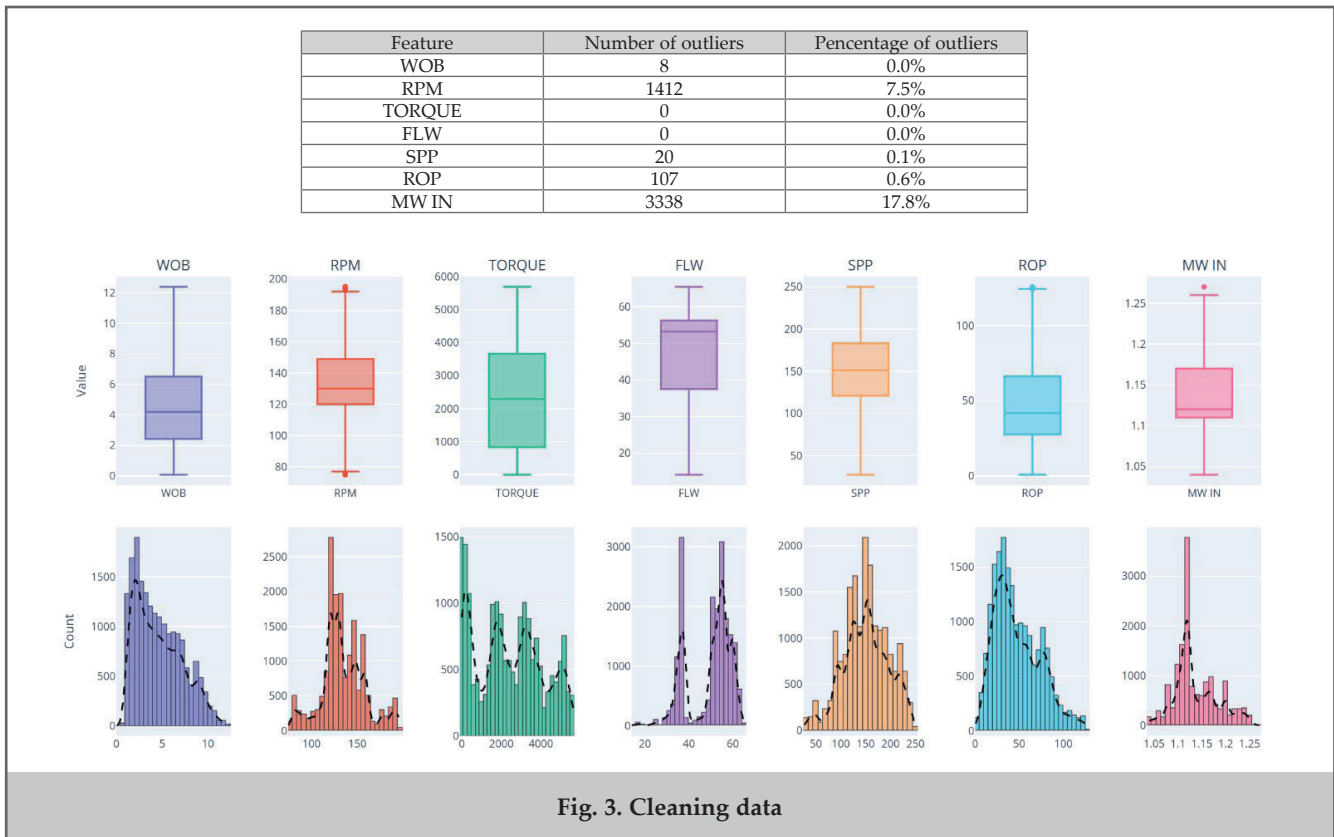


Fig. 2. Raw data



flagged as outliers and excluded from further analysis.

Beyond its robustness to non-normality, this method offers practical simplicity and adaptability, making it suitable for automated preprocessing workflows in this study. It outperforms parametric techniques in datasets with heavy-tailed distributions or significant noise, ensuring that outliers are reliably removed without compromising the dataset's core trends. Therefore, in this study, the authors employed this method to eliminate outliers, with the results presented in figure 3.

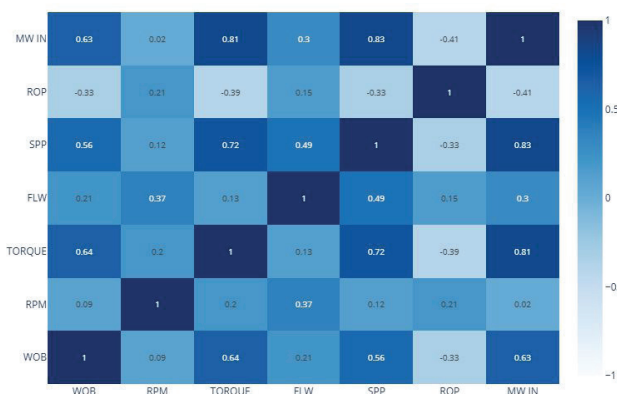
### 2.3. Feature selection

In the process of training a machine learning model, a fundamental principle is that the model must learn from all provided input data. Consequently, if the training dataset contains noise or irrelevant information, the accuracy of the output results will be compromised. To develop the lithology

prediction model, the authors utilized a comprehensive dataset comprising seven drilling parameters from wells. However, it is possible that certain components of this dataset do not significantly influence or contribute substantially to the accuracy of the prediction outcomes. Moreover, employing an excessive number of input parameters can slow down the training process, and the presence of abundant irrelevant data may introduce significant errors into the output results.

