UDC 550.822
http://doi.org/10.18799/24131830/2026/4/5066
Scientific paper

Young modulus prediction using drilling data at CT oil field, offshore Vietnam

Vu Hong Duong[1], Nguyen Tien Hung[1]✉, Vu Thiet Thach[1], Nguyen Minh Hoa[1], Nguyen Xuan Duy[2]

[1] *Hanoi University of Mining and Geology, Hanoi, Vietnam*
[2] *Vietnam Petroleum Institute, Hanoi, Vietnam*

✉nguyentienhung.dk@humg.edu.vn

Abstract. ***Relevance.*** Young modulus prediction from drilling data is highly relevant to enhancing the efficiency and safety of oil and gas operations, particularly in complex geological settings. This study addresses the pressing need for cost-effective, real-time geomechanical characterization methods, reducing dependence on costly traditional techniques. ***Aim.*** To develop a methodology employing machine learning and artificial intelligence to accurately predict the Young modulus using drilling parameter data from oil and gas wells at the CT oil field, offshore Vietnam. ***Objects.*** Encompass rock and soil formations encountered in oil and gas wells, focusing on drilling parameters such as weight on bit, torque, rotary speed, and rate of penetration. ***Methods.*** Advanced machine learning techniques, including deep neural networks and ensemble methods, to analyze and map the non-linear relationships between drilling data and Young modulus. Models are trained and validated using reference data from core testing and well logs, with data preprocessing applied to mitigate noise and enhance predictive accuracy. ***Results.*** By applying machine learning algorithms, the research team successfully developed models to directly predict Young modulus from drilling parameters measured in real-time during the drilling of wells at the CT oil field, offshore Vietnam. The model employing a backpropagation neural network demonstrated superior performance, achieving a correlation coefficient of up to 0.94 and an RMSE of only 0.483 when subjected to a blind test on a new well within the study area.

Keywords: Young modulus, geomechanics, drilling data, machine learning, CT field

Acknowledgements: This research is funded by Hanoi University of Mining and Geology under grant number T25-09. The authors express their profound gratitude for the invaluable sharing of academic knowledge and the enthusiastic technical support provided by the Vietnam Petroleum Institute.

For citation: Vu Hong Duong, Nguyen Tien Hung, Vu Thiet Thach, Nguyen Minh Hoa, Nguyen Xuan Duy. Young modulus prediction using drilling data at CT oil field, offshore Vietnam. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, 2026, vol. 337, no. 4, pp. 64–74. http://doi.org/10.18799/24131830/2026/4/5066

Conflict of interest: none.

УДК 550.822
http://doi.org/10.18799/24131830/2026/4/5066
Научная статья

Прогнозирование модуля Юнга по данным бурения на нефтяном месторождении СТ, шельф Вьетнама

Ву Хонг Зыонг[1], Нгуен Тиен Хунг[1]✉, Ву Тьет Тхак[1], Нгуен Минь Хоа[1], Нгуен Суан Зуи[2]

[1] *Ханойский университет горного дела и геологии, Вьетнам, г. Ханой*
[2] *Вьетнамский нефтяной институт, Вьетнам, г. Ханой*

✉nguyentienhung.dk@humg.edu.vn

Аннотация. ***Актуальность.*** Прогнозирование модуля Юнга по данным бурения имеет большое значение для повышения эффективности и безопасности нефтегазовых операций, особенно в сложных геологических условиях. Рассматривается насущная потребность в экономически эффективных методах определения геомеханических характеристик в режиме реального времени, снижающих зависимость от дорогостоящих традиционных методов. ***Целью*** исследования является разработка методологии, использующей машинное обучение и искусственный интеллект для точного прогнозирования модуля Юнга с применением данных параметров бурения скважин на месторождении СТ шельфа Вьетнама. ***Объектом*** исследования являются горные породы и грунты, встречающиеся в скважинах. Особое внимание уделяется параметрам бурения, таким как нагрузка на долото, крутящий момент, скорость вращения и механическая скорость бурения. ***Методы:*** передовые методы машинного обучения, включая глубокие нейронные сети и ансамблевые методы, для анализа и отображения нелинейных связей между данными бурения и модулем Юнга. Модели обучаются и проверяются с использованием эталонных данных, полученных в результате испытаний керна и геофизических исследований скважин, с предварительной обработкой данных для снижения шума и повышения точности прогнозирования. ***Результаты.*** Применяя алгоритмы машинного обучения, исследовательская группа успешно разработала модели для прямого прогнозирования модуля Юнга по параметрам бурения, измеряемым в режиме реального времени при бурении скважин на нефтяном месторождении СТ, шельф Вьетнама. Модель, использующая нейронную сеть обратного распространения, продемонстрировала превосходные результаты, достигнув коэффициента корреляции до 0,94 и среднеквадратичной ошибки всего 0,483 при слепом тестировании на новой скважине в пределах исследуемого участка.

Ключевые слова: модуль Юнга, геомеханика, данные бурения, машинное обучение, месторождение СТ

Благодарности: Работа выполнена при поддержке Ханойского университета горного дела и геологии, г. Ханой, код T25-09. Авторы выражают глубокую благодарность за бесценный обмен академическими знаниями и техническую поддержку, предоставленную Вьетнамским нефтяным институтом.

Для цитирования: Ву Хонг Зыонг, Нгуен Тиен Хунг, Ву Тьет Тхак, Нгуен Минь Хоа, Нгуен Суан Зуи. Прогнозирование модуля Юнга по данным бурения на нефтяном месторождении СТ, шельф Вьетнама. *Известия Томского политехнического университета. Инжиниринг георесурсов*, 2026, Т. 337, № 4, С. 64–74. http://doi.org/10.18799/24131830/2026/4/5066

Конфликт интересов: отсутствует.

Introduction

In the realm of geomechanics and subsurface engineering, understanding the mechanical properties of rock and soil formations is paramount for ensuring the safety, efficiency, and economic viability of oil and gas exploration and production operations. Among these properties, Young modulus stands out as a critical parameter that quantifies the stiffness or elasticity of geological materials under uniaxial stress [1]. Defined as the ratio of axial stress to axial strain within the elastic limit of a material, Young modulus (denoted as E) provides insights into how rocks and soils deform under applied forces, such as those encountered during drilling, hydraulic fracturing, or reservoir depletion. In the context of oil and gas wells, accurate determination of Young modulus is essential for designing stable boreholes, optimizing drilling parameters, predicting formation behavior, and mitigating geomechanical risks such as wellbore instability, sand production, and fault reactivation. Despite its importance, obtaining reliable estimates of Young modulus across heterogeneous subsurface formations remains a significant challenge, necessitating innovative approaches that leverage both traditional methodologies and cutting-edge technologies [2].

