Petroleum Science 21 (2024) 582–596

Contents lists available at ScienceDirect

Petroleum Science

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

Original Paper

Interpretation and characterization of rate of penetration intelligent prediction model

Zhi-Jun Pei [a], Xian-Zhi Song [a, b, *], Hai-Tao Wang [c], Yi-Qi Shi [a, b], Shou-Ceng Tian [a, b], Gen-Sheng Li [a, b]

[a] College of Petroleum Engineering, China University of Petroleum (Beijing), Beijing, 102249, China
[b] State Key Laboratory of Petroleum Resources and Prospecting, Beijing, 102249, China
[c] Kunlun Digital Technology Co., Ltd., Beijing, 102206, China

ARTICLE INFO

Article history:
Received 13 September 2022
Received in revised form
16 October 2023
Accepted 16 October 2023
Available online 16 October 2023

Edited by Jia-Jia Fei

Keywords:
Fully connected neural network
Explainable artificial intelligence
Rate of penetration
ReLU active function
Deep learning
Machine learning

ABSTRACT

Accurate prediction of the rate of penetration (ROP) is significant for drilling optimization. While the intelligent ROP prediction model based on fully connected neural networks (FNN) outperforms traditional ROP equations and machine learning algorithms, its lack of interpretability undermines its credibility. This study proposes a novel interpretation and characterization method for the FNN ROP prediction model using the Rectified Linear Unit (ReLU) activation function. By leveraging the derivative of the ReLU function, the FNN function calculation process is transformed into vector operations. The FNN model is linearly characterized through further simplification, enabling its interpretation and analysis. The proposed method is applied in ROP prediction scenarios using drilling data from three vertical wells in the Tarim Oilfield. The results demonstrate that the FNN ROP prediction model with ReLU as the activation function performs exceptionally well. The relative activation frequency curve of hidden layer neurons aids in analyzing the overfitting of the FNN ROP model and determining drilling data similarity. In the well sections with similar drilling data, averaging the weight parameters enables linear characterization of the FNN ROP prediction model, leading to the establishment of a corresponding linear representation equation. Furthermore, the quantitative analysis of each feature's influence on ROP facilitates the proposal of drilling parameter optimization schemes for the current well section. The established linear characterization equation exhibits high precision, strong stability, and adaptability through the application and validation across multiple well sections.

1. Introduction

Polycrystalline Diamond Compact (PDC) bit has been widely used in oil and gas drilling because of its high rock-breaking efficiency and strong formation applicability (Xiong et al., 2020). As the exploration of underground resources gradually advances to deep and hard formations, low rate of penetration (ROP) and high energy consumption have become the main factors affecting the drilling efficiency of PDC bit (Etesami et al., 2021; Mazen et al., 2021). The fundamental problem of increasing ROP is to crack the principle and method of efficient rock breaking and discover the influence law of multi-parameters and ROP (Ersoy and Waller, 1995; Garcia-Gavito and Azar, 1994).

As early as the 1960s, some scholars researched the drilling mechanism and established several ROP equations, such as the Bingham equation (Bingham, 1964), Warren equation (Warren, 1987), and Soares equation (Soares et al., 2020). The most widely used ROP equation is the modified Young's ROP equation (Bourgoyne et al., 1986; Young, 1969), which comprehensively considers the effects of weight on bit (WOB), rotation speed (RS), bit wear, bottom hole pressure difference, and hydraulic parameters on the ROP. In recent years, with the rapid development of artificial intelligence (AI) technology, information technology, data storage technology, and high-efficiency computing technology, a large number of scholars began to establish an intelligent prediction model of ROP based on artificial intelligence algorithm using big drilling data, to mine the complex nonlinear relationship between

* Corresponding author. College of Petroleum Engineering, China University of Petroleum (Beijing), Beijing, 102249, China.
E-mail address: songxz@cup.edu.cn (X.-Z. Song).

https://doi.org/10.1016/j.petsci.2023.10.011

multi parameters and ROP (Zhang et al., 2021). Many scholars widely use the machine learning algorithm because of its simplicity and strong interpretability (Hazbeh et al., 2021), such as random forest (Hegde et al., 2015), k-nearest neighbor (KNN) (Gupta et al., 2020), support vector regression (SVR) (Abdulmalek et al., 2018), extreme gradient boosting (XGB) (Zhou et al., 2022). Zhou et al. (2022) established the prediction model of ROP using KNN, SVR (Pei et al., 2022), and XGB algorithms, respectively, and found that the intelligent prediction model of ROP based on the XGB algorithm has the highest accuracy, 85%. Compared to the ROP equation, the accuracy of intelligent prediction models based on machine learning algorithms has greatly improved. However, when the amount of drilling data is particularly large, the machine learning algorithms are extremely prone to overfitting. Therefore, some scholars began to try to use deep learning algorithms to establish the prediction model of ROP, such as fully connected neural network (FNN) (Amadi et al., 2021; Bardhan et al., 2021; Elmgerbi et al., 2021; Lawal et al., 2021; Negara and Saad, 2020), recurrent neural network (RNN) (Li and Samuel, 2019), long short-term memory neural network (LSTM) (Gers et al., 2000; Mahmoodzadeh et al., 2022), gated recurrent neural network (GRU), etc. The deep learning models performs on big data significantly better than simple machine learning models, further improving the prediction accuracy of ROP (Pei et al., 2023). However, artificial neural networks have many hyperparameters, and how to find the optimal hyperparameters has become a new challenge. Some scholars attempt to use intelligent optimization algorithms to optimize the hyperparameters of neural networks to establish optimal prediction models. Mahmoodzadeh et al. (2022) and Encinas et al. (2022) established the prediction model of ROP based on antlion optimized FNN (ALO-FNN), multiple linear statistical model (MLSM), and multiple non-linear statistical model (MNLSM) respectively. The results show that ALO-FNN intelligent prediction model has the best stability and fitting with the highest prediction accuracy of 84.2%.

