



# Predicting Rate of Penetration in Ultra-deep Wells Based on Deep Learning Method

Chi Peng<sup>1</sup> · Jianyun Pang<sup>1</sup> · Jianhong Fu<sup>1</sup> · Quan Cao<sup>2</sup> · Jinhong Zhang<sup>3,4</sup> · Qingfeng Li<sup>1</sup> · Zhaoyong Deng<sup>1</sup> · Yun Yang<sup>2</sup> · Ziqiang Yu<sup>5</sup> · Danzhu Zheng<sup>6</sup>

Received: 3 February 2023 / Accepted: 8 June 2023 / Published online: 5 July 2023  
© King Fahd University of Petroleum & Minerals 2023

## Abstract

The accurate prediction of the rate of penetration (ROP) is crucial for optimizing drilling parameters and enhancing drilling efficiency in ultra-deep wells. However, this task is challenging due to the harsh geological conditions, complex drilling processes, voluminous drilling data, and nonlinear relationships between drilling parameters and rock-breaking. In this study, a comprehensive intelligent model is proposed that combines clustering and deep residual neural network to address these challenges. Specifically, relevant feature parameters are selected for ROP prediction and the Savitzky–Golay filter is employed to reduce noise in the field data. Formations with similar rock characteristics are clustered using well logging parameters, including sonic logging and natural gamma ray logging, which indicate the formation rock properties. A deep residual neural network is then used to develop the prediction model, with the clustering results and 13 mud logging parameters serving as inputs. The model is trained and tested using field data from an ultra-deep reservoir in northwest China, and its performance is evaluated. The impact of data noise reduction, formation clustering, and deep residual neural network on the prediction accuracy is analyzed through ablation experiments. The proposed model achieves high accuracy in predicting ROP, with relative errors ranging from 11.34 to 11.44% and  $R^2$  values from 0.92 to 0.94. Compared to traditional machine learning models, the approach demonstrates superior performance and is suitable for real-time drilling applications. This study provides a promising solution for accurate ROP prediction in ultra-deep wells, helping to optimize drilling parameters and improve drilling efficiency.

**Keywords** Rate of penetration · Machine learning · Big data · Clustering · Deep residual network

## 1 Introduction

Although there is a growing interest for sustainable energy, for the next decade, oil and gas will still remain as the basis for global energy consumption. With the depletion of shallow oil and gas resources, the exploration and development of deep oil and gas resources become popular [1–5]. The increasing well depth greatly raises the drilling costs, which brings about the necessity to improve drilling efficiency and rate of penetration (ROP). Accurate prediction of ROP is the first step for optimizing drilling parameters and improving drilling efficiency [6–8]. ROP prediction has been the subject of drilling engineering for more than half a century, which can be divided into two categories: physics-based models and data-driven models.

---

✉ Chi Peng  
pengchiswpu@swpu.edu.cn

<sup>1</sup> Petroleum Engineering School, Southwest Petroleum University, Chengdu 610500, Sichuan, China

<sup>2</sup> Engineering Technology Research Institute of CNPC Southwest Oil and Gas Field Company, Chengdu 610017, Sichuan, China

<sup>3</sup> Bazhou Branch, CNPC XIBU Drilling Engineering Company, Korla 841000, Xinjiang, China

<sup>4</sup> CNPC CCDC Drilling & Production Technology Research Institute, Xi'an 710021, Shanxi, China

<sup>5</sup> Southwest Petroleum University Science and Technology Development Company, Chengdu 610500, Sichuan, China

<sup>6</sup> The University of Tulsa, Tulsa, OK 74104, USA



The traditional physics-based ROP models were widely used because of their simplicity and clear physical background. Through extensive experimental data analysis, Bingham [9] discovered that weight on bit (WOB), rotation per minute (RPM), and wellbore diameter had substantial impacts on ROP and put forward a basic ROP equation. Later researchers worked to refine this equation and proposed a number of modified models [10–17]. More factors were considered, including formation strength, formation compaction, bit wear, hydraulic parameters, bottom hole pressure, inclination angle, dogleg degree, cuttings bed thickness and concentration, etc. Physics-based ROP models are devoted to establishing an explicit mathematical relationship between drilling parameters and ROP through mechanism analysis and laboratory experiments. However, the physical-based ROP models suffer from certain limitations. Due to the high reciprocity and nonlinearity among various relevant parameters, it is difficult to accurately predict ROP with traditional polynomial equations. In addition, the models have undetermined coefficients that are related to bit wear, wellbore cleaning, and rock mechanics. In ultra-deep wells, these deficiencies become more intolerable and make the physics-based ROP models poorly applicable.

In recent years, machine learning technology has advanced rapidly. The data-driven models do not rely on theoretical analysis and show the advantages such as high flexibility for input parameters and strong ability of fitting complex nonlinear relationships. ROP prediction using machine learning is receiving increasing attention [18–22]. Since Bilgesu et al. [23] published the first paper on ROP prediction applying artificial neural network (ANN) in 1997, researchers have done a lot of works on ROP prediction with ANN [24]. ANN is an information processing system that imitates the structure and function of neural networks in human brain. For a data set, ANN can learn and capture the unique relationship between input and target parameters, and has the characteristics of fault tolerance, high efficiency, and great adaptability [25]. The powerful nonlinear mapping ability of ANN provides a good solution for ROP prediction. Multilayer perceptron (MLP) and extreme learning machine (ELM) neural networks were widely adopted [26, 27]. In order to boost ROP prediction accuracy, researchers started to optimize the ANN structure (the number of hidden layers, the number of neurons in each layer, etc.) [28–30]. In addition, other popular machine learning methods, such as random forest (RF) and support vector machine (SVM), are also utilized to build ROP prediction models [31–33]. Christine et al. [31] compared five machine learning methods (RF, ANN, SVM, ridge regression, and gradient elevator) in ROP prediction.

From the perspective of machine learning, ROP prediction is a regression task. Although neural networks are competitive for regression tasks, there are challenges to build

high-performance ROP prediction models [19, 34–36]: (1) Due to the complex underground condition and instrumental errors, there are deviations and noise data in the original data set. The quality of the original field data set is poor, so it is difficult to train an accurate and reliable model with the original data. In previous publications,  $R^2 > 0.8$  could be regarded as a good accuracy [24, 37, 38]. (2) As a result of the local uplift and subsidence, the corresponding depth of the same formation in different wells varies. To complete the ROP prediction of the whole well, the traditional solution needs to build multiple models according to the geological formations, which is extremely time-consuming [30]. It is necessary to build a single model that both considers geological heterogeneity and can predict ROP of the whole well sections. (3) With the continuous automation of drilling engineering, the big data generated by drilling operations are more abundant in diversity and quantity. As the data size greatly increases, there are more hidden information that needs to be elucidated. The traditional neural networks have simple structure, which cannot fully perceive the latent relationships in the big data.

To deal with above challenges, this paper constructs a new intelligent ROP prediction model. Savitzky–Golay (SG) algorithm is adopted to filter the data, which ameliorates the data quality. The logging data are clustered. The geological formations with similar rock characteristics are classified into the same cluster. The cluster results replace the well logging data and act as a single feature parameter for formation characteristics. To fully dig the hidden information in drilling big data and achieve high-precision prediction, a 23-layer deep residual neural network (ResNet) is built for the final ROP prediction.

For the remaining contents, in Sect. 2, the frame of the proposed model is introduced. The sub-processes of noise reduction, formation clustering, and ResNet are presented in detail. The model evaluation indexes are also described. Field drilling data of ultra-deep reservoir in Xinjiang are collected to train and test the model. Section 3 displays the results and discussion, including the test results, ablation experiment, performance comparison with other machine learning methods, and a case study of real-time application. After that, a conclusion is drawn.

## 2 Material and Methods

### 2.1 Input Parameters for ROP Prediction

There are many mechanic, hydraulic and chemical parameters that affect ROP. Generally, field parameters can be divided as mud logging parameters and well logging parameters based on data source. According to previous researches [24, 37, 38], a total of 20 parameters that are closely related to ROP are selected, as shown in Table 1. Since all the ultra-

**Table 1** Input parameters for ROP prediction

Number	Feature parameter	Source
1	Well depth, WD/m	Mud logging parameter
2	Weight on bit, WOB/kN	Mud logging parameter
3	Rotation per minute, RPM/(r/min)	Mud logging parameter
4	Standpipe pressure, SPP/Mpa	Mud logging parameter
5	Flow rate, Q/(L/s)	Mud logging parameter
6	Torque, TOR/(kN m)	Mud logging parameter
7	Bit diameter, BD/mm	Mud logging parameter
8	Mud density, MD/(g/cm <sup>3</sup> )	Mud logging parameter
9	Equivalent mud density, ECD/(g/cm <sup>3</sup> )	Mud logging parameter
10	Hook load, HL/kN	Mud logging parameter
11	Pump stroke, PS/spm	Mud logging parameter
12	Mud pit volume, MPV/m <sup>3</sup>	Mud logging parameter
13	DC exponent, DC	Mud logging parameter
14	Interval transit time, AC/(μs/m)	Well logging parameter
15	Gamma ray, GR/(Ur/h)	Well logging parameter
16	Spontaneous potential, SP/mV	Well logging parameter
17	Compensated neutron logging, CNL/ev	Well logging parameter
18	Density logging, DEN/(g/cm <sup>3</sup> )	Well logging parameter
19	Resistivity, RD/(Ω m)	Well logging parameter
20	Photoelectric absorption cross-section index, PE	Well logging parameter

deep wells in the target reservoirs are vertical, deviation angle and azimuth are not included in Table 1.

