Petroleum Science 23 (2026) 4009–4033

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

A novel method for drilling rate of penetration prediction and transfer generalization based on physics-informed neural networks

Miao He [a,b,c,d,e,*], Qing-Rui Yuan [a,c,d,e,**], Jun Li [b], Chang-Cheng Zhou [f], Jing-Wei Liu [a,c,d,e], Xin Chen [a,c,d,e], Jia-Hao Cao [a,c,d,e]

[a] School of Petroleum Engineering, Yangtze University, Wuhan, 430100, Hubei, China
[b] State Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing), Beijing, 102249, China
[c] State Key Laboratory of Deep Oil and Gas, China University of Petroleum (East China), Qingdao, 266580, Shandong, China
[d] State Key Laboratory of Low Carbon Catalysis and Carbon Dioxide Utilization, Yangtze University, Wuhan, 430100, Hubei, China
[e] Key Laboratory of Drilling and Production Engineering for Oil and Gas, Hubei Province, Wuhan, 430100, Hubei, China
[f] CNOOC Hainan Branch Company, Haikou, 570312, Hainan, China

ARTICLE INFO

Article history:
Received 23 September 2025
Received in revised form
27 February 2026
Accepted 4 March 2026
Available online 9 March 2026

Edited by Jia-Jia Fei

Keywords:
Rate of penetration
Deep learning
Physics-informed neural networks
Adaptive loss
Model generalization

ABSTRACT

The rate of penetration (ROP) is a critical statistic for evaluating drilling efficiency, and its accurate prediction is essential for improving drilling performance and reducing operational costs. Conventional physical ROP models are theoretically sound but exhibit limited predictive accuracy, while machine learning algorithms can achieve higher precision yet often lack sufficient interpretability and generalization capability. Therefore, this study presents a physics-informed neural network (PINN)-based hybrid model for ROP prediction and transfer generalization. The proposed approach integrates the mechanical specific energy and Soares ROP models as physical constraints and establishes a hybrid deep learning architecture that combines a squeeze-and-excitation residual network (SE-ResNet), bidirectional long short-term memory (BiLSTM), and self-attention mechanism (SAM). Within this framework, an adaptive loss weighting strategy is introduced into a unified loss function to dynamically balance the contributions of data fitting and physical constraints during training. The experimental results show that compared with traditional physical models and mainstream intelligent algorithms such as support vector regression (SVR), back propagation neural network (BP), LSTM, the PINN model demonstrates superior predictive performance, with a root mean square error (RMSE) of 6.214, mean absolute error (MAE) of 3.483, mean absolute percentage error (MAPE) of 6.102%, and coefficient of determination (R^2) of 0.953. Furthermore, to evaluate the model's cross-domain generalization capability, zero-shot and few-shot transfer learning approaches are investigated. The few-shot transfer method achieves a MAPE of 10.413%, confirming the model's adaptability and robustness across different drilling datasets. This study helps to advance the deep integration of artificial intelligence, big data, and other technologies with traditional drilling mechanism models, and has significant theoretical and engineering application value.

1. Introduction

As oil and gas exploration advances into complex environments such as deepwater and ultra-deep formations, drilling operations face increasing challenges related to safety, cost, and efficiency (Wang et al., 2022; Zheng et al., 2022). The geological complexity, poor drillability, and high abrasiveness of deep formations significantly hinder drilling acceleration, becoming major constraints on efficient hydrocarbon development (Etesami et al., 2021;

* Corresponding author.
** Corresponding author.
E-mail addresses: hemiao@yangtzeu.edu.cn (M. He), 2023720476@yangtzeu.edu.cn (Q.-R. Yuan).

Peer review under the responsibility of China University of Petroleum (Beijing).

https://doi.org/10.1016/j.petsci.2026.03.010

Gooneratne et al., 2020). As a core component of modern intelligent drilling systems, rate of penetration (ROP) prediction plays a crucial role in improving drilling efficiency and reducing operational costs (Pei et al., 2024). Accurate ROP prediction enables better control of drilling parameters, enhances time efficiency, and ensures safe and effective operations, thereby supporting the digital and intelligent transformation of drilling technologies (Wan et al., 2024; Lee et al., 2025).

The development of drilling ROP prediction research can be broadly divided into two stages: the first stage involves physics-based ROP modeling, while the second is characterized by data-driven prediction using intelligent algorithms. Early ROP prediction efforts were primarily grounded in physical and empirical modeling. The foundation of mechanical ROP modeling was laid by Maurer (1962), who established a quantitative relationship among ROP, weight on bit (WOB), and revolutions per minute (RPM). Building on this, Bingham (1965), Bourgoyne Jr and Young Jr (1974) extended the framework through empirical calibration and multiple regression to account for the combined effects of drilling parameters and formation characteristics. Later studies enhanced the physical realism of ROP formulations by introducing various rock mechanical and operational factors, including internal friction, porosity, compressive strength, and bit wear effects (Warren, 1981; Walker et al., 1986; Hareland and Rampersad, 1994). Motahhari et al. (2010) developed a quantitative ROP model for polycrystalline diamond compact (PDC) bits that explicitly accounted for bit wear and rock strength effects. In a subsequent study, Al-AbdulJabbar et al. (2019) enhanced the predictive accuracy of regression-based ROP models by applying advanced statistical optimization techniques.

The emergence of big data and artificial intelligence has transformed ROP prediction modeling into a data-driven paradigm (Li et al., 2024b; Akkuş et al., 2025). Numerous researchers have developed a variety of machine learning-based ROP prediction models, such as artificial neural networks (ANN), random forest (RF), support vector regression (SVR), time convolutional network (TCN), and multi-layer perceptron (MLP), which have demonstrated superior predictive performance compared to conventional physics-based models (Elkatatny et al., 2020; Yin et al., 2024; Ahmed et al., 2019; Fan et al., 2025a; Brenjkar and Delijani, 2022). In recent studies, researchers have employed advanced techniques such as temporal modeling, ensemble algorithms, feature correlation analysis, improved optimization algorithms, dynamic time warping, and data stream learning frameworks to predict ROP. Compared with single-method approaches, these methods have effectively improved model prediction accuracy and computational efficiency (Gan et al., 2022; Shaygan and Jamshidi, 2023; Ghanitoos et al., 2024; Su et al., 2024; Yang et al., 2025; Junior and Lavi, 2025). However, machine learning models are often characterized by their “black box” nature. Lacking the constraints of physical prior knowledge, these models face challenges in achieving strong generalization capability and interpretability while maintaining high predictive precision (Allawi et al., 2025; Tu et al., 2024).

As a result, hybrid modeling strategies that combine physical mechanisms with intelligent algorithms have gained increasing attention. Ren et al. (2023) and Jiao et al. (2024) established systematic hybrid frameworks that integrate machine learning with physical equations through residual correction, weighted aggregation, and ensemble coupling, effectively reducing prediction bias and improving model consistency. Zhou et al. (2023) further refined hybrid modeling by introducing a data similarity-based fusion approach, where Euclidean distance was used to enhance the adaptability of machine learning models to empirical data. Fan et al. (2025b) advanced this direction by proposing a correlation-constrained ensemble (CCE) model that fuses geological and engineering information, combining multiple learners through gradient boosting and correlation-based regularization.

In conclusion, research combining intelligent algorithm-based models with traditional physical ROP equations remains at an early stage. Conventional physical equations are diverse and complex, containing empirical coefficients that strongly depend on geological conditions, making it difficult to evaluate their coupling with data-driven models. Existing hybrid approaches, such as residual modeling, weighted fusion, and ensemble learning, also face inherent limitations. They typically optimize physical and data-driven components separately, leading to inconsistent convergence between physical and data residuals, and their fixed or manually tuned weighting strategies limit adaptability during training, reducing generalization under varying drilling conditions.

In contrast, the physics-informed neural network (PINN) framework provides an end-to-end way to embed physical constraints directly into model training. Introduced by Raissi et al. (2019), PINN incorporate governing equations by minimizing differential residuals within the loss function, ensuring physical consistency during learning. They have been successfully applied in fluid dynamics (Raissi et al., 2020; Sun et al., 2020), heat transfer (Cai et al., 2021), and geosciences and engineering modeling (Liu et al., 2023; Li et al., 2024a; Ishitsuka and Lin, 2023). However, the PINN model is prone to getting stuck in local optima during the optimization process and has limited generalization ability under unknown conditions. Overall, the PINN model shows significant potential for modeling complex nonlinear coupled systems in engineering problems, offering a promising approach to enhancing the reliability and interpretability of ROP prediction models.

Thus, this study develops a PINN-based ROP prediction model that combines physical modeling with deep learning. The improved mechanical specific energy and Soares ROP models are incorporated as physical constraints, while a hybrid deep learning architecture—SE-ResNet-BiLSTM-SAM—is constructed to capture multi-scale spatial-temporal features of drilling data. A unified loss function combines regression and physical residuals, with adaptive weights optimized jointly with network parameters to balance data and physical constraints. Furthermore, the transfer learning is utilized to evaluate the model’s generalization capability across different drilling datasets. By integrating physical priors into a data-driven framework, this study aims to establish a PINN-based ROP prediction model with high prediction accuracy and a certain degree of generalization across different drilling blocks. The overall technical workflow is illustrated in Fig. 1.

2. Theory and methodology

2.1. Physics model

Mechanical specific energy, first proposed by Teale (1965), refers to the amount of energy required to break a unit volume of rock. Mechanical specific energy represents the energy required to break a unit volume of rock and reflects the rock-breaking efficiency of the drill bit, making it a key indicator of drilling performance. The mechanical specific energy can be calculated from parameters such as ROP, WOB, RPM, drill string torque (torque), and bit diameter. A higher mechanical specific energy value indicates lower drilling efficiency and poorer compatibility between the bit and the formation (Dupriest and Koederitz, 2005; Zhou et al., 2017).

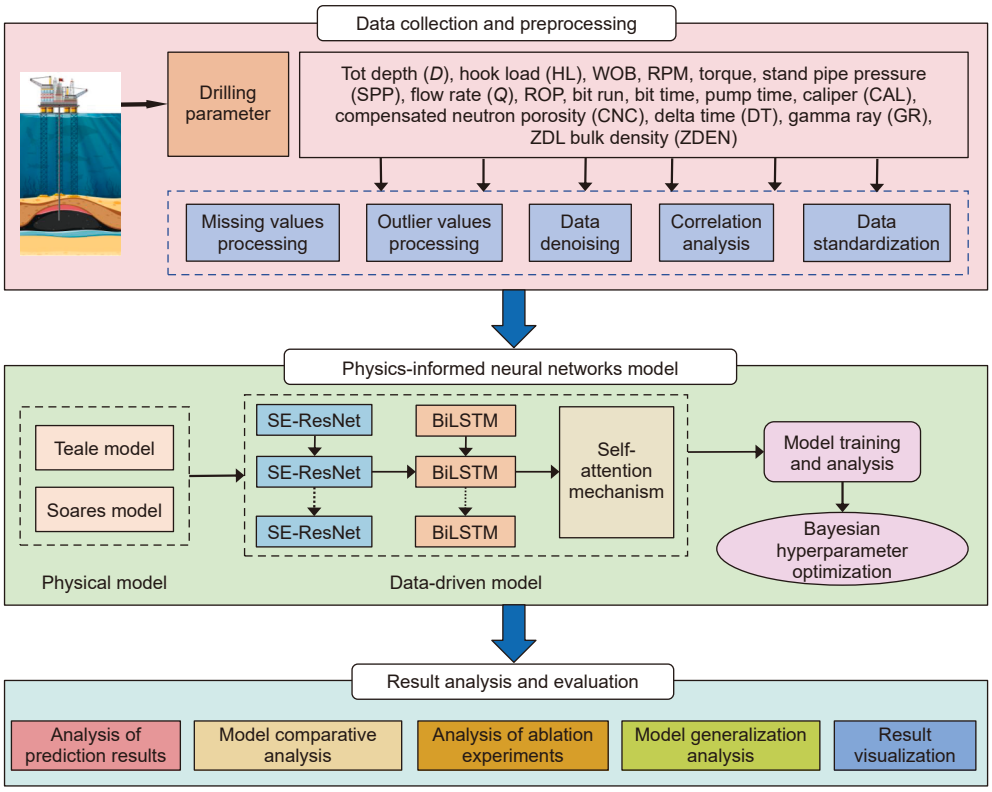

Fig. 1. The overall technical route for prediction rate of penetration.

$$\mathrm{mse} = \frac{4W}{\pi d_B^2} + \frac{480RT}{d_B^2 \cdot \mathrm{ROP}} \tag{1}$$

During actual drilling operations, the energy utilization rate is usually only between 30% and 40% due to factors such as friction and vibration. Some subsequent researchers revised Teale's mechanical specific energy model by introducing the effective energy utilization rate of the drill bit, denoted as E_f, and reconstructed the energy balance equation for the mechanical rock-breaking process of the drilling tools (Hammoutene and Bits, 2012; Chen et al., 2014).

$$\mathrm{mse} = E_f\left(\frac{4W}{\pi d_B^2} + \frac{480RT}{d_B^2 \cdot \mathrm{ROP}}\right) \tag{2}$$

Based on the sliding friction coefficient of the bit and the WOB, and through double integral, the torque expression for the drilling process can be derived.

$$T = \frac{1}{1000}\int_0^{\frac{d_B}{2}}\int_0^{2\pi} l^2 \frac{4\mu W}{\pi d_B^2}\,\mathrm{d}l\mathrm{d}\theta = \frac{\mu W d_B}{3000} \tag{3}$$

Ultimately, the mechanical specific energy model can be expressed as a function of the WOB and the RPM.

$$\mathrm{mse} = E_f W\left(\frac{4}{\pi d_B^2} + \frac{0.16\mu R}{d_B \cdot \mathrm{ROP}}\right) \tag{4}$$

where, mse is the mechanical specific energy, W is the WOB, R is the RPM, d_B is the bit diameter, T is the torque, E_f is generally taken as 0.35, l is the length of the infinitesimal radius of the drill bit, μ is the sliding friction coefficient of the drill bit (typically 0.25 for roller cone bit and 0.50 for PDC bit), ROP is the rate of penetration.

The B&Y drilling rate equation incorporates multiple influencing factors, each with specific coefficients that are often difficult to determine directly. In addition, the equation strongly depends on normalization constants for parameters such as depth (D), WOB, and RPM.

To address these limitations, Soares and colleagues proposed a modified ROP prediction model based on the B&Y equation. The revised model removes normalization factors and fixed constants while retaining the key variables, D, WOB, RPM, and flow rate (Q), and redefines the coefficient ranges accordingly (Soares and Gray, 2019).

$$\mathrm{ROP} = a_1 D^{a_2} W^{a_3} R^{a_4} Q^{a_5} \tag{5}$$

where, ROP is the rate of penetration; D is the well depth, Q is the flow rate. The coefficient range determined by the Cesar Soares ROP prediction model is shown in Table 1.

2.2. Data model

2.2.1. SE-ResNet

The residual network (ResNet) is a classic and widely used deep convolutional neural network architecture, which was proposed by Kaiming He and coworkers in 2016 (Alaeddine and Jihene, 2021; He et al., 2016).

The core of this method lies in the design of the residual block. By introducing an identity mapping through skip connections, feedforward information can bypass several convolutional layers and directly reach deeper network nodes. This mechanism effectively alleviates the problems of gradient vanishing and model degradation that arise with increasing network depth. As a result, the model's stability and computational efficiency are significantly improved.

Table 1
The coefficient range of the Cesar Soares ROP prediction model.

Coefficient	Lower bound	Upper bound
a_1	0.0001	10,000
a_2	−2	0
a_3	0.5	2
a_4	0.4	1
a_5	0.3	2

$$\boldsymbol{y}_l = h(\boldsymbol{x}_l) + F(\boldsymbol{x}_l, \boldsymbol{W}_l) \tag{6}$$

$$\boldsymbol{x}_{l+1} = f(\boldsymbol{y}_l) \tag{7}$$

where, $\boldsymbol{y}_l$ is the output feature of the residual block, $\boldsymbol{x}_l$ is the input feature, h is identity mapping, F is the feature mapping function, used to perform nonlinear transformation on the input feature, $\boldsymbol{W}_l$ is the set of parameters of the convolutional layer, $\boldsymbol{x}_{l+1}$ is the input feature of the next layer, f is the nonlinear activation function.

Squeeze-and-excitation ResNet (SE-ResNet) is an enhanced variant of ResNet that integrates the squeeze-and-excitation networks (SENet) module into its architecture. By employing one-dimensional convolution to model the interdependencies among feature channels, it effectively strengthens the ability of convolutional neural networks to capture cross-channel information (Wu et al., 2024; Madani et al., 2021).

Fig. 2 illustrates the architecture of the SE-ResNet. The SE module consists of two operations: squeeze and excitation. In the squeeze stage, global average pooling is applied to each convolutional channel, compressing the two-dimensional feature maps into one-dimensional global descriptors that capture the overall statistical characteristics of each channel. In the excitation stage, a two-layer fully connected network performs nonlinear transformation and normalization on these descriptors. The first layer uses a rectified linear unit (ReLU) activation for feature mapping, followed by a sigmoid activation in the second layer to generate channel-wise weights. These weights adaptively recalibrate channel importance, enhancing informative features while suppressing redundant ones (Hu et al., 2018; Jin et al., 2022; Zhu et al., 2024).

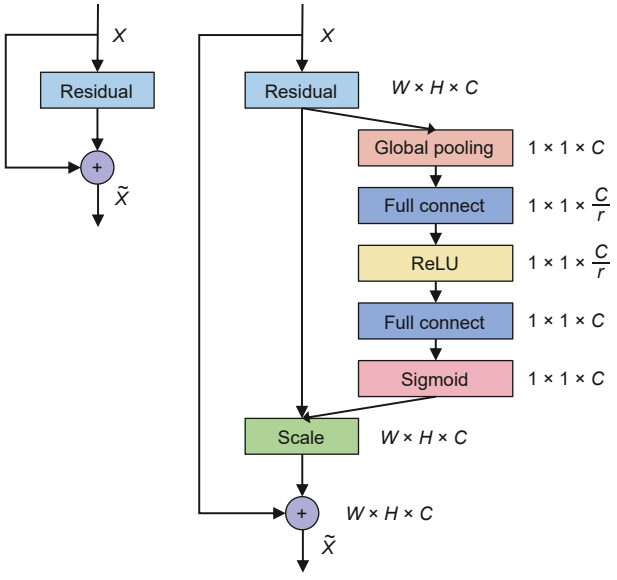

Fig. 2. SE-ResNet model structure diagram.

$$z_c = F_s(\boldsymbol{X}_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i,j) \tag{8}$$

$$\boldsymbol{s} = F_e(\boldsymbol{z}, \boldsymbol{W}) = \sigma(\boldsymbol{W}_2 \delta(\boldsymbol{W}_1 \boldsymbol{z})) \tag{9}$$

$$\tilde{\boldsymbol{X}}_c = s_c \cdot \boldsymbol{X}_c \tag{10}$$

where, $\boldsymbol{X}_c$ is the two-dimensional convolutional feature, z_c is the scalar obtained after global average pooling, F_s denotes the squeeze operation, H is the height of the convolutional feature, W is the width of the convolutional feature, $\frac{1}{H \times W}$ is the normalization coefficient, $\sum_{i=1}^{H} \sum_{j=1}^{W}$ represents the aggregation of all spatial information across the channel, $\boldsymbol{s}$ is the channel weight vector, F_e denotes the excitation operation, $\boldsymbol{z}$ is the global description generated by the squeeze operation, $\boldsymbol{W}_1$ is the weight matrix of the first fully connected layer, $\boldsymbol{W}_2$ is the weight matrix of the second fully connected layer, δ denotes the ReLU activation function, σ denotes the sigmoid activation function, $\tilde{\boldsymbol{X}}_c$ is the weighted feature mapping, $\boldsymbol{X}_c$ is the input convolutional feature.