Young modulus, fundamentally, reflects the ability of a material to resist deformation under stress, making it a cornerstone of rock mechanics and engineering geology. Mathematically, it is expressed as $E=\sigma/\varepsilon$, where σ represents the stress (force per unit area) and ε denotes the strain (relative deformation) [3]. For rocks and soils, Young modulus varies widely depending on factors such as mineral composition, porosity, cementation, fluid saturation, and in-situ stress conditions. In sedimentary basins targeted for hydrocarbon extraction, Young modulus can range from a few megapascals (MPa) for unconsolidated sands to tens of gigapascals (GPa) for highly compacted carbonates or crystalline basement rocks [4]. This variability underscores the need for precise measurement or estimation techniques tailored to specific geological settings. Traditionally, Young modulus has been determined through laboratory testing of core samples or inferred from wireline logging data, such as acoustic or sonic logs [5, 6]. However, these methods are often limited by cost, availability, and spatial resolution, prompting researchers and engineers to explore alternative data sources, such as drilling parameters, and advanced computational tools, including machine learning and artificial intelligence, to enhance prediction accuracy.

The urgency of developing reliable methods to predict Young modulus stems from the increasing complexity of modern oil and gas projects. As exploration moves toward deeper reservoirs, unconventional resources (e. g., shale gas and tight oil), and geologically challenging environments, the demand for real-time, cost-effective geomechanical characterization has surged. Drilling operations generate vast amounts of data, including measurements of weight on bit (WOB), rotary speed (RPM), rate of penetration (ROP), torque, and mud pressure, collectively referred to as drilling parameters. These data, historically used to monitor and optimize drilling performance, are now recognized as potential proxies for inferring subsurface mechanical properties. The ability to predict Young modulus directly from drilling data could revolutionize geomechanical analysis by reducing reliance on expensive core sampling or extensive logging campaigns, thereby accelerating decision-making and lowering operational costs. Moreover, such predictions could enable continuous, high-resolution profiling of formation stiffness along the well trajectory, offering a dynamic understanding of subsurface behavior that static measurements cannot provide.

Historically, the determination of Young modulus in the oil and gas industry has relied on a combination of direct and indirect methods. Direct measurement involves laboratory testing of core samples retrieved from boreholes [2, 7]. In these tests, cylindrical rock specimens are subjected to controlled uniaxial or triaxial compression, and the resulting stress-strain relationships are analyzed to calculate Young modulus. While this approach provides high accuracy, it is constrained by several limitations. Core retrieval is expensive, time-consuming, and often infeasible in highly fractured or unconsolidated formations. Additionally, core samples represent only discrete points along the wellbore, failing to capture the continuous variability of subsurface properties. To address these shortcomings, indirect methods based on well logging become standard practice [8, 9]. Sonic logs, which measure the travel time of compressional (P-wave) and shear (S-wave) acoustic waves through the formation, are commonly used to estimate dynamic Young modulus via empirical correlations with rock density. However, converting dynamic modulus (derived from wave velocities) to static modulus (relevant to engineering applications) introduces uncertainties, as the two properties can differ significantly due to strain amplitude effects and formation heterogeneity. Furthermore, well logging requires specialized tools and is typically performed post-drilling, limiting its utility for real-time applications.

Given the limitations of traditional approaches, researchers have increasingly turned to drilling data as an alternative source of geomechanical information. Drilling parameters are continuously recorded during the drilling process, providing a real-time, high-resolution dataset that reflects the interaction between the drill bit and the formation. The underlying hypothesis is that variations in drilling metrics such as ROP, WOB, and torque are affected by the mechanical properties of the rock, including its stiffness. For instance, harder, stiffer formations (with higher Young modulus) typically result in lower penetration rates and higher torque, while softer, more deformable rocks exhibit the opposite behavior. Early attempts to correlate drilling parameters with Young modulus relied on analytical models, such as the specific energy concept, which relates the energy required to break rock to its mechanical strength. While these models offered a theoretical framework, their predictive power was limited by simplifying assumptions and the complex, non-linear nature of rock–bit interactions. As a result, conventional methods struggled to account for confounding factors such as bit wear, mud properties, and formation anisotropy, highlighting the need for more sophisticated approaches.

The advent of machine learning (ML) and artificial intelligence (AI) opened new avenues for predicting Young modulus from both well log and drilling data. Unlike traditional methods, which depend on predefined physical models or empirical correlations, ML algorithms can identify complex patterns and relationships within large, multidimensional datasets without requiring explicit assumptions about the underlying physics [10, 11]. In the context of well logging, supervised learning techniques – such as artificial neural networks (ANNs), support vector machines (SVMs), and random forests – were widely applied to predict Young modulus from inputs like sonic velocities, density, and porosity logs [12, 13]. These models are typically trained on datasets where laboratory-measured Young modulus values serve as ground truth, allowing the algorithms to learn the mapping between log-derived properties and rock stiffness. Studies demonstrated that ML-based predictions can outperform traditional correlations, particularly in heterogeneous formations where empirical relationships break down. Moreover, ML approaches can incorporate additional contextual data, such as depth, lithology, and fluid content, to enhance accuracy and generalizability.

Building on this foundation, the application of ML and AI to drilling data represents a frontier in geomechanical research. Drilling parameters offer a unique advantage over well logs in that they are collected in real time during the drilling process, potentially enabling on-the-fly estimation of Young modulus. However, the relationship between drilling metrics and rock properties is inherently non-linear and affected by operational variables (e. g., bit type, drilling fluid rheology) and environmental conditions (e. g., temperature, pressure). To address these challenges, advanced ML techniques, such as deep learning and ensemble methods, were employed to model the complex interplay between drilling parameters and formation stiffness [14, 15]. For example, recurrent neural networks (RNNs) and convolutional neural networks (CNNs) can capture temporal and spatial dependencies in drilling data, while gradient boosting algorithms excel at handling noisy, high-dimensional inputs [16]. Recent studies shown promising results, with AI models achieving reasonable accuracy in predicting Young modulus when calibrated against core or log-derived reference data. These advancements suggest that drilling data, when paired with intelligent algorithms, could serve as a standalone or complementary tool for geomechanical characterization.

In summary, the prediction of Young modulus from drilling parameter data in oil and gas wells represents a transformative opportunity for geomechanical analysis. By addressing the shortcomings of traditional methods and harnessing the power of machine learning and AI, this research seeks to unlock the potential of drilling data as a reliable indicator of rock stiffness. The significance of this work lies not only in its technical innovation but also in its practical implications for optimizing drilling operations, enhancing reservoir management, and reducing exploration risks. The following sections of this article will delve into the methodology, including data preprocessing, model development, and validation, as well as the results and implications of applying AI-driven techniques to predict Young modulus in real-world drilling scenarios.