### 2.2 Model Framework

Figure 1 summarizes the structure and workflow of the model, which is composed of data noise reduction, formation clustering (K-means), and ResNet. It is abbreviated as DKR model. First, the input drilling data undergo the noise reduction process to eliminate the outliers and missing values. Then, the well logging parameters are clustered using K-means algorithm. The formations with similar rock characteristics are classified into the same cluster. The cluster results replace the well logging parameters and provide a single feature parameter for formation characteristics. It helps to enhance the correlation between formation characteristics and ROP. Last, a ResNet with optimized structure is trained to correlate feature parameters with ROP.

### 2.3 Data Noise Reduction

Data noise reduction aims to improve data quality and ensure the reliability of the model. Drilling data are collected from multiple sources (mud logging and well logging in this work). Due to facility or human factors, field drilling data are inevitably interfered by noise signals, which reduces the model accuracy and increases the training time [39]. Appropriate noise reduction should be performed to improve the consistency and integrity of the data.

In this paper, SG filter is adopted, which is a filtering algorithm that realizes polynomial fitting of local interval through least square convolution [40]. It has been widely used for data smoothing and noise reduction [41, 42]. It can maintain the shape and width of the original signal while removing the noise. The process of SG filter is shown in Fig. 2. Within a sliding window of  $2m + 1$  continuous data points (width), the least square fitting with a certain fitting order  $k$  is conducted, and the fitting curve value at the center of the sliding window is taken as the filtered value. The window continues to move, and the above process is repeated to fulfil the filtering of all data points.

The basic principle of SG filter is as follows:

$$s_j^* = \frac{\sum_{i=-m}^m C_i S_{j+1}}{N} \tag{1}$$

where  $S$  is the original signal;  $s^*$  is the signal after noise reduction;  $C_i$  is the noise reduction coefficient for the  $i$ -th time;  $N$  is the sliding window width for  $2m + 1$  data points;  $j$  is the  $j$ -th point in the data set. Two parameters need to be determined when applying SG filter: sliding window width  $N$  and local polynomial fitting order  $k$ . Reasonable selection of  $N$  and  $k$  can reduce signal distortion and ensure filtering quality [43].

### 2.4 Formation Clustering

The purpose of formation clustering is to arrange geological formations that share similar rock properties. Due to the local geological movement, the same formation may appear at various vertical depth or even disappear for different wells. To guarantee the accuracy, previous machine learning method has to establish multiple ROP models for all the formations (Fig. 3). In this work, well logging data are used to cluster the formations. After dividing the formations with similar rock characteristics into the same cluster, the well logging parameters are replaced by cluster results as the input parameters for ROP prediction. It helps to strengthen the correlation between formation feature and ROP, and facilitate the model training. Most importantly, the ROP prediction of the whole well can be realized with a single model.

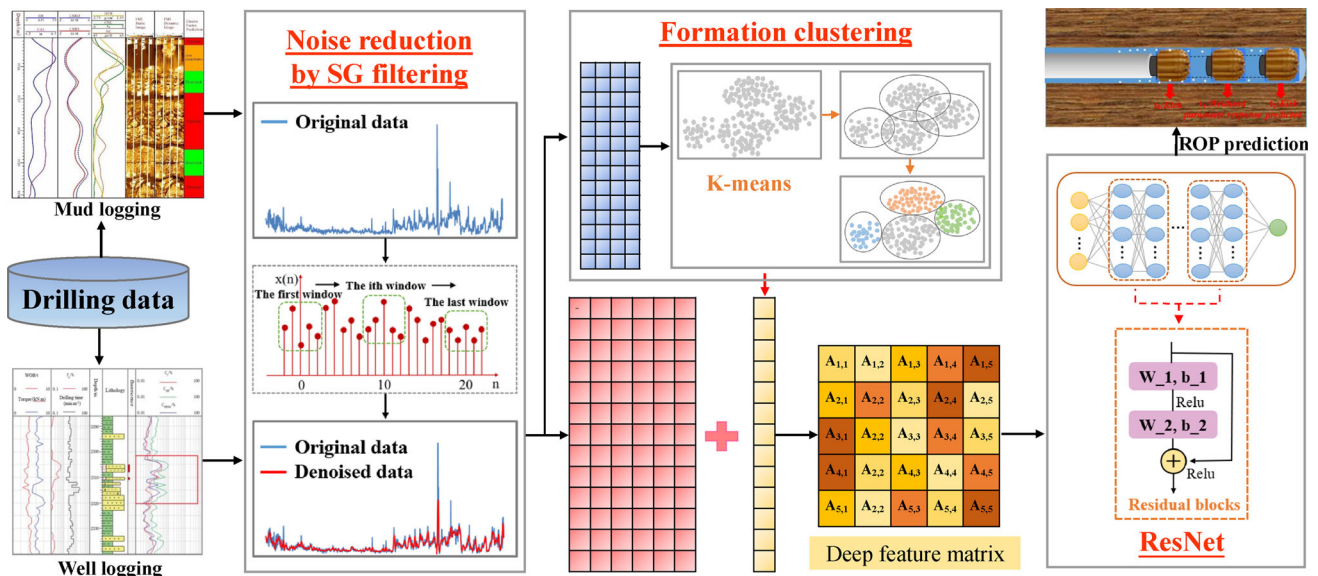
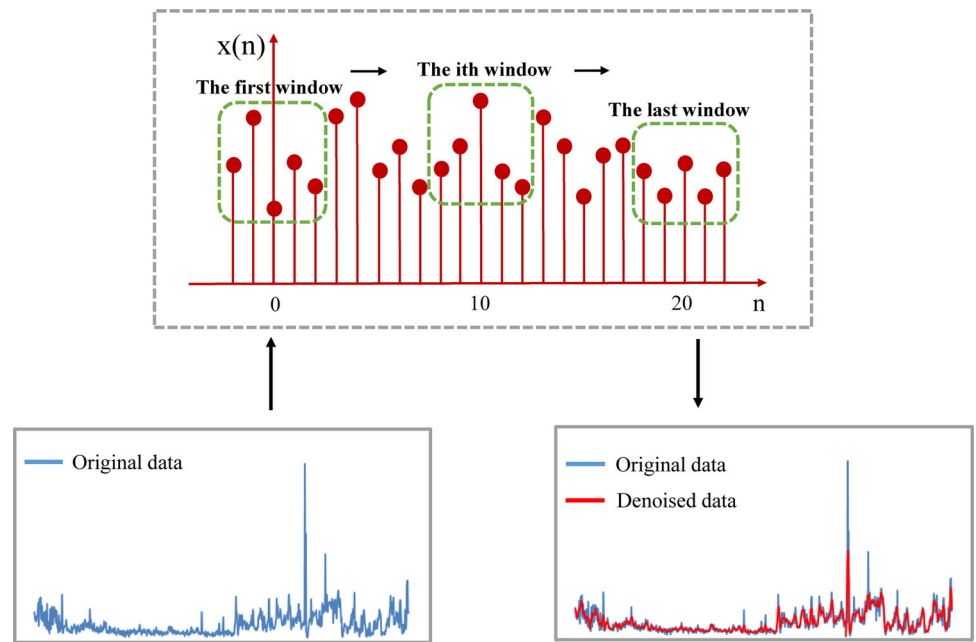


Fig. 1 Framework of DKR model

Fig. 2 The process of SG filter



The process of formation clustering (unsupervised learning) is shown in Fig. 4. Specifically, K-means algorithm is applied. The algorithm requires to specify the number of clusters  $k$  and initial cluster centers in advance. Euclidean distance is used to evaluate the similarity between data points. The similarity is inversely proportional to the distance. The location of each cluster center is updated according to the similarity between the data points and the cluster center.

The essence of K-means algorithm is to divide the unlabeled data set  $X$  (Eq. 2) into  $k$  ( $k < m$ ) clusters  $C = c^{(1)}, c^{(2)}, \dots, c^{(k)}$ .

$$X = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(m)} \end{bmatrix}, x^{(i)} \in R^n \quad (2)$$

where  $x^{(i)}$  is the data points in the data set.