2.2.2. Bi-directional long short-term memory

The bidirectional long short-term memory network (BiLSTM) is an extension and improvement of the standard long short-term memory (LSTM) architecture. Its structure is illustrated in Fig. 3.

BiLSTM consists of two LSTM layers, one forward and one backward, that jointly perform bidirectional learning and optimization on sequential data. Specifically, the forward LSTM processes information from the beginning to the current time step, while the backward LSTM propagates information from the end of the sequence to the current time step (Schuster and Paliwal, 1997; Michael et al., 2024).

In the analysis of sequential data, BiLSTM effectively integrates information from both forward and backward directions, thereby taking into account both historical data and future trends at each time step. This enhances the understanding of contextual relationships within the sequence. Consequently, it augments the model's capacity to comprehend and anticipate the data at the current time step (Xiong et al., 2024).

$$\boldsymbol{h}_t^{\mathrm{f}} = \mathrm{LSTM_f}\left(\boldsymbol{x}_t, \boldsymbol{h}_{t-1}^{\mathrm{f}}\right) \tag{11}$$

$$\boldsymbol{h}_t^{\mathrm{b}} = \mathrm{LSTM_b}\left(\boldsymbol{x}_t, \boldsymbol{h}_{t+1}^{\mathrm{b}}\right) \tag{12}$$

$$\boldsymbol{h}_t = \left[\boldsymbol{h}_t^{\mathrm{f}}, \boldsymbol{h}_t^{\mathrm{b}}\right] \tag{13}$$

where, $\boldsymbol{h}_t^{\mathrm{f}}$ is the hidden state of the forward LSTM at time step t, $\boldsymbol{x}_t$ is the input feature vector at time step t, $\boldsymbol{h}_{t-1}^{\mathrm{f}}$ is the hidden state of the forward LSTM at the previous time step, $\boldsymbol{h}_t^{\mathrm{b}}$ is the hidden state of the backward LSTM at time step t, $\boldsymbol{h}_{t+1}^{\mathrm{b}}$ is the hidden state of the backward LSTM at the subsequent time step, $\boldsymbol{h}_t$ is the final hidden state of the bidirectional LSTM at time step t.

2.2.3. Self-attention mechanism

The attention mechanism (AM) is a computational model inspired by human cognitive attention. This mechanism has been widely applied in domains such as natural language processing, statistical learning, and image recognition (Niu et al., 2021).

The self-attention mechanism (SAM) is a special form of the attention mechanism in which the query ($\boldsymbol{Q}$) and key ($\boldsymbol{K}$) vectors are identical. As shown in Fig. 4, the SAM computes the similarity

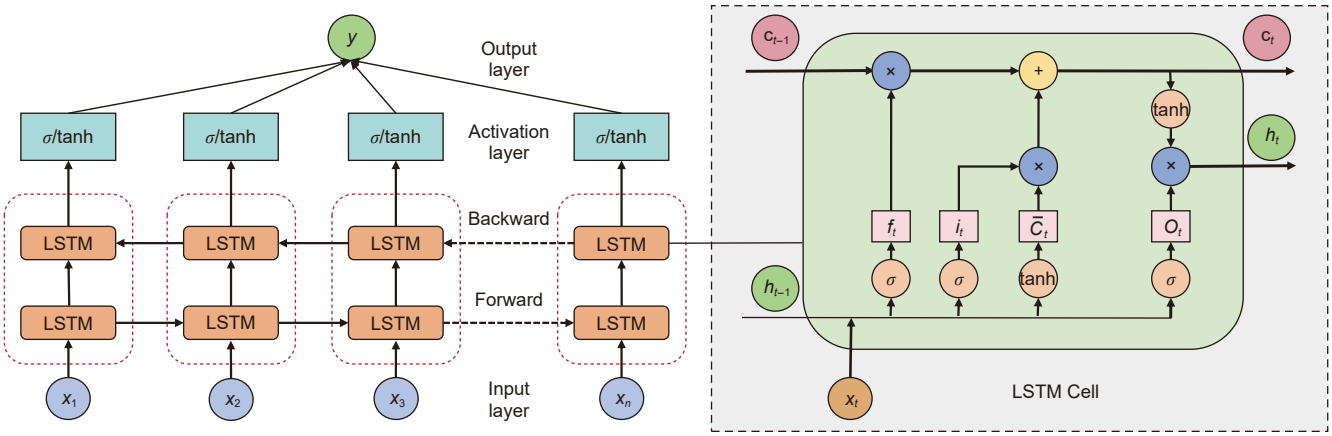

Fig. 3. Structure of BiLSTM network.

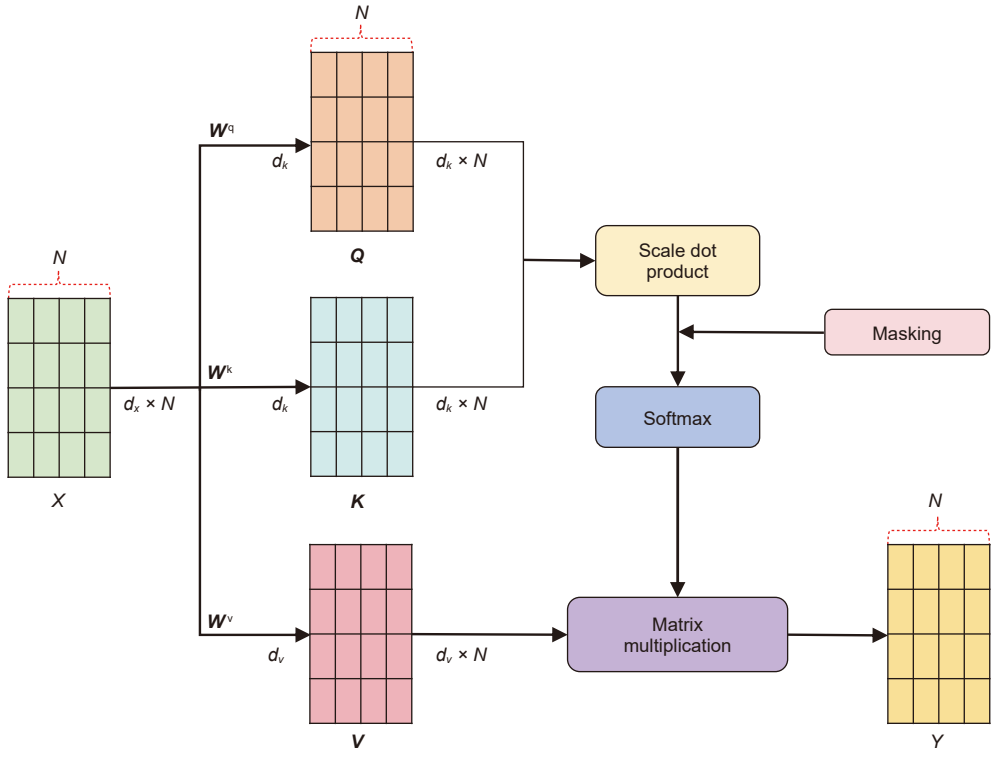

Fig. 4. Structure diagram of self-attention mechanism.

between different parts of the input sequence, using the sequence's own information to capture internal dependencies. This mechanism can flexibly handle variable-length sequences and greatly enhances the model's ability to capture long-range dependencies (Vaswani et al., 2017; Fu et al., 2022; Pan et al., 2023).

$$\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{W}^{q}, \boldsymbol{K} = \boldsymbol{X}\boldsymbol{W}^{k}, \boldsymbol{V} = \boldsymbol{X}\boldsymbol{W}^{v} \tag{14}$$

$$\text{Attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{T}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{15}$$

where, $\boldsymbol{W}^{q}$, $\boldsymbol{W}^{k}$ and $\boldsymbol{W}^{v}$ are learnable weight matrices, $\boldsymbol{X}$ is the input sequence with feature dimension d_x, $\boldsymbol{Q}$ is the query vector, $\boldsymbol{K}$ is the key vector, $\boldsymbol{V}$ is the value vector with dimension d_v, d_k is the dimension of the $\boldsymbol{Q}$ and $\boldsymbol{K}$ vectors, used to prevent the dot-product values from becoming excessively large.

2.3. Bayesian optimization algorithm

Bayesian optimization (BO) is an efficient global optimization method, suitable for scenarios where the objective function is unknown and the cost of parameter exploration is high (Wang et al., 2023; Du et al., 2022).

In hyper-parameter optimization, a probabilistic surrogate model of the objective function is constructed, and the posterior distribution is inferred from historical observations. Based on this distribution, an acquisition function selects the next evaluation point, enabling rapid approximation of the global optimum with as

few objective function evaluations as possible (Victoria and Maragatham, 2021; Alibrahim and Ludwig, 2021).

Bayesian optimization is essentially a sequential model-based optimization (SMBO) strategy that balances exploration and exploitation, thereby enabling efficient tuning of model hyperparameters.

The Bayesian optimization algorithm consists of two core components: a probabilistic surrogate model and an acquisition function. The surrogate model approximates the original objective function and continuously updates the prior distribution to refine this approximation. In most cases, a Gaussian process (GP) is employed to represent the prior distribution of the objective function $f(x)$ (Binois and Wycoff, 2022).

$$f(\boldsymbol{x}) \sim \mathrm{gp}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}^{'})) \tag{16}$$

where, $f(\boldsymbol{x})$ is the objective function, $m(\boldsymbol{x})$ is the mean function, $k(\boldsymbol{x}, \boldsymbol{x}')$ is the kernel function. Based on the observed dataset, the GP can predict the mean value $\mu(\boldsymbol{x})$ and variance $\sigma(\boldsymbol{x})$ of the objective function at any given point.

The acquisition function selects the next evaluation point based on the posterior probability distribution, prioritizing regions that are likely to contain the optimum as well as unexplored areas (Chen et al., 2024). In this study, the expected improvement (EI) function is adopted as the acquisition function. When the current best function value is v^*, the EI function is defined for $\sigma_t(\boldsymbol{x}) > 0$ as:

$$\alpha_t(\boldsymbol{x}; \boldsymbol{D}_{1:t}) = (v^* - \mu_t(\boldsymbol{x}))\Phi\left(\frac{v^* - \mu_t(\boldsymbol{x})}{\sigma_t(\boldsymbol{x})}\right) + \sigma_t(\boldsymbol{x})\phi\left(\frac{v^* - \mu_t(\boldsymbol{x})}{\sigma_t(\boldsymbol{x})}\right) \tag{17}$$

where, $\alpha_t(\boldsymbol{x}; \boldsymbol{D}_{1:t})$ denotes the acquisition function value at iteration t, based on the observed dataset $\boldsymbol{D}_{1:t}$, $\mu_t(\boldsymbol{x})$ and $\sigma_t(\boldsymbol{x})$ represent the predictive mean and standard deviation of the GP model at input $\boldsymbol{x}$, respectively, $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution function and probability density function of the standard normal distribution, respectively.

The purpose of Bayesian optimization is to identify the hyperparameter configuration X that yields the global optimum of the objective function $f(x)$. The algorithmic flow is illustrated in Fig. 5 (Zulfiqar et al., 2022; Chen et al., 2025).

2.4. Physics-informed neural networks model

Deep learning algorithms typically adopt the mean squared error as the loss function to quantify errors in each iteration. In this study, two physics-based constraint terms, the residual of the mechanical specific energy model L_{mse} and the residual of Soares ROP equation L_{Soares}, are incorporated in a soft-constraint manner. Together with the standard regression loss of the data-driven model, they constitute the new loss function of the PINN. By substituting the known drill bit effective energy utilization rate E_{f} into Eq. (4), Eq. (20) can be derived.

$$L_{\mathrm{data}} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathrm{ROP}_{\mathrm{pred},i} - \mathrm{ROP}_{\mathrm{true},i}\right)^2 \tag{18}$$

$$L_{\mathrm{mse}} = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{1.4W_i}{\pi d_{\mathrm{B},i}^2} + \frac{0.056\mu R_i W_i}{d_{\mathrm{B},i}\cdot \mathrm{ROP}_{\mathrm{pred},i}}\right)^2 - \left(\frac{1.4W_i}{\pi d_{\mathrm{B},i}^2} + \frac{0.056\mu R_i W_i}{d_{\mathrm{B},i}\cdot \mathrm{ROP}_{\mathrm{true},i}}\right)\right] \tag{19}$$

$$L_{\mathrm{Soares}} = \frac{1}{N}\sum_{i=1}^{N}\left(a_1 D_i^{a_2} W_i^{a_3} R_i^{a_4} Q_i^{a_5} - \mathrm{ROP}_{\mathrm{true},i}\right)^2 \tag{20}$$

$$\mathrm{Loss} = L_{\mathrm{data}} + L_{\mathrm{mse}} + L_{\mathrm{Soares}} \tag{21}$$

where, N is the sample size, $\mathrm{ROP}_{\mathrm{pred},i}$ denotes the ROP value predicted directly by the PINN model, $\mathrm{ROP}_{\mathrm{true},i}$ is the true ROP for the i-th sample, μ is taken as 0.25/0.50 depending on the bit type. Additionally, it should be noted that the predicted values output by the neural network are only used to compute the physical residuals and are not reused as inputs during training.

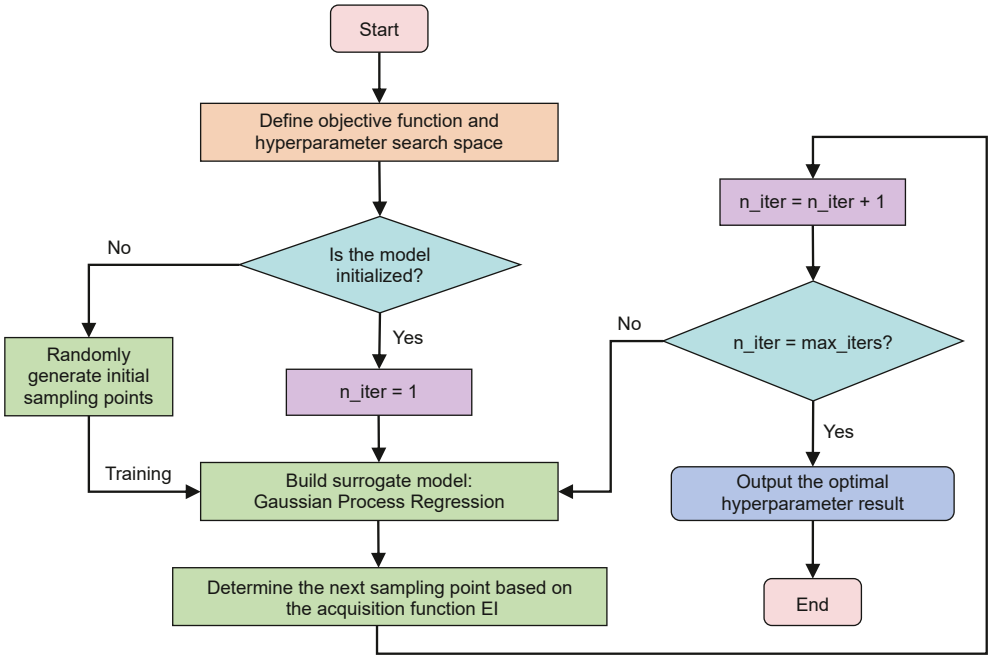

Fig. 5. Bayesian algorithm parameter optimization process.

As defined above, the total loss function of the proposed PINN model consists of multiple terms that jointly constrain the learning process. However, during gradient descent, different loss terms often exhibit disparities in magnitude and gradient scale, leading to an imbalance during training. This imbalance can cause unstable convergence, biased optimization, and degradation of the model's predictive performance (Jahani-Nasab and Bijarchi, 2024).

To mitigate this issue, this study employs an adaptive loss strategy that dynamically adjusts the weighting of different components within the loss function through trainable coefficients. Unlike methods that rely on fixed or manually assigned weights, the adaptive loss weighting mechanism automatically optimizes the weight distribution and continuously balances the contribution of each loss term during training. This approach mitigates gradient-scale imbalance, prevents any single component from dominating optimization, and directs the model to focus on the most effective convergence paths at various training stages. Consequently, the adaptive weighting strategy enhances both convergence efficiency and prediction accuracy.

The Softmax normalization is applied to constrain these coefficients, ensuring that they remain non-negative and sum to 1 (Xiang et al., 2022). During training, the PINN model uses the adaptive moment estimation (Adam) optimizer to jointly update both the network parameters and the weight coefficients. Similar to the network weights, the coefficients are updated through backpropagation in each iteration, thereby achieving adaptive dynamic balance and improved convergence among the different loss terms (Wang et al., 2021; McClenny and Braga-Neto, 2023).

The weighted loss function can be expressed as:

$$\mathrm{Loss}(\omega,\lambda_{\mathrm{data}},\lambda_{\mathrm{mse}},\lambda_{\mathrm{Soares}}) = L_{\mathrm{data}}(\omega,\lambda_{\mathrm{data}}) + L_{\mathrm{mse}}(\omega,\lambda_{\mathrm{mse}}) + L_{\mathrm{Soares}}(\omega,\lambda_{\mathrm{Soares}}) \tag{22}$$

$$L_{\mathrm{data}}(\omega,\lambda_{\mathrm{data}}) = \frac{1}{N}\sum_{i=1}^{N}\left(\lambda_{\mathrm{data}}^{i}\right)^{2}\left[\left(\mathrm{ROP}_{\mathrm{pred},i} - \mathrm{ROP}_{\mathrm{true},i}\right)^{2}\right] \tag{23}$$

$$L_{\mathrm{mse}}(\omega,\lambda_{\mathrm{mse}}) = \frac{1}{N}\sum_{i=1}^{N}\left(\lambda_{\mathrm{mse}}^{i}\right)^{2}\left[\left(\frac{1.4W_i}{\pi d_{\mathrm{B},i}^{2}} + \frac{0.056\mu R_i W_i}{d_{\mathrm{B},i}\cdot\mathrm{ROP}_{\mathrm{pred},i}}\right)^{2} - \left(\frac{1.4W_i}{\pi d_{\mathrm{B},i}^{2}} + \frac{0.056\mu R_i W_i}{d_{\mathrm{B},i}\cdot\mathrm{ROP}_{\mathrm{true},i}}\right)\right] \tag{24}$$

$$L_{\mathrm{Soares}}(\omega,\lambda_{\mathrm{Soares}}) = \frac{1}{N}\sum_{i=1}^{N}\left(\lambda_{\mathrm{Soares}}^{i}\right)^{2}\left[\left(a_1 D_i^{a_2} W_i^{a_3} R_i^{a_4} Q_i^{a_5} - \mathrm{ROP}_{\mathrm{true},i}\right)^{2}\right] \tag{25}$$

During training, the adaptive weighting mechanism operates in a min-max framework, where the network parameters are optimized to minimize the overall loss, while the weight coefficients are updated to maximize it. This formulation enables collaborative optimization between the network parameters and adaptive weight terms.

$$\min_{\boldsymbol{\omega}}\max_{\lambda_{\mathrm{data}},\lambda_{\mathrm{mse}},\lambda_{\mathrm{Soares}}} \mathrm{Loss}(\boldsymbol{\omega},\lambda_{\mathrm{data}},\lambda_{\mathrm{mse}},\lambda_{\mathrm{Soares}}) \tag{26}$$

The weights are updated according to the following formula:

$$\begin{aligned}
\boldsymbol{\omega}^{k+1} &= \boldsymbol{\omega}^{k} - \eta_k\nabla_{\boldsymbol{\omega}}\mathrm{Loss}\left(\boldsymbol{\omega}^{k},\lambda_{\mathrm{data}}^{k},\lambda_{\mathrm{mse}}^{k},\lambda_{\mathrm{Soares}}^{k}\right)\\
\lambda_{\mathrm{data}}^{k+1} &= \lambda_{\mathrm{data}}^{k} + \eta_k\nabla_{\lambda_{\mathrm{data}}}\mathrm{Loss}\left(\boldsymbol{\omega}^{k},\lambda_{\mathrm{data}}^{k},\lambda_{\mathrm{mse}}^{k},\lambda_{\mathrm{Soares}}^{k}\right)\\
\lambda_{\mathrm{mse}}^{k+1} &= \lambda_{\mathrm{mse}}^{k} + \eta_k\nabla_{\lambda_{\mathrm{mse}}}\mathrm{Loss}\left(\boldsymbol{\omega}^{k},\lambda_{\mathrm{data}}^{k},\lambda_{\mathrm{mse}}^{k},\lambda_{\mathrm{Soares}}^{k}\right)\\
\lambda_{\mathrm{Soares}}^{k+1} &= \lambda_{\mathrm{Soares}}^{k} + \eta_k\nabla_{\lambda_{\mathrm{Soares}}}\mathrm{Loss}\left(\boldsymbol{\omega}^{k},\lambda_{\mathrm{data}}^{k},\lambda_{\mathrm{mse}}^{k},\lambda_{\mathrm{Soares}}^{k}\right)
\end{aligned} \tag{27}$$

where, $\boldsymbol{\omega}$ is the trainable parameters of the neural network, λ_{data}, λ_{mse}, $\lambda_{\mathrm{Soares}}$ is the weight coefficients of different loss components in the weighted loss function, η_k is the learning rate at iteration k, ∇ is the gradient operator, k is the iteration index.