Methodology and data set

The objective of this paper is to develop a machine learning model for the direct prediction of Young modulus during drilling operations at the CT oil and gas field, offshore Vietnam. The model utilizes a dataset comprising drilling parameters as input data, with Young modulus values calculated from density logs and sonic wave velocities serving as reference data to validate the model accuracy; this problem can be classified as a regression task, a typical problem effectively addressed by supervised machine learning algorithms such as Artificial Neural Networks (ANN), Random Forest and Support Vector Machine (SVM).

Artificial Neural Networks

Artificial Neural Networks (ANN) are a machine learning algorithm inspired by the structure and function of biological neural networks in the human brain, consisting of interconnected nodes organized in layers to process information [17]. The theoretical foundation of ANN relies on employing non-linear activation functions to learn complex relationships between inputs and outputs through a training process, where weights are adjusted using backpropagation and gradient optimization techniques. In the context of predicting Young modulus from well log geophysical data or drilling data, ANN excels due to its ability to model non-linear relationships between parameters such as acoustic wave velocities, density, WOB, and ROP with rock stiffness. Applications of ANN in this domain typically involve training on reference data from core tests or well logs, enabling continuous prediction of Young modulus along the well trajectory. For instance, Smith et al. [18] demonstrated that ANN could predict Young modulus from sonic log data with an error margin below 10%, outperforming traditional empirical correlations. Similarly, when applied to drilling data, ANN integrates factors like torque and mud pressure to enhance prediction accuracy. However, ANN requires substantial data volumes and computational resources, which may limit its practical deployment in some scenarios [18]. This makes ANN a powerful tool for geomechanical analysis where high predictive precision is paramount.

Random Forest

Random Forest (RF) is a machine learning algorithm rooted in ensemble learning, aggregating multiple decision trees to improve predictive accuracy and stability [19]. The theoretical basis of RF involves averaging outputs from independent decision trees, each trained on a random subset of data and features using the bagging (bootstrap aggregating) technique to mitigate overfitting. In predicting Young modulus from well log geophysical or drilling data, RF proves effective due to its capacity to handle multidimensional, noisy, and heterogeneous datasets, such as those combining acoustic velocities, porosity, and drilling parameters like ROP and RPM. RF applications in this field include forecasting Young modulus from well logs, where it can quantify the importance of input variables, aiding in the mechanical characterization of formations. Research [20] revealed that RF achieved a correlation coefficient above 0.9 when predicting Young modulus from drilling data, particularly in heterogeneous formations like shales. RF offers advantages such as ease of implementation and reduced sensitivity to parameter tuning compared to more complex models, though it may lag behind ANN in capturing deep non-linear relationships. Consequently, RF is well-suited for scenarios with diverse inputs and a need for rapid, reliable predictions, making it a practical choice in geomechanical studies.

Support Vector Machine

Support Vector Machine (SVM) is a machine learning algorithm grounded in classification and regression theory, utilizing the concept of an optimal hyperplane to separate or map data in high-dimensional space [21]. The theoretical framework of SVM hinges on maximizing the margin between classes or employing kernel functions, such as the Gaussian (RBF) kernel, to address non-linear relationships between inputs and outputs. In predicting Young modulus from well log geophysical or drilling data, SVM is applied to map parameters like shear wave velocity, density, or WOB and ROP to rock stiffness values, proving particularly effective for small to medium-sized datasets. SVM has been used to forecast Young modulus from well logs, excelling in handling features with low correlation and delivering accurate predictions despite noisy data. Lee et al. [22] reported that SVM outperformed linear methods by over 15% in predicting Young modulus from drilling data in carbonate formations. Its application extends to integrating drilling and geophysical data, leveraging its strong generalization ability to minimize errors in complex geological settings. Nevertheless, SVM faces challenges in scaling to large datasets and depends heavily on appropriate kernel selection, which can constrain its use. SVM thus offers a robust option for precise geomechanical predictions where data volume is moderate.

Data set

The CT oil field, spanning an area of 5,559 km², is located within Block 09-3/12 at the southeastern margin of the Cuu Long basin, offshore Vietnam continental shelf, approximately 160 km southeast of Vung Tau city (Fig. 1).

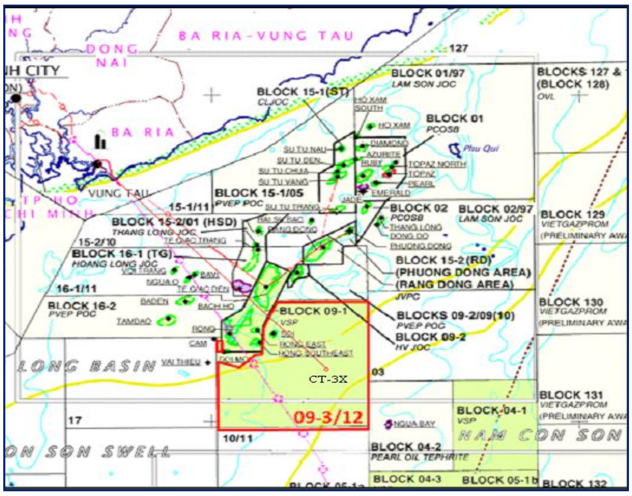

Fig. 1. *Study area (red rectangle) [23]*
Рис. 1. *Район исследования (красный прямоугольник) [23]*

It is bordered by Block 09-1 to the northwest, Block 09-2/09 to the north, Blocks 03 and 04-2 to the east, Block 10 to the south, and Block 17 to the west. The drilling targets of the CT field are the Lower Miocene and Upper Oligocene formations, characterized by highly complex geological structures, where sandstone layers are interbedded with highly reactive clay zones, posing a significant risk of drilling complications and incidents related to wellbore instability. It is imperative to develop solutions for predicting the mechanical properties of the drilled rock layers, thereby enabling the selection of appropriate technological methods and parameters to mitigate risks of complications and incidents, enhance operational efficiency, and maximize savings in both time and cost.

In this study, the dataset utilized to construct the Young modulus prediction model comprises 1061 samples from two wells, A and B, at the CT oil field, offshore Vietnam, incorporating six drilling parameters: ROP, RPM, WOB, standpipe pressure (SPP), pump flow rate (FLW), and torque (TQ) (Table). The output of the machine learning model, Young modulus, will be compared with the Young modulus curve derived from well logging data to assess the model accuracy and its potential to substitute well logging data with direct drilling data for predicting Young modulus during drilling.

