Journal of Mining and Earth Sciences

Website: https://jmes.humg.edu.vn

Data-driven analysis of well logging data for the coal mining

Duong Hong Vu, Hung Tien Nguyen*, Vinh The Nguyen

Hanoi University of Mining and Geology, Hanoi, Vietnam

ARTICLE INFO

Article history:
Received 27th Mar. 2025
Revised 30th Jun. 2025
Accepted 15th Jul. 2025

Keywords:
Coal mining,
Data-driven,
Machine learning,
Well logging.

ABSTRACT

Coal remains one of the most widely utilized fossil fuels globally, playing a crucial role in energy production and industrial processes. As global energy demands continue to rise, the efficient and sustainable exploitation of coal resources has become increasingly important. Efficiency can be significantly enhanced through the application of geological and geophysical methods, among which well-logging holds particular significance due to its ability to provide detailed subsurface information. Well-logging data, when properly analyzed and interpreted, offer critical insights into the geological and stratigraphic characteristics of coal-bearing formations. These insights are essential for constructing accurate geological models, which, in turn, ensure that coal extraction is conducted safely, efficiently, and within planned timelines. In recent years, the integration of artificial intelligence (AI) and machine learning (ML) techniques into geoscientific workflows has opened new avenues for data-driven decision-making. These technologies are particularly valuable in handling the vast and complex datasets generated during coal assessment, exploration, and discovery. By identifying patterns and relationships within the data, ML models can enhance predictive accuracy and reduce the reliance on manual interpretation. This study applied several machine learning algorithms to predict coal seam depth and thickness using well-logging data collected from the X mine site in Quảng Ninh Province. The final model demonstrated consistently strong predictive performance when validated against actual well data, accurately identifying lithological boundaries and coal-bearing intervals. These encouraging outcomes highlight the potential of advanced computational techniques to significantly enhance coal seam characterization, offering more efficient, accurate, and cost-effective alternatives to traditional exploration methods.

*Corresponding author
E - mail: nguyentienhung.dk@humg.edu.vn
DOI: 10.46326/JMES.2025.66(4).06

1. Introduction

At most coal mines in Vietnam-including Mine X in Quảng Ninh Province-coal seam thickness has traditionally been estimated through exploratory drilling and core sampling. However, field challenges such as active mining galleries, water loss, protruding rock formations, and excessively thin coal seams often hinder core recovery (Nguyen et al., 2024). In situations where core samples are absent or fragmented, well-logging techniques emerge as the most viable method for determining seam depth and thickness. These geophysical measurements provide consistent, objective data independently of core quality, offering a reliable solution where conventional sampling is compromised.

Well-logging data not only support seam detection but also allows engineers to assess drilling efficiency, monitor physical changes during drilling, and delineate subsurface lithostratigraphy. In the broader context of coal resource development-particularly for deeper seams-geological and geophysical investigations play a foundational role in exploration, assessment, and reserve estimation. Among these methods, well-logging geophysics (WLG) stands out as particularly crucial due to its non-destructive nature and detailed subsurface characterization (Chatterjee and Paul, 2012; Miller and Mackey, 1980).

The accurate estimation of coal reserves, which depends heavily on seam thickness in each borehole, relies on well-logging measurements. Therefore, determining the depth and thickness of coal seams based on geophysical parameters measured in boreholes is of critical importance for mine planning and safety (Hatherly, 2013).

On average, annual coal exploration drilling projects in Vietnam's northeastern region involve a substantial workload. The number of boreholes continues to rise; however, well-logging geophysical data are often processed and interpreted by different engineers, leading to subjective results and time-consuming efforts. Consequently, researching new technologies to shorten processing times and enhance the accuracy and objectivity of well-logging data interpretation is highly necessary. Such advancements would significantly improve the determination of coal seam depth and thickness, contributing to more effective coal exploration and exploitation.

Well logging is a geophysical technique used to record measurements that reflect the physical properties of subsurface formations. When combined, these logs create a comprehensive dataset that reflects the lithological and petrophysical characteristics of coal-bearing sequences (Srinaiah et al., 2018). Traditionally, coal seam identification relied on manual thresholding and cross-plotting of log responses, guided by geological expertise. For instance, gamma ray logs measure natural radioactivity, which is typically low in coal due to its organic composition-making it a useful indicator for identifying coal seams (Jalil et al., 2015). Density logs, which record bulk density, help distinguish coal from denser surrounding rocks such as sandstone or shale (Xianjie et al., 2013). However, these methods are often subjective, struggle with noisy or complex data, and are inefficient when applied to large datasets from multiple wells.

The emergence of machine learning (ML) offers a transformative solution. ML techniques enhance the accuracy, efficiency, and automation of coal seam classification by identifying patterns that may be overlooked by conventional methods. This advancement addresses the limitations of traditional approaches and opens new possibilities for subsurface geological interpretation. With the increasing volume of geological and well-logging data collected from coal mines, scientists worldwide, particularly in developed countries-have been researching the application of artificial intelligence and machine learning to fully leverage this vast dataset for coal assessment, exploration, and discovery (Zhou et al., 2016; Wood et al., 1983; Wood & Cai, 2022; McLean, 2015). Numerous studies have focused on developing new processing algorithms and automated models for detecting coal seams based on geophysical parameters measured in exploration boreholes (Shi et al., 2020; Zhou & Guo, 2020; Keskinsezer, 2019; Maxwell et al., 2021). From these recent global publications, it is evident that the application of machine learning and artificial intelligence techniques to determine the depth, thickness, and characteristics of coal seams from geophysical well-logging data is a

highly regarded research direction. The results of these studies highlight the superior effectiveness of these methods, significantly enhancing the accuracy of predictive models while saving considerable time in data processing and interpretation. Overall, the application of machine learning to classify coal seams from well-logging data holds great promise for the Vietnamese mining industry. As the country continues to expand its coal exploration efforts, incorporating data-driven methods will not only reduce manual workload and subjective bias but also improve interpretation accuracy. Given the success of such approaches in international studies, this research direction is both timely and relevant for improving coal exploration practices in Vietnam.