In this context, prior to constructing the petrophysical prediction model, a data analysis is essential to select the most appropriate drilling parameters for the training set of the forecasting model. Based on the heatmap (fig. 4) and pair plot (fig. 5), it is evident that the correlation between the drilling parameters ranges from moderate to weak, except SPP-MW IN (0.83) and TORQUE-MW IN (0.81). This observation aligns with practical experience, as mud weight (MW IN) typically exerts a significant influence on both standpipe pressure (SPP) and torque, suggesting that including all three parameters as input features for the predictive model is likely to result in data redundancy. In contrast, most other parameter pairs (e.g., RPM and WOB, TORQUE and WOB) exhibit weak correlations, indicating that they capture distinct aspects of the drilling process.

To further evaluate the suitability of drilling parameters for classifying lithologies (Sand, Shale, Claystone), a comprehensive statistical analysis was conducted on seven features – TORQUE, ROP, SPP, WOB, MW IN, RPM, and FLW – based on their minimum, maximum, mean, and standard deviation values across lithology classes (table 1). Utilizing lithological data from three wells, the rock formations were categorized into three primary lithological types: Sand (54.4%), Claystone (26%), and Shale, which constitutes the smallest proportion (19.6%) (fig. 6).



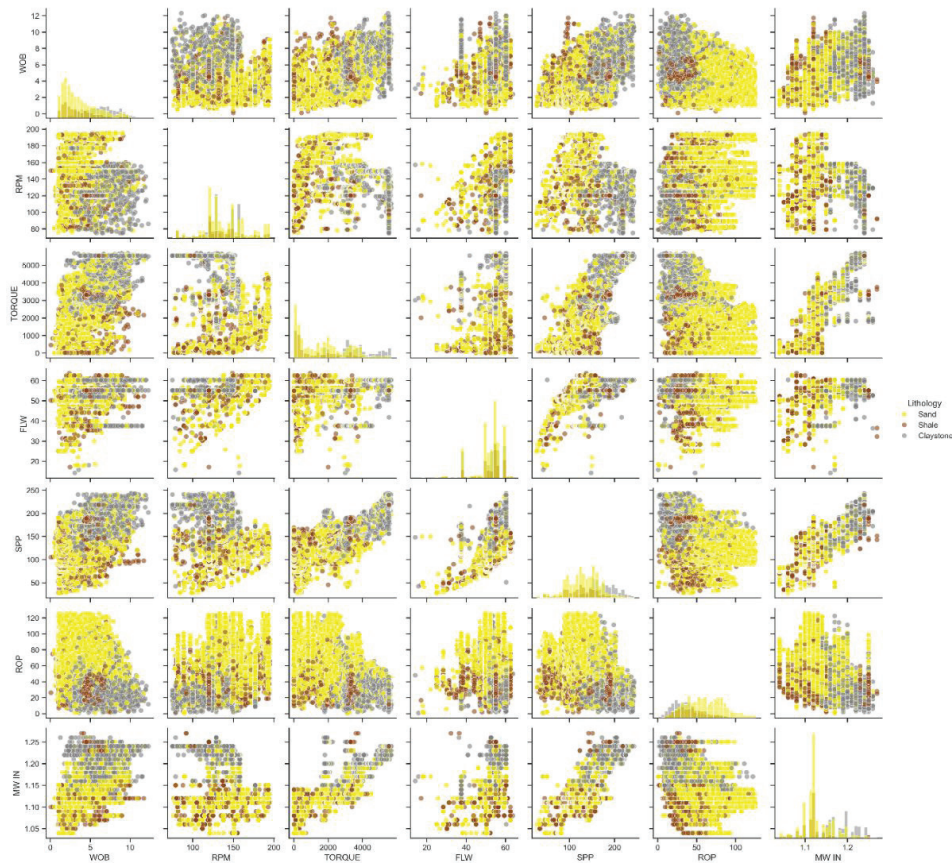


Fig. 5. Pair plot of dataset

Based on the statistical analysis presented in table 1, the evaluation of individual features yields the following observations:

- TORQUE exhibits strong discriminative capability, with Claystone displaying significantly higher mean (3642.41) and minimum (592.25) values compared to Sand (1716.65, 1.04) and Shale (1161.74, 0.84), establishing it as a critical indicator for identifying Claystone.
- ROP demonstrates a clear decreasing trend in mean values across lithologies (Sand: 61.44, Shale: 45.32, Claystone: 34.73), confirming its effectiveness in distinguishing all three classes.
- SPP shows higher mean (175.59) and minimum (42.2) values for Claystone compared to Sand (131.41, 27.7) and Shale (122.05, 28.7), though its ability to differentiate between Sand and Shale is limited.
- WOB presents a higher mean for Claystone (6.32) compared to Sand (3.67) and Shale (3.55), offering moderate discriminative power, particularly for identifying Claystone.