The weight coefficients are updated in the opposite direction of the loss gradients, thereby balancing the contributions of different loss components and stabilizing the optimization process.

The framework of the PINN consists of four main modules, as illustrated in Fig. 6. The original input data are simultaneously fed into multiple SE-ResNet submodules, where high-level features are extracted and recalibrated through residual connections, global pooling, and activation functions.

The recalibrated feature sequences are then passed to a BiLSTM network, where the forward and backward layers capture bidirectional temporal dependencies. After ReLU activation, the hidden states from both directions are fused and sent to the self-attention module, which performs $\boldsymbol{Q}$, $\boldsymbol{K}$, and $\boldsymbol{V}$ mapping, scaled dot-product computation, Softmax normalization, and matrix multiplication to model global feature correlations. The output of the self-attention mechanism is flattened and passed through a linear layer to generate the final prediction.

During training, an adaptive weighted loss function that combines data loss, mechanical specific energy residual, and Soares model residual is employed. The model performs forward prediction and backward optimization simultaneously, where both network parameters and loss weights are jointly updated to achieve dynamic balance and improved training stability.

3. Data processing

3.1. Data overview

The dataset used in this study is derived from the well data of the BZ block. The BZ 19-X structural area is located in an anticlinal structural belt characterized by an intra-depression uplift, sandwiched between the southwestern sub-sag and the southern sub-sag of the BZ Sag. Its northeastern part is adjacent to the main depression of the BZ Sag, the eastern part borders the Bonan Low Uplift, the southern part is connected to the Huanghekou Sag, the southwestern part to the Chengbei Low Uplift, and the northwestern part to the Shaleitian Uplift. Stratigraphically, from top to bottom, the sequence consists of the Quaternary Pingyuan Group, the Neogene (Minghuazhen Formation, Guantao Formation), the Paleogene (Dongying Formation, Shahejie Formation, Kongdian Formation), and the Archean (Zhang et al., 2024; Xu et al., 2024).

The oil and gas resources in the Archean buried-hill strata are buried at relatively great depths. The Archean metamorphic rocks underwent varying degrees of migmatization after vein intrusion, forming migmatites represented by migmatitic granite. These rocks have complex lithological compositions and high strength, which constrain the speed and efficiency of mechanical drilling (Xue et al., 2021; Yi et al., 2022). The distribution and development characteristics of the buried-hill formation in BZ 19-X structure are shown in Fig. 7.

The dataset comprises a total of 42,674 records, including drilling engineering parameters, geological parameters, among

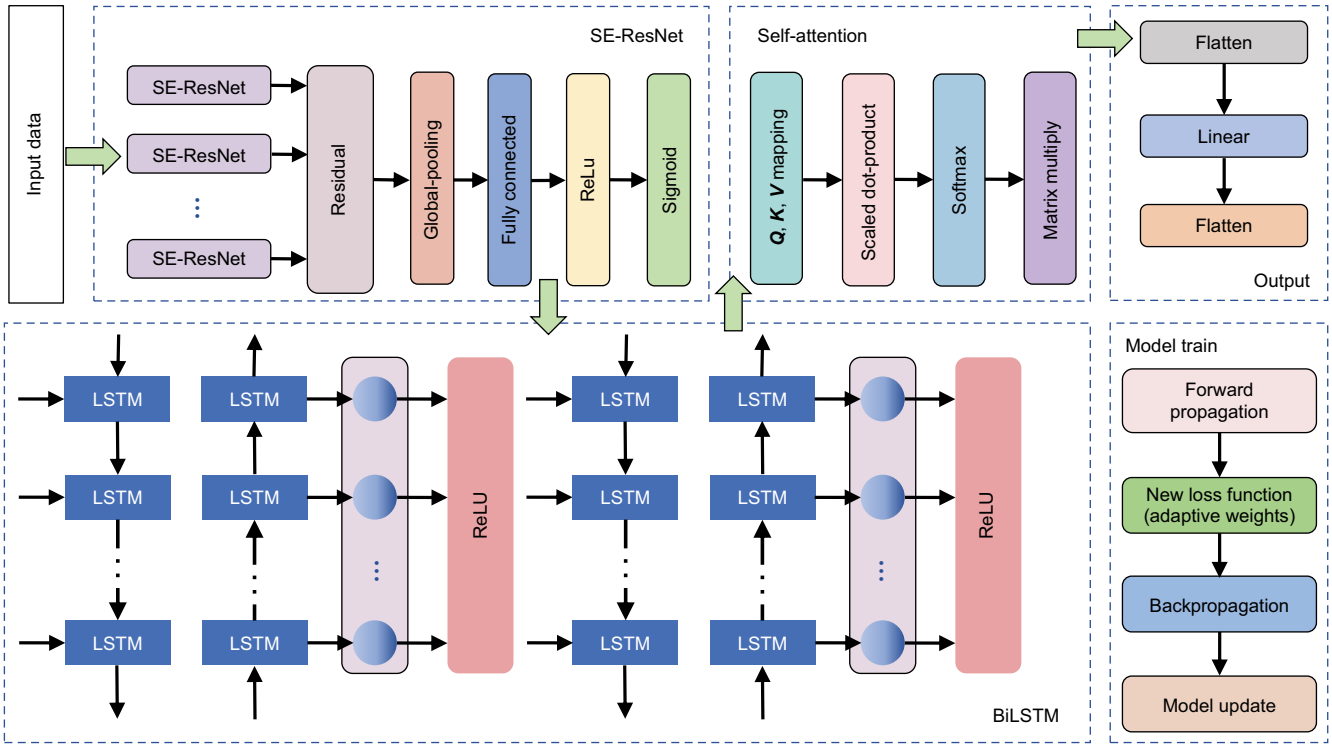

Fig. 6. The framework structure of physics-informed neural networks.

others. The parameters are diverse in type and numerous in quantity, with the main ones being *D*, hook load (HL), WOB, RPM, torque, stand pipe pressure (SPP), *Q*, ROP, bit run, bit time, pump time, caliper (CAL), compensated neutron porosity (CNC), delta time (DT), gamma ray (GR), and ZDL bulk density (ZDEN). The descriptive statistical analysis of the raw data for well A is shown in Table 2.

Fig. 8 shows the distribution of the original data for some parameters. There are significant differences in the distribution characteristics of the parameters. The HL, WOB, RPM, torque, and *Q* exhibit approximately normal distributions. The ROP, however, shows a right-skewed distribution, with most of the data concentrated in the low-value range and only a few samples extending into the higher value range, forming a long tail.

Table 3 presents the bit usage data for well B. Five types of bits were used during the drilling process, with diameters of 660.4, 406.4, 311.2, 215.9, and 152.4 mm, respectively. It should be noted that the bit diameter was directly incorporated as a calculation factor in the mechanical specific energy, and therefore was not treated as an input feature of the model.

3.2. Data cleaning and filtering for noise reduction

Due to factors such as data acquisition methods, sensing devices, and complex environments, datasets often contain missing values, anomalies, and considerable noise. Accordingly, mean imputation is applied to handle missing values, and the Isolation Forest (iForest) algorithm is applied to detect anomalies, thereby improving data quality and model reliability. Fig. 9 shows the underlying principle of the iForest algorithm.

The iForest is an unsupervised anomaly detection algorithm. It builds multiple random trees and detects anomalies by using the path length required to isolate each data point (Al Farizi et al., 2021). The algorithm defines anomalies as 'points that can be readily isolated', that is, as points with low data density and located far from the main data cluster. The anomaly score is obtained by normalizing the path length, with the original algorithm identifying points with higher scores as anomalies (Xu et al., 2023; Zhang and Liu, 2022).

This study implements the iForest anomaly detection algorithm using Scikit-learn. It is worth noting that the decision_function() method in Scikit-learn changes the way anomaly scores are represented: negative scores correspond to anomalies, whereas positive scores indicate normal observations.

Taking WOB and ROP as examples, Fig. 10 presents the scatter plots of anomaly detection results and the histograms of the corresponding anomaly scores.

In the scatter plot, blue '*' markers represent normal samples, while red '+' markers denote detected anomalies. The iForest algorithm effectively differentiates normal data from anomalous points. Anomalies are mainly concentrated in regions with significantly lower or higher values, whereas normal samples exhibit a continuous and stable trend as the well depth increases.

The histogram of anomaly scores shows that each sample is assigned an anomaly score by the iForest algorithm, with lower scores indicating a higher likelihood of being classified as anomalies. A distinct boundary is observed in the score distribution between normal and anomalous samples, where negative scores correspond to the detected anomalies.

After detecting and analyzing anomalous data, the wavelet transform is applied to suppress noise within the signal. The principle of this method is based on the distinct differences in the wavelet coefficients of the signal and noise across multiple scales. According to the characteristics of the signal, an appropriate wavelet basis and decomposition level are first selected, after which an optimal thresholding strategy is applied to reconstruct the denoised signal (Zhang, 2019; Zhang et al., 2023).

This approach effectively reduces the influence of random noise while capturing sudden changes and non-stationary features within the data. It also offers strong smoothing capability and frequency-band separation performance, enabling the reconstructed signal to preserve the essential features of the original

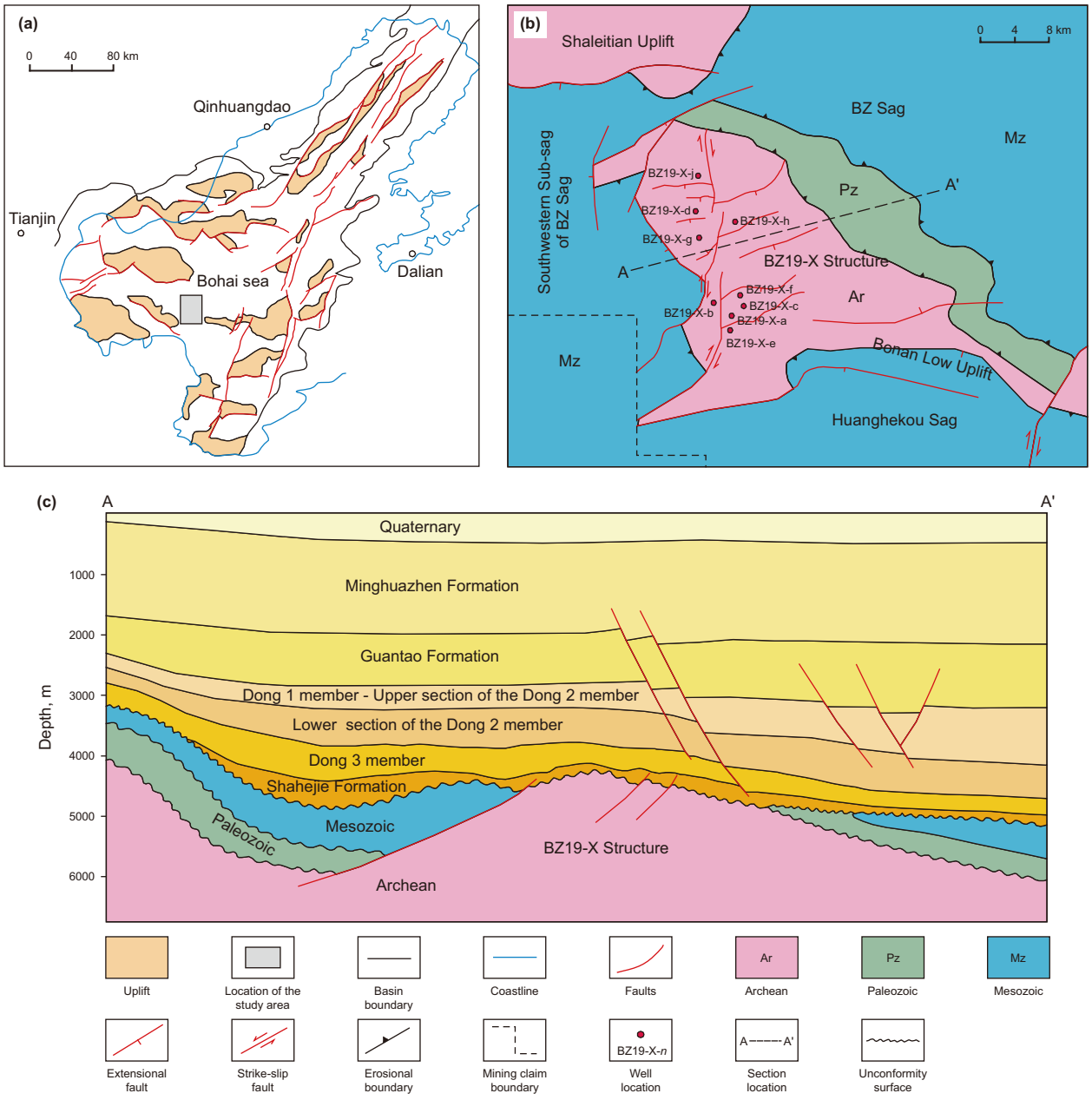

Fig. 7. Structural location and geological cross-section of the BZ19-X structure. (**a**) Regional location of the Bohai Bay Basin; (**b**) Structural position of the BZ19-X area; (**c**) Geological cross-section of the BZ19-X structure along line (A–A').

Table 2
Descriptive statistical analysis table of raw data for Well A.

	D, m	HL, t	WOB, t	RPM, rpm	Torque, kN·m	*Q*, L/min	SPP, MPa	ROP, m/h	Bit run, m	Bit time, h	Pump time, h	CAL, in	CNC, PU	DT, μs/ft	GR, API	ZDEN, g/cm^3
mean	2881.00	133.03	5.79	74.01	12.87	3228.22	16.96	48.09	471.89	13.00	19.38	12.13	22.86	85.39	89.40	2.39
std	1634.33	38.78	3.62	20.8	4.15	1338.39	6.05	47.83	416.89	11.44	13.82	1.27	6.34	10.85	18.05	0.14
min	51.00	41.30	0.00	13.00	0.01	53.00	0.24	0.30	0.10	0.00	0.03	2.63	6.42	56.26	49.79	1.65
25%	1466.00	102.80	2.20	60.00	10.26	1939.00	13.11	14.00	124.20	4.20	8.95	12.03	18.86	76.48	74.43	2.30
50%	2881.00	138.50	6.30	61.00	12.75	3617.00	19.88	34.00	339.10	9.70	16.18	12.18	21.73	83.75	88.04	2.40
75%	4296.00	166.20	8.90	80.00	15.49	4378.00	22.04	68.50	735.70	18.9	27.79	12.35	25.68	94.43	103.95	2.49
max	5711.00	190.80	15.80	132.00	26.96	5428.00	24.85	338.00	1561.80	54.60	62.43	21.14	106.15	125.29	140.35	2.75

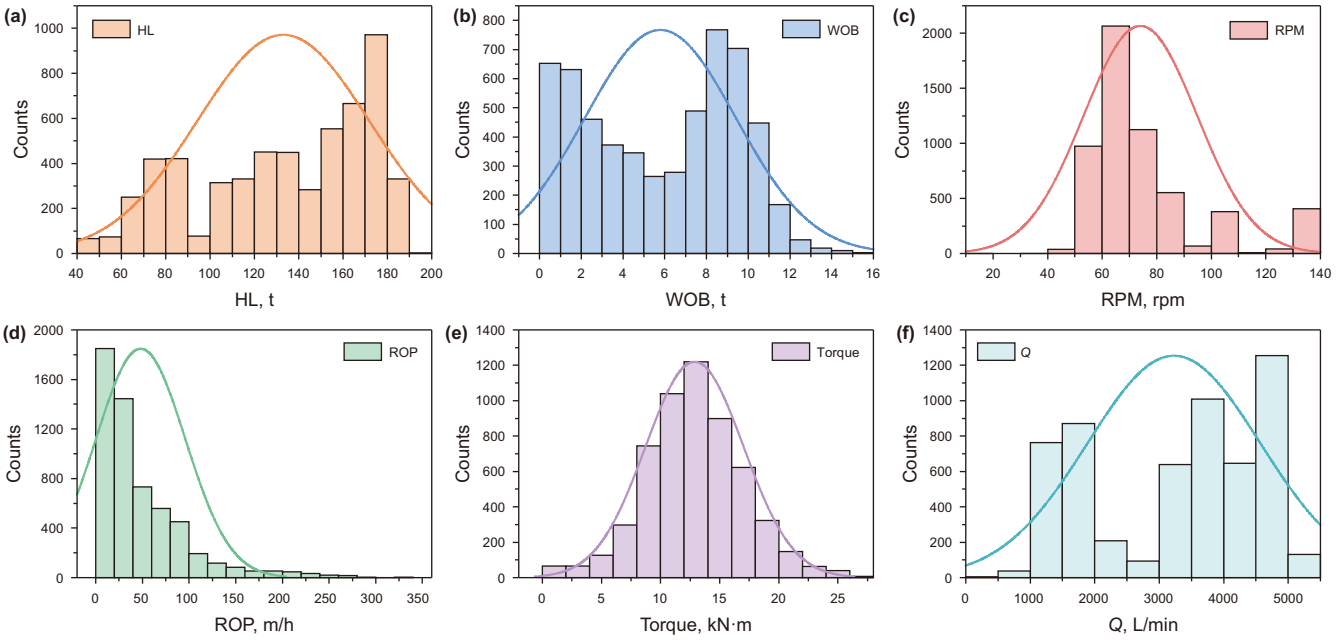

Fig. 8. Distributions of original drilling parameters. (**a**) HL; (**b**) WOB; (**c**) RPM; (**d**) ROP; (**e**) Torque; (**f**) *Q*.

Table 3
Summary of the drill bit usage of Well B.