2. Database

2.1. Overview of data

The database in this study consists of well-logging data from two wells at the X coal mine with various parameters such as Natural Gamma, Resistivity, Caliper, and Density. The amount of this data is not overly abundant, but it ensures the availability of the basic log curves necessary to address the task of determining the depth of coal seams. The quality of these log curves is high, with clear signals and minimal noise, allowing for reliable interpretation and analysis. The consistency in data acquisition methods and the stability of measurement tools across both wells contribute to a uniform and reliable dataset. This consistency is essential when applying machine learning techniques, as it minimizes biases caused by variations in data quality or instrumentation. Moreover, the good quality of the logs enhances the performance of classification models by providing distinct and interpretable signals corresponding to different lithologies.

The limited number of logging curves in this study is primarily due to the nature of solid mineral exploration. Unlike oil and gas wells, which are often deep, heavily funded, and equipped with a wide range of sophisticated logging tools, solid mineral exploration wells-especially those drilled onshore and targeting shallow depths of only a few hundred meters-tend to focus on acquiring only the most essential measurements. These basic logs are usually sufficient for identifying key lithological features, particularly coal seam boundaries, without the need for more complex or costly logging techniques. Budget constraints, equipment availability, and practical field conditions all contribute to this focused and efficient logging strategy commonly adopted in coal exploration projects.

Despite the relatively modest size of the dataset, the coverage of essential geophysical logging parameters makes it suitable for detailed subsurface characterization. Natural Gamma logs provide valuable information about the shale content and help in identifying lithological boundaries. Resistivity logs, on the other hand, are crucial in differentiating between coal and surrounding rock formations due to their contrasting electrical properties. Caliper logs offer insights into borehole conditions and help assess the reliability of other measurements. Density logs are particularly important in identifying coal seams, as coal typically has a lower density compared to most sedimentary rocks.

In conclusion, although the dataset is limited in quantity, its completeness in terms of key logging parameters and high signal quality make it a strong foundation for developing and testing models aimed at identifying coal seams. This data plays a critical role in enhancing the understanding of subsurface geological structures and supports more accurate and efficient exploration and resource assessment.

According to the sample data, three main lithological groups were identified in the wells: sand, shale, and coal, with thicknesses ranging from several centimeters to meters (Figure 1). It is apparent that the distribution of rock layers in the two wells is quite complex, with rock layers tending to interweave and having relatively thin thicknesses. Therefore, it can be concluded that although these wells are in the same mining area, the depth intervals of the rock layers are quite different, making it difficult to predict the lithological column of a new well without sample data.

2.2. Detect and remove the outliers in database

Figure 2 presents the frequency histogram and boxplot of the five well-logging parameters. These charts provide a visual representation of