Its framework is as follows:

- 1) Randomly select  $k$  cluster center points  $\mu_1, \mu_2, \dots, \mu_k$ ;

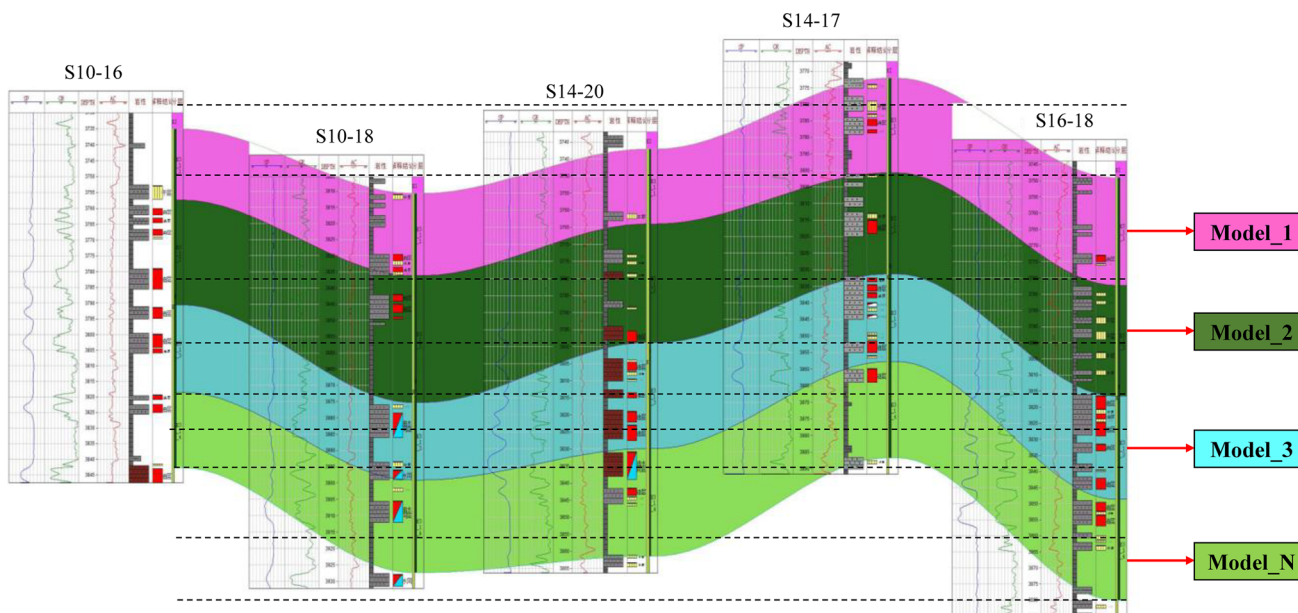


Fig. 3 Traditional ROP prediction strategy for different formations in the same well

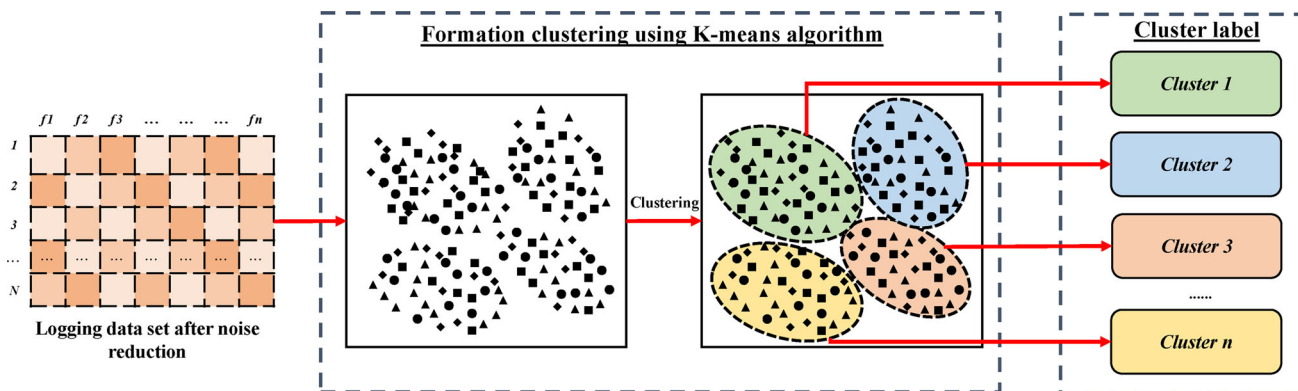


Fig. 4 Formation clustering with K-means algorithm

2) For each data point  $x^{(i)}$ , decide the cluster it belongs to: ensure consistency.

$$c^{(i)} := \arg \min_j \|x^{(i)} - \mu_j\|^2 \tag{3}$$

$$J(c, \mu) = \sum_{i=1}^m \|x^{(i)} - \mu_{c^{(i)}}\|^2 \tag{5}$$

where  $c^{(i)}$  is the cluster closest to  $x^{(i)}$ ,  $i < k$ ;

3) For each cluster  $j$ , recalculate the center  $\mu_j$ ;

$$\mu_j := \frac{\sum_{i=1}^m 1\{c^{(i)} = j\}x^{(i)}}{\sum_{i=1}^m 1\{c^{(i)} = j\}} \tag{4}$$

4) Repeat step 2) and 3) until convergence. Distortion function Eq. (5) represents the square sum of the distance from each data point to its cluster center. When  $J$  reach the smallest value, clustering is convergent. Because of the non-convexity of Eq. (5), local convergence may happen. So the clustering process should be conducted for multiple times to

In this paper, the elbow method [44] is used to determine a proper  $k$  value.

### 2.5 Deep Residual Neural Network

For big data of drilling, simple neural networks are inadequate in learning ability and generalization ability, and more complicated networks have been proposed. A deeper and larger network is believed to better capture the complex nonlinear relationship between feature parameters and ROP. With the increase in network layers, it can extract more features and capture intricate relationships. However, the

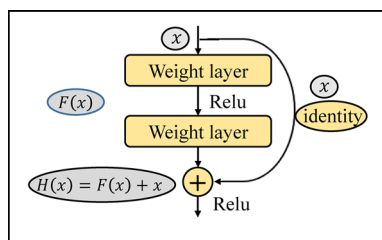


Fig. 5 ResNet residual learning

increased complexity of network is not always in favor of the prediction accuracy. On the other hand, there are growing risks of over-fitting, gradient vanishing, and gradient explosion. Traditional neural networks are unable to deal with the big data in ROP prediction.

In view of the successful applications of ResNet in solving complicated regression problems with big data [45–47], this work uses ResNet [48] to meet the challenges. ResNet adds shortcut connections in the network, which directly transfer the output of the previous layer to a subsequent nonadjacent layer through identity mapping. By modifying the network structure, ResNet can assure model stability and accuracy with the increased network layers.

The residual learning in ResNet is shown in Fig. 5. Suppose that the input of a neural network is  $x$  and the expected output is  $H(x)$ . It is difficult to directly train the model to find the relationship between  $x$  and  $H(x)$ . In the residual learning scenario, the input  $x$  is directly transferred to the output as the initial result through shortcut connection, and the output becomes  $H(x) = F(x) + x$ . If  $F(x) = 0$ , then  $H(x) = x$ , which is an identity mapping. Thus, the learning objective of ResNet is no longer a complete output, but the difference between the target value  $H(x)$  and  $x$ , residual  $F(x) = H(x) - x$ . The training target is to minimize the residual to 0. This jump structure of residual breaks the convention that in the neural network the output of one layer can only be used as the input for the next neighboring layer. The output of previous layer can cross several layers and directly serve as the input of a subsequent layer. Therefore, with the increase in network layers, the accuracy and stability of model do not decline. The workflow of ResNet is demonstrated in Fig. 6. The input includes mud logging parameters and formation cluster label.

## 2.6 Evaluation Index

The model performance is evaluated by mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination ( $R^2$ ). MAE is the absolute error

between the predicted ROP  $y^{\text{pre}}$  and the actual ROP  $y$ :

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i^{\text{pre}} - y_i| \quad (6)$$

where  $n$  is the number of data points;  $y_i^{\text{pre}}$  is the  $i$ -th predicted ROP, m/h;  $y_i$  is the  $i$ -th real ROP, m/h.

MAPE measures the relative error between  $y^{\text{pre}}$  and  $y$ :

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i^{\text{pre}} - y_i|}{y_i} \quad (7)$$

For a high-performance model, MAE and MAPE should be small.

$R^2$  evaluates the fitting performance of the regression model. The closer  $R^2$  is to 1, the higher the degree of model fitting.  $R^2$  is calculated by:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i^{\text{pre}} - y_i)^2}{\sum_{i=1}^n (y_i^{\text{pre}} - \frac{1}{n} \sum_{i=1}^n y_i)^2} \quad (8)$$

## 2.7 Field Data

This work collects 243,000 sets of data from 40 ultra-deep wells in NY reservoir, Xinjiang Province, China as the original data set, which covers all the main geological formations for the reservoir. The typical geological stratifications and well structure of the target reservoir are illustrated in Fig. 7. Part of the data set are presented in Tables 2 and 3. To avoid data leakage, the division of training data set and test data set is conducted for the wells instead of the data points. The data from 32 wells are used as training data set, while the data of remaining 8 wells comprise test data set.

## 3 Results and Discussion

### 3.1 Model Performance and Analysis

The structure and main parameters of DKR model are listed in Table 4.  $N$  and  $k$  (SG filter) are decided by trial and error. The optimal  $K$  (formation clustering) is determined to be 23 by elbow method. The data value is obtained after several tests. In ResNet, a total of 7 residual blocks are constructed, and each residual block contains three weight layers. This is the optimal network structure obtained through multiple tests.