Number	Bit size, in	Bit model	Bit diameter, mm	Start depth, m	End depth, m	Bit run, m	Drilling time, h
1	26	TS1951SS	660.4	51.00	238.00	187	2.50
2	16	CKS605	406.4	238.00	1296.00	1058.00	12.00
3	16	CKS605	406.4	1296.00	1803.00	507.00	11.25
4	12 1/4	STS935M	311.2	1803.00	3128.00	1325.00	43.25
5	12 1/4	STS935RS	311.2	3128.00	3830.00	702.00	45.00
6	8 1/2	MS1951RS	215.9	3830.00	4337.00	507.00	31.50
7	8 1/2	MS1951RS	215.9	4337.00	4765.00	428.00	25.50
8	8 1/2	TP1654	215.9	4765.00	4907.00	142.00	27.50
9	6	U613M	152.4	4907.00	5050.00	143.00	24.00
10	6	U613M	152.4	5050.00	5335.00	285.00	54.75
11	6	MD637HDY	152.4	5335.00	5341.00	6.00	6.00
12	6	U613M	152.4	5341.00	5448.00	107.00	20.75
13	6	U713M	152.4	5448.00	5649.00	201.00	46.00
14	6	U713M	152.4	5649.00	5711.00	62.00	13.50

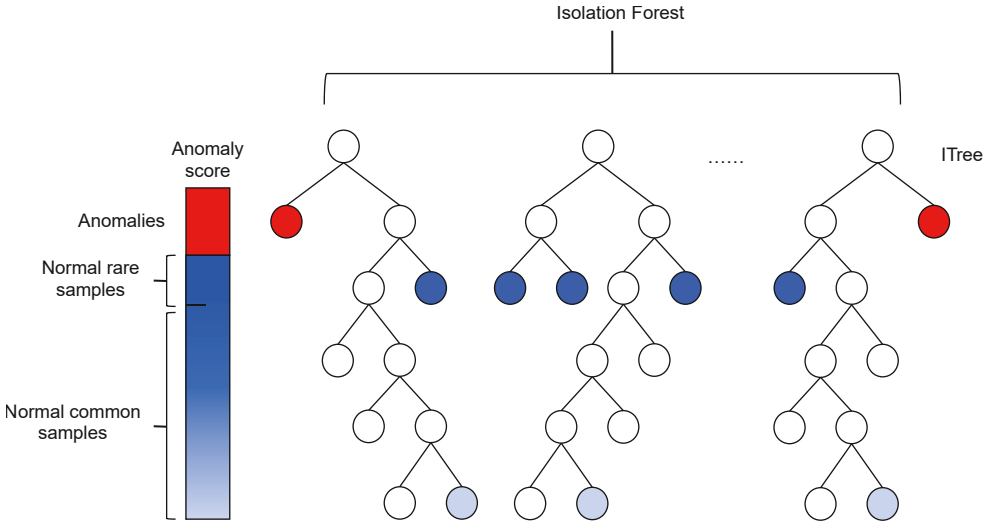

Fig. 9. Isolation forest algorithm schematic.

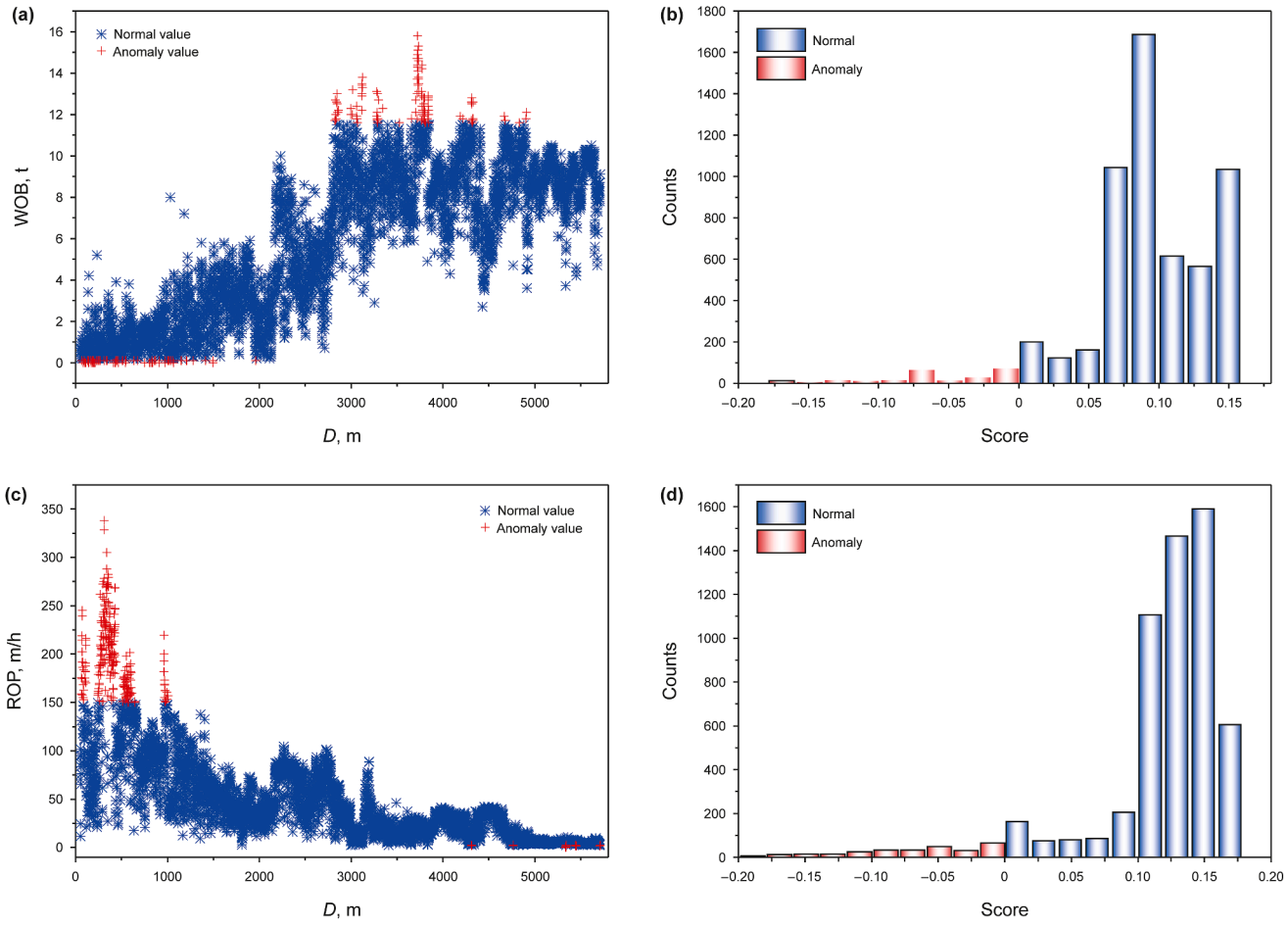

Fig. 10. Scatter plots of anomaly detection results and histograms of anomaly scores. (**a**) Scatter plot of anomaly detection results for WOB data; (**b**) Histogram of anomaly scores for WOB data; (**c**) Scatter plot of anomaly detection results for ROP data; (**d**) Histogram of anomaly scores for ROP data.

data to the greatest extent possible (Harrou et al., 2024; Gan et al., 2019).

$$W_{\psi}\ (f,a,b) = \int_{-\infty}^{+\infty} f(t)\psi\left(\frac{t-b}{a}\right)\mathrm{d}t \tag{28}$$

where, $W_{\psi}(f, a, b)$ is the transform coefficient at scale a and position b, $f(t)$ is the signal, $\psi(\cdot)$ is the wavelet basis function.

In this study, Daubechies 10 was chosen as the wavelet basis function, and the soft thresholding method based on the squared logarithmic rule $\sqrt{2\log n}$ was applied. The denoising procedure was repeated four times. A periodic extension mode was employed to minimize boundary effects while maintaining the continuity and integrity of the overall signal structure (Nigam and Srivastava, 2023).

Fig. 11 presents the denoised curves of selected parameters. As shown in the figure, by applying the wavelet transform with multi-scale decomposition and reconstruction, the high-frequency random noise in the data was effectively suppressed, while the low-frequency key features were well preserved. Consequently, the processed data exhibit smoother and more stable trends, with enhanced robustness and resistance to noise interference. This denoising process ensures high-quality input data for subsequent feature extraction and model training, thereby improving the overall reliability of the predictive modeling framework.

3.3. Feature correlation analysis

In the feature correlation analysis, the first step is to examine the distribution of each feature to identify its type and statistical characteristics, ensuring that an appropriate correlation method is selected. By quantifying the relationships among features, the correlation analysis helps uncover the intrinsic dependencies between variables and identify the key features most strongly associated with the target variable. This process effectively filters redundant or irrelevant inputs, thereby enhancing the interpretability and predictive performance of the model.

The Shapiro-Wilk test is a statistical method used to determine whether a data sample follows a normal distribution. It works by comparing the sample values with the expected values under a normal distribution to evaluate whether the data conforms to the normality assumption (González-Estrada et al., 2022; Monter-Pozos and González-Estrada, 2024).

The test statistic W reflects the degree of deviation of the data from a normal distribution. A value of W closer to 1 indicates that the sample more closely follows normality, whereas lower values suggest departures from the normal distribution.

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{29}$$

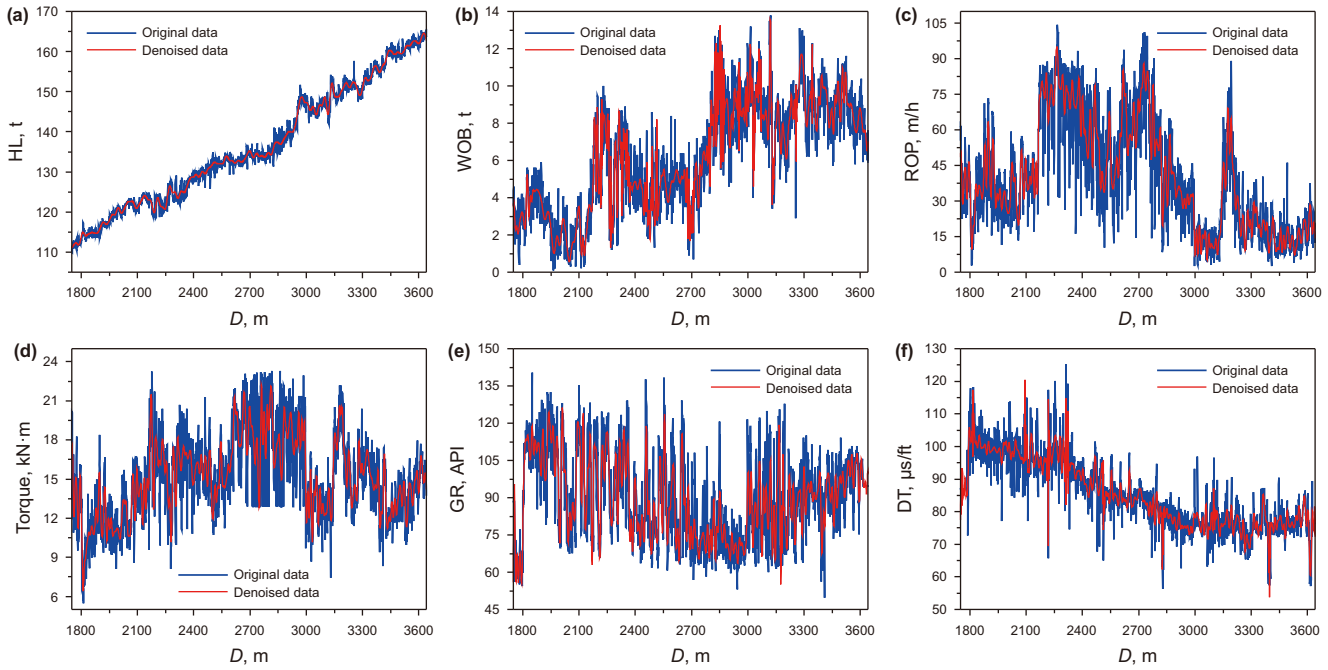

Fig. 11. Comparison of original and denoised curves for selected parameters. (**a**) HL; (**b**) WOB; (**c**) ROP; (**d**) Torque; (**e**) GR; (**f**) DT.

where, W is the test statistic, a_i denotes the constants weights derived from the normal distribution, $x_{(i)}$ represents the ordered sample data, $\bar{x}$ is the sample mean.

When the assumption of normality holds, the p-value, representing the probability of observing the sample difference from the normal distribution under the null hypothesis, is used to assess normality. The significance level is set at 0.05. If the p-value is less than the significance level, the data are considered to deviate from normality. If the p-value is greater, the data can be regarded as approximately normally distributed (Arnastauskaitė et al., 2021). The results show that the p-values of all parameters are well below the significance level, indicating that the distributions of these variables do not conform to normality.

The correlations between drilling engineering parameters and geological parameters were analyzed using the Spearman's rank correlation coefficient (Ali Abd Al-Hameed, 2022; Yu and Hutson, 2024), and the correlation heat map is shown in Fig. 12.

In the upper triangular region, ellipses are employed to represent the magnitude and direction of the Spearman's correlation coefficient. Elongated ellipses indicate stronger correlations, whereas shapes closer to a circle denote weaker correlations. The orientation of the ellipses reflects the relationship type, with ellipses inclined from the lower left to the upper right representing positive correlations, and those from the upper left to the lower right representing negative correlations.

In Spearman's correlation, a value of 0 indicates no correlation, -1 indicates a negative correlation, and 1 indicates a positive correlation. The closer the absolute value is to 1, the stronger the monotonic correlation. It can be observed that parameters such as D, WOB, HL, torque, and DT exhibit strong monotonic correlations with ROP, whereas parameters such as RPM and bit run show relatively weak monotonic correlations with ROP.

However, according to established mechanical rock-breaking theory in drilling engineering, parameters such as D, WOB, RPM, torque, and bit run are known to be important influencing factors of ROP. Spearman's rank correlation primarily evaluates linear monotonic and nonlinear monotonic relationships between variables. In cases involving complex multivariate nonlinear coupling, Spearman correlation alone may not fully capture such intricate dependencies.

To further explore potential nonlinear dependencies among variables, the maximal information coefficient (MIC) was utilized to quantify the maximum degree of nonlinear statistical association between each input parameter and the target variable. As an improvement over the traditional mutual information method, MIC can effectively capture both linear and non-linear relationships between two variables by partitioning the data space into multi-scale grids and identifying the optimal division that maximizes mutual information (Kinney and Atwal, 2014; Chen et al., 2023).

$$I(X;Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{30}$$

$$\mathrm{MIC}(X,Y) = \max_{G_{xy}} \frac{I(X;Y)}{\log(\min\{m,n\})} \tag{31}$$

where, $I(X;Y)$ is the mutual information, $p(x, y)$ is the joint probability distribution of X and Y, $p(x)$ and $p(y)$ are the marginal probability distributions of X and Y, respectively, G_{xy} denotes the two-dimensional grid partitions used to capture different dependency patterns between X and Y, $\log(\min\{m,n\})$ is the normalization factor that constrains the MIC value between 0 and 1, where m and n represent the sample sizes of X and Y, max indicates the maximum normalized mutual information score obtained across all possible grid partitions.

Fig. 13 presents the importance scores of all features. After ranking by MIC values, the five features most strongly correlated with the ROP are D, Q, HL, WOB, and bit run, while features such as SPP, pump time, CNC, and ZDEN show relatively weak correlations. With a threshold of MIC > 0.40, 11 effective input features were obtained: D, Q, HL, WOB, bit run, GR, RPM, torque, DT, bit time, CAL.

Considering the sequential characteristics of the drilling process, the lagged feature ROP($t-1$) was also included to enhance the

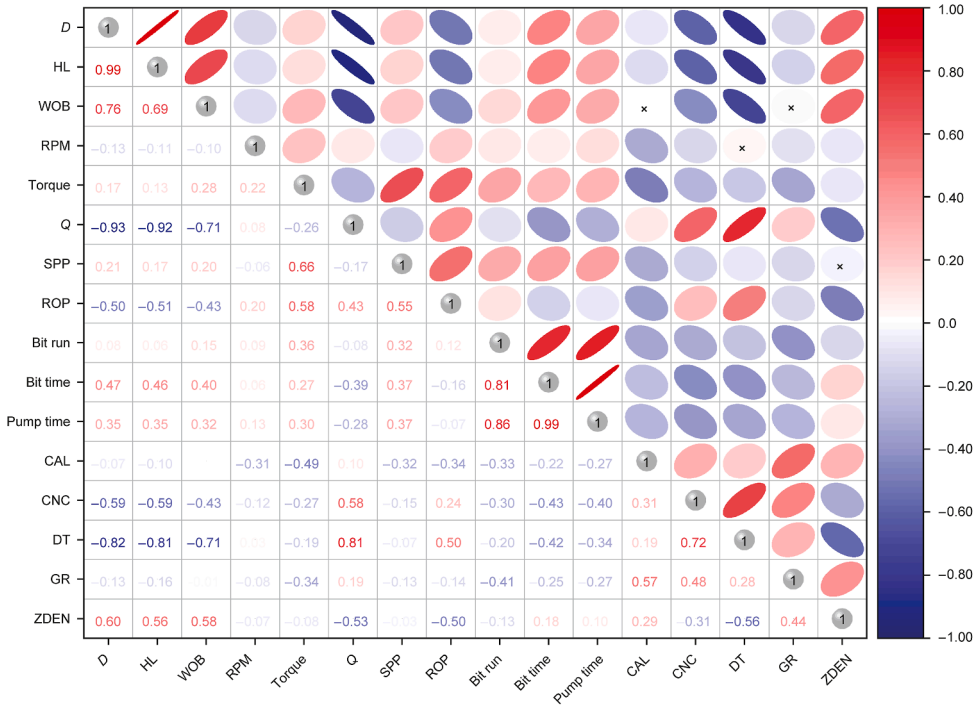

Fig. 12. Heatmap of spearman's rank correlation coefficients among parameters.

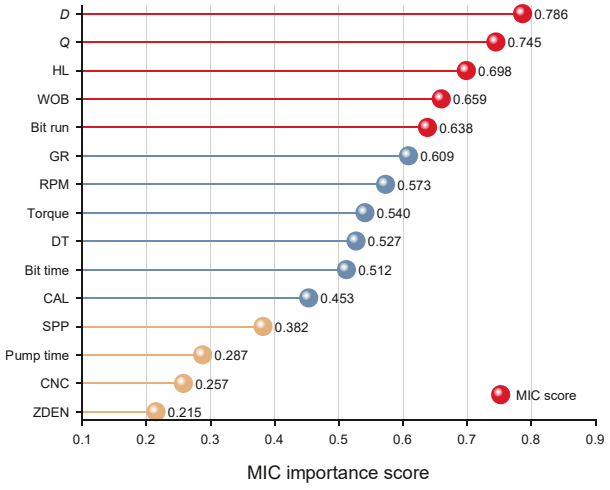

Fig. 13. MIC importance scores of parameters.

model's ability to capture dynamic variations. Finally, 12 parameters were selected as the input parameters for the ROP prediction model, as shown in Table 4.

3.4. Data standardization

Due to the substantial differences in the numerical ranges of drilling parameters, the *z*-score standardization method was applied to normalize the data. This transformation scales the parameters to have a mean of 0 and a standard deviation of 1, thereby eliminating the effects of differing dimensions across features and improving both the convergence speed and performance of the model (Al-Faiz et al., 2018; Owen et al., 2022).

Table 4
The final input parameters of the model.

Parameter type	Parameter name
Geological parameters	GR, DT, CAL
Engineering parameters	*D*, *Q*, HL, WOB, RPM, torque, ROP($t-1$)
Drill bit parameters	Bit run, bit time

$$z = \frac{x - \mu}{\sigma} \tag{32}$$

where, x is the original data value, μ is the mean of the original data, σ is the standard deviation of the original data, z is the standardized value.

4. Experimental design and analysis

4.1. Model training and optimization

This paper presents a systematic process of model training and hyper-parameter optimization to ensure optimal model performance. The process consists of four stages: initial model training, Bayesian hyper-parameter optimization, controlled variable analysis, and training with optimal parameters combined with dynamic learning rate scheduling.

Considering that the drilling data are arranged sequentially with respect to depth and exhibit continuity, the dataset was treated as a depth-based sequence and partitioned into training, validation, and test sets with a ratio of 6:2:2. This approach allows the training set to be used for model learning, the validation set for hyper-parameter tuning and overfitting control, and the test set for evaluating model generalization performance, while effectively preventing data leakage between subsets.