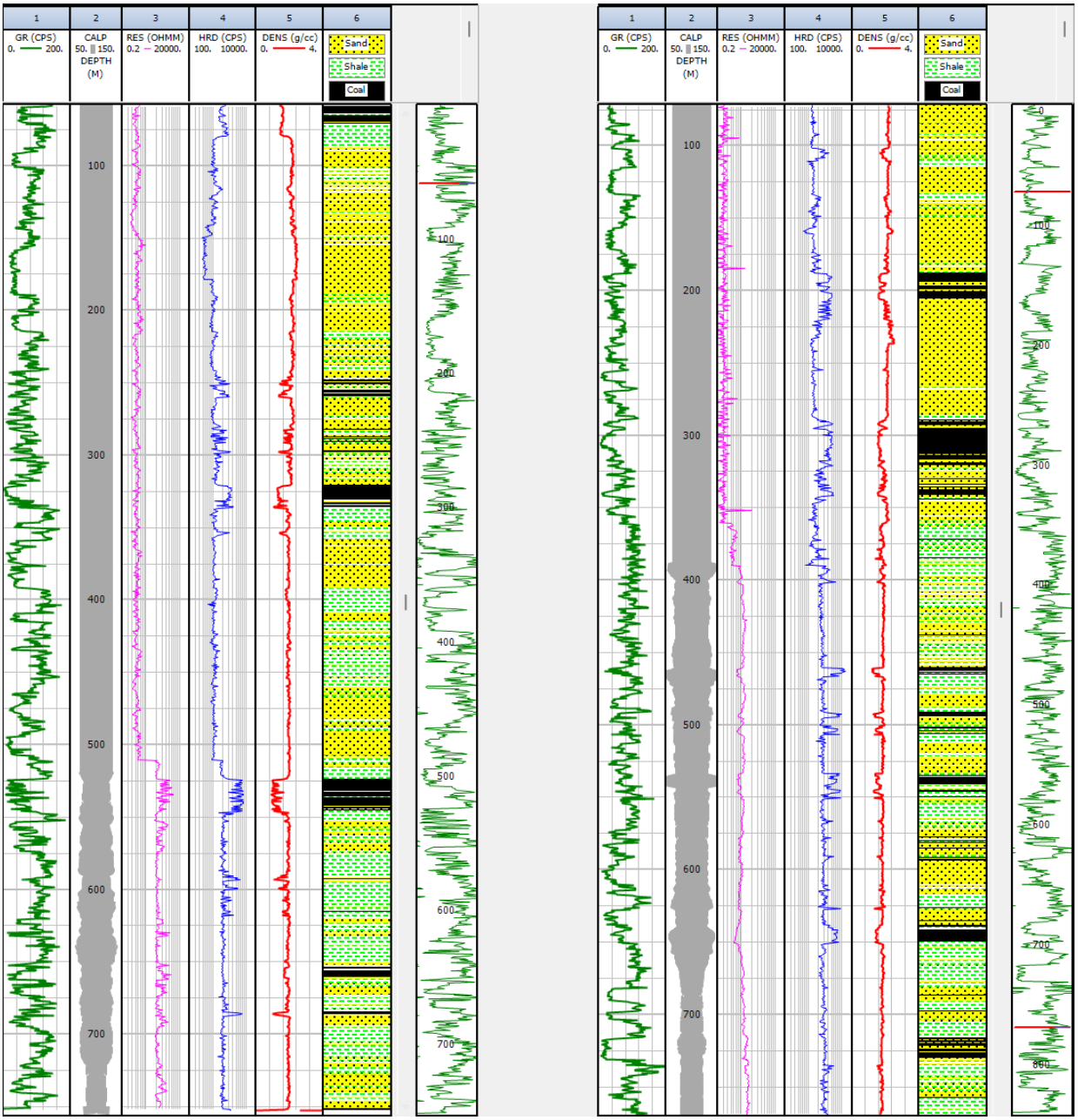


Figure1. Well logging data of Well A and Well B.

the distribution characteristics of the dataset. The horizontal axis (X-axis) illustrates the range of the logging parameter values, while the vertical axis (Y-axis) describes the frequency of data points within a range. In the chart, the Y-axis quantifies the number of occurrences for specific value ranges in the dataset.

In the sphere of geophysical data analysis, it can be said that well logging datasets prove to be quite problematic due to their vulnerability to noise and measurement errors, and the distributions often exhibit a skew. For example, gamma ray intensity, density, or resistivity may display rather abnormal behaviors due to certain geological heterogeneities, borehole conditions, or equipment malfunctions. Traditional statistical methods dependent on means and (especially) standard deviations lose much of their efficiency under such conditions because of the considerable influence of the extremes and

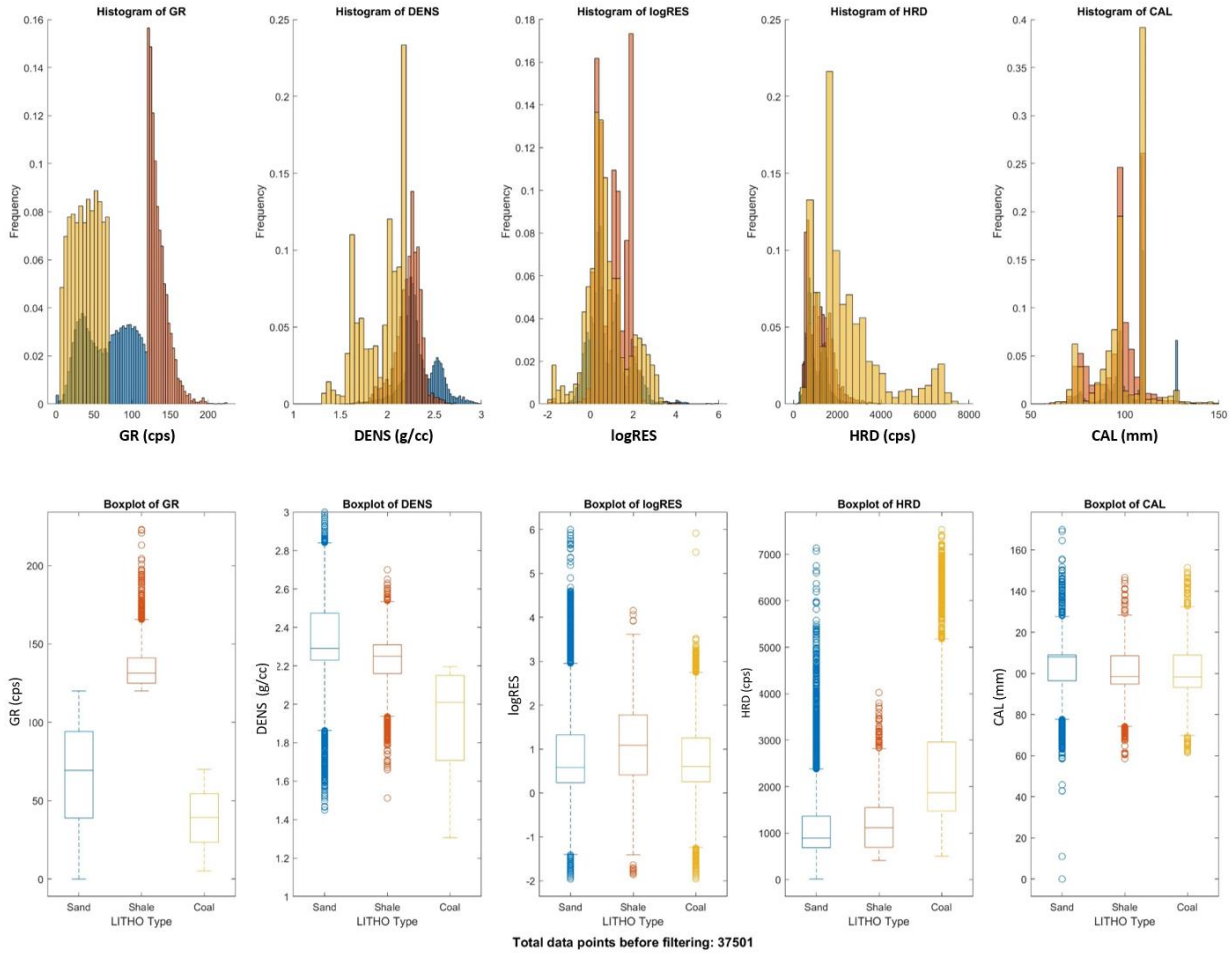


Figure 2. Raw dataset.