Figure 8 shows an example of noise reduction using SG filter. It can be seen that the general trends of the processed curves are consistent with the original data. The abnormal

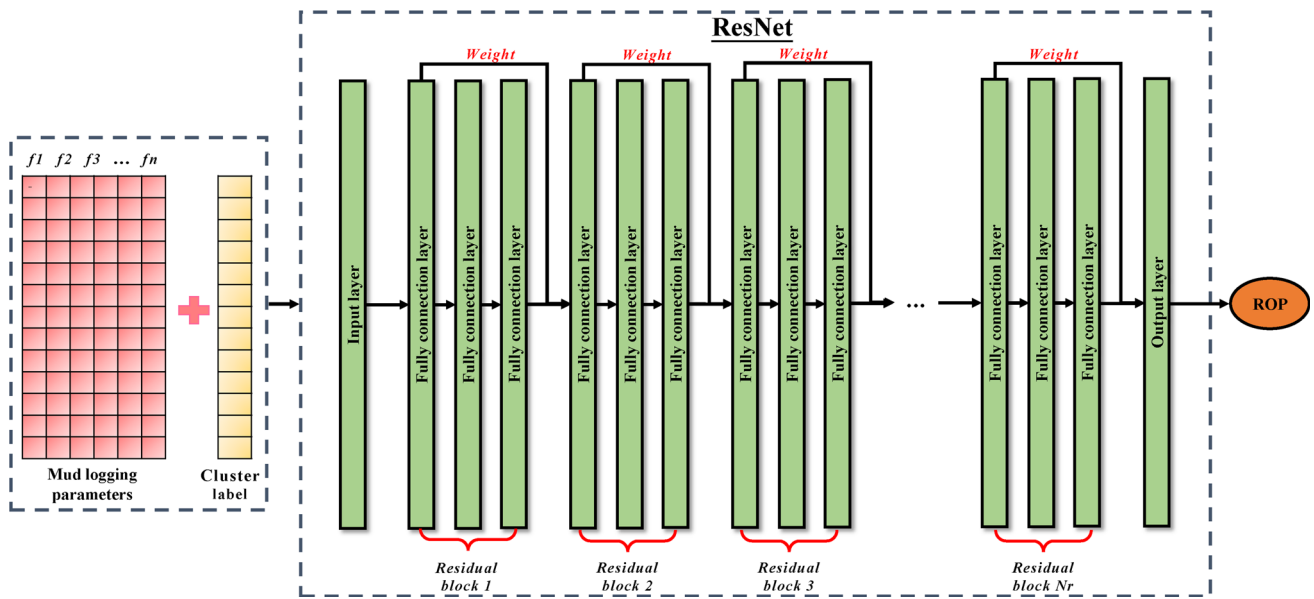
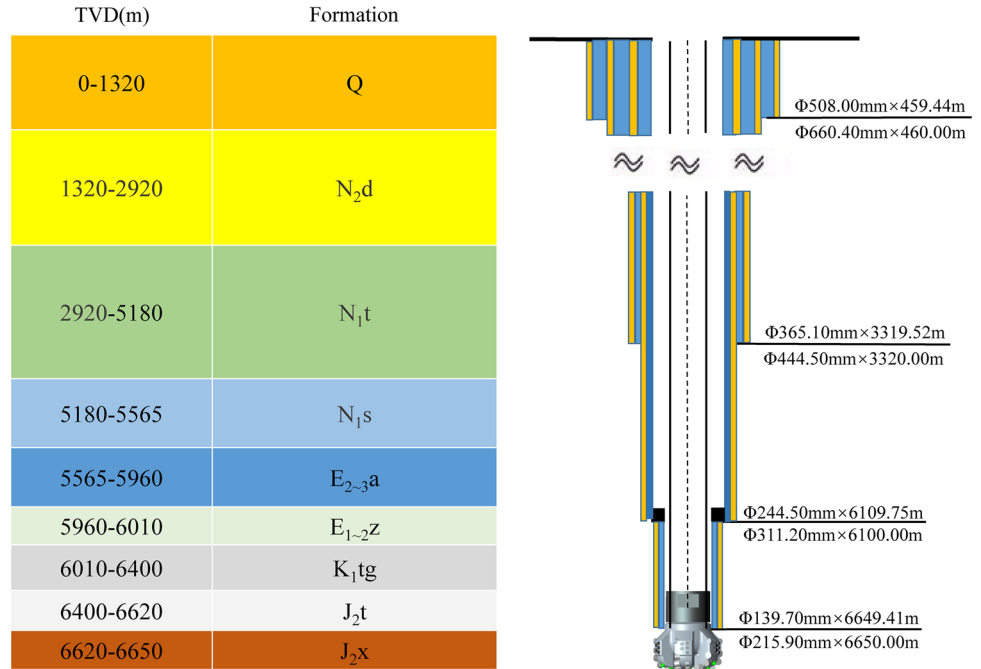


Fig. 6 Workflow of ResNet ROP prediction

Fig. 7 Typical geological stratifications and well structure in the target reservoir



values are significantly reduced, and the curves are smoothed. This is beneficial for model training.

To evaluate the effect of formation clustering, the standard deviation and variation coefficient of rock compressive strength are calculated for 4 continuous major formations (K<sub>1</sub>l, K<sub>1</sub>h, K<sub>1</sub>q, and J<sub>3</sub>k) in the reservoir. The standard deviation and coefficient of variation indicate the dispersion degree of data. Greater standard deviation/coefficient of variation means a more disperse data distribution. The results are presented in Table 5. The 4 formations are classified as 6

different clusters (Cluster 1, 3, 4, 6, 8, and 9). Before clustering, the variation coefficients of 4 formations are all larger than 0.1 (weighted average 0.1253), which is statistically moderate-variant. After clustering, the variation coefficients are all less than 0.1 (the weighted average 0.079), which shows a weak variation. The weighted average of variation coefficient is lowered by 36.95%. The data in the same cluster are more homogeneous in terms of rock property. The results of remaining formations are similar. In this case, the influence of rock characteristics on ROP can be regarded

**Table 2** Part of the drilling data collected from the target reservoir

WD (m)	ROP (m/h)	WOB (kN)	RPM (r/min)	SPP (MPa)	Q (L/s)	...	TOR (kN/m)	BD (mm)	MD (g/cm <sup>3</sup> )
...	...	...	...	...	...	...	...	...	...
3810	7.16	87	54	32.63	25.2	...	10.34	311.2	2.1
3811	5.25	100	54	32.98	25.3	...	7.03	311.2	2.1
3812	4.96	115	54	32.75	24.5	...	11.62	311.2	2.1
3813	5.42	94	54	32.68	25.3	...	8.84	311.2	2.1
3814	5.66	119	54	32.61	23.6	...	8.32	311.2	2.1
3815	8.94	103	54	33.43	25.2	...	13.07	311.2	2.1
3816	9.43	106	54	33.15	24.6	...	10.18	311.2	2.1
3817	6.2	86	54	32.05	24.5	...	8.6	311.2	2.1
3818	6.89	85	54	32.21	24.8	...	8.98	311.2	2.1
3819	8.45	103	54	32.09	24.6	...	6.43	311.2	2.1
3820	9.79	104	54	32.78	24.6	...	9.01	311.2	2.1
...	...	...	...	...	...	...	...	...	...

**Table 3** Part of well logging data collected from the target reservoir

WD (m)	SP (mV)	GR (Ur/h)	AC (μs/m)	...	CNL (ev)	DEN (g/cm <sup>3</sup> )
...	...	...	...	...	...	...
3810	36.345	53.120	229.381	...	7.782	2.583
3811	37.872	44.250	211.906	...	5.752	2.592
3812	48.363	60.089	238.488	...	15.445	2.517
3813	65.256	119.977	260.765	...	19.789	2.462
3814	71.389	150.572	264.012	...	20.515	2.486
3815	70.778	117.411	249.534	...	25.157	2.384
3816	70.142	117.899	266.082	...	27.363	2.313
3817	69.045	130.283	277.094	...	27.625	2.354
3818	69.214	124.869	265.173	...	23.597	2.366
3819	69.664	107.098	248.831	...	19.004	2.386
3820	68.495	111.681	250.341	...	20.567	2.303
...	...	...	...	...	...	...

**Table 4** Key parameters and network structure for DKR model

Sub-process	Key parameters/network structure
Data noise reduction	$N = 25, k = 3$
Formation clustering	$K = 23$
ResNet	Input.layers(14) Residual block 1 Weight.layers(128, activation = ReLU) Weight.layers(256, activation = ReLU) Weight.layers(128, activation = ReLU) ... Residual block 7 Weight.layers(256, activation = ReLU) Weight.layers(256, activation = ReLU) Weight.layers(128, activation = ReLU) Out.layers(1, activation = ReLU)