In contrast, MW IN, RPM, and FLW exhibit minimal variation across lithologies, indicating limited discriminative value. Furthermore, MW IN is excluded due to its high correlation with both SPP and TORQUE (higher than 0.81), suggesting redundancy. In conclusion, TORQUE, ROP, SPP, and WOB are recommended as input features for the classification model due to their distinct statistical properties and robust discriminative capabilities across lithology classes. Excluding MW IN, RPM, and FLW minimizes redundancy, thereby enhancing model performance and computational efficiency.

**Table 1**  
Detailed statistics of the well-logging dataset

| Lithology | Feature | Min    | Max     | Mean    | Standard Deviation |
|-----------|---------|--------|---------|---------|--------------------|
| Sand      | TORQUE  | 1.04   | 5653.47 | 1716.65 | 1427.53            |
|           | ROP     | 0.79   | 125.71  | 61.44   | 24.54              |
|           | SPP     | 27.7   | 242.7   | 131.41  | 36.08              |
|           | WOB     | 0.53   | 11.76   | 3.67    | 2.2                |
|           | MW IN   | 1.04   | 1.26    | 1.13    | 0.04               |
|           | RPM     | 76     | 195     | 139.75  | 25.65              |
|           | FLW     | 17.63  | 62.52   | 52.28   | 6.81               |
| Shale     | TORQUE  | 0.84   | 5541.84 | 1161.74 | 1200.06            |
|           | ROP     | 1.71   | 122.08  | 45.32   | 20.17              |
|           | SPP     | 28.7   | 218.1   | 122.05  | 35.13              |
|           | WOB     | 0.08   | 11.73   | 3.55    | 2.1                |
|           | MW IN   | 1.04   | 1.27    | 1.12    | 0.03               |
|           | RPM     | 77     | 195     | 134.15  | 26.99              |
|           | FLW     | 17.05  | 63.64   | 52.33   | 7.09               |
| Claystone | TORQUE  | 592.25 | 5679.73 | 3642.41 | 1114.05            |
|           | ROP     | 0.75   | 122.24  | 34.73   | 18.07              |
|           | SPP     | 42.2   | 243.3   | 175.59  | 28.14              |
|           | WOB     | 1.26   | 12.3    | 6.32    | 2.08               |
|           | MW IN   | 1.06   | 1.26    | 1.19    | 0.03               |
|           | RPM     | 75     | 177     | 136.83  | 17.33              |
|           | FLW     | 14.12  | 60.36   | 52.7    | 7.49               |