Table. *Statistical analysis of drilling parameters in the dataset*
Таблица. *Статистический анализ параметров бурения в наборе данных*

Parameters Параметры	Min Мин	Max Макс	Mean Среднее	Standard deviation Стандартное отклонение
ROP (m/hr) Механическая скорость бурения (м/ч)	6.32	95.76	38.1	21.25
RPM (rpm) Обороты в минуту (об/мин.)	54.85	141.6	136.09	5.34
SPP (atm) Давление в стояке (атм)	87.02	200.92	172.92	18.76
Torque (Kg*m) Крутящий момент (кг*м)	789.91	2227.25	1767.95	182.01
WOB (tons) Нагрузка на долото (т)	1.04	6.85	4.15	1.11
FLW (l/s) Расход насоса (л/с)	27.09	39.02	37.92	0.91
Young modulus (GPa) Модуль Юнга (ГПа)	10.24	27.17	18.39	2.59

Fig. 2 presents frequency distribution histograms and box plots of the six drilling parameters and Young modulus data, offering a visual insight into the distributional characteristics of the dataset. The horizontal axis represents the range of values for the drilling parameters, while the vertical axis indicates the number of data points falling within each value interval. In these plots, the Y-axis reflects the frequency of occurrence of specific value ranges within the dataset. By analyzing the variation and distribution of different drilling regime parameters, such as the mechanical ROP and others, the predictive algorithms can be better refined. Furthermore, analyzing these distributions helps identify potential outliers or irregularities that may impact the model accuracy. As a result, visualizing raw data distributions acts as a preliminary quality check, ensuring that the data used accurately reflect real-world drilling conditions while remaining unbiased to preserve the model predictive reliability. Comprehensive visualization of these parameters also plays a critical role in selecting an appropriate parameter set for training the forecasting model. It can be observed that, for all six drilling parameters, the frequency distribution histograms exhibit a bell-shaped curve, characteristic of a normal distribution commonly encountered in probability statistics. The initial box plot, before removing outliers, highlights five critical statistical markers within the dataset: the minimum value, first quartile (Q1), median, third quartile (Q3), and maximum value. The central box represents the interquartile range (IQR), which contains the middle 50% of the data, with a line inside indicating the median. The whiskers extend to the smallest and largest values, excluding any outliers. Detecting and managing outliers data points that significantly differ from the majority is crucial for maintaining the accuracy and reliability of the analysis. These anomalies can result from various factors, including measurement errors, incorrect data entry, or rare but valid occurrences. If not addressed, they can skew statistical interpretations, resulting in misleading conclusions and unreliable predictive models.

From the box plots in Fig. 2, the presence of outliers is readily apparent in the datasets of two parameters: SPP and FLW. These values may compromise the accuracy of the prediction results and are thus considered noise, warranting their removal from the dataset used to construct the forecasting model. To address this issue, the Interquartile Range (IQR) method was employed to detect and eliminate outliers. For each parameter column, the first (Q1) and third quartiles (Q3) were computed, and the interquartile range (IQR) was obtained by subtracting Q1 from Q3. Data points that fell outside the range of (Q1–1.5×IQR) or (Q3+1.5×IQR) were classified as outliers and subsequently removed. Utilizing the IQR method allowed for effective detection and elimination of outliers while retaining the majority of data that accurately represents the overall distribution. This process enhances the accuracy and reliability of the analysis, ensuring that meaningful insights can be drawn. After removing outliers, the dataset exhibits a more structured and interpretable distribution, as illustrated in subsequent visualizations (Fig. 3).

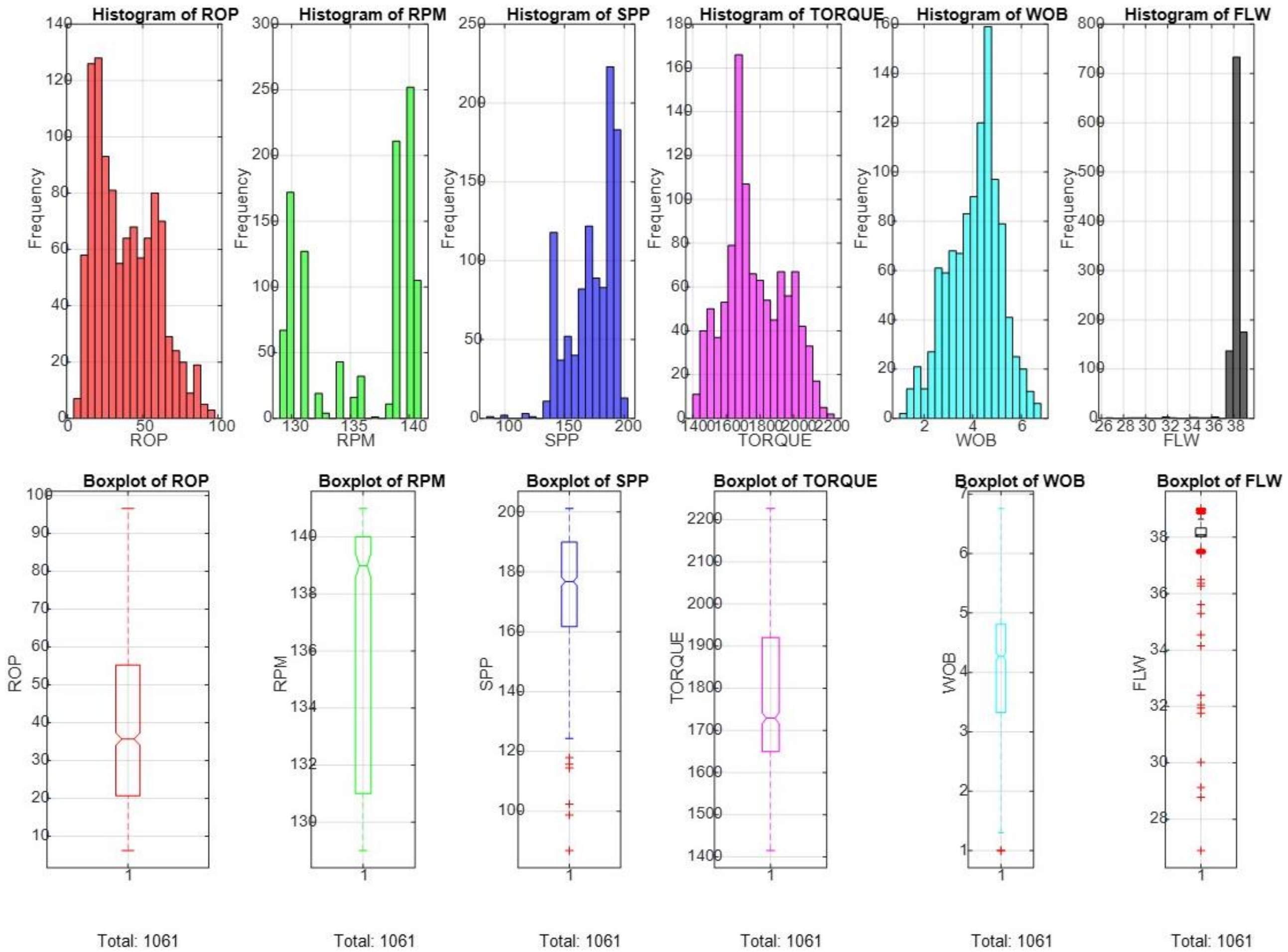

Fig. 2. *Frequency distribution histograms and box plots of the raw dataset*
Рис. 2. *Гистограммы распределения частот и диаграммы ящиков исходного набора данных*