The PINN model was trained and predicted using the PyTorch framework, and the training and optimization were performed using NVIDIA GPU parallel computing, which can accelerate the convergence process and improve computational efficiency. In the initial model training phase, empirically preset hyper-parameters were employed for baseline training to establish the model's performance benchmark. To mitigate the risk of overfitting, the dropout mechanism was introduced during training, randomly dropping some neurons with a certain probability during forward propagation to reduce the model's complexity. The model's initial hyper-parameter settings are shown in Table 5.

To optimize the network structure and improve model performance, the minimization of root mean square error (RMSE) was defined as the optimization objective. A Bayesian hyper-parameter optimization algorithm was implemented within the Optuna framework to jointly search the learning rate, the number of LSTM layers, the number of hidden units, and the dropout rate, thereby identifying the optimal initial hyper-parameter configuration.

During sampling, the algorithm employs the tree-structured parzen estimator (TPE) strategy, which constructs two probability density functions to represent superior and average hyper-parameter distributions. Parameters are preferentially selected from the superior distribution to accelerate the optimization process and improve search efficiency (Zhou et al., 2024; Hanifi et al., 2024). The hyper-parameter optimization settings and value ranges are shown in Table 6.

The results of the Bayesian optimization process indicate that the optimal hyper-parameter configuration consists of a learning rate of 1e-4, 3 LSTM layers, 64 hidden units, and a dropout rate of 0.4. Under this configuration, the model achieves the minimum error.

To further evaluate the independent effects of different hyper-parameters on model performance, the control variable method was adopted. With all other parameters fixed, each hyper-parameter was varied individually, and the corresponding changes in RMSE were recorded. Fig. 14 illustrates the performance response curves of the model under different hyper-parameter settings.

The results indicate that the learning rate, the number of LSTM layers, the number of hidden units, and the dropout rate all exert significant influence on model performance. The model is highly sensitive to the learning rate, an appropriately chosen value accelerates convergence, whereas values that are too large or too small cause noticeable degradation in performance. The number of LSTM layers plays a decisive role in feature extraction, where an optimal depth enables more effective pattern learning. Similarly, the number of hidden units determines the model's representational capacity and directly affects its learning ability. The dropout rate acts as an important regularization factor. When set to 0.4, it effectively suppresses overfitting and enhances generalization, achieving the best overall performance.

Table 5
The initial hyper-parameter settings of the model.

Parameter	Setting	Parameter	Setting
Input dimension	12	Output dimension	1
Activation function	ReLU	Optimizer	Adam
Loss function	Loss	Training epochs	100
Learning rate	0.0001	LSTM layers	3
Hidden units	64	Dropout rate	0.3
Batch size	64	Residual blocks	2
Input channels	1	Convolution kernel size	3
Output channels	128	Attention heads	8

Table 6
The hyper-parameter optimization settings and value ranges.

Parameter	Value range
Learning rate	[1e-6, 1e-5, 1e-4, 1e-3, 1e-2, 1e-1]
LSTM layers	[1, 2, 3, 4, 5, 6]
Hidden units	[32, 64, 96, 128, 160, 192, 224, 256]
Dropout rate	[0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6]

A composite strategy combining static hyper-parameter optimization with dynamic learning rate scheduling is adopted. After the optimal hyper-parameter configuration is determined, the OneCycleLR scheduler is introduced to adaptively adjust the learning rate according to training progression, thereby improving training efficiency. During the early training phase, the scheduler gradually increases the learning rate to its maximum value through either a linear or cosine function, and subsequently decreases it to a minimum. This scheduling mechanism enables the model to escape local optima and expedites the convergence process (Smith and Topin, 2019). In this study, the core parameters of the OneCycleLR learning rate scheduler are shown in Table 7.

Fig. 15 shows the loss function curves of the model along with the corresponding variations in the learning rate. The step size of the learning rate scheduler is defined as the product of the number of mini-batches per epoch and the total number of training epochs, which is simplified to the number of epochs for visualization purposes.

At the beginning of training, both the training and validation losses exhibit a clear downward trend, and their trajectories remain largely consistent. As training proceeds, the training loss decreases rapidly and then stabilizes, while the validation loss shows mild fluctuations but ultimately remains at a relatively low level. These patterns indicate that the model achieves fast and stable convergence, successfully learning the feature dependencies within the data without significant overfitting.

The learning rate follows a typical single-cycle cosine schedule, gradually increasing during the initial epochs to help the model escape local optima and enhance its global search ability. After approximately 30 epochs, it reaches the maximum value and then gradually decreases, enabling finer parameter adjustments in later stages. This scheduling strategy effectively balances the needs for rapid convergence and stable optimization.

In addition, the evolution of the adaptive loss weighting reveals that, at the early stage of training, the three loss components maintained approximately balanced weights, representing the learning phase dominated by physical constraints. As the model parameters and gradients were continuously updated and optimized, the weight of the data-driven term L_{data} gradually increased and stabilized at around 0.65 by the end of training, while the two physics-based constraint losses, L_{mse} and L_{Soares}, converged to approximately 0.15 and 0.20, respectively.

To balance the gradient magnitudes among different loss components, the adaptive mechanism automatically increases the weight of the data-driven term as the model progressively learns and internalizes the underlying physical relationships. This adjustment guides the optimization process toward a more efficient convergence direction and ensures the stability of the training process. Although the final numerical weights of the physics-based terms are relatively small, their gradient signals continuously play a regularizing and constraining role throughout training, ensuring the physical consistency of the model predictions.

Overall, the adaptive weighting strategy enables a natural dynamic transition during training, evolving from an initial stage of physics-guided learning to a later stage of data-driven optimization. This evolution illustrates the model's ability to autonomously shift learning focus based on convergence feedback, ensuring both

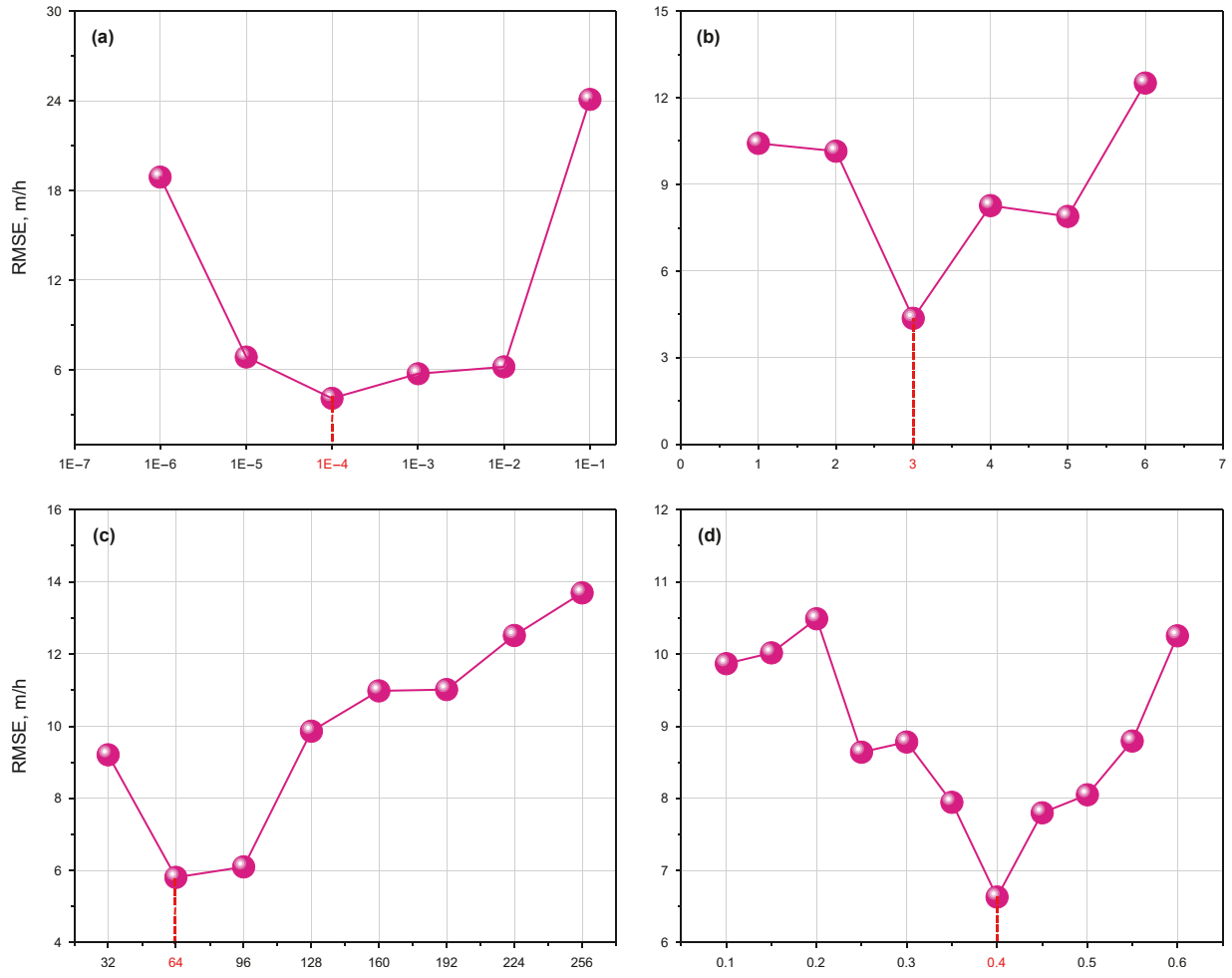

Fig. 14. Performance response curves under different model hyper-parameters. (**a**) Learning rate; (**b**) LSTM layers; (**c**) Hidden units; (**d**) Dropout rate.

Table 7
The core parameter settings of OneCycleLR learning rate scheduler.

Parameter	Setting	Parameter	Setting
Maximum learning rate	0.001	Annealing strategy	Cosine
Initial learning rate factor	25	Final learning rate factor	1000

empirical accuracy and physical fidelity, thereby achieving a dynamic balance between physics-informed guidance and data-driven learning.

4.2. Prediction results analysis

This study uses RMSE, mean absolute error (MAE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2) as evaluation metrics to assess model performance from multiple perspectives (Bai et al., 2024).

To comprehensively evaluate the performance of the proposed PINN model, three classical physical models for ROP prediction, the Bourgoyne and Young model (B&Y), the Bingham model, and

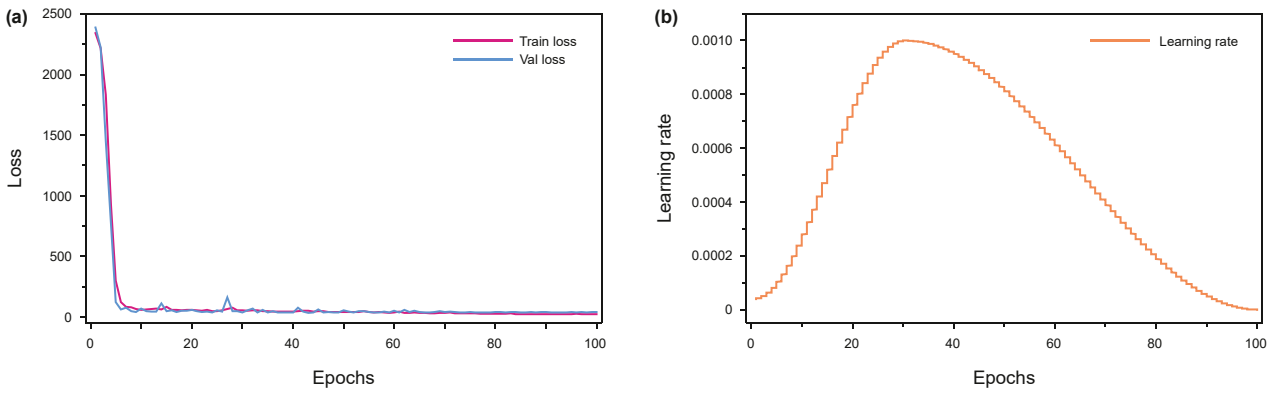

Fig. 15. Loss function and learning rate variation. (**a**) Model loss function curve; (**b**) Learning rate change curve.

the Soares model, were selected to perform ROP prediction on the same dataset.

The B&Y model describes the relationship between the ROP and eight influencing factors from multiple perspectives.

$$\mathrm{ROP} = \exp\left(a_1 + \sum_{i=2}^{8} a_i x_i\right) = \prod_{i=1}^{8} f_i \tag{33}$$

where, a_1 is a proportional coefficient, x_i represents the influencing parameters of ROP, and f_1–f_8 denote the corresponding functional relationships of the influencing factors. Specifically, f_1 represents formation strength, f_2 is the depth influence, f_3 corresponds to formation compaction, f_4 accounts for the pressure differential, f_5 represents the WOB effect, f_6 denotes RPM, f_7 describes bit wear, and f_8 reflects hydraulic effects.

However, due to the absence of several parameters required in the original B&Y model within the collected dataset, a simplified six-factor version of the B&Y equation was developed in this study. The simplified form considers the effects of formation drillability, D, WOB, RPM, Q, and bit diameter.

$$\mathrm{ROP} = K \cdot \mathrm{e}^{a_1 \frac{D}{1000}} \cdot \mathrm{e}^{a_2 \left(\frac{D-D_0}{1000}\right)^2} \cdot \left(\frac{W}{W_0}\right)^{a_3} \cdot \left(\frac{R}{R_0}\right)^{a_4} \cdot \left(\frac{Q}{Q_0}\right)^{a_5} \cdot \left(\frac{d_\mathrm{B}}{d_{\mathrm{B}0}}\right)^{a_6} \tag{34}$$

where, K is the formation drillability coefficient, D_0, W_0, R_0, Q_0, $d_{\mathrm{B}0}$ are the reference parameters used for normalization, a_1 and a_2 represent the compaction and undercompaction coefficients, and a_3–a_6 are the exponents of WOB, RPM, Q, and bit diameter, respectively.

Bingham established a function relating ROP to WOB and RPM.

$$\mathrm{ROP} = K\left(\frac{W}{d_\mathrm{B}}\right)^a R \tag{35}$$

where, K is a proportional constant, which takes into account the influence of rock strength, and a is the drill bit index.

Based on the principle of least squares, the curve_fit function in the Scipy library was used to perform nonlinear regression fitting for the three drilling rate equations respectively, and the dimensionless parameters in the three equations were obtained.

The evaluation indicators of the physical model are shown in Table 8. It can be seen that the prediction accuracy of the physical model is relatively low, and the error between the predicted results and the true values is large. Among the three physical models, the Soares model has the lowest MAE and MAPE, which are 16.54 and 58.89% respectively, and the R^2 is 0.46.

The prediction results of the physical models are illustrated in Fig. 16. In each scatter plot, the horizontal axis represents the true ROP values, while the vertical axis represents the predicted ROP values. Points that lie closer to the diagonal line, or are evenly distributed along both sides of it, indicate higher prediction accuracy. Conversely, larger deviations from the diagonal imply lower predictive accuracy.

Table 8
Comparison of performance evaluation indicators for physical model.

	B&Y	Bingham	Soares
RMSE	26.22	49.63	21.75
MAE	19.35	30.08	16.54
MAPE, %	62.92	93.49	58.89
R^2	0.31	0.06	0.46

It can be observed that the prediction results of all three physical models show considerable deviations from the measured values, indicating relatively low accuracy. In contrast, machine learning models can flexibly capture nonlinear correlations and hidden coupling effects among multiple drilling parameters by learning directly from historical drilling data, without relying on explicit physical equations.

Therefore, to further verify the predictive performance and overall superiority of the proposed PINN model, several commonly used machine learning algorithms, SVR, BP, LSTM, and convolutional neural network-bi-directional long short-term memory (CNN-BiLSTM), were selected for comparative analysis against the PINN model. All models were tested on the same dataset to ensure consistency.

The core parameters of the SVR model include the radial basis function (RBF) as the kernel, a penalty coefficient $C = 1.0$, and a kernel coefficient γ set to "scale". For the BP neural network, the key settings consist of a ReLU activation function and three hidden layers with 128, 64, and 32 units, respectively. In both the LSTM and CNN-BiLSTM models, the main hyper-parameters, such as the number of LSTM layers, hidden units, learning rate, dropout rate, and convolution kernel size, are consistent with the optimal configuration used in the SE-ResNet-BiLSTM-SAM architecture, which constitutes part of the proposed PINN framework.

Table 9 presents the performance evaluation metrics of the different models, providing a direct comparison across RMSE, MAE, MAPE, and R^2. Fig. 17 further depicts the radar chart of model performance, which highlights the overall disparities among the five models across all evaluation dimensions.

For the proposed PINN model, the RMSE is 6.214, the MAE is 3.483, the MAPE is 6.102%, and the R^2 is 0.953. Compared with the SVR, BP, LSTM, and CNN-BiLSTM models, the PINN model outperforms across all four metrics. Specifically, MAPE decreases from 21.193% in the SVR model to 6.102% in the PINN model, while R^2 increases from 0.777 to 0.953. These results demonstrate that the PINN model not only reduces prediction errors but also achieves higher goodness of fit, while additionally exhibiting superior stability and generalization ability.

In order to further reveal the differences in error distribution characteristics among the models and to verify their stability and predictive precision, a comprehensive analysis of the cumulative error distribution and error satisfaction was performed for each ROP prediction model.

Fig. 18 illustrates the cumulative error distribution and statistical comparison of different models. In subplots (a–e), the cumulative error curves depict the discrepancies between the predicted and measured ROP values. The metrics d_{10}, d_{50}, and d_{90} represent the error magnitudes corresponding to the 10th, 50th, and 90th percentiles, respectively, reflecting the model's prediction accuracy across different confidence levels.

The results indicate that the PINN model achieves the highest prediction precision and stability. Its error distribution is sharply concentrated around zero, forming a narrow and steep peak, which signifies that the vast majority of prediction residuals are minimal. In contrast, the SVR and BP models exhibit wide error dispersion with heavy tails, indicating a significant number of large deviations. The LSTM and CNN-BiLSTM models perform moderately better, showing a more compact distribution with reduced volatility compared to the traditional machine learning models. Among them, CNN-BiLSTM shows a more compact distribution than LSTM, approaching the performance of the PINN model.

Additionally, subplot (f) presents the boxplot comparison of prediction errors, which further highlights the better stability of the PINN model. Its interquartile range (IQR) and variance are the

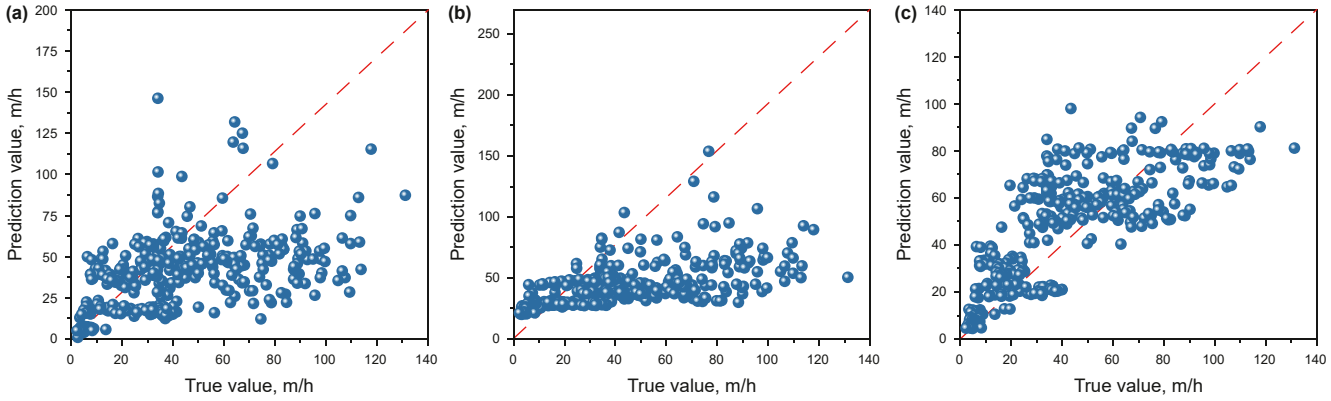

Fig. 16. Scatter plot of the prediction results of the physical model. (**a**) B&Y model; (**b**) Bingham model; (**c**) Soares model.