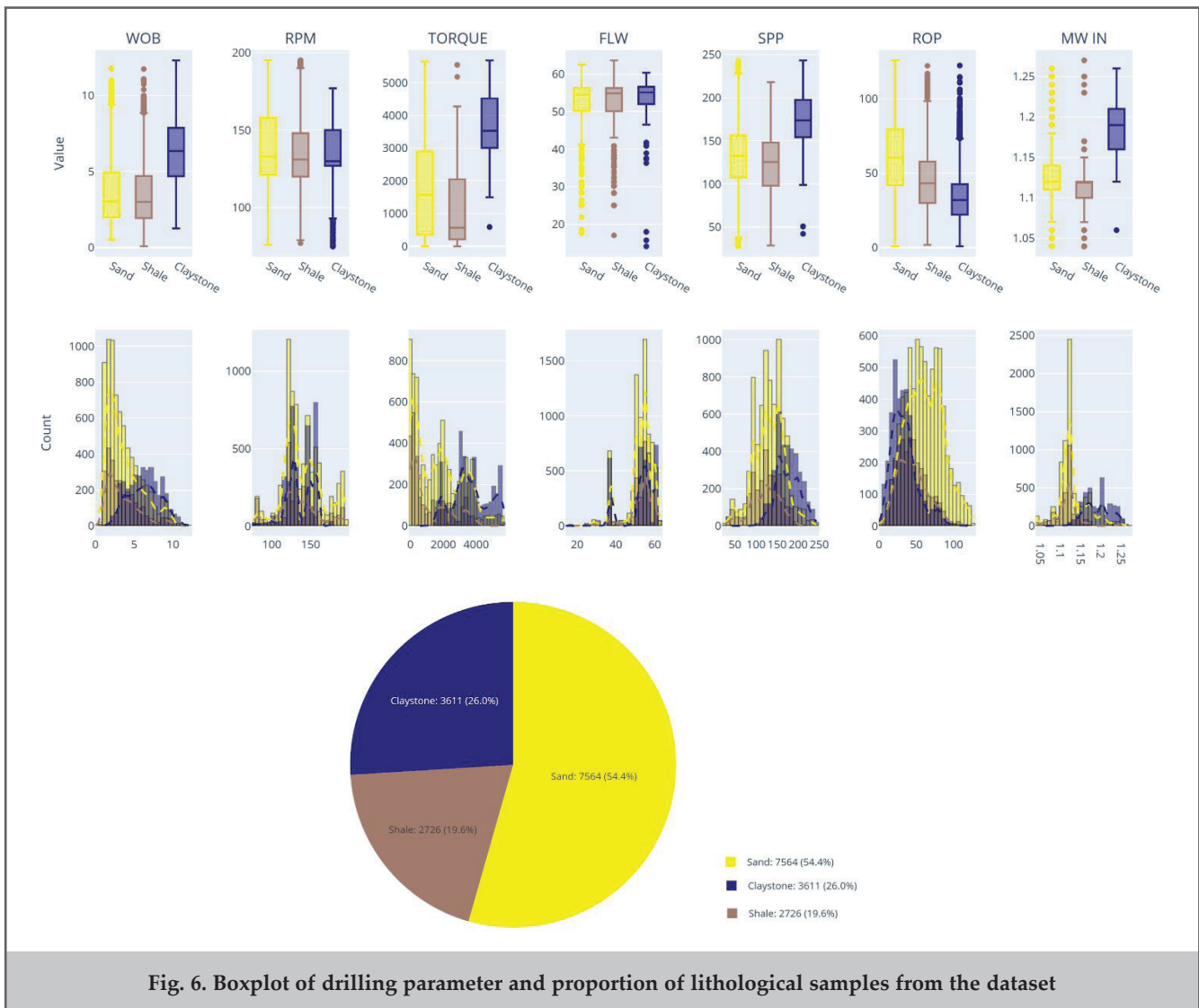


Fig. 6. Boxplot of drilling parameter and proportion of lithological samples from the dataset

### 3. Methodology and result

In this paper, the authors aim to evaluate the effectiveness of machine learning techniques in predicting lithological layers for three wells in the Ca Tam oil field. Using drilling parameters – ROP, RPM, torque, flow-in, mud weight (MW), standpipe pressure (SPP), and weight on bit (WOB) – alongside lithological group distributions (sand, shale, and claystone) by depth, the task is framed as a supervised classification problem. Four machine learning algorithms were applied: Artificial Neural Networks (ANN), Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost), each selected for its distinct theoretical strengths and application suitability.

#### Artificial Neural Network (ANN)

ANNs are deep learning models inspired by the human brain, composed of interconnected layers of neurons (input, hidden, and output). They learn to map input features to class labels by minimizing a loss function through backpropagation. ANNs are well-suited for modeling complex, nonlinear relationships in large datasets. However, they require significant computational resources, large training datasets, and careful tuning to avoid overfitting. Common applications include image recognition and natural language processing.

#### Random Forest (RF)

Random Forest is an ensemble learning method that

builds multiple decision trees using bootstrapped data samples and random feature selection. Final predictions are made via majority voting. RF is robust to overfitting, handles imbalanced data well, and provides insights through feature importance metrics. Its main drawback is slower prediction times on large datasets due to the number of trees involved.

#### Support Vector Machine (SVM)

SVM is a supervised learning algorithm that finds the optimal hyperplane to separate classes, maximizing the margin between support vectors. It handles both linear and non-linear data using kernel functions (e.g., RBF, linear). SVMs are effective in high-dimensional spaces and resilient to noise, but they can be computationally intensive and require careful parameter tuning for optimal performance.

#### Extreme Gradient Boosting (XGBoost)

XGBoost is a high-performance gradient boosting algorithm that builds decision trees sequentially, with each tree correcting the errors of the previous ones. It offers fast computation, strong performance on imbalanced datasets, and flexible hyperparameter tuning. XGBoost also provides feature importance metrics. However, it can overfit if not properly regularized and is more complex to configure than Random Forest.

During the development of classification models, model

Table 2

| Optimum set of parameters for models |                                                                                                                                                    |
|--------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| Model                                | Hyper parameter                                                                                                                                    |
| ANN                                  | 'learning_rate_init': 0.01, 'hidden_layer_sizes': (128, 64, 32), 'alpha': 0.0001, 'activation': 'relu'                                             |
| Random forest                        | 'n_estimators': 100, 'min_samples_split': 10, 'min_samples_leaf': 8, 'max_features': log2, 'max_depth': 50, 'bootstrap': False                     |
| SVM                                  | 'kernel': 'rbf', 'gamma': 'auto', 'C': 10                                                                                                          |
| XGB                                  | 'subsample': 0.8, 'reg_lambda': 0, 'reg_alpha': 1, 'n_estimators': 200, 'max_depth': 7, 'learning_rate': 0.05, 'gamma': 0, 'colsample_bytree': 1.0 |

optimization (fine-tuning) was conducted by adjusting hyperparameters throughout the training process to identify the best-performing model aligned with the predefined objective function. The Grid Search algorithm was applied to identify the optimal model by evaluating the predictive performance of different models as hyperparameters were adjusted. The final optimal hyperparameters for each model are listed in table 2.

Model performance was assessed using the following evaluation metrics. Based on the confusion matrix (True Positive - TP, True Negative - TN, False Positive - FP, False Negative - FN), key metrics include:

- $\text{Accuracy}_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}$
- $\text{Precision}_i = \frac{TP_i}{TP_i + FP_i}$
- $\text{Recall}_i = \frac{TP_i}{TP_i + FN_i}$
- $\text{F1}_i = 2 \frac{\text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}$

where:

$TP_i$ : True positives, i.e., data points correctly classified as belonging to class «i»

$FP_i$ : False positives, i.e., data points incorrectly classified as class «i»

$FN_i$ : False negatives, i.e., data points of class «i» incorrectly classified as belonging to other classes.