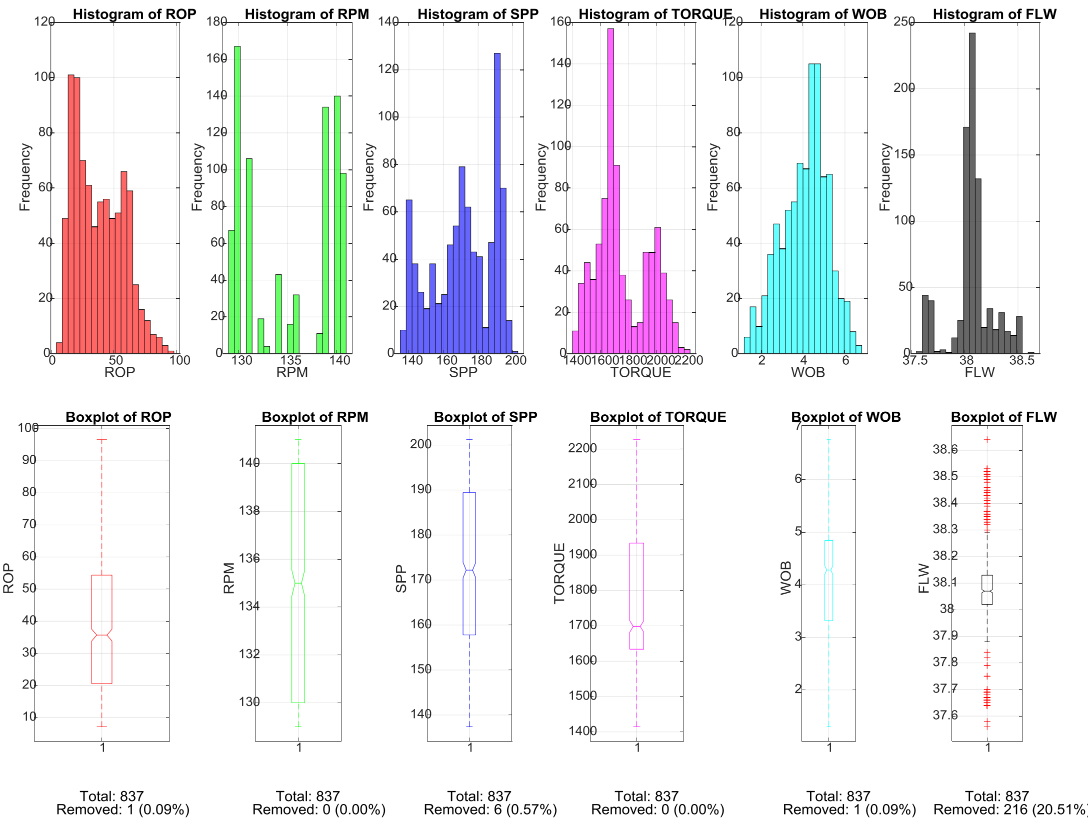

Fig. 3. *Frequency distribution histograms and box plots of the dataset after outlier removal*
Рис. 3. *Гистограммы распределения частот и диаграммы ящиков набора данных после удаления выбросов*

Selection of input features for the model

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 Young modulus prediction model, the authors utilized a comprehensive dataset comprising six drilling parameters from wells A and B. However, it is possible that certain components of this dataset do not significantly affect 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), it is evident that the correlation between the drilling parameters and the Young modulus of the rock ranges from moderate to weak (0.16–0.63). Thus, it can be inferred that relying on a single parameter is insufficient for accurately predicting Young modulus; instead, a combination of multiple parameters from the drilling dataset is necessary. Each parameter serves as an independent feature, utilized as input for the prediction model without mutual interference, thereby avoiding a reduction in accuracy or an increase in model complexity, provided the number of input features remains manageable (in this study, six parameters). This approach ensures an effective balance between predictive power and computational efficiency.

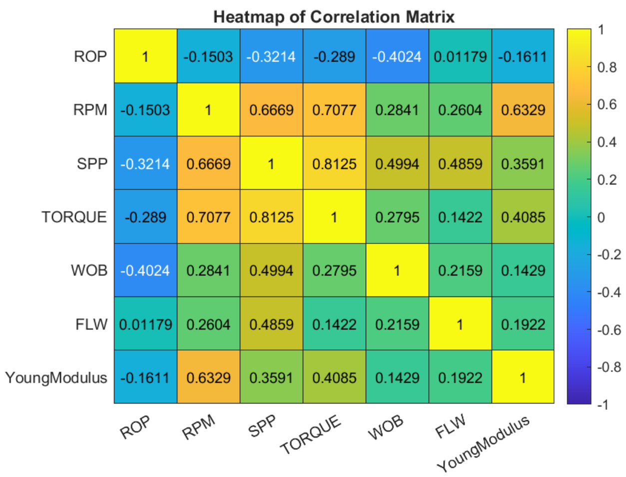

Fig. 4. *Heatmap of correlation coefficients between parameters in the dataset*
Рис. 4. *Тепловая карта коэффициентов корреляции между параметрами в наборе данных*

To address the proposed task, the researcher applied three distinct machine learning algorithms: RF, ANN, and SVM to predict Young modulus from the dataset of drilling parameters. Fine-tuning regression machine learning models, such as ANN, SVM, and RF, is a critical step to optimize their performance for specific tasks like predicting Young modulus from drilling data. For ANN, fine-tuning involves adjusting hyperparameters such as the number of hidden layers, neurons per layer, learning rate, and activation functions to enhance the model ability to capture non-linear relationships within the data. In this paper, the RandomSearchTuner, utilizing the Keras library for the Neural Network, were selected for the fine-tuning. In the case of SVM, fine-tuning focuses on optimizing the kernel type, the regularization parameter, and the kernel-specific parameter, which control the trade-off between model complexity and generalization. Cross-validation was used to balance this trade-off, ensuring the model performs well on unseen data. For RF model, fine-tuning entails adjusting parameters like the number of trees, maximum depth, and minimum samples per split or leaf, which affect the ensemble robustness and predictive power and the RandomizedSearchCV technique, implemented using the scikit-learn library for the RF. Potential configurations were generated through random combinations of the hyperparameters corresponding to each model. Ultimately, this fine-tuning enhances model accuracy, reliability, and applicability in real-time geomechanical predictions.