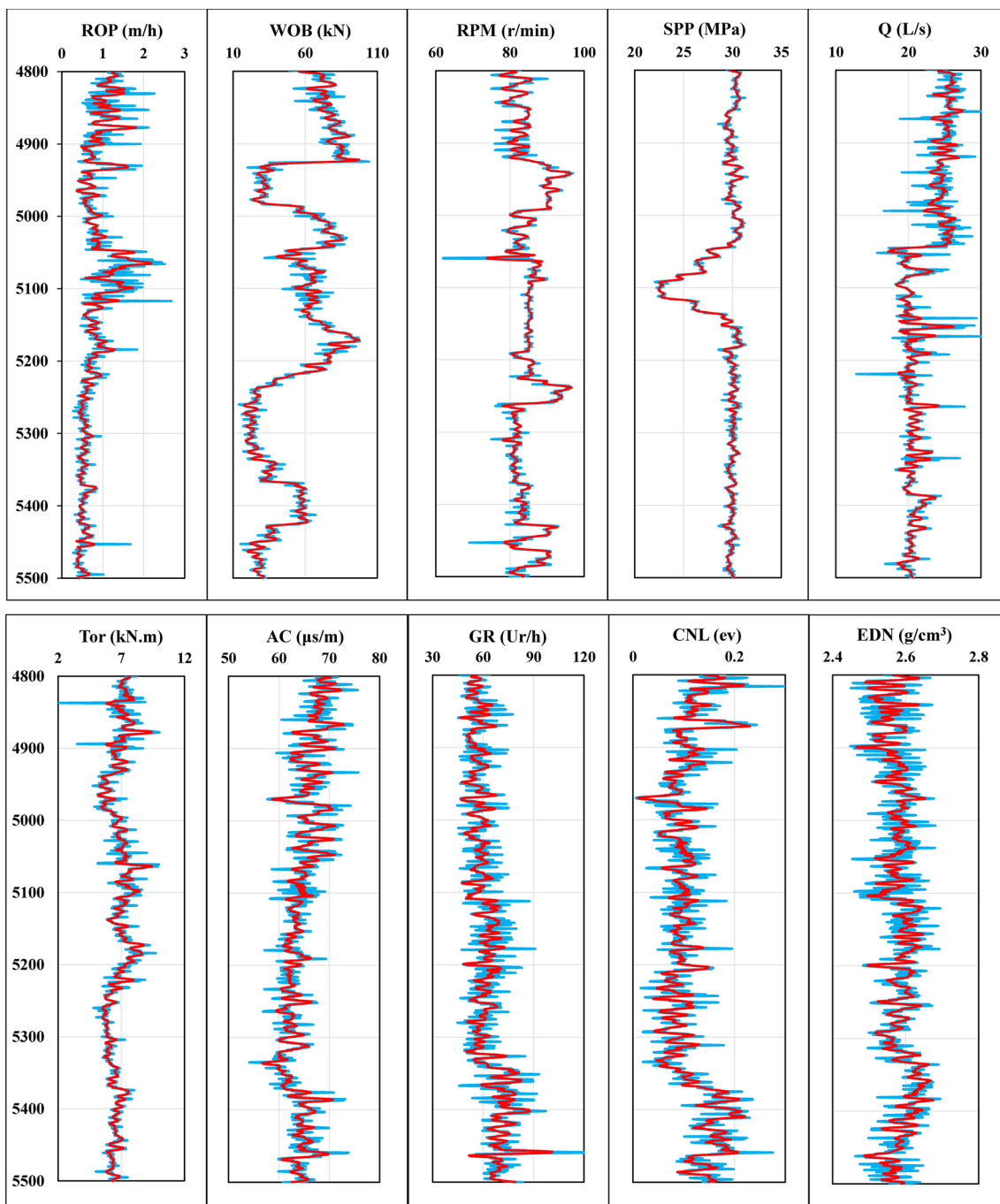


Fig. 8 Result of noise reduction

as identical in the same formation cluster. The cluster label substitutes well logging parameters as the input for ResNet, which effectively reduces the input parameters for ROP production, and facilitates model training.

In order to validate the generalization ability of DKR model, 10 repetition of training and test are conducted. In each experiment, the training wells and test wells are re-divided. The prediction deviations vary in each repetition test, and the 10 test results are averaged for the wells in the

test set, which are shown in Table 6. Note well1 to well8 only represent the sequence in test set, not a certain well. The maximum MAE of the DKR model is 0.66 m/h, the minimum MAE is 0.48 m/h, and the average MAE is 0.55 m/h. The maximum MAPE is 12.77%, the minimum is 9.82%, and the average is 11.34%. The absolute deviation and relative deviation are both tolerable. The maximum  $R^2$  is 0.94, the minimum  $R^2$  is 0.89, and the average  $R^2$  is 0.92. The overall

**Table 5** Standard deviation and variation coefficient of compressive strength before and after clustering

	Formation	Numbers of data	Mean compressive strength (MPa)	Standard deviation (MPa)	Coefficient of variation
Before clustering	K <sub>1</sub> l	19,268	137.2	15.85	0.1155
	K <sub>1</sub> h	29,650	132.1	14.94	0.1131
	K <sub>1</sub> q	20,750	138.9	21.28	0.1532
	J <sub>3</sub> k	16,744	174.6	21.61	0.1238
After clustering	Cluster 1	12,317	140.34	11.91	0.0849
	Cluster 3	14,905	188.13	13.04	0.0693
	Cluster 4	16,096	103.98	9.29	0.0893
	Cluster 6	14,535	152.48	12.83	0.0841
	Cluster 8	13,094	132.79	11.95	0.0900
	Cluster 9	15,465	170.22	10.03	0.0589

**Table 6** Average test results of DKR model

	MAE (m/h)	MAPE (%)	R <sup>2</sup>
Well1	0.54	11.54	0.92
Well2	0.50	10.48	0.93
Well3	0.59	12.01	0.91
Well4	0.66	12.77	0.89
Well5	0.51	11.41	0.93
Well6	0.57	11.60	0.91
Well7	0.48	9.82	0.94
Well8	0.55	11.10	0.92
Average	0.55	11.34	0.92

fitting accuracy is good. The repetition experiments demonstrate that DKR model can achieve accurate and stable ROP predictions. Despite of small fluctuations, MAE, MAPE, and R<sup>2</sup> are maintained at satisfactory values.

Violin plots (Figs. 9 and 10) are applied to analyze the error distribution of DKR model. Violin Plot combines the characteristics of box plot and density plot. It displays data distribution and its probability density. The AEs (absolute error) generated by DKR model during the tests are shown in Fig. 9. The box plots indicate the 25 and 75% quantiles. For the test wells, most of the AEs are less than 1 m/h, which concentrate around the medians (0.50–0.71 m/h). The distribution density decreases from the peak value near the median. The occurrence of large AE is rare. The maximum AE is between 3.01 and 3.68 m/h. The APEs (absolute percentage error) of DKR model are shown in Fig. 10. The medians of relative error range from 9.8 to 13.1%, validating a fair model accuracy. Most of the APEs are distributed in between 5 and 15%. The maximum APE ranges from 34 to 41%. There are no extreme outliers in the plots, indicating that the model has good generalization ability and stability.

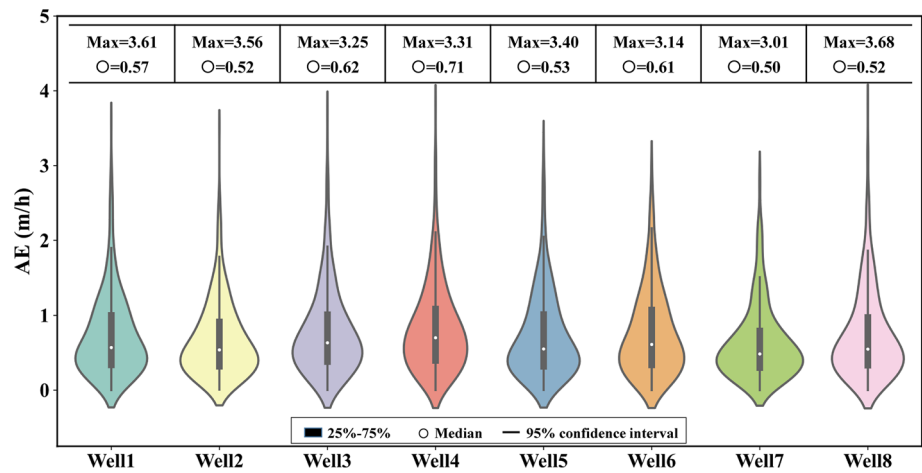
Figure 11 depicts the relationship between predicted ROP and real ROP values. In the DKR model, the determination coefficients are in the range of 0.89 ~ 0.94, indicating that the proposed model precisely fits the nonlinear relationship between the ROP and the feature parameters. In conclusion, the repetition test results prove that DKR model has satisfactory performance for ROP prediction in ultra-deep wells, with high prediction accuracy, good robustness, and strong generalization ability.