To identify the most effective model for lithology classification, a comparative analysis was conducted using four algorithms – Artificial Neural Network (ANN), Random Forest, Support Vector Machine (SVM), and XGBoost – based on their performance across accuracy, precision, recall, and F1 score on both training and test sets (fig. 7).

- *Training Performance*: Random Forest model achieved the highest performance across all metrics, with an accuracy of 0.881, precision of 0.881, recall of 0.865, and F1 score of 0.87, significantly outperforming ANN (accuracy: 0.736, F1: 0.705), SVM (accuracy: 0.741, F1: 0.696), and XGBoost (accuracy: 0.856, F1: 0.844).
- *Test Performance*: Random Forest model exhibited the best test performance with an accuracy of 0.753, precision of 0.735, recall of 0.723, and F1 score of 0.723, slightly outperforming XGBoost and significantly surpassing ANN (accuracy: 0.727, F1: 0.692) and SVM (accuracy: 0.730, F1: 0.684).

Therefore, Random Forest is the optimal model for this classification task due to its superior and consistent performance across both datasets, balanced evaluation metrics, and strong generalization capability (fig. 8). Specifically, RF demonstrates several key advantages that make it particularly well-suited for lithology prediction from drilling data: First, its ensemble learning approach effectively handles the heterogeneous nature of drilling parameters, seamlessly processing both continuous variables (such as ROP, WOB, and torque) and categorical variables (such as mud type and drill-

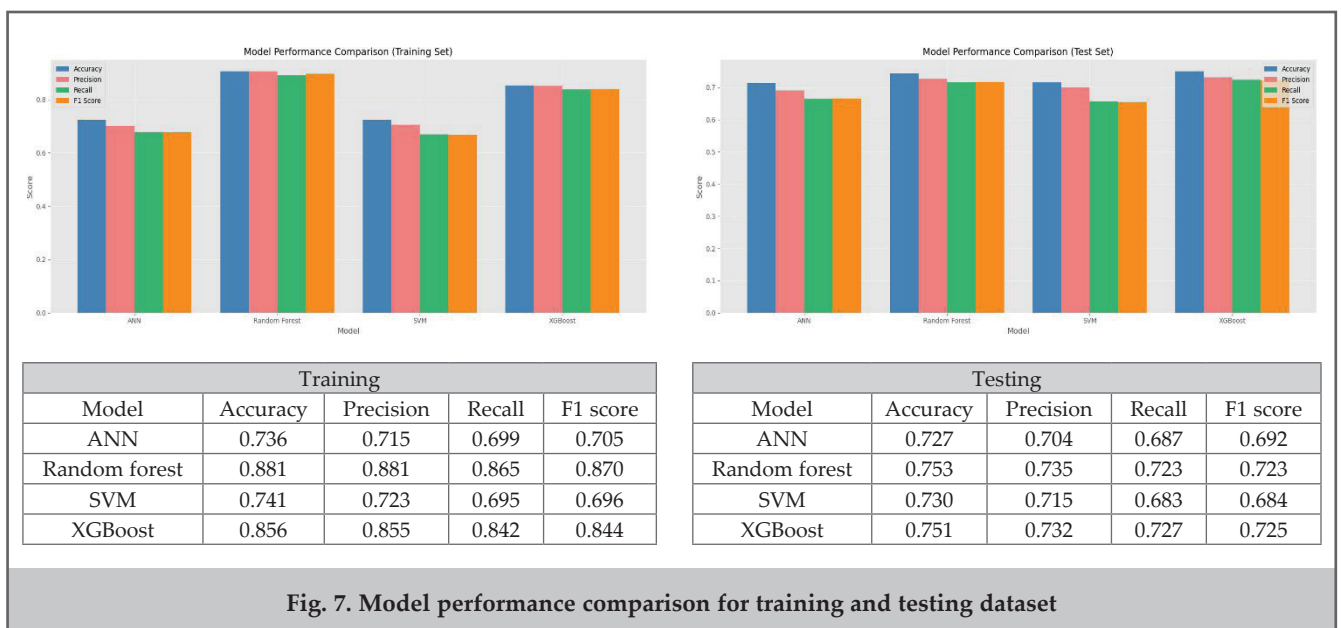
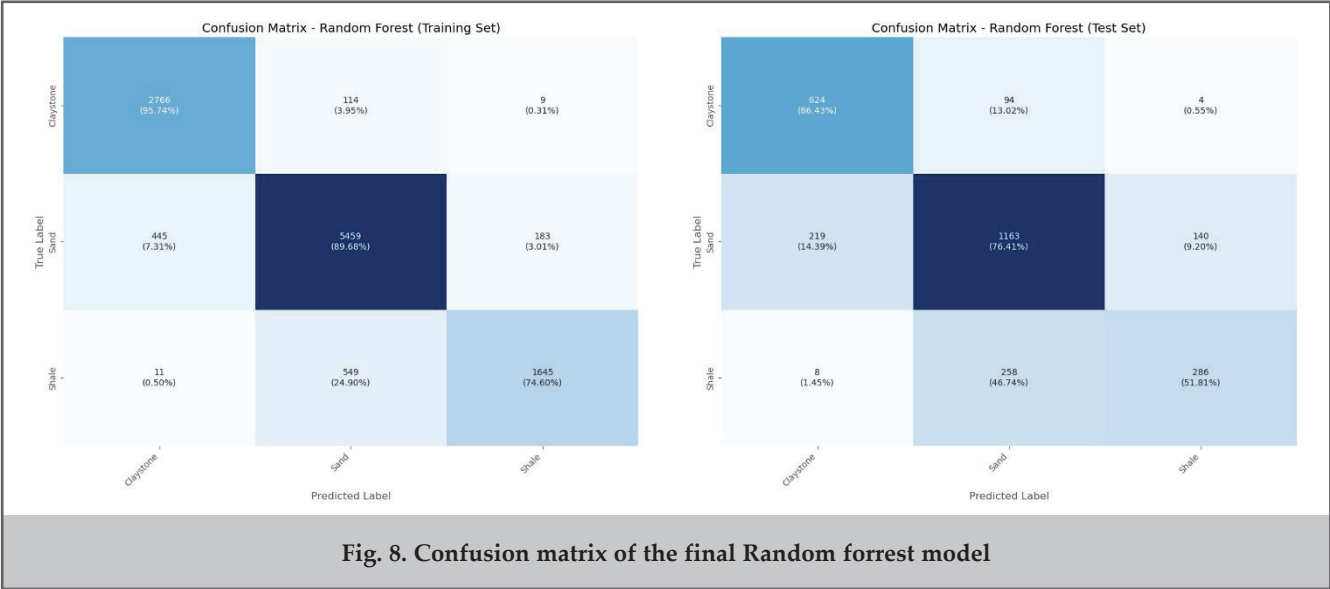