All three aforementioned models were evaluated using a cross-validation approach. The training dataset was divided into N subsets, with N-1 subsets used for model construction and the remaining subset reserved for model evaluation. It should be noted that this subset is entirely distinct from the testing dataset. Upon determining the optimal set of hyperparameters, the final model was retrained using the entire training dataset.

Results and discussions

It is evident that the models employing the three aforementioned algorithms all demonstrate strong capabilities in predicting Young modulus, exhibiting high correlation coefficients and very low error rates (Fig. 5).

When benchmarked against RF and SVM models, the ANN demonstrated superior predictive accuracy for Young modulus across all three datasets training, validation, and test. Consequently, the chosen model is an ANN designed to predict Young modulus based on six drilling parameters as input features. It consists of a single hidden layer with 100 neurons, employing the ReLU activation function to capture non-linear relationships in the data. The network is trained with a maximum of 1000 iterations and applies L2 regularization with a strength of 0.01 to mitigate overfitting. The six input parameters, representing drilling conditions, are processed by the wide hidden layer to extract relevant features, which are then transformed by ReLU for sparsity and efficiency. The output layer, likely using a linear activation function for regression, provides a continuous prediction of Young modulus, making this model well-suited for geomechanical applications in drilling engineering.

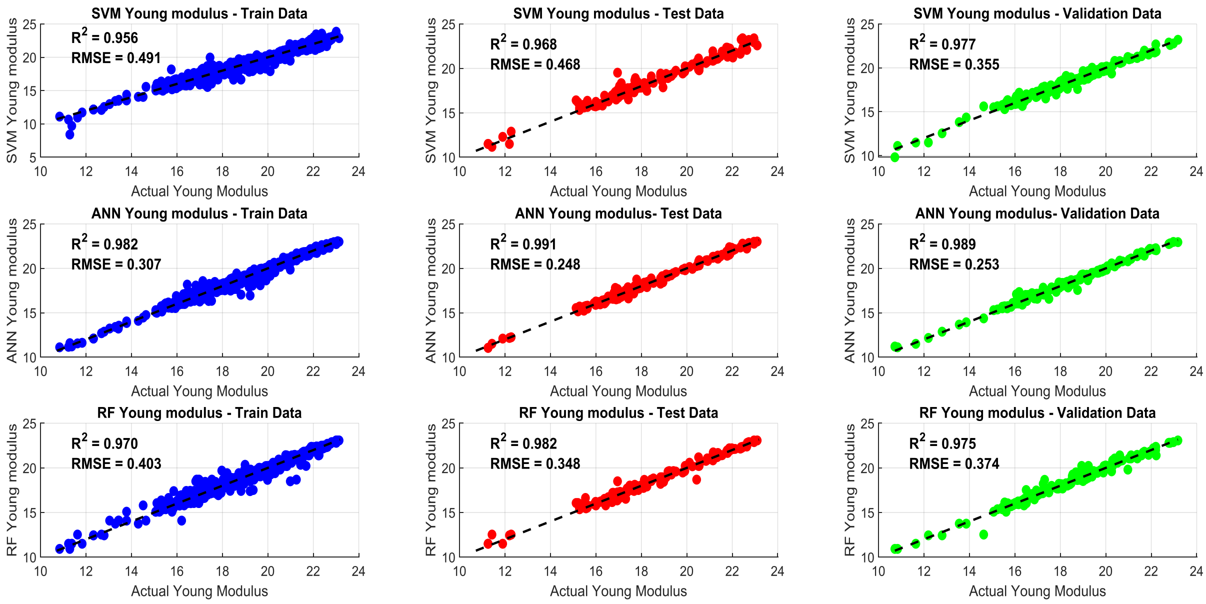

Fig. 5. *Prediction results of Young modulus from drilling data using the SVM, ANN, and RF algorithms*
Рис. 5. *Результаты прогнозирования модуля Юнга по данным бурения с использованием алгоритмов: метода опорных векторов (SVM), искусственной нейронной сети (ANN) и случайного леса (RF)*

To further evaluate the accuracy and practical applicability of the prediction model, beyond merely noting the alignment between the predicted results and the dataset used to develop the model, the model was also employed to forecast Young modulus for an additional well in the CT oil field Well C whose drilling data had not been utilized during the model development. Comparing the model predicted outcomes with the actual data from this blind test well provides a more objective and qualitative assessment. The proposed ANN model achieved a correlation coefficient of 0.94 and RMSE of 0.483 when compared to the actual Young modulus derived from well logs in well C (Fig. 6). This high accuracy underscores the ANN ability to effectively capture the complex, non-linear relationships between the six drilling parameters ROP, RPM, WOB, SPP, FLW, and TQ and rock stiffness. A blind test with a new well, which has not been included in the training or validation datasets, provides an unbiased assessment of the model ability to generalize to unseen data. This is critical in heterogeneous area like the CT oil field, where geological variability (e. g., sand, shale interbeds, and fractured granite) can significantly affect drilling parameters. By testing on a new well, the model performance is evaluated under real-world conditions, avoiding overfitting to known wells.

These results affirm the reliability of the ANN model for real-time geomechanical predictions and highlight its potential as a practical tool for enhancing drilling operations in the absence of traditional logging data, offering both precision and operational efficiency.