### 3.2 Ablation Experiment

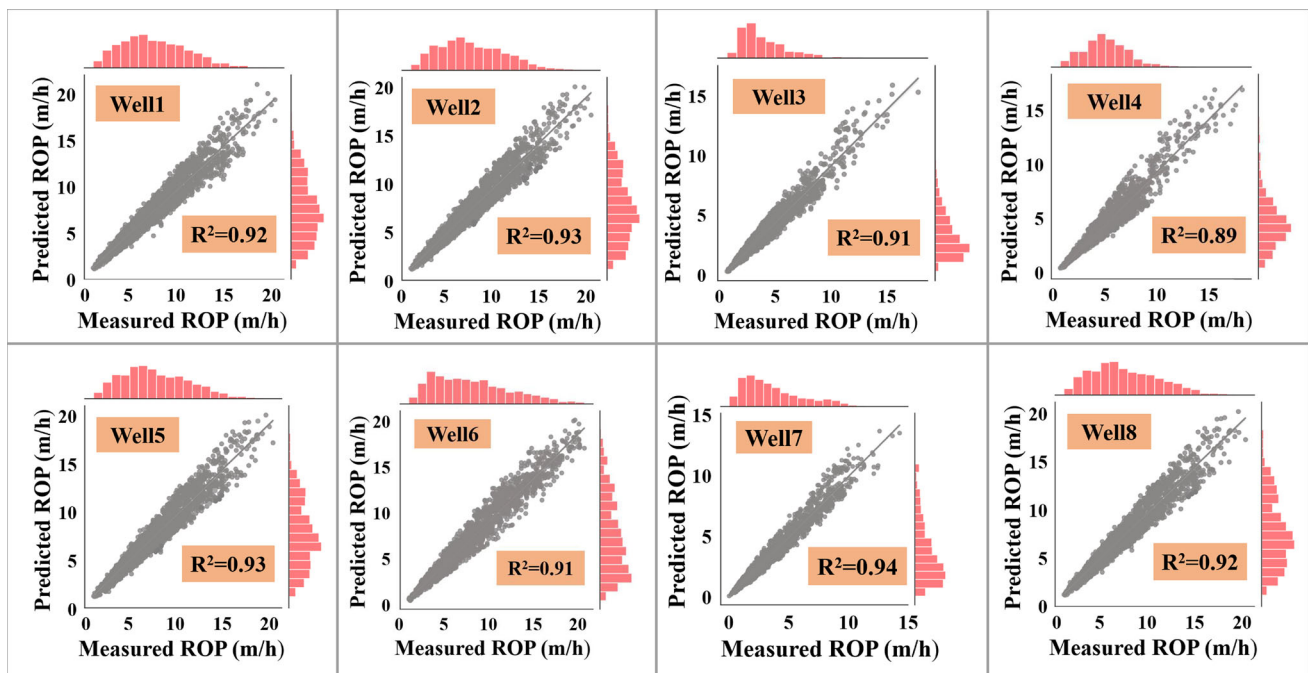
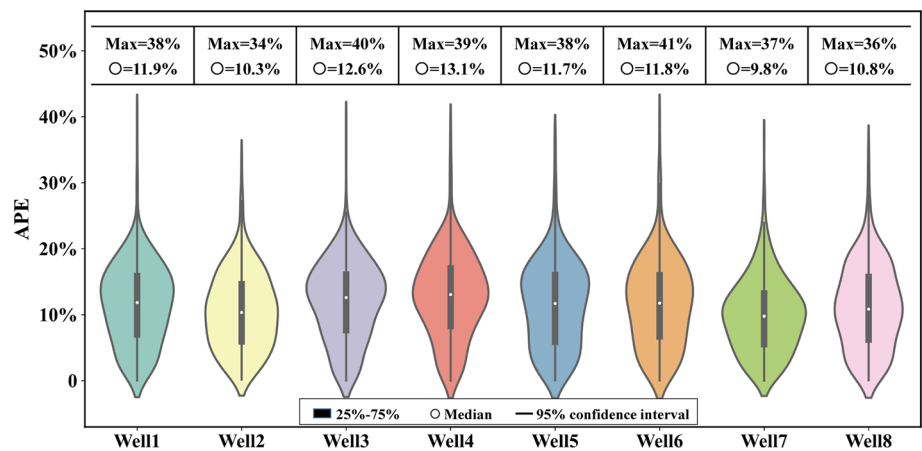
An ablation experiment in machine learning involves removing certain elements from the dataset or model in order to better understand its behavior. In this section, the ablation experiments of DKR model is performed to evaluate the contribution of each sub-process. Specifically, following combinations are tested: (1) K-means + ResNet, where the data noise reduction is removed; (2) Data noise reduction + ResNet, where the formation clustering is removed; (3) Data noise reduction + K-means + ANN, where ResNet is replaced by a traditional 23-layer ANN. Similarly, 10 repetition tests are conducted for each combination, and the average test results are shown in Table 7. It is demonstrated that the removal of any part of DKR model leads to certain deterioration, while the substitution of ResNet with ANN results in the largest impact.

Figure 12 shows the AE and APE of ROP prediction with and without noise reduction. Without noise reduction, the median, quartile range, and 95% confidence of the AE and APE obviously grow. MAE and MAPE increase by 12.73 and 9.0%. R<sup>2</sup> is reduced from 0.92 to 0.90. It reveals that data noise reduction plays an important role in building a high-precision ROP model. By data noise reduction, the accuracy of the ROP prediction can be improved to a certain extent.

**Fig. 9** AE distribution of DKR model



**Fig. 10** APE distribution of DKR model



**Fig. 11** Comparison of predicted ROP and real ROP

**Table 7** Results of ablation experiment

Model	MAE(m/h)	MAPE (%)	$R^2$
Data noise reduction + K-means + ResNet (DKR)	0.55	11.34	0.92
K-means + ResNet (KR)	0.62	12.36	0.90
Data noise reduction + ResNet (DR)	0.69	13.47	0.87
Data noise reduction + K-means + ANN (DKA)	0.81	15.55	0.84

Figure 13 shows the AE and APE of ROP prediction with and without formation clustering. Both of the models use SG filter for noise reduction and the ResNet with same structure. Similar to Fig. 12, the median, quartile range, and 95% confidence interval of errors have significantly increased when formation clustering is removed.  $R^2$  reduces to 0.87. The AE and APE of DR model are 0.69 m/h and 13.47%. In comparison, the AE and APE of DKR model are 0.55 m/h and 11.34%. It proves that formation clustering is a crucial contributor for the accuracy of DKR model.

The effect of replacing ResNet with an equal-layer ANN is shown in Fig. 14. The AE and APE distribution of the DKA model are significantly larger than those of DKR model. The AE and APE of DKA model are 0.81 m/h and 15.55%, increased by 47.27 and 37.13%.  $R^2$  declines to 0.84. It demonstrates that conventional ANN cannot solve the problem that the model accuracy decreases with the increased

network layers. Otherwise, ResNet successfully overcomes this challenge.

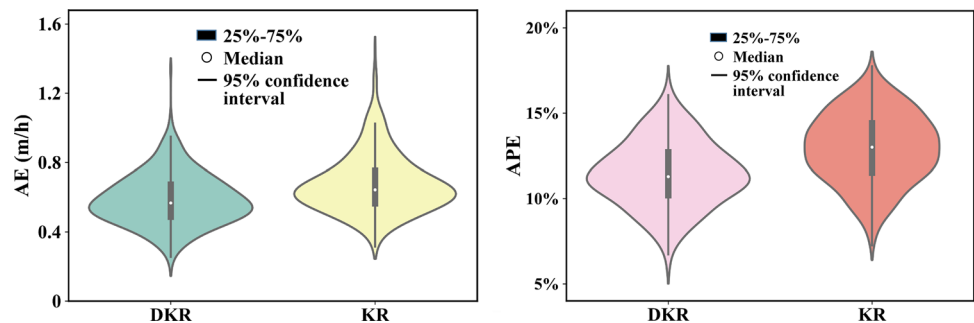
### 3.3 Comparison with Other Machine Learning Methods

The performance of the proposed ROP prediction model (DKR) is compared with three benchmark machine learning models (Back-Propagation neural network (BP), SVM, and RF) [24, 49–54]. To ensure consistency, noise reduction and formation clustering are also performed for the three models. And the model parameters are the optimized values after trial and error. Each model has been trained and tested for 10 times, and the results are averaged. It can be seen from Fig. 15 that DKR model has better performance than BP, SVM, and RF. The AE and APE of DKR are the lowest, while its  $R^2$  is the highest. BP shows a decent accuracy, but SVM and RF perform poorly on ROP prediction in ultra-deep wells.

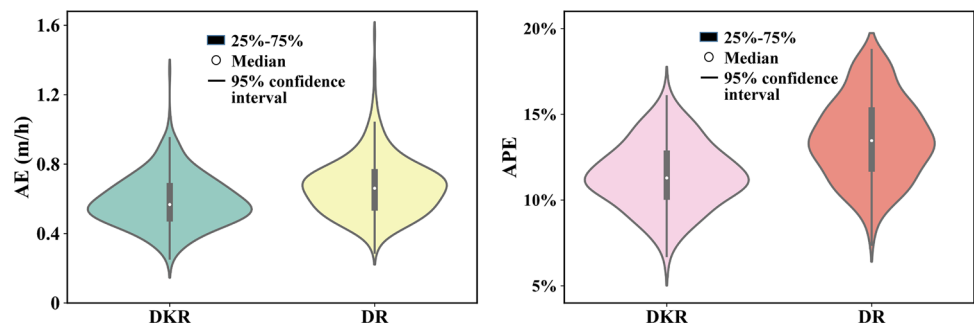
### 3.4 Real-time Field Application

Recently, the trained DKR model was applied for a real-time ROP prediction of a new ultra-deep well (HT-X) in the reservoir. The prediction started from a measure depth of 5700 m, and ended at 7600 m. The test results are shown in Fig. 16. Figure 16a displays measured ROP and predicted ROP along the measure depth. It can be seen that the predicted ROP curve closely follows the measured ROP curve, despite that the peak and bottom values are less prominent in the predicted curve. The smooth change of predicted ROP curve is the result of SG filtering. Figure 16b shows the AE and