Table 9
Comparison of performance evaluation indicators for different models.

Model	RMSE	MAE	MAPE, %	R^2
SVR	13.939	8.544	21.193	0.777
BP	11.402	6.978	17.788	0.851
LSTM	10.002	5.877	15.008	0.905
CNN-BiLSTM	7.839	4.548	10.503	0.929
PINN	6.214	3.483	6.102	0.953

smallest among all models, confirming that the proposed PINN approach effectively suppresses large deviations and ensures consistent predictive performance.

In engineering practice, predictive results are generally allowed to deviate within an acceptable error range. To address the limitations of conventional evaluation metrics such as RMSE and MAPE, which are highly sensitive to individual extreme errors, Fan et al. (2025b) proposed the "error satisfaction" metric.

This metric evaluates the reliability of overall predictions by calculating the relative error between predicted and true values and by determining the proportion of sample points that fall within a predefined error range. Compared with RMSE and MAPE, the error satisfaction metric is less sensitive to individual anomalies and offers a method of greater engineering value for the reassessment and improvement of model predictive performance.

$$S_j = \frac{T_j}{T} \times 100\% \tag{36}$$

where, j denotes the permissible error range, S_j represents the error satisfaction when the error range is set to j, T is the total number of samples, T_j is the number of samples with prediction errors falling within the error range. The error ranges used for calculating error satisfaction are shown in Table 10.

Fig. 19 shows the error satisfaction bubble charts for the different models. The results indicate that as the error range increases, the error satisfaction of all models improves markedly. Among the five models, the PINN model achieves the highest error satisfaction. When the error range is set to S_{15}, the PINN model attains an error satisfaction of 89.71%, which further increases to 97.62% at S_{25}. This means that 97.62% of the predicted values deviate from the true observations by less than 25%. The CNN-BiLSTM model ranks second, showing higher error satisfaction than the LSTM, BP, and SVR models under the same conditions. In contrast, the SVR, BP, and LSTM models exhibit lower overall error satisfaction, indicating larger deviations between predictions and true values.

Fig. 20 compares the prediction curves of different models. The results indicate that although all models can capture the overall ROP trend, notable differences exist in local fitting accuracy and error control.

The SVR and BP models exhibit significant overall deviations and insufficient fitting between predicted and true values, with certain predictions showing clear discrepancies. This highlights that an accurate overall trend does not necessarily guarantee superior error metrics, as large deviations in individual samples can still elevate the global error level. In contrast, the LSTM and CNN-BiLSTM models demonstrate closer alignment with the true values, effectively capturing the overall trend but showing localized fluctuations in certain intervals. The PINN model, however, exhibits the best performance, closely matching the true values and accurately representing both global trends and local variations.

These results suggest that by integrating physical constraints with an adaptive loss-weighting mechanism, the PINN model mitigates the "black box" nature of traditional data-driven approaches, achieving enhanced predictive accuracy and stability.

4.3. Ablation experiment and feature analysis

To further investigate the internal mechanism of the proposed PINN model, a series of ablation experiments were conducted to quantify the contributions of different structural and physical components to the overall prediction performance.

The ablation study, a widely adopted analytical approach in artificial intelligence and machine learning research, involves removing or replacing specific modules, features, or training strategies to examine how such changes affect model performance. This helps to identify the relative importance of individual components and clarify their functional roles within the overall framework.

In this study, six ablation experiments were systematically designed to evaluate the effects of the physical constraints, hybrid deep-learning architecture, and adaptive loss weighting strategy in the PINN model. Apart from the removal of individual components, all other experimental settings, parameters, and training conditions remained consistent with those of the full model. The specific design of the ablation experiment is as follows:

(1) Experiment 1 removes all physical constraints, reverting the framework to a purely data-driven model for ROP prediction;
(2) Experiment 2 eliminates the mechanical specific energy constraint while retaining the Soares ROP model as a single physical term;

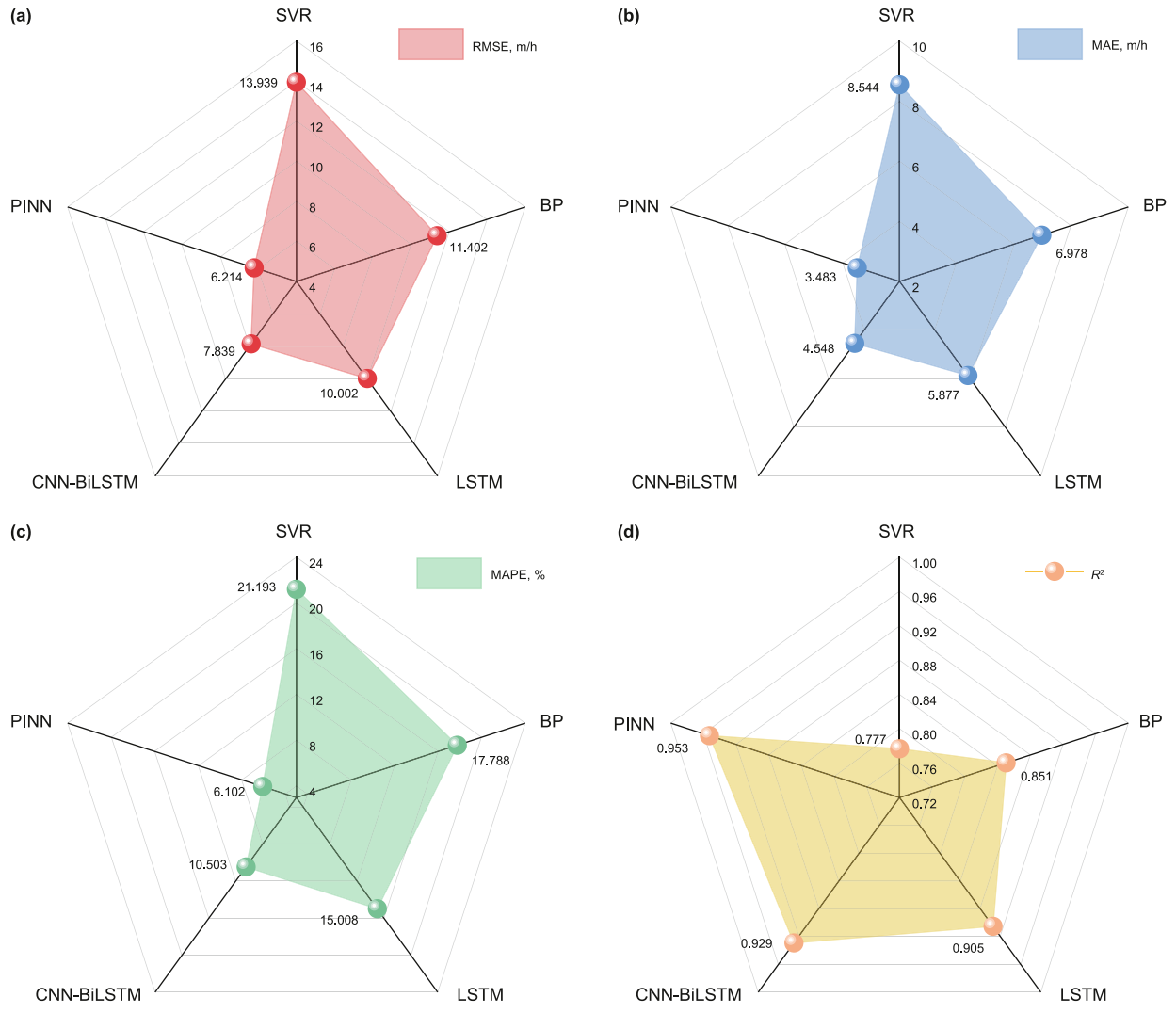

Fig. 17. Radar chart of performance indicators for different models. (**a**) RMSE; (**b**) MAE; (**c**) MAPE; (**d**) R^2.

(3) Experiment 3 removes the Soares model constraint and keeps only the mechanical specific energy constraint;
(4) Experiment 4 excludes the SAM module from the hybrid architecture;
(5) Experiment 5 removes the SE-ResNet module while keeping the remaining architecture unchanged;
(6) Experiment 6 replaces the adaptive loss-weighting mechanism with fixed weights for the different loss components, where the data-driven term L_{data} is set to 1.0, and the two physics-based constraint terms L_{mse} and L_{Soares} are each set to 0.1.

As shown in Fig. 21, the performance metrics of different ablation experiments indicate that removing various components leads to different degrees of degradation in model accuracy, suggesting that each module plays an essential role in the PINN framework.

In Experiment 1, removing all physical constraints increased the RMSE by 48% and reduced the R^2 to 0.903. Without physical guidance, the model generated extreme, non-physical predictions, illustrating that physical constraints are essential for maintaining prediction stability and physical plausibility.

In Experiments 2 and 3, removing the mechanical specific energy and Soares constraints led to RMSE increases of 28% and 22%, respectively. The former contributed more strongly because of its closer relationship to the energy balance in rock-breaking, effectively suppressing non-physical fluctuations.

In Experiments 4 and 5, excluding the SAM and SE-ResNet components raised the RMSE by 38% and 56%, and reduced the R^2 to 0.925 and 0.892, respectively. The SAM module mainly captures global feature interactions and has a moderate influence, whereas the SE-ResNet module is crucial for extracting nonlinear features, with its removal severely weakened feature representation capability.

In Experiment 6, replacing the adaptive loss weighting with fixed weights increased the RMSE by 24%. Although the model still achieved reasonable predictive accuracy under the dominance of the data-driven term, several predictions deviated from the physically reasonable range. Moreover, the convergence became slower and the overall training stability was reduced.

Overall, embedding physical information loss restricts model predictions within a physically reasonable range, while the SE-ResNet-BiLSTM-SAM architecture effectively captures the multi-scale and long-sequence dependencies among drilling parameters. Therefore, the proposed PINN model not only achieves higher predictive accuracy but also demonstrates improved interpretability and physical consistency.

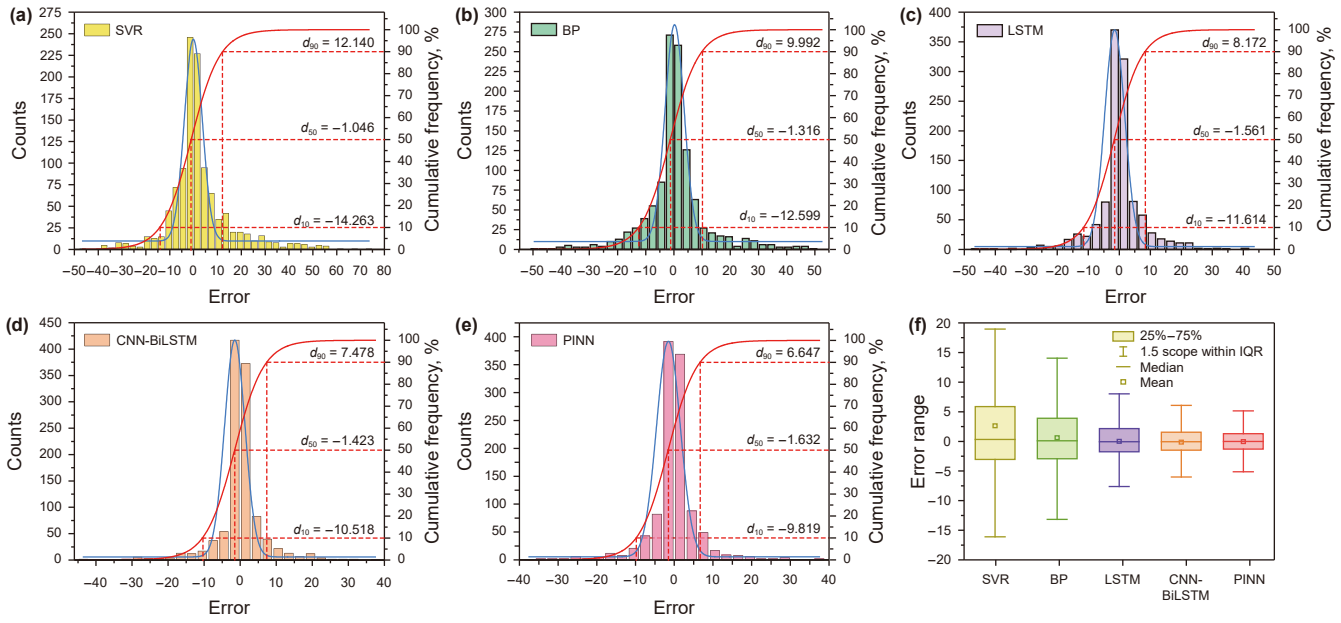

Fig. 18. Error distribution and comparative analysis of different models. (**a**) SVR; (**b**) BP; (**c**) LSTM; (**d**) CNN-BiLSTM; (**e**) PINN; (**f**) Boxplot of prediction errors.

Table 10
Error range of error satisfaction.

Number	Type	Error range
1	S_5	Relative error ≤ 5%
2	S_{10}	Relative error ≤ 10%
3	S_{15}	Relative error ≤ 15%
4	S_{20}	Relative error ≤ 20%
5	S_{25}	Relative error ≤ 25%

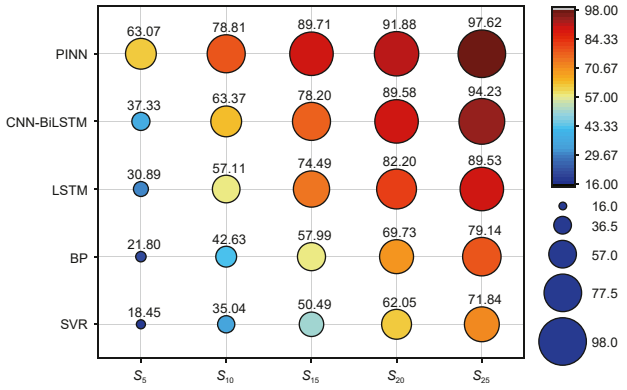

Fig. 19. Error satisfaction bubble chart for each model.

For the inherently non-transparent "black box" models, the self-attention mechanism and the SHAP method were jointly employed to quantitatively analyze the contribution of each feature to the ROP prediction in the proposed PINN framework.

Within the PINN model, the SAM is primarily designed to capture the global dependencies and interaction weights among input features, thereby enhancing the model's ability to represent complex coupled relationships among drilling parameters. By integrating the SHAP post-hoc explainability approach, which is based on a game-theoretic allocation principle to compute each feature's marginal contribution to the model output, the internal decision process of the model can be interpreted in a statistically meaningful way. This enables a deeper understanding of how different drilling features influence the predicted ROP. The feature importance dependencies are illustrated in Fig. 22.

The horizontal axis in the figure represents the contribution magnitude of each input feature to the predicted ROP. Each point corresponds to an individual sample, and the color indicates the feature value, where red represents higher values and blue represents lower values. The vertical dispersion of points reflects the variability of each feature's influence across all samples, while the horizontal direction indicates whether the feature has a positive or negative effect on the predicted ROP.

Overall, ROP($t-1$), torque, WOB, and D are the main factors affecting the predictions of the PINN model. ROP($t-1$) has the highest SHAP value, showing that the sequential dependency of drilling rate strongly influences prediction accuracy. torque and WOB have clear positive effects, where higher torque and bit load lead to higher predicted ROP. D shows a bidirectional effect. ROP is higher in shallow intervals, while it exhibits a negative contribution at greater depths. In addition, Q, bit run, and RPM have a certain degree of positive influence on ROP. In contrast, the effects of GR, DT, and CAL are relatively small, mainly reflecting the indirect influence of formation properties on ROP. HL and bit time have low SHAP values, indicating that these two features are not persistent dominant factors.

The SHAP-based analysis confirms that the relationships identified by the PINN model between drilling parameters and ROP are consistent with established drilling mechanics. The results show that the physics-constrained hybrid model effectively enhances the explainability and stability of traditional machine learning approaches. It also avoids the occurrence of non-physical prediction results and improves the physical and knowledge consistency of the proposed prediction framework.

4.4. Model generalization analysis

Although PINN can effectively integrate physical constraints into the learning process, their generalization performance often degrades when applied to data from different physical or geological regimes. This limitation becomes particularly evident in

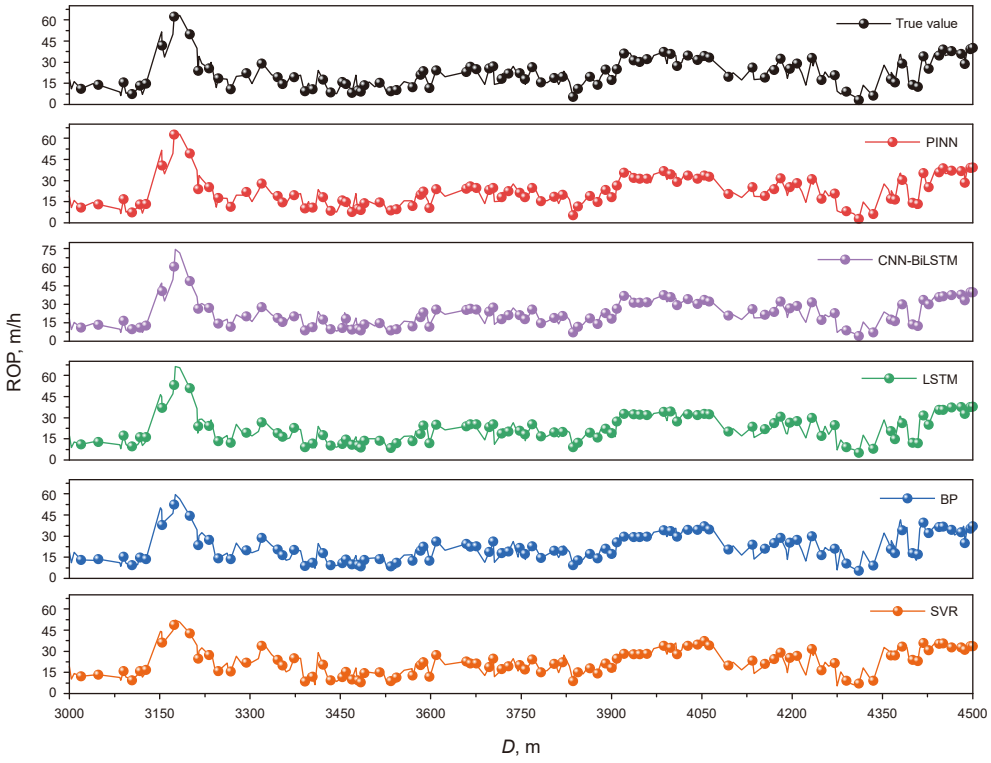

Fig. 20. Comparison curves of prediction results for different models.

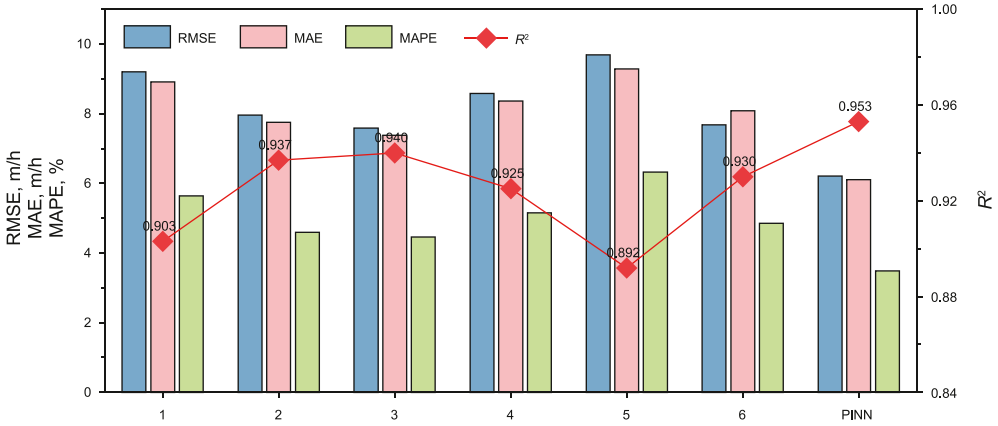

Fig. 21. Comparison of evaluation metrics for ablation experiments.

drilling engineering, where variations in geological conditions, drilling tools, and sensor acquisition lead to significant differences in parameter distributions across regions.