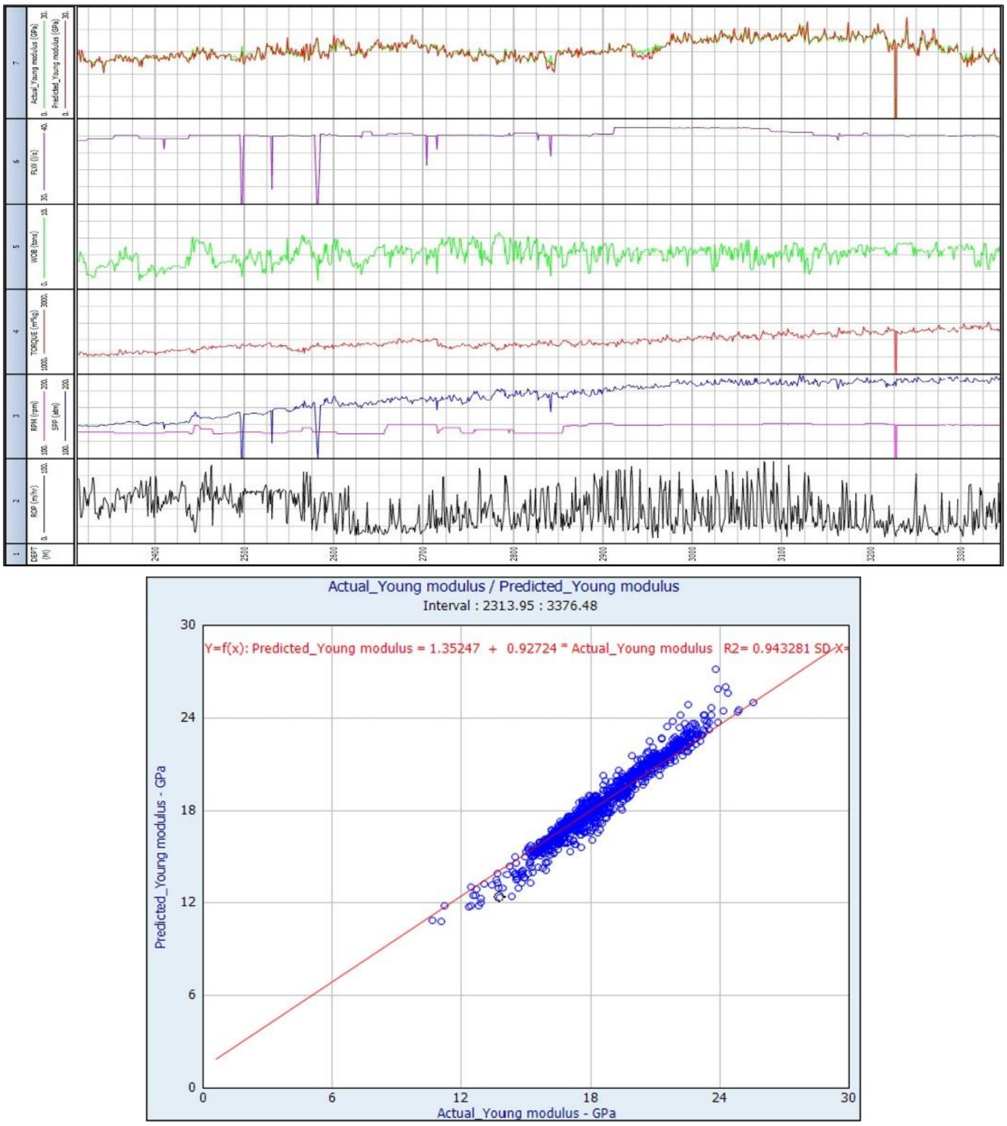

Fig. 6. *Comparison of the predicted Young modulus curve from the model with the actual curve for Well C in the CT field*

Рис. 6. *Сравнение прогнозируемой кривой модуля Юнга, полученной с помощью модели, с фактической кривой для скважины C на месторождении CT*

Conclusions

This research presents a comprehensive investigation into the application of machine learning techniques specifically ANN, SVM, and RF to predict Young modulus directly from drilling parameters at the CT oil field, located offshore Vietnam in Block 09-3/12 at the southeastern margin of the Cuu Long basin. The study utilized a dataset comprising 1061 samples collected from wells A and B, incorporating six key drilling parameters: ROP, RPM, WOB, SPP, FLW, and TQ. These parameters were meticulously preprocessed, including outlier removal using the IQR method, to ensure data quality and representativeness of real-world drilling conditions in the complex Lower Miocene and Upper Oligocene formations. The models were trained on 70% of the dataset, with 15% allocated for validation and 15% for testing, followed by a blind test on Well C, an independent well excluded from the training phase, to assess their practical applicability.

The results affirm the robustness of the machine learning approach, with all three models demonstrating high predictive accuracy. Notably, the ANN model, fine-tuned using the RandomSearch-Tuner within the Keras library, outperformed its counterparts, achieving an impressive correlation coefficient of 0.94 and RMSE of 0.483 during the blind test on Well C. This superior performance highlights ANN capability to capture the non-linear relationships between drilling parameters and rock stiffness, a critical factor given the moderate-to-weak correlations (ranging from 0.16 to 0.63) observed in the heatmap analysis (Fig. 3). The SVM and RF models, optimized via kernel parameter tuning and RandomizedSearchCV respectively, also exhibited commendable accuracy, though slightly lower than ANN, reinforcing the value of ensemble and kernel-based methods in handling heterogeneous drilling data.

The successful prediction of Young modulus from drilling data offers a transformative alternative to traditional methods reliant on costly and time-intensive well logging or core testing. By validating the predicted values against well log-derived Young modulus curves, this study demonstrates the potential to integrate real-time drilling data into geomechanical workflows, thereby reducing operational costs and enhancing decision-making during drilling operations. The multi-parameter approach, necessitated by the limited predictive power of individual parameters, was effectively managed by the machine learning models, ensuring computational efficiency without compromising accuracy, as the six-feature input set remained manageable.

These findings carry significant implications for the oil and gas industry, particularly in optimizing drilling strategies, improving wellbore stability, and mitigating risks in geologically challenging environments like CT. The ability to forecast Young modulus instantaneously during drilling operations could revolutionize reservoir characterization and operational planning. However, limitations such as the dataset confinement to a single field suggest avenues for future research, including expanding the model scope to incorporate additional drilling parameters (e. g., mud properties), testing across diverse geological settings, or refining calibration techniques to further minimize prediction errors.