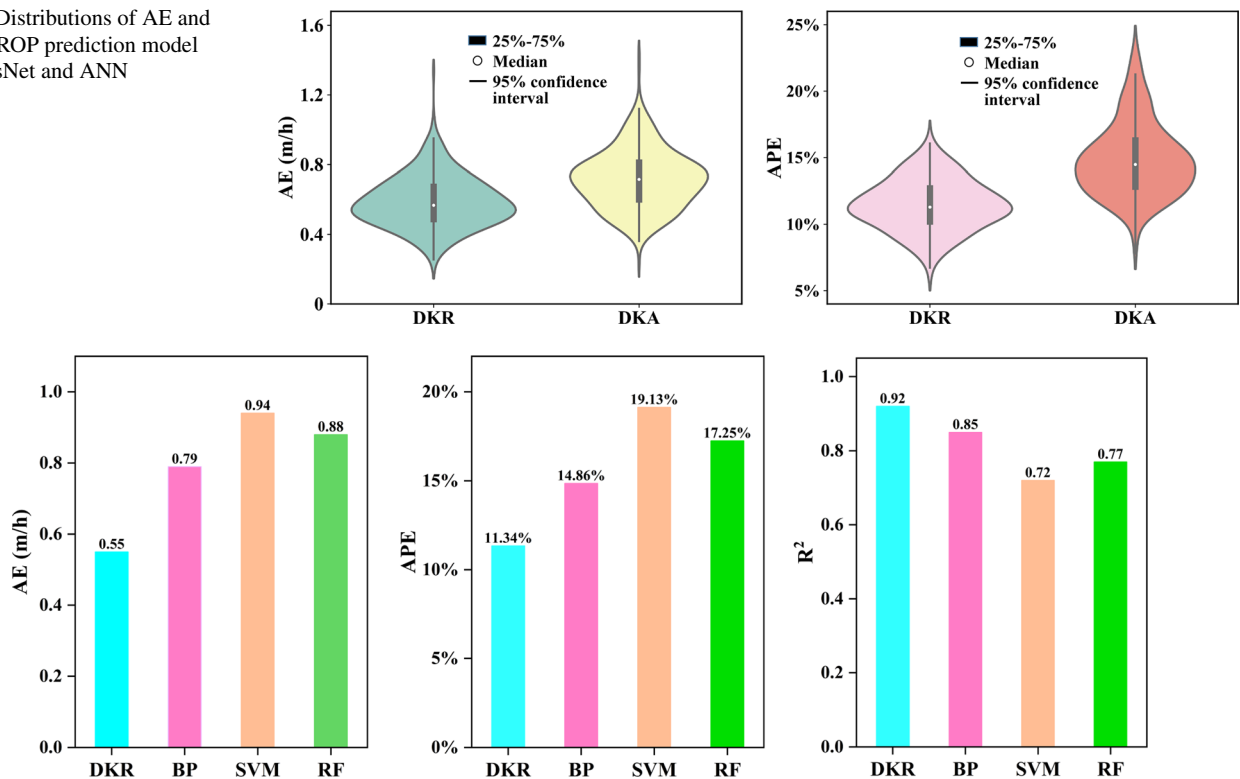
**Fig. 12** Distributions of AE and APE of ROP prediction model with and without data noise reduction



**Fig. 13** Distributions of AE and APE of ROP prediction model with and without formation clustering



**Fig. 14** Distributions of AE and APE of ROP prediction model with ResNet and ANN



**Fig. 15** Comparison of DKR and other machine learning models

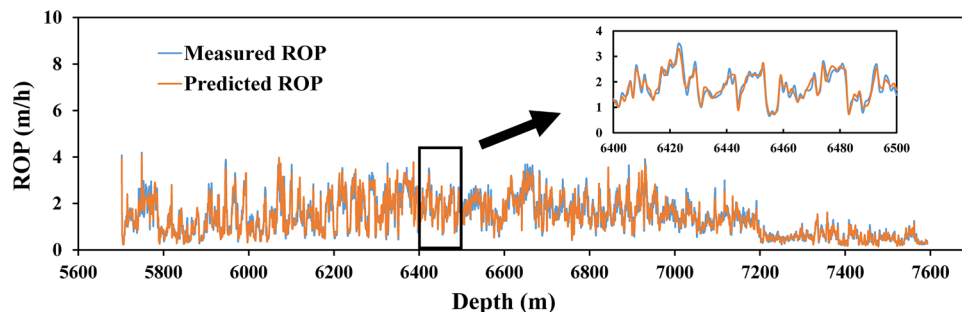
APE distributions of ROP prediction. The AE is mostly distributed in the range of 0–0.2 m/h, while the APE is generally located in the range of 5–15%. The maximum and average AEs of the DKR model are, respectively, 0.77 and 0.17 m/h, corresponding to APEs of 30.3 and 11.2%. There are only a few of unsatisfactory large errors in the whole ROP prediction process. Figure 16c depicts the relationship between predicted and measured ROPs. The high  $R^2$  of 0.91 indicates the prediction accuracy of DKR model is great. The field test validates the feasibility of applying DKR model to real-time ROP prediction in drilling ultra-deep wells.

### 4 Conclusion

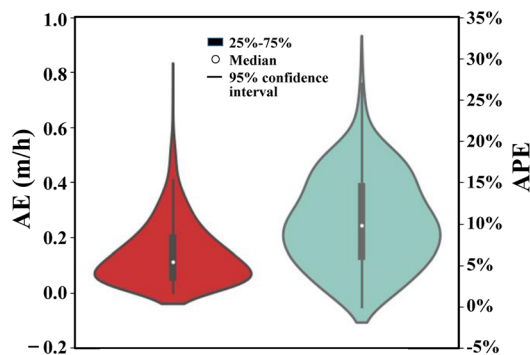
This paper proposes a new ROP intelligent prediction model for ultra-deep wells. The model utilizes mud logging and well logging parameters that are closely related to ROP as inputs and incorporates several innovative techniques, including data noise reduction using SG filter, formation clustering using K-means algorithm, and a ResNet-based

neural network for prediction. The model is tested on field data from an ultra-deep reservoir in northwest China and is found to have high prediction accuracy, good robustness, and strong generalization ability. The average MAE, MAPE, and  $R^2$  are 0.55 m/h, 11.34%, and 0.92. Ablation experiments demonstrate the importance of each sub-process, while comparison with other mainstream machine learning models confirms the superiority of the proposed model. Notably, the proposed model with ResNet outperforms the conventional ANN model by successfully overcoming the challenge of decreased accuracy with increased network layers. Furthermore, a field test validates the model’s feasibility for real-time ROP prediction in drilling ultra-deep wells (MAPE = 11.2%,  $R^2 = 0.91$ ). Overall, the proposed ROP prediction model based on formation clustering and ResNet shows great potential for further field application in ultra-deep wells, making it a highly novel and competitive approach to ROP prediction.

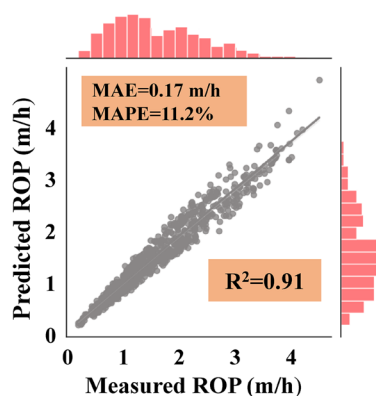
**Fig. 16** Results of real-time ROP prediction in HT-X well using DKR model



(a) Measured and predicted ROPs vs. measure depth.



(b) AE and APE distributions of ROP prediction.



(c) Relationship between predicted ROP and measured ROP.

**Acknowledgements** This work is financially supported by National Natural Science foundation of China (Grant No. 52104006), and Science and Technology Cooperation Project of the CNPC-SWPU Innovation Alliance (Grant No. 2020CX040202). The authors would also like to thank Professor Michael C. Sukop of Florida International University for his suggestions and help.

## References

- Xiao, D.; Hu, Y.; Wang, Y., et al.: Wellbore cooling and heat energy utilization method for deep shale gas horizontal well drilling. *Appl. Therm. Eng.* **213**, 118684 (2022)
- Shi, X.; Luo, C.; Cao, G., et al.: Differences of main enrichment factors of S1111-1 sublayer shale gas in southern Sichuan Basin. *Energies* **14**, 5472 (2021)
- Guo, T.; Tang, S.; Liu, S., et al.: Physical simulation of hydraulic fracturing of large-sized tight sandstone outcrops. *SPE J.* **26**(1), 372–393 (2021)
- Al-Shargabi, M.; Davoodi, S.; Wood, D.A., et al.: A critical review of self-diverting acid treatments applied to carbonate oil and gas reservoirs. *Pet. Sci.* (2022). <https://doi.org/10.1016/j.petsci.2022.10.005>
- Fang, Y.J.; Yang, E.L.; Guo, S.L., et al.: Study on micro remaining oil distribution of polymer flooding in Class-II B oil layer of Daqing Oilfield. *Energy* **254**, 124479 (2022)
- Ashena, R.; Rabiei, M.; Rasouli, V., et al.: Drilling parameters optimization using an innovative artificial intelligence model. *J. Energy Res. Technol.* **143**, 50902 (2021)
- Bajolvand, M.; Ramezanzadeh, A.; Mehrad, M.; Roohi, A.: Optimization of controllable drilling parameters using a novel geomechanics-based workflow. *J. Pet. Sci. Eng.* **218**, 111004 (2022)