Gaussian distribution assumptions. Another technique, Robust Statistics, offers a much stronger alternative in such a situation to the extent that their estimators use quantities like the median and Median Absolute Deviation (MAD) to signal and remove the outliers so that the data remains clean for the geological interpretations that follow (Hampel et al., 1986). This method computes the Modified Z-score:

$$Mi = 0.6745 \times (xi - median(x))/MAD \quad (1)$$

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 flagged as outliers and excluded from further analysis.

In this study, the authors employed the Robust Statistics method to eliminate outliers, with the results presented in Figure 3.

Statistical analyses would play a prime role in the present work by explaining the dataset before any model development. They would detect outliers and help establish feature distributions of such factors as DENS, GR, HRD, and RES. For minimum values, maximum values, standard deviations, and percentiles (P10, P50, P90) of the data, the results have to be cross-checked for consistency and reliability (Table 1).

Based on sample data from wells A and B, the rock layers here are classified into three main lithological types: predominantly sand (52.5%), shale (25.2%), and coal, which has a relatively small proportion (22%) (Figure 4).

2.3. Feature selection

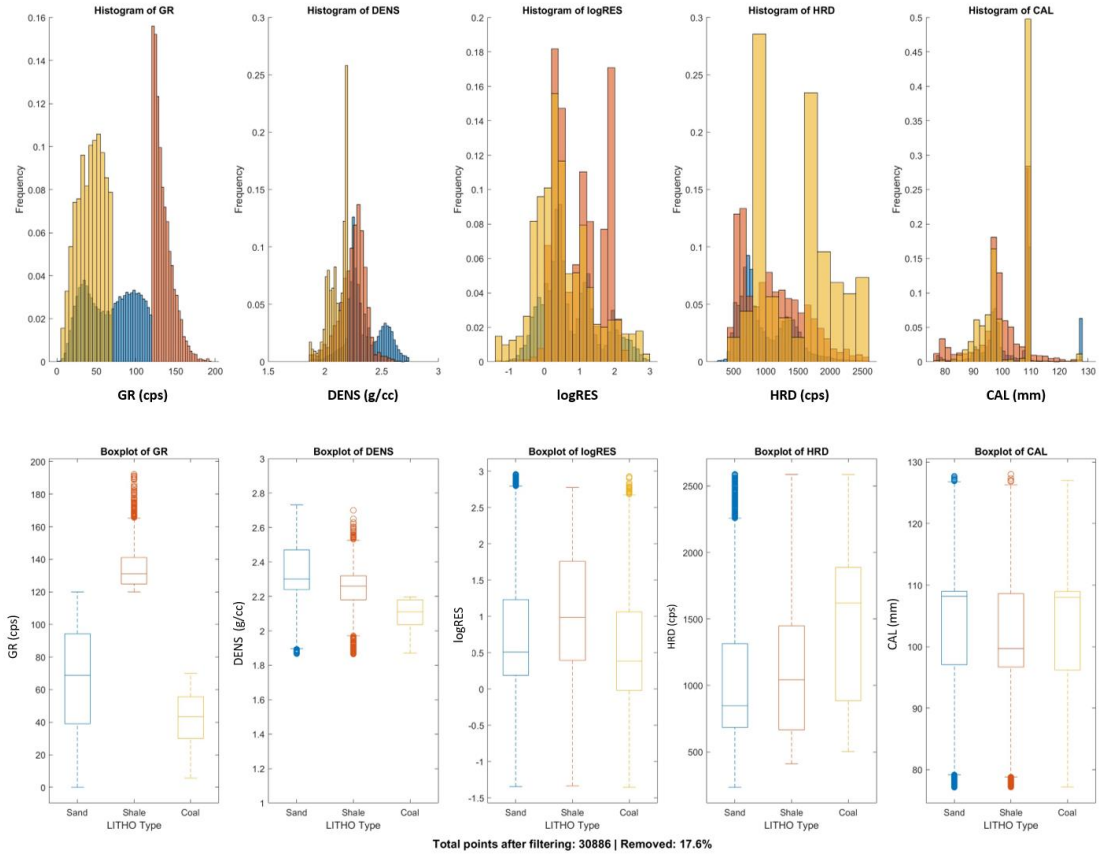


Figure 3. Dataset after removing outliers.

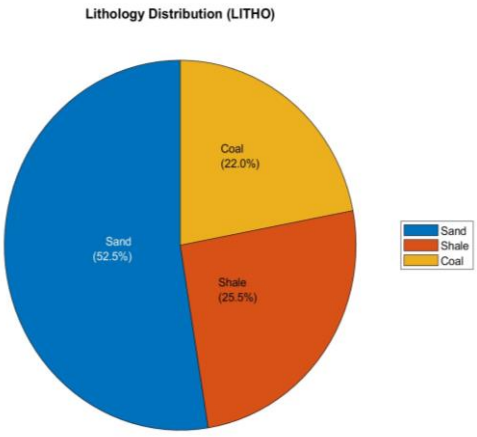


Figure 4. Proportion of lithological samples from the dataset.

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

Statistic	CAL	DENS	HRD	GR	log10RES
Min	77.19	1.863	232.94	2.34	-1.3565
Max	127.69	2.735	2589.6	192.16	2.9585
Mode	127	2.25	517.65	100	-0.16749
Std Dev	9.5863	0.16265	439.62	36.199	0.79507
P10	92.062	2.13	564.22	27.05	-0.17393
P50	108	2.29	889.09	72.58	0.52114
P90	109.41	2.56	1623.8	121.8	1.9522

In this study, feature selection was conducted using the Minimum Redundancy Maximum Relevance (MRMR) algorithm to identify the most significant well log parameters for classifying Sand, Shale, and Coal formations. The MRMR algorithm selects features that are highly relevant to the target variable while minimizing redundancy among selected features, ensuring an optimal balance between information gain and feature independence. The MRMR-based feature

ranking revealed that Gamma Ray (GR), Density (DENS), and Resistivity (RES) were the most influential features for lithology classification (Figure 5).