In response to this limitation, this study employs a transfer learning strategy by applying the pretrained model to other drilling blocks and evaluating its cross-regional adaptability and robustness (Zhuang et al., 2020; Hosna et al., 2022).

Transfer learning is a machine learning method that leverages knowledge learned in the source domain to improve learning performance in the target domain (Fang et al., 2021). Instead of retraining a model for new tasks, transfer learning only requires appropriate adjustments to a pretrained model to enable ROP prediction in the target domain. The principle of transfer learning is illustrated in Fig. 23.

In terms of the transferred content, the approach employed in this study is classified as feature-based transfer learning, where knowledge transfer is achieved by learning and extracting shared feature representations between the source and target domains. With respect to tasks and data distributions, the transfer learning in this study is classified as isomorphic transfer learning, since the source and target domains share the same feature spaces and task types, although their data distributions may differ (Li et al., 2021; Huang et al., 2024).

To evaluate the cross-domain generalization capability of the proposed model, two transfer learning schemes were designed: zero-shot transfer and few-shot transfer. In the zero-shot transfer scheme, the model is directly applied to the target domain after data preprocessing, without using any target-domain training samples. In contrast, the few-shot transfer scheme introduces a small portion of target-domain samples to fine-tune the pretrained model, thereby improving its adaptability to new geological conditions (Guo et al., 2019; Soleimani-Babakamali et al.,

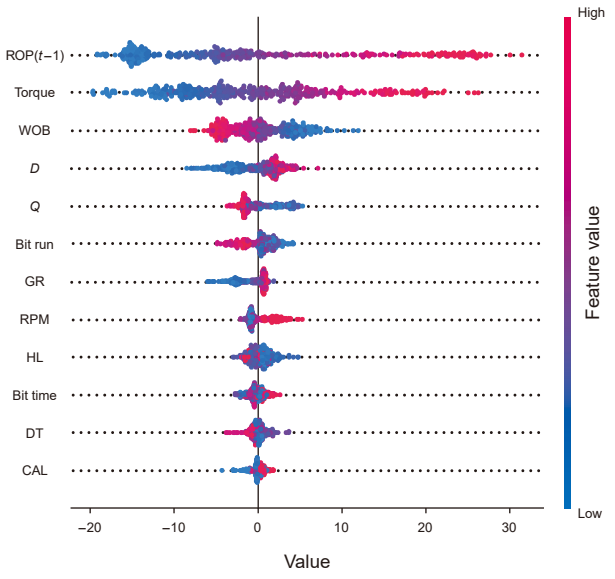

Fig. 22. Importance and influence of input parameters.

2023). In this study, the fine-tuning dataset in the few-shot transfer setting accounts for 10% of the total target-domain data.

For clarity and comparison, the source-domain test is referred to as Model A, the zero-shot transfer as Model B, and the few-shot transfer as Model C. The BY block in the eastern South China Sea was selected as the target domain for transfer learning. Compared with the source domain, the BY block was developed at a later stage and has relatively limited data availability.

Two wells, BY5-X and BY5-Y, were chosen as representative target-domain wells for model transfer experiments. The descriptive statistical analysis of the original data for well BY5-X is shown in Table 11.

By comparing Table 11 with Tables 2, it can be seen that there are significant statistical differences in the drilling data of the two blocks. The mean values of most parameters in the BY block are higher than those of the corresponding parameters in the BZ block. The mean value differences of some parameters reach more than 200%, more than 100%, more than 50%, and more than 40% respectively. At the same time, there are obvious differences in the standard deviations of the data sets of the two blocks. The parameter variability of the BY block is significantly greater than that of the BZ block. These statistical differences reflect the essential differences in the drilling engineering characteristics and geological characteristics of the two different blocks, providing a data-based basis for the transfer learning of the BY block.

The feature parameters, data preprocessing, feature correlation analysis, and data standardization procedures used in the target domain were kept consistent with those employed in the source domain to ensure comparability and reliability of the transfer results.

Comparisons of model performance across the source domain, zero-shot transfer, and few-shot transfer, along with comparisons between true and predicted values under the two transfer schemes, were plotted, as shown in Figs. 24 and 25.

The figures reveal distinct differences between the two approaches in both performance metrics and predictive outcomes. Under zero-shot transfer, the model performs substantially worse than the source-domain model, showing higher RMSE, MAE, and MAPE, a significantly lower R^2, and larger deviations between predictions and true values. Specifically, it yields a MAPE of 13.673% and an R^2 of only 0.734. In contrast, the few-shot transfer scenario demonstrates notable improvements after fine-tuning with a small number of target-domain samples. The model achieves an RMSE of 12.037, an MAE of 9.311, a MAPE of 10.413%, and an R^2 of 0.852, with predicted values more closely aligned along the diagonal line representing the true values.

Although zero-shot transfer has limitations in adapting to the differences in geological and data distributions between the source and target domains, few-shot transfer effectively mitigates performance degradation and adaptability issues across different blocks. Fine-tuning with a small amount of data allows the model

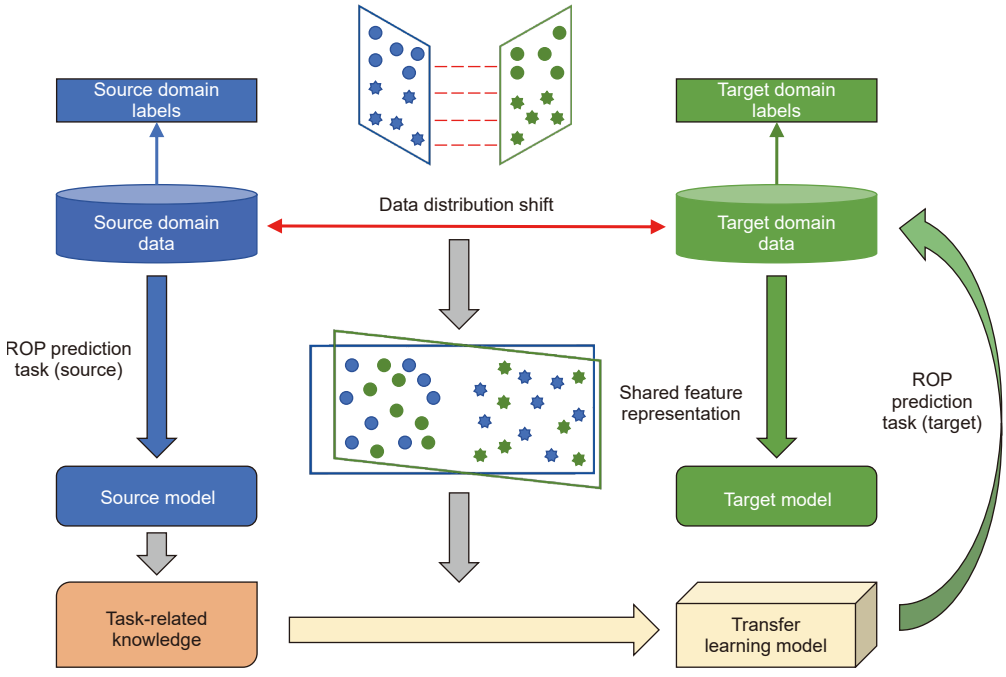

Fig. 23. Schematic diagram of transfer learning method.

Table 11
Descriptive statistical analysis table of raw data for Well BY5-X.

	D, m	HL, t	WOB, t	RPM, rpm	Torque, kN·m	*Q*, L/min	SPP, MPa	ROP, m/h	Bit run, m	Bit time, h	Pump time, h	CAL, in	CNC, PU	DT, μs/ft	GR, API	ZDEN, g/cm^3
mean	2614.00	360.92	19.49	108.58	7.05	3707.37	18.33	43.80	409.70	24.39	38.70	15.78	35.20	103.80	128.35	2.41
std	1021.77	53.59	9.44	23.72	2.89	1010.45	4.83	27.36	328.00	22.70	30.23	3.25	14.64	17.35	17.83	0.18
min	845.00	243.72	0.20	0.00	0.00	443.00	0.23	2.66	0.00	0.01	0.12	3.41	6.59	46.76	50.22	1.24
25%	1729.50	311.50	14.10	100.00	5.45	2215.00	14.12	16.56	142.00	7.22	13.98	12.66	24.44	91.19	119.62	2.30
50%	2614.00	359.92	18.60	111.00	6.62	4100.00	19.23	40.79	334.94	17.11	29.51	17.24	34.03	105.08	129.84	2.44
75%	3498.50	407.62	24.10	121.00	8.51	4123.00	22.79	63.46	577.68	34.90	57.10	17.6	41.88	113.55	138.46	2.54
max	4383.00	461.25	132.30	140.00	15.25	5211.00	26.54	166.53	1269.87	100.08	130.49	21.87	97.84	163.18	211.40	2.73

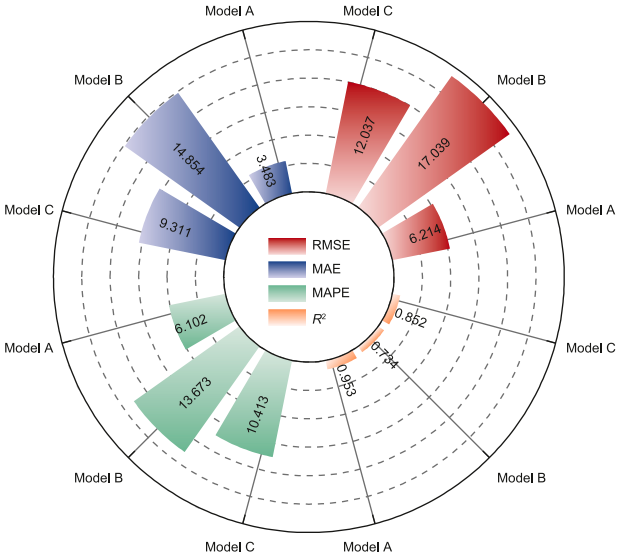

Fig. 24. Comparison of model performance across the source domain (Model A), zero-shot transfer (Model B), and few-shot transfer (Model C).

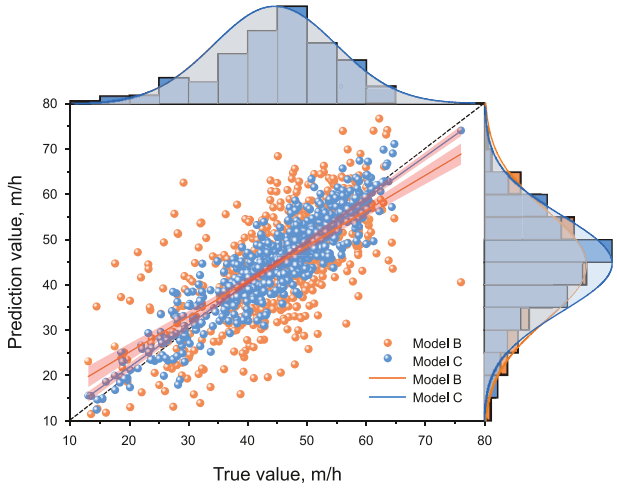

Fig. 25. Comparison of true and predicted values under zero-shot transfer (Model B) and few-shot transfer (Model C).

to better adapt to the feature distribution of the target domain, thereby improving prediction accuracy.

These results demonstrate that the PINN model exhibits strong cross-domain generalization capability. By integrating physical constraints with data-driven learning, the model maintains good robustness and predictive performance in transfer learning contexts.

The performance discrepancy between the source and target domains is a common phenomenon in transfer learning, often referred to as domain adaptation. Although few-shot transfer learning can enhance the model's adaptability and prediction accuracy in the target domain, there remain inherent geological heterogeneities and differences in data feature distributions between the two domains. Such disparities cannot be fully compensated for through limited fine-tuning samples.

Moreover, the calibration of the physical model coefficients strongly depends on the quality and representativeness of the available drilling data, and these calibrated values may differ across geological blocks. In addition, the physical constraints embedded in the PINN are constructed based on the physical laws derived from the source-domain data. While these physical principles are generally applicable, certain key parameters within the equations, such as the energy factor E_f, which is assumed constant, and the friction coefficient μ, which varies depending on the drill bit type, may require recalibration for the target domain.

5. Conclusions

This study proposed a novel approach for ROP prediction and transfer generalization based on PINN. The traditional drilling mechanism models were embedded into a deep learning framework as physical constraints in the loss function. Through the regularization and constraint of physical prior knowledge, the model not only improved prediction accuracy but also mitigated the limitations of purely data-driven approaches in interpretability and generalization. Based on this study, the following conclusions can be drawn:

(1) An enhanced mechanical specific energy model and the Soares ROP model were integrated into the SE-ResNet-BiLSTM-SAM architecture, while an adaptive loss weighting strategy was applied to dynamically balance physical constraints and data-driven residuals. A Bayesian optimization algorithm based on the Optuna framework was further employed to systematically evaluate the influence of hyper-parameters such as learning rate, LSTM layers, hidden units, and dropout rate, thereby improving model stability and convergence.

(2) Experimental results show that the PINN model is significantly superior to traditional physical models and intelligent prediction methods such as SVR, BP, and LSTM. The RMSE, MAE, MAPE and R^2 values of the PINN model are 6.214, 3.483, 6.102% and 0.953 respectively. Furthermore, when the error range is 25%, the model's error satisfaction is 97.62%, indicating that the engineering applicability and error tolerance of the PINN model are better.

(3) The proposed PINN model also demonstrated strong transfer generalization capability. Zero-shot transfer learning is difficult to directly accommodate the differences in geological conditions and data distribution between different regions. However, few-shot transfer learning has better prediction performance in the target domain, with a prediction accuracy of 89.59%, demonstrating good adaptability and generalization ability.
(4) In future research, an unsupervised domain adaptation approach will be explored to further improve the model's applicability across diverse geological environments. At the same time, a drilling parameter optimization method that integrates multiple physical models based on the PINN framework will be developed to enable real-time and dynamic optimization of drilling parameters.

CRediT authorship contribution statement

Miao He: Writing – review & editing, Resources, Funding acquisition, Conceptualization. **Qing-Rui Yuan:** Writing – original draft, Visualization, Methodology, Investigation, Conceptualization. **Jun Li:** Supervision, Resources, Project administration. **Chang-Cheng Zhou:** Investigation, Data curation. **Jing-Wei Liu:** Formal analysis, Data curation. **Xin Chen:** Validation, Data curation. **Jia-Hao Cao:** Data curation.

Declaration of competing interest

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

Acknowledgments

This paper is financially supported by the National Natural Science Foundation of China (No. 51904034), the State Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing) (No. PRE/open-2506), the Open Research Fund of State Key Laboratory of Deep Oil and Gas (No. SKLDOG2024-KFYB-11), Xinjiang Key Laboratory of intelligent Petroleum Exploration and Engineering (No. XJIPEE2025008), the Guangdong Provincial Key Laboratory of Deep Earth Sciences and Geothermal Energy Exploitation and Utilization (No. DESGEEU-2022-02).

References

Ahmed, O.S., Adeniran, A.A., Samsuri, A., 2019. Computational intelligence based prediction of drilling rate of penetration: A comparative study. J. Petrol. Sci. Eng. 172, 1–12. https://doi.org/10.1016/j.petrol.2018.09.027.
Akkuş, M.Ş., Abbasov, A., Merey, Ş., 2025. Predicting rate of penetration in large-diameter offshore drilling using artificial neural networks. Acta Geotech. 1–21. https://doi.org/10.1007/s11440-025-02815-4.
Al-AbdulJabbar, A., Elkatatny, S., Mahmoud, M., et al., 2019. A robust rate of penetration model for carbonate formation. J. Energy Resour. Technol. 141 (4), 042903. https://doi.org/10.1115/1.4041840.
Alaeddine, H., Jihene, M., 2021. Deep residual network in network. Comput. Intell. Neurosci. 2021 (1), 6659083. https://doi.org/10.1155/2021/6659083.
Al-Faiz, M.Z., Ibrahim, A.A., Hadi, S.M., 2018. The effect of Z-Score standardization (normalization) on binary input due the speed of learning in back-propagation neural network. Iraq. J. Inform. Commun. Technol. 1 (3), 42–48. https://doi.org/10.31987/ijict.1.3.41.
Al Farizi, W.S., Hidayah, I., Rizal, M.N., 2021. Isolation forest based anomaly detection: A systematic literature review. In: 2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE). IEEE, pp. 118–122. https://doi.org/10.1109/ICITACEE53184.2021.9617498.
Ali Abd Al-Hameed, K., 2022. Spearman's correlation coefficient in statistical analysis. Int. J. Nonlinear Anal. Appl. 13 (1), 3249–3255. https://doi.org/10.22075/ijnaa.2022.6079.
Alibrahim, H., Ludwig, S.A., 2021. Hyperparameter optimization: Comparing genetic algorithm against grid search and bayesian optimization. In: 2021 IEEE Congress on Evolutionary Computation (CEC). IEEE, pp. 1551–1559. https://doi.org/10.1109/CEC45853.2021.9504761.
Allawi, R.H., Al-Mudhafar, W.J., Abbas, M.A., et al., 2025. Leveraging boosting machine learning for drilling rate of penetration (ROP) prediction based on drilling and petrophysical parameters. Artif. Intell. Geosci., 100121 https://doi.org/10.1016/j.aiig.2025.100121.
Arnastauskaitė, J., Ruzgas, T., Bražėnas, M., 2021. An exhaustive power comparison of normality tests. Mathematics 9 (7), 788. https://doi.org/10.3390/math9070788.
Bai, K., Jin, S., Zhang, Z., et al., 2024. Drilling rate of penetration prediction based on CBT-LSTM neural network. Sensors 24 (21), 6966. https://doi.org/10.3390/s24216966.
Bingham, M.G., 1965. A New Approach to interpreting–rock Drillability. Petroleum Publishing Company.
Binois, M., Wycoff, N., 2022. A survey on high-dimensional Gaussian process modeling with application to Bayesian optimization. ACM Transactions on Evolutionary Learning and Optimization. Optim. 2 (2), 1–26. https://doi.org/10.1145/3545611.
Bourgoyne Jr, A.T., Young Jr, F.S., 1974. A multiple regression approach to optimal drilling and abnormal pressure detection. Soc. Petrol. Eng. J. 14 (4), 371–384. https://doi.org/10.2118/4238-PA.
Brenjkar, E., Delijani, E.B., 2022. Computational prediction of the drilling rate of penetration (ROP): A comparison of various machine learning approaches and traditional models. J. Petrol. Sci. Eng. 210, 110033. https://doi.org/10.1016/j.petrol.2021.110033.
Cai, S., Wang, Z., Wang, S., et al., 2021. Physics-informed neural networks for heat transfer problems. J. Heat Tran. 143 (6), 060801. https://doi.org/10.1115/1.4050542.
Chen, J., Peng, T., Qian, S., et al., 2025. An error-corrected deep autoformer model via Bayesian optimization algorithm and secondary decomposition for photovoltaic power prediction. Appl. Energy 377, 124738. https://doi.org/10.1016/j.apenergy.2024.124738.
Chen, W., Cai, Y., Li, A., et al., 2023. EEG feature selection method based on maximum information coefficient and quantum particle swarm. Sci. Rep. 13 (1), 14515. https://doi.org/10.1038/s41598-023-41682-5.
Chen, X., Fan, H., Guo, B., et al., 2014. Real-time prediction and optimization of drilling performance based on a new mechanical specific energy model. Arabian J. Sci. Eng. 39, 8221–8231. https://doi.org/10.1007/s13369-014-1376-0.
Chen, Z., Mak, S., Wu, C.J., 2024. A hierarchical expected improvement method for Bayesian optimization. J. Am. Stat. Assoc. 119 (546), 1619–1632. https://doi.org/10.1080/01621459.2023.2210803.
Du, L., Gao, R., Suganthan, P.N., et al., 2022. Bayesian optimization based dynamic ensemble for time series forecasting. Inf. Sci. 591, 155–175. https://doi.org/10.1016/j.ins.2022.01.010.
Dupriest, F.E., Koederitz, W.L., 2005. Maximizing drill rates with real-time surveillance of mechanical specific energy. In: SPE/IADC Drilling Conference and Exhibition. SPE 92194. https://doi.org/10.2118/92194-MS.
Elkatatny, S., Al-AbdulJabbar, A., Abdelgawad, K., 2020. A new model for predicting rate of penetration using an artificial neural network. Sensors 20 (7), 2058. https://doi.org/10.3390/s20072058.
Etesami, D.G., Shirangi, M., Zhang, W.J., 2021. A semiempirical model for rate of penetration with application to an offshore gas field. SPE Drill. Complet. 36 (1), 29–46. https://doi.org/10.2118/202481-PA.
Fan, Y.D., Pang, H.W., Jin, Y., et al., 2025a. A new method for the rate of penetration prediction and control based on signal decomposition and causal inference. Pet. Sci. 22 (6), 2414–2437. https://doi.org/10.1016/j.petsci.2025.04.005.
Fan, Y., Jin, Y., Pang, H., et al., 2025b. A physically-constrained ensemble learning rate of penetration prediction model based on multi-source data fusion. Appl. Intell. 55 (3), 226. https://doi.org/10.1007/s10489-024-05922-z.
Fang, X., Gong, G., Li, G., et al., 2021. A hybrid deep transfer learning strategy for short term cross-building energy prediction. Energy 215, 119208. https://doi.org/10.1016/j.energy.2020.119208.
Fu, E., Zhang, Y., Yang, F., et al., 2022. Temporal self-attention-based Conv-LSTM network for multivariate time series prediction. Neurocomputing 501, 162–173. https://doi.org/10.1016/j.neucom.2022.06.014.
Gan, C., Cao, W., Wu, M., et al., 2019. Prediction of drilling rate of penetration (ROP) using hybrid support vector regression: A case study on the shennongjia area, Central China. J. Petrol. Sci. Eng. 181, 106200. https://doi.org/10.1016/j.petrol.2019.106200.
Gan, C., Cao, W.H., Liu, K.Z., et al., 2022. A novel dynamic model for the online prediction of rate of penetration and its industrial application to a drilling process. J. Process Control 109, 83–92. https://doi.org/10.1016/j.jprocont.2021.12.002.
Ghanitoos, H., Goharimanesh, M., Akbari, A., 2024. Prediction of drill penetration rate in drilling oil wells using mathematical and neurofuzzy modeling methods. Energy Rep. 11, 145–152. https://doi.org/10.1016/j.egyr.2023.11.006.
González-Estrada, E., Villaseñor, J.A., Acosta-Pech, R., 2022. Shapiro-Wilk test for multivariate skew-normality. Comput. Stat. 37 (4), 1985–2001. https://doi.org/10.1007/s00180-021-01188-y.