8. Zang, C.; Lu, Z.; Ye, S.; Xu, X.; Xi, C.; Song, X.; Guo, Y.; Pan, T.: Drilling parameters optimization for horizontal wells based on a multiobjective genetic algorithm to improve the rate of penetration and reduce drill string drag. *Appl. Sci.* **12**, 11704 (2022)
9. Bingham, G.: How rock properties are related to drilling. *Oil Gas J.* **62**, 94–101 (1964)
10. Young, F.S.: Computerized drilling control. *J. Petrol. Technol.* **4**, 483–496 (1969)
11. Bourgoyne, A.T.; Young, F.S.: A multiple regression approach to optimal drilling and abnormal pressure detection. *Soc. Pet. Eng. J.* **14**(4), 1452 (1974)
12. Warren, T.M.: Penetration-rate performance of roller-cone bits. *SPE Drill. Eng.* **2**(1), 9–18 (1987). <https://doi.org/10.2118/13259-PA>
13. Hareland, G; Hoberock, L.L.: Use of drilling parameters to predict in-situ stress bounds. Paper presented at the SPE/IADC drilling conference, Amsterdam, The Netherlands, 22–25 Feb. SPE-25727-MS (1993). <https://doi.org/10.2118/25727-MS>
14. Rommetveit, R.; Bjrkevoll, K.; Halsey, G.W., et al.: Drilltronics: An Integrated System for Real-Time Optimization of the Drilling Process. Croom Helm (2004)
15. Osgouei, R.E.: Rate of penetration estimation model for directional and horizontal wells. Middle East Technical University, Ankara, Turkey (2007)
16. Etesami, D.; Shirangi, M.G.; Zhang, W.J.: A semiempirical model for rate of penetration with application to an offshore gas field. *SPE Drill. Compl.* **36**(1), 29–46 (2021). <https://doi.org/10.2118/202481-PA>
17. Wiktorski, E.; Kuznetsov, A.; Sui, D.: ROP optimization and modeling in directional drilling process. Paper presented at the SPE Bergen one day seminar, Bergen, Norway, 5 April. SPE-185909-MS (2017). <https://doi.org/10.2118/185909-MS>
18. Soares, C.; Gray, K.: Real-time predictive capabilities of analytical and machine learning rate of penetration (ROP) models. *J. Pet. Sci. Eng.* **172**, 934–959 (2019)
19. Elkatatny, S.: Real-time prediction of rate of penetration while drilling complex lithologies using artificial intelligence techniques. *Ain Shams Eng. J.* (2020)
20. Zhou, Y.; Chen, X.; Zhao, H., et al.: A novel rate of penetration prediction model with identified condition for the complex geological drilling process. *J. Process Control* **100**(4), 30–40 (2021)
21. Brenjkar, E.; Delijani, E.B.: Computational prediction of the drilling rate of penetration (ROP): a comparison of various machine learning approaches and traditional models (2021)
22. Liu, N.; Gao, H.; Zhao, Z., et al.: A stacked generalization ensemble model for optimization and prediction of the gas well rate of penetration: a case study in Xinjiang. *J. Pet. Explor. Prod. Technol.* **12**, 1595–1608 (2022)
23. Bilgesu, H.I.; Tetrick, L.T.; Altmis, U.; et al. A new approach for the prediction of rate of penetration (ROP) values
24. Zhong, R.; Salehi, C.; Johnson, R., Jr.: Machine learning for drilling applications: a review. *J. Nat. Gas Sci. Eng.* **108**, 104807 (2022)
25. Chen, S.; Tan, D.: A SA-ANN-based modeling method for human cognition mechanism and the PSACO cognition algorithm. *Complexity* **2018**, 1–21 (2018)
26. Sabah, M.; Talebkeikhah, M.; Wood, D.A., et al.: A machine learning approach to predict drilling rate using petrophysical and mud logging data. *Earth Sci. Inform.* **12**, 319–339 (2019)
27. Shi, X.; Liu, G.; Gong, X., et al.: An efficient approach for real-time prediction of rate of penetration in offshore drilling. *Math. Probl. Eng.* **100**, 200 (2016). <https://doi.org/10.1155/2016/3575380>
28. Anemangely, M.; Ramezanzadeh, A.; Tokhmechi, B., et al.: Drilling rate prediction from petrophysical logs and mud logging data using an optimized multilayer perceptron neural network. *J. Geophys. Eng.* **4**, 15 (2018)
29. Diaz, M.B.; Kim, K.Y.; Shin, H.S.; Zhuang, L.: Predicting rate of penetration during drilling of deep geothermal well in Korea using artificial neural networks and real-time data collection. *J. Nat. Gas Sci. Eng.* **67**, 225–232 (2019)
30. Zhang, C.; Song, X.; Su, Y.; Li, G.: Real-time prediction of rate of penetration by combining attention-based gated recurrent unit network and fully connected neural networks. *J. Pet. Sci. Eng.* **213**, 110396 (2022)
31. Noshi, C.I.; Schubert, J.J.: Application of data science and machine learning algorithms for ROP prediction: turning data into knowledge (Conference Paper). In: Proceedings of the Annual Offshore Technology Conference, pp. 12–15 (2019)
32. Alsaihati, A.; Elkatatny, S.; Gamal, H.: Rate of penetration prediction while drilling vertical complex lithology using an ensemble learning model. *J. Petrol. Sci. Eng.* **208**, 109335 (2022)
33. Hashemizadeh, A.; Bahonar, E.; Chahardowli, M., et al.: Analysis of rate of penetration prediction in drilling using data-driven models based on weight on hook measurement. *Earth Sci. Inform.* **15**, 2133–2153 (2022)
34. Hegde, C.; Daigle, H.; Millwater, H., et al.: Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models. *J. Pet. Sci. Eng.* **159**, 295–306 (2017)
35. Mehrad, M.; Bajolvand, M.; Ramezanzadeh, A., et al.: Developing a new rigorous drilling rate prediction model using a machine learning technique. *J. Pet. Sci. Eng.* **192**, 107338 (2020)
36. Ashrafi, S.B.; Anemangely, M.; Sabah, M.; et al.: Application of hybrid artificial neural networks for predicting rate of penetration (ROP): a case study from Marun oil field. *J. Pet. Sci. Eng.* (2018)
37. Olukoga, T.A.; Feng, Y.: Practical machine-learning applications in well-drilling operations. *SPE Drill. Complet.* **36**(04), 849–867 (2021)
38. Barbosa, L.F.F.M.; Nascimento, A.; Mathias, M.H., et al.: Machine learning methods applied to drilling rate of penetration prediction and optimization-A review. *J. Petrol. Sci. Eng.* **183**, 106332 (2019)
39. Gupta, S.; Gupta, A.: Dealing with noise problem in machine learning data-sets: a systematic review. *Procedia Comput. Sci.* **161**, 466–474 (2019)
40. Savitzky, A.; Golay, M.J.E.: Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**(8), 1627–1639 (1964)
41. Acharya, D.; Rani, A.; Agarwal, S., et al.: Application of adaptive Savitzky-Golay filter for EEG signal processing. *Perspect. Sci.* **8**, 677–679 (2016)
42. Kordestani, H.; Zhang, C.: Direct use of the Savitzky-Golay filter to develop an output-only trend line-based damage detection method. *Sensors* **20**(7), 1983 (2020)
43. Liu, Y.; Dang, B.; Li, Y., et al.: Applications of Savitzky-Golay filter for seismic random noise reduction. *Acta Geophys.* **64**, 101–124 (2016)
44. Ikotun, A.M.; Ezugwu, A.E.; Abualigah, L., et al.: K-means clustering algorithms: a comprehensive review, variants analysis, and advances in the era of big data. *Inf. Sci.* (2022). <https://doi.org/10.1016/j.ins.2022.11.139>
45. Zuo, C.; Qian, J.; Feng, S., et al.: Deep learning in optical metrology: a review. *Light Sci. Appl.* **11**(1), 1–54 (2022)
46. Khan, S.; Naseer, M.; Hayat, M., et al.: Transformers in vision: a survey. *ACM Comput. Surv. (CSUR)* **54**(10s), 1–41 (2022)
47. Guo, M.H.; Xu, T.X.; Liu, J.J.; et al.: Attention mechanisms in computer vision: a survey. *Comput. Vis. Media* 1–38 (2022)
48. He, K.; Zhang, X.; Ren, S.; et al.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2016)
49. Azimi, H.; Shiri, H.; Malta, E.R.: A non-tuned machine learning method to simulate ice-seabed interaction process in clay. *J. Pipeline Sci. Eng.* **1**(4), 379–394 (2021)



50. Azimi, H.; Shiri, H.; Mahdianpari, M.: Iceberg-seabed interaction evaluation in clay seabed using tree-based machine learning algorithms. *J. Pipeline Sci. Eng.* **2**(4), 100075 (2022)
51. Lin, Q.T.; Liu, J.J.; Pang, H., et al.: Application of MIPSO-XGBoost algorithm in prediction of gasoline yield. *Acta Pet. Sin. (Petroleum Processing Section)* **39**(3), 19–19 (2023)
52. Sang, K.-H.; Yin, X.-Y.; Zhang, F.-C.: Machine learning seismic reservoir prediction method based on virtual sample generation. *Pet. Sci.* **18**, 1662–1674 (2021). <https://doi.org/10.1016/j.petsci.2021.09.034>
53. Zhang, Y.-Y.; Xi, K.-L.; Cao, Y.-C.; Bao-Hai, Yu.; Wang, H.; Lin, M.-R.; Li, Ke.; Zhang, Y.-Y.: The application of machine learning under supervision in identification of shale lamina combination types—A case study of Chang 73 sub-member organic-rich shales in the Triassic Yanchang Formation, Ordos Basin, NW China. *Pet. Sci.* **18**, 1619–1629 (2021). <https://doi.org/10.1016/j.petsci.2021.09.033>
54. Luo, S.-H.; Xiao, L.-Z.; Jin, Y.; Liao, G.-Z.; Bin-Sen, Xu.; Zhou, J.; Liang, C.: A machine learning framework for low-field NMR data processing. *Pet. Sci.* **19**, 581–593 (2022). <https://doi.org/10.1016/j.petsci.2022.02.001>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