GR effectively differentiates shale from sand and coal, while DENS is crucial for identifying coal due to its significantly lower density. RES and HRD can provide additional information to distinguish between different rock formations. Since the importance score of Caliper (CAL) was relatively low, it was excluded from the final model to enhance computational efficiency and prevent the inclusion of less informative features.

3. Research Methods and Results

The objective set by the authors is to try applying and evaluating the effectiveness of machine learning techniques in solving the problem of predicting lithological layers for wells A and B at the X coal mine based on well-logging parameters, including Natural Gamma (GR), Density Gamma (DENS), Resistivity (RES), CALIPER (CAL), along with the distribution information of lithological groups (sand, shale, and coal) by vertical depth. This can be classified as a classification problem, which can be effectively handled by supervised machine learning techniques. To solve the task, the authors applied two algorithms-K-Nearest Neighbors (KNN) and Random Forest then evaluated their performance before proposing a final model (Figure 6).

K-Nearest Neighbors (KNN) and Random Forest have very stark differences in both theoretical background and practical operation. The comparison will, therefore, highlight these differences, all from the perspective of classification for geophysical well logging data in coal exploration. KNN method is an instance-based, lazy learning algorithm that predicts the label of a new data point by finding the k-nearest neighbors in the feature space and assigning the majority class among them-usually through majority voting. It is unlike most conventional algorithms in that it does not have a distinct training phase, requiring the algorithm to store the entire training dataset for real-time computation. On the other hand, Random Forest is an ensemble learning model-based method. It will create a number of decision trees during training based on the bootstrap sampling and random feature selection process, and then take the average/mean of these predictions. The method needs a strong training phase to build a forest of trees.

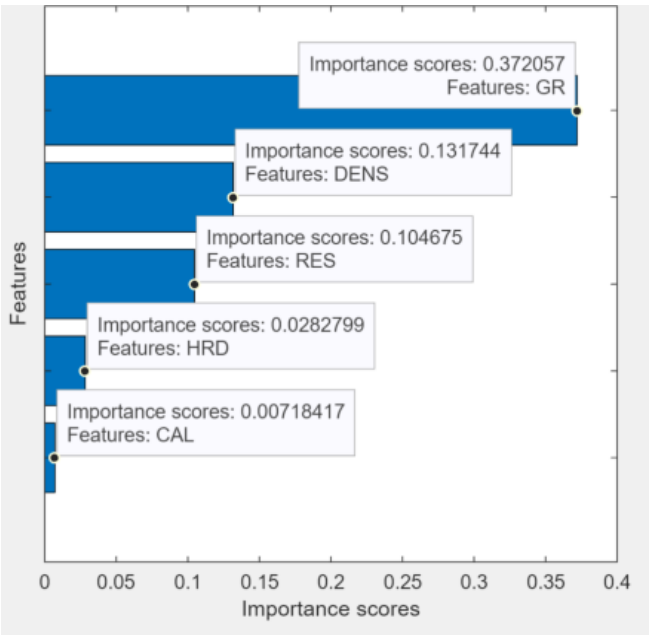


Figure 5. Feature importance score sorted.

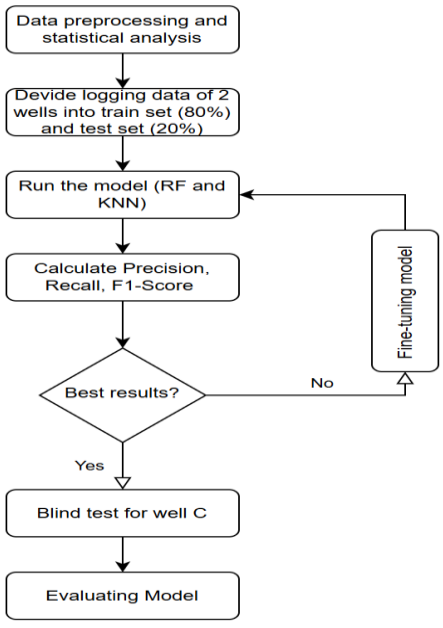


Figure 6. Workflow of this study.

During the development of classification models, they were optimized (fine-tuned) by varying approaches of changing hyperparameters in the process of training to track the best model that has the best performance according to the predefined objective function. In this paper, the Grid Search algorithm was used to figure out the best model based on how well different models predict the occurrence of default when adjusting hyperparameters. The final optimum hyperparameters for each model are summarized in Table 2.

Model performance was estimated using the following evaluation metrics. From the confusion matrix (True Positive - TP, True Negative - TN, False Positive - FP, False Negative - FN), the main estimates are as follows:

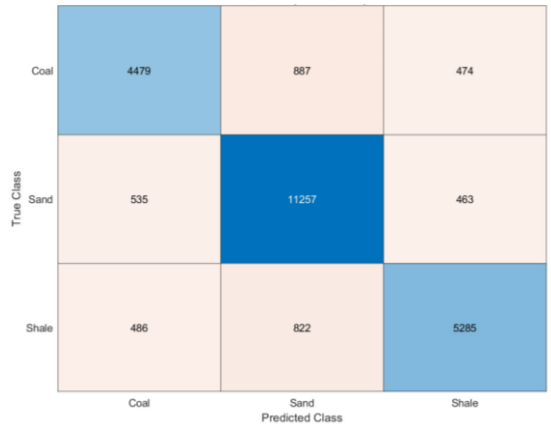
Table 2. Optimum set of parameters for models.