Fig. 7. Model performance comparison for training and testing dataset



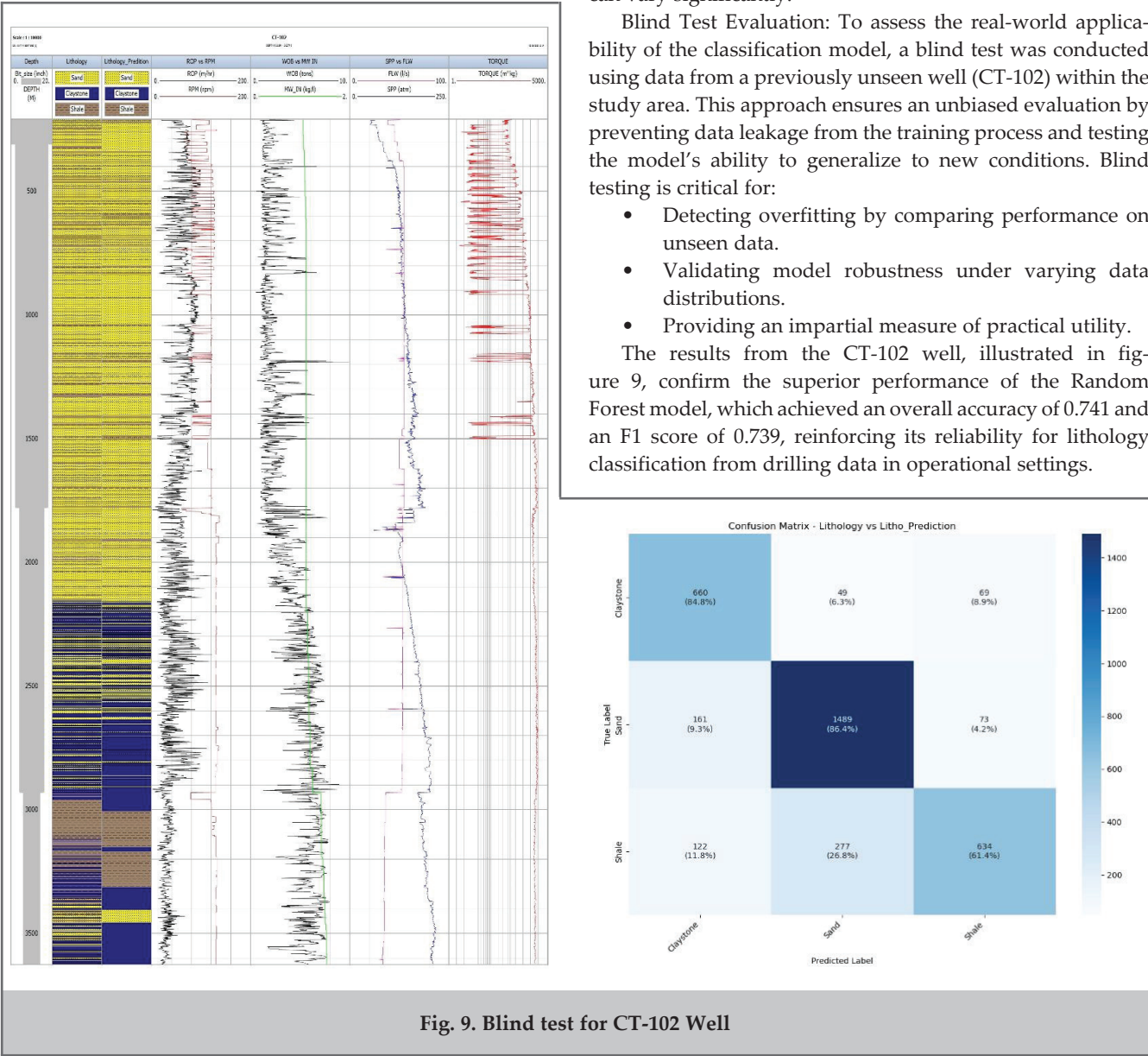
ing mode) without requiring extensive data preprocessing. Secondly, RF reduced sensitivity to outliers and its capability to maintain stable performance across different geological