REFERENCES/СПИСОК ЛИТЕРАТУРЫ

1. Brady B.H.G., Brown E.T. *Rock mechanics for underground mining.* 3rd ed. with corrections. Springer, 2004. 628 p.
2. Malkowski P., Ostrowski L. The methodology for the Young modulus derivation for rocks and its value. *Procedia Engineering*, 2017, vol. 191, pp. 134–141.
3. Hoek E., Brown E.T. Empirical strength criterion for rock masses. *Journal of the Geotechnical Engineering Division*, 1980, vol. 106 (9), pp. 1013–1035.
4. Castagna J.P., Batzle M.L., Kan T.K., Backus M.M. Rock physics – the link between rock properties and AVO response. *Offset-dependent reflectivity – Theory and practice of AVO analysis*, SEG, 1993, pp. 135–171.
5. Zoveidavianpoor M. An integrated approach in determination of elastic rock properties from well log data in a heterogeneous carbonate reservoir. *Journal of Petroleum Science and Engineering*, 2017, vol. 153, pp. 314–324.
6. Palmstrom A., Singh R. The deformation modulus of rock masses: comparisons between in situ tests and indirect estimates. *Tunnelling and Underground Space Technology*, 2001, vol. 16, pp. 115–131.
7. Mogi K. *Experimental rock mechanics.* 1st ed. CRC Press, 2006. 375 p.
8. Tariq Z., Elkatatny S., Mahmoud M., Abdulazeez A. A holistic approach to develop new rigorous empirical correlation for static Young's modulus. *Abu Dhabi International Petroleum Exhibition & Conference*. Abu Dhabi, UAE, 2016. SPE-183545-MS. https://doi.org/10.2118/183545-MS
9. Sharifi J., Nooraiepour M., Amiri M., Mondol N.H. Developing a relationship between static Young's modulus and seismic parameters. *Journal of Petroleum Exploration and Production Technology*, 2023, vol. 13, pp. 203–218.
10. Tu X., He Z., Huang Y., Zhang Z., Yang M., Zhao J. An overview of large AI models and their applications. *Visual intelligence*, 2024, vol. 2 (34). DOI: https://doi.org/10.1007/s44267-024-00065-8
11. Almeida J.S. Predictive non-linear modeling of complex data by artificial neural networks. *Current Opinion in Biotechnology*, 2002, vol. 13 (1), pp. 72–76.
12. Gong Y., Mehana M., Xiong F., Xu F., El-Monier I. Towards better estimations of rock mechanical properties integrating machine learning techniques for application to hydraulic fracturing. *SPE Annual Technical Conference and Exhibition*. Calgary, Alberta, Canada, 2019. SPE-195884-MS.
13. Ajah N.J., Dosunmu A., Akaolisa C.C.Z., Dagogo T. Analysis of elastic geomechanical properties derived from well log and seismic data, using Artificial Intelligence (ANN): a case study of "AJAH" field offshore Niger Delta. *IOSR Journal of Applied Geology and Geophysics*, 2020, vol. 8 (1), pp. 19–27.

14. Mahmoud A.A., Elkatatny S., Ali A., Moussa T. Estimation of static Young's modulus for sandstone formation using artificial neural networks. *Energies*, 2019, vol. 12 (11), 2125.
15. Mahmoud A.A., Gamal H., Elkatatny S., Chen W. Real-time evaluation of the dynamic Young's modulus for composite formations based on the drilling parameters using different machine learning algorithms. *Frontiers in Earth Science*, 2022, vol. 10, 1034704.
16. Elkatatny S., Tariq Z., Mahmoud M., Abdulraheem A., Mohamed I. An integrated approach for estimating static Young's modulus using artificial intelligence tools. *Neural Computing and Applications*, 2019, vol. 31 (12), pp. 4123–4135.
17. Krenker A., Bešter J., Kos A. Introduction to the artificial neural networks. *Artificial Neural Networks: Methodological Advances and Biomedical Applications*. Ed. by K. Suzuki. InTech, 2011. pp. 1–18.
18. Smith J. Application of neural networks in geomechanical property prediction. *Journal of Petroleum Science and Engineering*, 2020, vol. 185, pp. 106–115.
19. Fawagreh K., Gaber M.M., Elyan E. Random forests: from early developments to recent advancements. *Systems Science & Control Engineering*, 2014, vol. 2 (1), pp. 602–609.
20. Johnson L. Random Forest applications in rock mechanics prediction. *Geophysics*, 2021, vol. 86 (3), pp. 45–56.
21. Al-Anazi A., Gates I.D. A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs. *Engineering Geology*, 2010, vol. 114 (3–4), pp. 267–277.
22. Lee K. Support vector machines for geomechanical parameter estimation. *Journal of Rock Mechanics and Geotechnical Engineering*, 2019, vol. 11 (4), pp. 789–799.
23. Vu H.D., Nguyen M.H., Nguyen T.H., Nguyen T.V. Prediction of penetration rate and optimization of weight on a bit using artificial neural networks. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, 2024, vol. 335, no. 3, pp. 192–203. (In Russ.). DOI: https://doi.org/10.18799/24131830/2024/3/4376

Information about the authors

Vu Hong Duong, MSc, Lecturer, Hanoi University of Mining and Geology, 18, Vien street, Duc Thang Ward, Bac Tu Liem District, Hanoi, 11900, Vietnam; vuhongduong@humg.edu.vn. https://orcid.org/ 0009-0007-0923-0880
Nguyen Tien Hung, Cand. Sci. (Techn.), Lecturer, Hanoi University of Mining and Geology, 18, Vien street, Duc Thang Ward, Bac Tu Liem District, Hanoi, 11900, Vietnam; nguyentienhung.dk@humg.edu.vn. https://orcid.org/0009-0001-4849-7058
Vu Thiet Thach, MSc, Lecturer, Hanoi University of Mining and Geology, 18, Vien street, Duc Thang Ward, Bac Tu Liem District, Hanoi, 11900, Vietnam; vuthietthach@humg.edu.vn
Nguyen Minh Hoa, Cand. Sci. (Geol.-Mineral.), Lecturer, Hanoi University of Mining and Geology, 18, Vien street, Duc Thang Ward, Bac Tu Liem District, Hanoi, 11900, Vietnam; nguyenminhhoa@humg.edu.vn. https://orcid.org/0000-0002-6555-662X
Nguyen Xuan Duy, MSc, Trainee, Vietnam Petroleum Institute, 167, Trung Kinh street, Hanoi, 10000, Vietnam; duynx@vpi.pvn.vn

Received: 20.03.2025
Revised: 25.03.2025
Accepted: 16.03.2026

Информация об авторах

Ву Хонг Зыонг, магистр технических наук, преподаватель, Ханойский университет горного дела и геологии, Вьетнам, 11900, г. Ханой, ул. Вьен, 18; vuhongduong@humg.edu.vn. https://orcid.org/ 0009-0007-0923-0880
Нгуен Тиен Хунг, кандидат технологических наук, преподаватель, Ханойский университет горного дела и геологии, Вьетнам, 11900, г. Ханой, ул. Вьен, 18; nguyentienhung.dk@humg.edu.vn. https://orcid.org/0009-0001-4849-7058
Ву Тьет Тхак, магистр технических наук, преподаватель, Ханойский университет горного дела и геологии, Вьетнам, 11900, г. Ханой, ул. Вьен, 18; vuthiethach@humg.edu.vn
Нгуен Минь Хоа, кандидат геолого-минералогических наук, преподаватель, Ханойский университет горного дела и геологии, Вьетнам, 11900, г. Ханой, ул. Вьен, 18; nguyenminhhoa@humg.edu.vn. https://orcid.org/0000-0002-6555-662X
Нгуен Суан Зуи, магистр технических наук, стажер, Вьетнамский нефтяной институт, Вьетнам, 10000, г. Ханой, ул. Чунг Кинь, 167; duynx@vpi.pvn.vn

Поступила в редакцию: 20.03.2025
Поступила после рецензирования: 25.03.2025
Принята к публикации: 16.03.2026