Model	Hyper parameter
KNN	- n_neighbors= 14 - Weights= distance - Metric= Cosine
Random forest	- n_estimators = 50 - max_depth = 5 - min_samples_split = 2 - min_samples_leaf = 1

- $\textbf{Precision}_i = \frac{TP_i}{TP_i+FP_i}$
- $\textbf{Recall}_i = \frac{TP_i}{TP_i+FN_i}$
- $\textbf{F1i} = 2\frac{Precision*Recall}{Precision + Recall}$

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.

The classification results indicate that the Random Forest (RF) model performs better than the K-Nearest Neighbors (KNN) algorithm (Figures 7,8). In this multi-class lithology classification problem, the confusion matrix is a 3×3 table, represents the number of lithology classes (Shale, Sand, Coal). This matrix serves as a valuable tool for evaluating the performance of classification models by comparing actual labels with predicted ones. Each row of the matrix corresponds to the actual class, while each column represents the predicted class. The diagonal elements indicate the number of correctly classified instances for each class, whereas the off-diagonal elements reflect misclassifications, where instances are



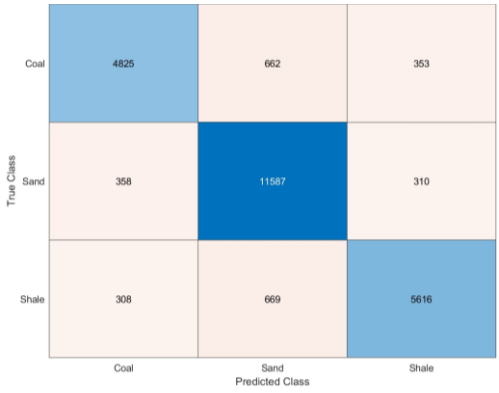
Train data

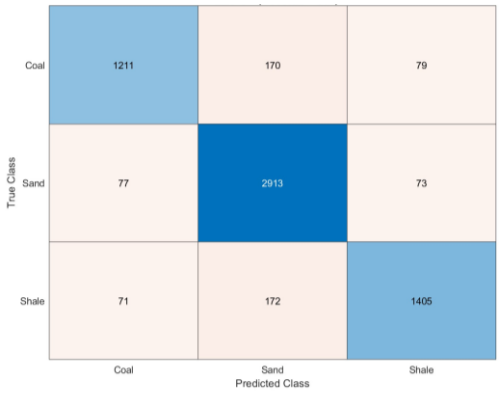
Lithology	Precision	Recall	F1 score
Sand	0.86	0.92	0.89
Shale	0.85	0.80	0.82
Coal	0.81	0.77	0.79

Test data

Lithology	Precision	Recall	F1 score
Sand	0.87	0.91	0.89
Shale	0.86	0.81	0.83
Coal	0.82	0.78	0.8

Figure 7. Confusion matrix of KNN model.





Train data

Lithology	Precision	Recall	F1 score
Sand	0.9	0.94	0.92
Shale	0.89	0.85	0.87
Coal	0.88	0.83	0.85

Test data

Lithology	Precision	Recall	F1 score
Sand	0.89	0.95	0.92
Shale	0.9	0.85	0.87
Coal	0.89	0.83	0.86

Figure 8. Confusion matrix of Random forrest model.

incorrectly assigned to other classes. Conversely, higher values in the off-diagonal cells suggest confusion between classes. The confusion matrix not only provides a comprehensive overview of model performance but also helps identify specific class-level misclassifications. This insight is instrumental in refining model accuracy and is often used in conjunction with evaluation metrics such as precision, recall, and F1-score. From the evaluation of the confusion matrices pertaining to both the training set and the test set, RF excels in terms of precision, recall and F1-score for all lithology classes, especially in the identification of the coal layer.

In the case of the training set, RF has a much lower misclassification rate than KNN, with the F1 score of coal to be 0.85 in RF and 0.79 in KNN, indicating that Precision and Recall are better with RF. This improvement is due to RF's use of multiple decision tree aggregations, which manage complex decision boundaries and high-dimensional data, therefore reducing noise and increasing the prediction accuracy. On the other hand, KNN always depends on distances, which makes it scale sensitive and occupies the class distributions, resulting in more misclassification, particularly those lying at the borders between shale and coal. On the test set, RF contrarily does not lose generalization performance; coal layer recall remains RF at 0.86 and KNN at 0.8. The confusion matrix also indicates that KNN makes more mistakes than RF in assigning coal layers to the wrong class of shale, RF being set at low gamma ray and density overlaps. On the other hand, RF's feature importance weighting allows it to differentiate coal from surrounding lithologies more effectively. These findings emphasize the dominating dependability of RF regarding coal seam classification, thus reinforcing the use of RF in geological interpretation and resource evaluation.

The last stage in assessing how this classification model can be used in practice is the Blind Test, which consists of providing data from a new well (well C) in the study area. This well's data has not been introduced in any form throughout the model building exercise, thus guaranteeing non-partial evaluation of the model in practice. Blind testing is very critical in the evaluation and validation of a machine learning model's performance. It guarantees that the model is not simply reproducing the training data, but can correctly predict new data. The procedure helps identify overfitting, which is when a model performs exceptionally well on training data but does not do well on real data. Blind testing also helps provide impartial evaluation by preventing pretest data exposure bias. It further checks the

performance of the model in practical situations, which can be the case where the distribution of data is different from the training set. The results of the blind test on Well C's data are shown in Figure 9.