settings demonstrate its strong generalization ability, making it highly suitable for deployment in diverse drilling environments where formation characteristics and drilling conditions can vary significantly.

**Blind Test Evaluation:** To assess the real-world applicability of the classification model, a blind test was conducted using data from a previously unseen well (CT-102) within the study area. This approach ensures an unbiased evaluation by preventing data leakage from the training process and testing the model's ability to generalize to new conditions. Blind testing is critical for:

- Detecting overfitting by comparing performance on unseen data.
- Validating model robustness under varying data distributions.
- Providing an impartial measure of practical utility.

The results from the CT-102 well, illustrated in figure 9, confirm the superior performance of the Random Forest model, which achieved an overall accuracy of 0.741 and an F1 score of 0.739, reinforcing its reliability for lithology classification from drilling data in operational settings.



## Conclusions

1. This study introduces an innovative methodology for real-time lithology prediction through the analysis of drilling parameter measurements, implemented in wells at the Cá Tầm oil field, offshore Vietnam. This approach offers distinct advantages over conventional techniques reliant on well logging and core sampling. By utilizing real-time data – including weight on bit (WOB), revolutions per minute (RPM), torque, standpipe pressure (SPP), rate of penetration (ROP), mud weight in (MW IN), and flow rate (FLW) – the method facilitates immediate lithology identification during drilling operations. This capability significantly reduces operational delays and minimizes downtime. Furthermore, the incorporation of machine learning enables the model to adapt dynamically to evolving geological conditions, thereby enhancing predictive accuracy without requiring invasive sampling procedures. In contrast, traditional well logging necessitates post-drilling analysis and specialized equipment, while core sampling introduces risks to well integrity and is constrained by depth and accessibility limitations. The proposed data-driven methodology provides a scalable and operationally efficient alternative, advancing the precision and timeliness of lithology characterization in complex geological settings.
2. Through Grid Search optimization, the Random Forest model emerged as the top-performing classifier, demonstrating significantly higher accuracy and reliability than alternative approaches. To assess its practical applicability, a blind test was conducted using data from a previously unseen well (CT-102), excluded from the training process. The results confirm Random forest model's ability to maintain high classification accuracy, demonstrating strong generalization to new data.
3. This study underscores the transformative potential of machine learning in lithology identification using real-time drilling data, offering an objective, efficient, and precise alternative to conventional techniques. Future work will focus on expanding the dataset, incorporating additional well log parameters, and exploring deep learning methods to further improve classification performance and adaptability across diverse geological settings.

*This research funded by Hanoi University of Mining and Geology under grant number T25-07. The authors would also like to express their profound gratitude for the invaluable sharing of academic knowledge and the enthusiastic technical support provided by the Vietnam Petroleum Institute.*