Gooneratne, C.P., Magana-Mora, A., Otalvora, W.C., et al., 2020. Drilling in the fourth industrial revolution—Vision and challenges. IEEE Eng. Manag. Rev. 48 (4), 144–159. https://doi.org/10.1109/EMR.2020.2999420.
Guo, Y., Shi, H., Kumar, A., et al., 2019. Spottune: transfer learning through adaptive fine-tuning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4805–4814.
Hammoutene, C., Bits, S., 2012. FEA modelled MSE/UCS values optimise PDC design for entire hole section. In: SPE North Africa Technical Conference and Exhibition. SPE 149372. https://doi.org/10.2118/149372-MS.
Hanifi, S., Cammarono, A., Zare-Behtash, H., 2024. Advanced hyperparameter optimization of deep learning models for wind power prediction. Renew. Energy 221, 119700. https://doi.org/10.1016/j.renene.2023.119700.
Hareland, G., Rampersad, P.R., 1994. Drag-bit model including wear. In: SPE Latin America and Caribbean Petroleum Engineering Conference. https://doi.org/10.2118/26957-MS.
Harrou, F., Zeroual, A., Kadri, F., et al., 2024. Enhancing road traffic flow prediction with improved deep learning using wavelet transforms. Results Eng. 23, 102342. https://doi.org/10.1016/j.rineng.2024.102342.
He, K., Zhang, X., Ren, S., et al., 2016. Identity mappings in deep residual networks. In: European Conference on Computer Vision. Springer International Publishing, pp. 630–645. https://doi.org/10.1007/978-3-319-46493-0_38.
Hosna, A., Merry, E., Gyalmo, J., et al., 2022. Transfer learning: A friendly introduction. J. Big Data 9 (1), 102. https://doi.org/10.1186/s40537-022-00652-w.
Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141.
Huang, Z., Zhu, L., Wang, C., et al., 2024. Intelligent prediction of rate of penetration using mechanism-data fusion and transfer learning. Processes 12 (10), 2133. https://doi.org/10.3390/pr12102133.
Ishitsuka, K., Lin, W., 2023. Physics-informed neural network for inverse modeling of natural-state geothermal systems. Appl. Energy 337, 120855. https://doi.org/10.1016/j.apenergy.2023.120855.
Jahani-Nasab, M., Bijarchi, M.A., 2024. Enhancing convergence speed with feature enforcing physics-informed neural networks using boundary conditions as prior knowledge. Sci. Rep. 14 (1), 23836. https://doi.org/10.1038/s41598-024-74711-y.
Jiao, S., Li, W., Li, Z., et al., 2024. Hybrid physics-machine learning models for predicting rate of penetration in the Halahatang oil field, Tarim Basin. Sci. Rep. 14 (1), 5957. https://doi.org/10.1038/s41598-024-56640-y.
Jin, X., Xie, Y., Wei, X.S., et al., 2022. Delving deep into spatial pooling for squeeze-and-excitation networks. Pattern Recogn. 121, 108159. https://doi.org/10.1016/j.patcog.2021.108159.
Junior, J.R.B., Lavi, B., 2025. Enhancing rate of penetration prediction in drilling operations: A data stream framework approach. Eng. Appl. Artif. Intell. 143, 110034. https://doi.org/10.1016/j.engappai.2025.110034.
Kinney, J.B., Atwal, G.S., 2014. Equitability, mutual information, and the maximal information coefficient. Proc. Natl. Acad. Sci. 111 (9), 3354–3359. https://doi.org/10.1073/pnas.1309933111.
Lee, C., Kim, J., Kim, N., et al., 2025. Evaluating the rate of penetration with deep-learning predictive models. Int. J. Energy Res. 2025 (1), 8872793. https://doi.org/10.1155/er/8872793.
Li, J., Lu, J., Chen, C., et al., 2021. Tool wear state prediction based on feature-based transfer learning. Int. J. Adv. Manuf. Technol. 113 (11), 3283–3301. https://doi.org/10.1007/s00170-021-06780-6.
Li, P., Liu, M., Alfarraj, M., et al., 2024a. Probabilistic physics-informed neural network for seismic petrophysical inversion. Geophysics 89 (2), M17–M32. https://doi.org/10.1190/geo2023-0214.1.
Li, Q., Li, J.P., Xie, L.L., 2024b. A systematic review of machine learning modeling processes and applications in ROP prediction in the past decade. Pet. Sci. 21 (5), 3496–3516. https://doi.org/10.1016/j.petsci.2024.05.013.
Liu, M., Vashisth, D., Grana, D., et al., 2023. Joint inversion of geophysical data for geologic carbon sequestration monitoring: A differentiable physics-informed neural network model. J. Geophys. Res. Solid Earth 128 (3), e2022JB025372. https://doi.org/10.1029/2022JB025372.
Madani, M., Lin, K., Tarakanova, A., 2021. DSResSol: A sequence-based solubility predictor created with dilated squeeze excitation residual networks. Int. J. Mol. Sci. 22 (24), 13555. https://doi.org/10.3390/ijms222413555.
Maurer, W.C., 1962. The “perfect-cleaning” theory of rotary drilling. J. Petrol. Technol. 14 (11), 1270–1274. https://doi.org/10.2118/408-PA.
McClenny, L.D., Braga-Neto, U.M., 2023. Self-adaptive physics-informed neural networks. J. Comput. Phys. 474, 111722. https://doi.org/10.1016/j.jcp.2022.111722.
Michael, N.E., Bansal, R.C., Ismail, A.A.A., et al., 2024. A cohesive structure of bi-directional long-short-term memory (BiLSTM)-GRU for predicting hourly solar radiation. Renew. Energy 222, 119943. https://doi.org/10.1016/j.renene.2024.119943.
Monter-Pozos, A., González-Estrada, E., 2024. On testing the skew normal distribution by using shapiro–wilk test. J. Comput. Appl. Math. 440, 115649. https://doi.org/10.1016/j.cam.2023.115649.
Motahhari, H.R., Hareland, G., James, J.A., 2010. Improved drilling efficiency technique using integrated PDM and PDC bit parameters. J. Can. Petrol. Technol. 49 (10), 45–52. https://doi.org/10.2118/141651-PA.
Nigam, H.K., Srivastava, H.M., 2023. Filtering of audio signals using discrete wavelet transforms. Mathematics 11 (19), 4117. https://doi.org/10.3390/math11194117.
Niu, Z., Zhong, G., Yu, H., 2021. A review on the attention mechanism of deep learning. Neurocomputing 452, 48–62. https://doi.org/10.1016/j.neucom.2021.03.091.
Owen, C.A., Dick, G., Whigham, P.A., 2022. Standardization and data augmentation in genetic programming. IEEE Trans. Evol. Comput. 26 (6), 1596–1608. https://doi.org/10.1109/TEVC.2022.3160414.
Pan, S., Yang, B., Wang, S., et al., 2023. Oil well production prediction based on CNN-LSTM model with self-attention mechanism. Energy 284, 128701. https://doi.org/10.1016/j.energy.2023.128701.
Pei, Z.J., Song, X.Z., Wang, H.T., et al., 2024. Interpretation and characterization of rate of penetration intelligent prediction model. Pet. Sci. 21 (1), 582–596. https://doi.org/10.1016/j.petsci.2023.10.011.
Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045.
Raissi, M., Yazdani, A., Karniadakis, G.E., 2020. Hidden fluid mechanics: learning velocity and pressure fields from flow visualizations. Science 367 (6481), 1026–1030. https://doi.org/10.1126/science.aaw4741.
Ren, C., Huang, W., Gao, D., 2023. Predicting rate of penetration of horizontal drilling by combining physical model with machine learning method in the China Jimusar oil field. SPE J. 28 (6), 2713–2736. https://doi.org/10.2118/212294-PA.
Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. IEEE Trans. Signal Process. 45 (11), 2673–2681. https://doi.org/10.1109/78.650093.
Shaygan, K., Jamshidi, S., 2023. Prediction of rate of penetration in directional drilling using data mining techniques. Geoenergy Sci. Eng. 221, 111293. https://doi.org/10.1016/j.petrol.2022.111293.
Smith, L.N., Topin, N., 2019. Super-convergence: Very fast training of neural networks using large learning rates. In: Artificial Intelligence and Machine Learning for multi-domain Operations Applications. SPIE, pp. 369–386. https://doi.org/10.1117/12.2520589.
Soares, C., Gray, K., 2019. Real-time predictive capabilities of analytical and machine learning rate of penetration (ROP) models. J. Petrol. Sci. Eng. 172, 934–959. https://doi.org/10.1016/j.petrol.2018.08.083.
Soleimani-Babakamali, M.H., Soleimani-Babakamali, R., Nasrollahzadeh, K., et al., 2023. Zero-shot transfer learning for structural health monitoring using generative adversarial networks and spectral mapping. Mech. Syst. Signal Process. 198, 110404. https://doi.org/10.1016/j.ymssp.2023.110404.
Su, K., Da, W., Li, M., et al., 2024. Research on a drilling rate of penetration prediction model based on the improved chaos whale optimization and back propagation algorithm. Geoenergy Sci. Eng. 240, 213017. https://doi.org/10.1016/j.geoen.2024.213017.
Sun, L., Gao, H., Pan, S., et al., 2020. Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data. Comput. Methods Appl. Mech. Eng. 361, 112732. https://doi.org/10.1016/j.cma.2019.112732.
Teale, R., 1965. The concept of specific energy in rock drilling. In: International journal of rock mechanics and mining sciences & geomechanics abstracts. Pergamon, 2 (1), pp. 57–73. https://doi.org/10.1016/0148-9062(65)90022-7.
Tu, B., Bai, K., Zhan, C., Zhang, W., 2024. Real-time prediction of ROP based on GRU-informer. Sci. Rep. 14 (1), 2133. https://doi.org/10.1038/s41598-024-52261-7.
Vaswani, A., Shazeer, N., Parmar, N., et al., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.
Victoria, A.H., Maragatham, G., 2021. Automatic tuning of hyperparameters using Bayesian optimization. Evol. Syst. 12 (1), 217–223. https://doi.org/10.1007/s12530-020-09345-2.
Walker, B.H., Black, A.D., Klauber, W.P., et al., 1986. Roller-bit penetration rate response as a function of rock properties and well depth. In: SPE Annual Technical Conference and Exhibition? https://doi.org/10.2118/15620-MS.
Wan, Y., Liu, X., Xiong, J., et al., 2024. Intelligent prediction of drilling rate of penetration based on method-data dual validity analysis. SPE J. 29 (5), 2257–2274. https://doi.org/10.2118/217977-PA.
Wang, H., Huang, H., Bi, W., et al., 2022. Deep and ultra-deep oil and gas well drilling technologies: Progress and prospect. Nat. Gas. Ind. B 9 (2), 141–157. https://doi.org/10.1016/j.ngib.2021.08.019.
Wang, S., Teng, Y., Perdikaris, P., 2021. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. SIAM J. Sci. Comput. 43 (5), A3055–A3081. https://doi.org/10.1137/20M1318043.
Wang, X., Jin, Y., Schmitt, S., et al., 2023. Recent advances in Bayesian optimization. ACM Comput. Surv. 55 (13s), 1–36. https://doi.org/10.1145/3582078.
Warren, T.M., 1981. Drilling model for soft-formation bits. J. Petrol. Technol. 33 (6), 963–970. https://doi.org/10.2118/8438-PA.
Wu, Z., He, L., Wang, W., et al., 2024. A fault prediction method for CNC machine tools based on se-resnet-transformer. Machines 12 (6), 418. https://doi.org/10.3390/machines12060418.
Xiang, Z., Peng, W., Liu, X., et al., 2022. Self-adaptive loss balanced physics-informed neural networks. Neurocomputing 496, 11–34. https://doi.org/10.1016/j.neucom.2022.05.015.
Xiong, M., Zheng, S., Liu, W., et al., 2024. A rate of penetration (ROP) prediction method based on improved dung beetle optimization algorithm and BiLSTM-SA. Sci. Rep. 14 (1), 25856. https://doi.org/10.1038/s41598-024-75703-8.
Xu, C., Yang, H., Wang, F., et al., 2024. Formation conditions of deep to ultra-deep large composite buried-hill hydrocarbon reservoirs in offshore Bohai Bay

Basin, China. Petrol. Explor. Dev. 51 (6), 1421–1434. https://doi.org/10.1016/S1876-3804(25)60550-7.
Xu, H., Pang, G., Wang, Y., et al., 2023. Deep isolation forest for anomaly detection. IEEE Trans. Knowl. Data Eng. 35 (12), 12591–12604. https://doi.org/10.1109/TKDE.2023.3270293.
Xue, Y., Zhao, M., Liu, X., 2021. Reservoir characteristics and controlling factors of the metamorphic buried hill of Bozhong Sag, Bohai Bay Basin. J. Earth Sci. 32 (4), 919–926. https://doi.org/10.1007/s12583-021-1415-1.
Yang, Y., Cen, X., Ni, H., et al., 2025. A highly accurate and robust prediction framework for drilling rate of penetration based on machine learning ensemble algorithm. Geoenergy Sci. Eng. 244, 213423. https://doi.org/10.1016/j.geoen.2024.213423.
Yi, J., Li, H.Y., Shan, X.L., et al., 2022. Division and identification of vertical reservoir units in Archaeozoic metamorphic buried hill of Bozhong Sag, Bohai Bay Basin, East China. Petrol. Explor. Dev. 49 (6), 1282–1294. https://doi.org/10.1016/S1876-3804(23)60349-0.
Yin, H., Zhao, X., Li, Q., 2024. Research on adaptive prediction model of rate of penetration under dynamic formation conditions. Eng. Appl. Artif. Intell. 133, 108281. https://doi.org/10.1016/j.engappai.2024.108281.
Yu, H., Hutson, A.D., 2024. A robust Spearman correlation coefficient permutation test. Commun. Stat. Theor. Methods 53 (6), 2141–2153. https://doi.org/10.1080/03610926.2022.2121144.
Zhang, D., 2019. Wavelet Transform. Fundamentals of Image Data Mining: Analysis, Features, Classification and Retrieval. Springer International Publishing, pp. 35–44. https://doi.org/10.1007/978-3-030-17989-2_3.
Zhang, G., Tong, D.J., Chen, K., et al., 2024. Tectonic evolution and source rocks development of the super oil-rich Bohai Bay Basin, East China. Petrol. Explor. Dev. 51 (5), 1165–1182. https://doi.org/10.1016/S1876-3804(25)60533-7.
Zhang, L., Liu, L., 2022. Data anomaly detection based on isolation forest algorithm. In: 2022 International Conference on Computation, Big-Data and Engineering (ICCBE). IEEE, pp. 87–89. https://doi.org/10.1109/ICCBE56101.2022.9888169.
Zhang, Z., Guo, D., Zhou, S., et al., 2023. Flight trajectory prediction enabled by time-frequency wavelet transform. Nat. Commun. 14 (1), 5258. https://doi.org/10.1038/s41467-023-40903-9.
Zheng, X., Shi, J., Cao, G., et al., 2022. Progress and prospects of oil and gas production engineering technology in China. Petrol. Explor. Dev. 49 (3), 644–659. https://doi.org/10.1016/S1876-3804(22)60054-5.
Zhou, F., Fan, H., Liu, Y., et al., 2023. Hybrid model of machine learning method and empirical method for rate of penetration prediction based on data similarity. Appl. Sci. 13 (10), 5870. https://doi.org/10.3390/app13105870.
Zhou, Y., Dong, Z., Bao, X., 2024. A ship trajectory prediction method based on an optuna–BILSTM model. Appl. Sci. 14 (9), 3719. https://doi.org/10.3390/app14093719.
Zhou, Y., Zhang, W., Gamwo, I., et al., 2017. Mechanical specific energy versus depth of cut in rock cutting and drilling. Int. J. Rock Mech. Min. Sci. 100, 287–297. https://doi.org/10.1016/j.ijrmms.2017.11.004.
Zhu, R., Tang, J., He, X., et al., 2024. SE-MAConvLSTM: A deep learning framework for short-term traffic flow prediction combining squeeze-and-excitation network and multi-attention convolutional LSTM network. PLoS One 19 (12), e0312601. https://doi.org/10.1371/journal.pone.0312601.
Zhuang, F., Qi, Z., Duan, K., et al., 2020. A comprehensive survey on transfer learning. Proc. IEEE 109 (1), 43–76. https://doi.org/10.1109/JPROC.2020.3004555.
Zulfiqar, M., Gamage, K.A., Kamran, M., et al., 2022. Hyperparameter optimization of bayesian neural network using bayesian optimization and intelligent feature engineering for load forecasting. Sensors 22 (12), 4446. https://doi.org/10.3390/s22124446.