To accurately identify and distinguish coal seams from surrounding lithological formations, geoscientists have traditionally relied on a suite of well log measurements, including gamma ray, density, resistivity, and sonic logs. Coal seams are characterized by distinct geophysical signatures: they generally exhibit low gamma ray readings (commonly below 50 API) due to their organic composition and minimal natural radioactivity.

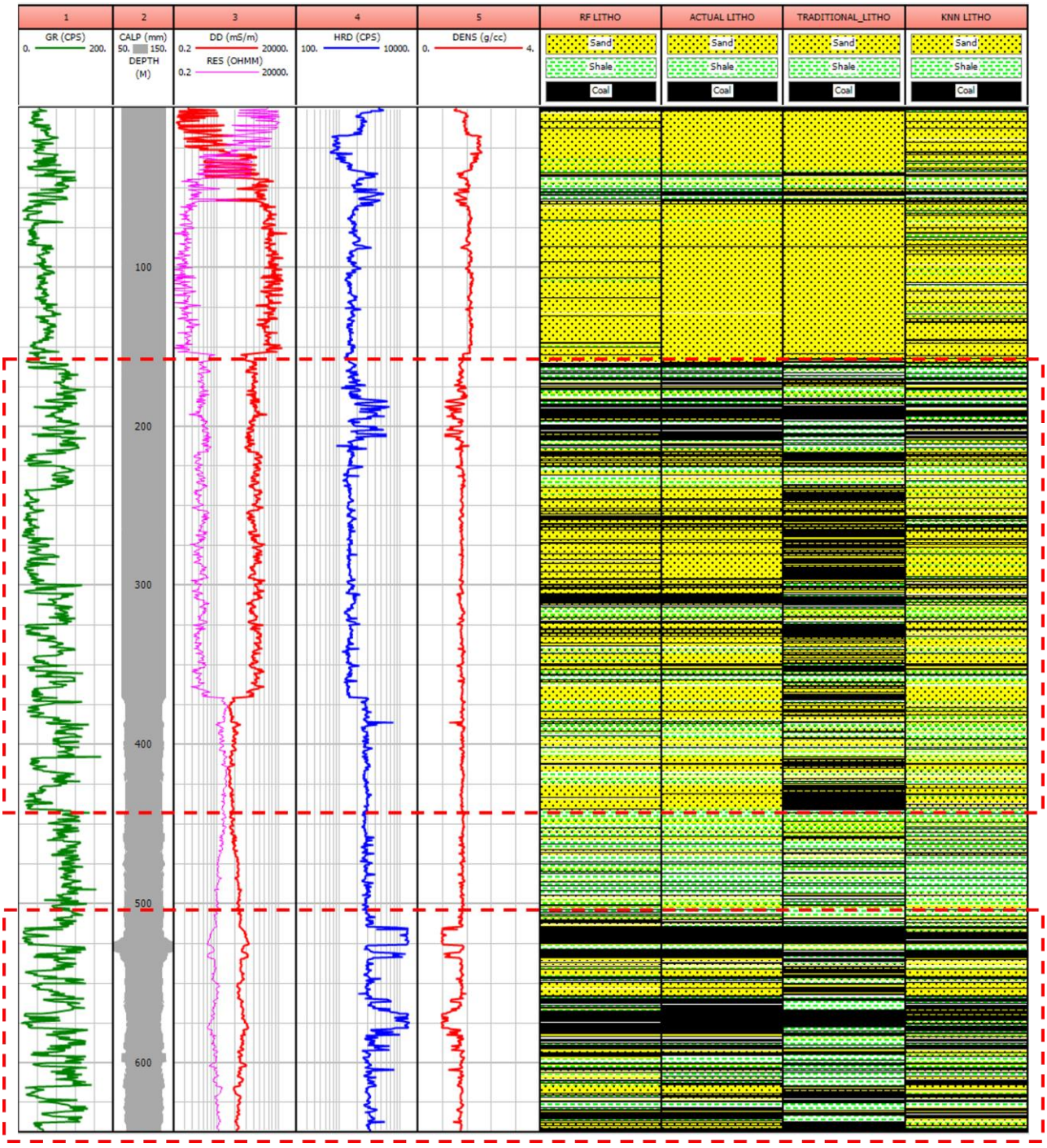


Figure 9. Blind test for Well C.

Additionally, coal has a low bulk density, typically ranging from 1.2 to 1.8 g/cm^3, and high resistivity, especially under dry conditions, owing to its poor electrical conductivity (Jalil et al., 2015). In contrast, sandstone formations display moderate gamma ray values (approximately 20÷80 API), higher bulk densities (around 2.0÷2.6 g/cm^3), and resistivity values that vary depending on fluid saturation. Sonic logs for sandstone usually indicate intermediate acoustic travel times, reflecting moderate seismic velocities. Shale formations, on the other hand, are identifiable by their high gamma ray responses (often exceeding 100 API), which are indicative of radioactive minerals such as potassium-bearing clays. Shale typically has a density higher than coal but lower than quartz-rich sandstone, and its resistivity is generally low due to its fine-grained texture and higher clay and water content (Xianjie et al., 2013). By integrating these diverse log measurements, geologists can construct robust lithological models and delineate boundaries between different rock types. Cross-plotting techniques-such as gamma ray versus density-further enhance lithological discrimination. In the case of the blind test well, traditional interpretation methods were employed to generate a lithological column representing the vertical distribution of coal, sandstone, and shale. This column was subsequently validated against core sample data and compared with lithology predictions produced by the machine learning models proposed in this study.