## References

1. dos Santos Barbosa, M., Duarte Gonzaga, I. M., da Silva Vilar, D., et al. (2022). Cost analysis of drilling operations of an offshore well with four phases. *Journal of Engineering Research*, 2(6).
2. Pessier, R., Fear, M. (1992). Quantifying common drilling problems with mechanical specific energy and a bit-specific coefficient of sliding friction. SPE-24584-MS. In: *SPE Annual Technical Conference and Exhibition, Washington, D.C., October. Society of Petroleum Engineers*.
3. Efendiyev, G., Mammadov, P., Piriverdiyev, I., Mammadov, V. (2018). Estimation of the lost circulation rate using fuzzy clustering of geological objects by petrophysical properties. *Visnyk of Taras Shevchenko National University of Kyiv. Geology*, 2(81), 28–33.
4. Chernikov, A. D., Eremin, N. A., Stolyarov, V. E., et al. (2020). Application of artificial intelligence methods for identifying and predicting complications during the construction of oil and gas wells: problems and main directions of solutions. *Georesources*, 22(3), 87–96.
5. Dmitrievsky, A. N., Eremin, N. A., Safarova, E. A., et al. (2020). Qualitative analysis of time-series geodata for preventing complications and emergency situations during the drilling of oil and gas wells. *SOCAR Proceedings*, 3, 31–37.
6. Cuddy, S. J., Glover, P. W. J. (2002). The application of fuzzy logic and genetic algorithms to reservoir characterization and modeling. *Soft Computing for Reservoir Characterization and Modeling*, 1, 219–241.
7. Efendiyev, G. M., Mammadov, P. Z., Piriverdiyev, I. A. (2019). Modeling and evaluation of rock properties based on integrated logging while drilling with the use of statistical methods and fuzzy logic. In: Aliev, R., Kacprzyk, J., Pedrycz, W., Jamshidi, M., Babanli, M., Sadikoglu, F. (eds) 10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions - ICSCCW-2019. ICSCCW 2019. Advances in Intelligent Systems and Computing, vol 1095. Springer, Cham.
8. Efendiyev, G. M., Karazhanova, M. K., Piriverdiyev, I. A., et al. (2024). B. Assessment of the characteristics of the geological section of wells based on probabilistic-fuzzy analysis of complex geophysical and geological-technological information. *SOCAR Proceedings*, 3, 3-8.
9. Bourgoyne, A. T., Millheim, K. K., Chenevert, M. E., Young, F. S. (1991). Applied drilling engineering. 2<sup>nd</sup> ed. *Society of Petroleum Engineers*.
10. Caers, J., Zhang, T. (2004). Multiple-point geostatistics: A quantitative vehicle for integrating geologic analogs into reservoir modeling. *AAPG Memoir*, 80, 383–394.

11. Rider, M. H. (2002). The geological interpretation of well logs. 2<sup>nd</sup> ed. Rider-French Consulting Ltd.
12. Serra, O. (2007). Well logging and reservoir evaluation. Editions Technip.
13. Asquith, G., Krygowski, D. (2004). Basic well log analysis. 2nd ed. American Association of Petroleum Geologists.
14. Gong, K., Ye, Zh., Chen, D., et al. (2018). Investigation on automatic recognition of stratigraphic lithology based on well logging data using ensemble learning algorithm. SPE-192006-MS. In: *The SPE Asia Pacific Oil and Gas Conference and Exhibition, Brisbane, Australia, October*.
15. Sun, J., Li, Q., Chen, M., et al. (2019). Optimization of models for a rapid identification of lithology while drilling - A winwin strategy based on machine learning. *Journal of Petroleum Science and Engineering*, 176, 321–341.
16. Gang, C. (2020). Study on realtime lithology identification method of LWD. *IOP Conference Series: Earth and Environmental Science*, 546, 052007.
17. Ren, X., Hou, J., Song, S., et al. (2019). Lithology identification using well logs: A method by integrating artificial neural networks and sedimentary patterns. *Journal of Petroleum Science and Engineering*, 182, 106336.
18. Efendiyeve, G. M., Mammadov, P. Z., Piriverdiyev, I. A. (2019, August). Modeling and evaluation of rock properties based on integrated logging while drilling with the use of statistical methods and fuzzy logic. In: *10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions, ICSCCW-2019*, 1095, 503-511.
19. Zhang, J., Liu, Y., Ma, Y., et al. (2024). Real-time lithology identification from drilling data with self & cross attention model and wavelet transform. *Geoenergy Science and Engineering*, 244, 213427.
20. Li, T.-T., Xiang, R.-Y., Jiao, T., Yan, X.-D. (2022). Research on intelligent lithology identification method based on real-time data of drilling wells. In: *2022 4th International Conference on Robotics, Intelligent Control and Artificial Intelligence (RICAI 2022), December 16–18, Dong-guan, China. ACM*.
21. Moazzeni, A., Haffar, M. A. (2015). Artificial intelligence for lithology identification through real-time drilling data. *Journal of Earth Science and Climatic Change*, 6(3), 1–4.
22. Agrawal, R., Malik, A., Samuel, R., Saxena, A. (2021). Real-time prediction of litho-facies from drilling data using an artificial neural network: A comparative field data study with optimizing algorithms. *Journal of Energy Resources Technology*, 144(4), 043003.
23. Mahmoud, A. A., Elkatatny, S., Al-AbdulJabbar, A. (2021). Application of machine learning models for real-time prediction of the formation lithology and tops from the drilling parameters. *Journal of Petroleum Science and Engineering*, 203, 108574.
24. Khalifa, H., Tomomewo, O. S., Ndulue, U. F., Berrehal, B. E. (2023). Machine learning-based real-time prediction of formation lithology and tops using drilling parameters with a web App Integration. *Eng*, 4(3), 2443–2467.
25. Elkatatny, S., Al-AbdulJabbar, A., Mahmoud, A. A. (2024). Automated real-time prediction of geological formation tops during drilling operations: an applied machine learning solution for the Norwegian Continental Shelf. *Journal of Petroleum Exploration and Production Technology*, 14, 1661-1703.
26. Hansen, T. F., Erharter, G. H., Liu, Z., Torresen, J. (2024). A comparative study on machine learning approaches for rock mass classification using drilling data. *Applied Computing and Geosciences*, 24, 100199.
27. Azar, J. J., Samuel, G. R. (2007). Drilling engineering. PennWell Books.