The blind test results from well C further confirm the superiority of the Random Forest (RF) model over both the K-Nearest Neighbors (KNN) algorithm and traditional lithological interpretation methods in identifying coal seams. The RF model achieved an accuracy of 87.9%, while KNN reached only 72.4%, and the traditional method yielded a significantly lower accuracy of just 54.7% when compared with core data. Two labeled depth intervals, highlighted in Figure 9, illustrate that while the traditional method tends to overestimate the thickness of coal seams, the KNN model frequently misclassifies coal as shale-likely due to overlapping gamma ray and density values. In contrast, the RF model effectively distinguishes all lithologies by leveraging feature importance and ensemble decision trees, thereby reducing data noise. These findings reinforce the reliability and robustness of the RF model for lithological classification in coal exploration.

4. Conclusions and/or Recommendations

This study proves that machine learning-based lithology classification is more effective than attempts for traditional coal seam identification. Classical methods, including well log manual interpretation or mere empirical threshold division, are often vague, subconscious, and faulty when faced with multilayered lithological changes. Contrary to this, the RF and KNN models developed in this study are more precise and can automatically serve the purpose of distinguishing Sand, Shale, and Coal formations in a scalable manner.

Following an optimization with Grid Search, the RF model outperformed other classifiers with exceptionally high accuracy and reliability, far surpassing KNN and traditional methods. Throughout the tests, RF was able to more accurately recognize coal seams (F1-score = 0.86 vs 0.80 for KNN) as well as classify both sand (F1-score = 0.92 vs 0.89) and shale (F1-score = 0.87 vs 0.83). Unlike KNN and more traditional approaches, which tend to fail in complicated geological settings, RF uses an ensemble of decision trees to detect complex structures in well log data that make it more powerful and robust. In order to test the practical usability of the method RF, a blind test was conducted on a new well (well C) that had no prior associated data from model training and the results prove that RF is able to generalize quite well to unseen data.

This research demonstrates the role of machine learning tailored towards seam recognition and marks a distinct shift from dependency on classical methods. Further works will focus on extending the scope of the dataset that includes more well log attributes, along with the application of deep learning for further refinement of classification precision and generalization to various geological settings.

Acknowledgement

This research is supported by Hanoi University of Mining and Geology under grant number T25-07.

Contributions of authors

Duong Hong Vu - methodology, writing, review and editing; Hung Tien Nguyen - writing, review and editing; Vinh The Nguyen – supervision.

References

Chatterjee, R., & Paul, S. (2012). Application of cross-plotting techniques for delineation of coal and non-coal litho-units from well logs. *Geomaterials*, *2*(4), 94-104.

Hatherly, P. (2013). Overview on the application of geophysics in coal mining. *International Journal of Coal Geology*, *114*, 74-84.

Shi, J., Zeng, L., Dong, S., Wang, J., & Zhang, Y. (2020). Identification of coal structures using geophysical logging data in Qinshui Basin, China: Investigation by kernel Fisher discriminant analysis. *International Journal of Coal Geology*, *217*, 103314.

Maxwell, K., Rajabi, M., & Esterle, J. (2021). Spatial interpolation of coal properties using geographic quantile regression forest. *International Journal of Coal Geology*, *248*, 103869. https://doi.org/10.1016/j.coal.2021.103869.

Keskinsezer, A. (2019). Determination of coal layers using geophysical well-logging methods for correlation of the Gelik-Zonguldak and Kazpınar-Amasra (Bartın) coalfields, Turkey. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, *5*, 223-235. https://doi.org/10.1007/s40948-019-00105-4.

Miller, M. S., & Moore, M. (1980, September). Geophysical logging and exploration techniques in the Appalachian coal fields. In *SPE Annual Technical Conference and Exhibition?* (pp. SPE-9466). SPE. https://doi.org/10.2118/9466-MS.

Nguyen, P., Nguyen, H. Hoang, Do, K. Xuan and Nguyen, D. Phuong (2024). The exploratory work of the Dong Bac coal basin - Current situation and solutions (in Vietnamese). Mining Industry Journal. XXXIII, 3 (Jun, 2024), 50-60.

Jalil, S., & Rashid, M. (2015). Analysis of natural radioactivity in coal and ashes from a coal fired power plant. *Chemical Engineering Transactions*, *45*, 1549-1554.

Xianjie, S. H. A. O., Yubo, S. U. N., Jingmin, S. U. N., Dazhen, T. A. N. G., Hao, X. U., Xinxiu, D. O. N. G., & Yumin, L. Ü. (2013). Log interpretation for coal petrologic parameters: A case study of Hancheng mining area, Central China. *Petroleum Exploration and Development*, *40*(5), 599-605. https://doi.org/10.1016/S1876-3804(13)60078-6.

Srinaiah, J., Raju, D., Udayalaxmi, G., & Ramadass, G. (2018). Application of well logging techniques for identification of coal seams: a case study of Auranga coalfield, Latehar District, Jharkhand state, India. *J Geol Geophys*, *7*(1), 1-11.

Zhou, B., & Guo, H. (2020). Applications of geophysical logs to coal mining—some illustrative examples. *Resources*, *9*(2), 11. https://doi.org/10.3390/resources9020011.

Zhou, B., & O'Brien, G. (2016). Improving coal quality estimation through multiple geophysical log analysis. *International Journal of Coal Geology*, *167*, 75-92.

Wood, D. A., & Cai, J. (2022). Coal-bed methane reservoir characterization using well-log data. In *Sustainable geoscience for natural gas subsurface systems* (pp. 243-274). Gulf Professional Publishing.

Wood Jr, G. H., Kehn, T. M., Carter, M. D., & Culbertson, W. C. (1983). *Coal resource classification system of the US Geological Survey* (No. 891). US Geological Survey.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). Robust statistics: The approach based on influence functions. NY: John Wiley & Sons. DOI:10.1002/9781118186435.

McLean, C. R. (2015). Pseudo proximate analysis: Methode using wireline logs to estimate component of coal bearing rock matrix without control data. *Minithesis submitted in University of the Western Cape*.