The intelligent ROP prediction model aims to capture the complex relationship between multiple parameters and ROP. By employing the FNN algorithm, the accuracy and stability of the intelligent ROP prediction model are significantly improved. However, as the hidden layers increase, the lack of explainability of the FNN becomes more pronounced. We can only input parameters into the artificial intelligence model and obtain the calculated results without fully understanding the calculation process, the relative importance of each feature, and the correlations between the parameters and ROP. The intelligent ROP prediction model may exhibit high accuracy. However, if the prediction results cannot be mathematically explained, it may lead to skepticism among construction personnel. This lack of explainability poses a key and core challenge for applying artificial intelligence technology in the petroleum engineering domains.

The term “Explainable AI” was first coined by Van Lent et al. (2004) to describe the ability of a training system developed for the US Army to explain its AI-driven decisions. After years of development, it has been successfully applied in medical treatment, finance, the chemical industry, image recognition, natural language processing, and other fields. Zhang et al. (2022) introduced Explainable Artificial Intelligence (XAI) technology to improve the transparency and interpretability of artificial intelligence applications applied to audit tasks. Tsoka et al. (2022) classify the building energy performance certificate (EPC) tags based on the artificial neural network algorithm. With the help of XAI, some less important input features of the FNN classification model can be removed without seriously affecting the accuracy of the model. Brito et al. (2022) established a fault detection and diagnosis method for rotating machinery based on various machine learning algorithms. The black box model is interpreted by Shapley additive explanation method to obtain the importance ranking of features and perform fault diagnosis.

In oil exploration and development, most scholars focus on combining artificial intelligence algorithms and industrial scenarios and less on the interpretability of models. Nasiri et al. (2021) successfully illustrate intercorrelations between rock properties (porosity, point load index, P-wave velocity, and Schmidt hammer rebound number) and their representative Uniaxial compressive strength and Young's modulus through the SHAP method. However, although there are few interpretability studies in the field of smart drilling, there is no doubt that the interpretability of the intelligent drilling model plays a vital role in promoting the application of intelligent drilling.

This paper transforms the fully connected neural network (FNN) calculation process into a linear representation by analyzing and deriving the Rectified Linear Unit (ReLU) activation function. The authors establish an optimal intelligent ROP prediction model based on field drilling data. Statistical methods are then utilized to analyze neurons' activation state and weight parameters within the intelligent ROP prediction model. The FNN ROP prediction model is represented linearly within the data similarity layer. By utilizing the linear characterization equation, the paper provides an intuitive display of the contribution of input features to the ROP. Moreover, a corresponding drilling parameter optimization scheme is proposed based on these findings.

2. Methodology

2.1. Fully connected neural network

FNN is a kind of artificial neural network, mainly composed of three parts: input layer, hidden layer, and output layer. The number of neurons in the input layer and output layer of FNN depends on the number of input parameters and output parameters. The structure of the common three-layer neural network is shown in Fig. 1. Generally, the number of hidden layers of the FNN model is at least 1, and the number of hidden layers can even exceed 10 when the data is complex. The three-layer neural network calculation process shown in Fig. 1 is as Eq. (1). The calculation of each neuron in the hidden layer mainly includes linear combination and activation calculation. Normally, the output layer does not perform an

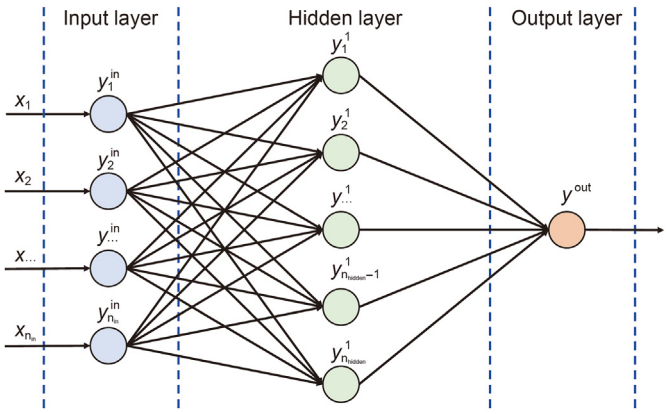

Fig. 1. Schematic diagram of the FNN model.

activation calculation.

$$\begin{cases} y_i^{\text{in}} = x_i, i \in [1, n_{\text{in}}] \\ y_i^1 = \varphi_{\text{active}}\left(\left(\sum_{k=1}^{n_{\text{in}}} w_{k,i}^1 \times y_k^{\text{in}}\right) + b_i^1\right), i \in [1, n_{\text{hidden}}] \\ y^{\text{out}} = \left(\sum_{k=1}^{5} w_k^{\text{out}} \times y_k^1\right) + b_i^{\text{out}} \end{cases} \tag{1}$$

In Eq. (1), n_{in} is the number of input parameters, n_{hidden} is the number of hidden layer neurons. $w_{k,i}^1$ is the weight of the output of the neuron k of the input layer when neuron i of the hidden layer performs linear combination, and b_i^1 is the bias. φ_{active} is the activation function.

The most important function of the activation function is to convert the linear combination of neurons into nonlinear calculation so that complex data can be fitted. There are three common activation functions: ReLU, tanh, and sigmoid. Compared with the tanh and sigmoid activation functions, the ReLU function has the characteristics of fast convergence. It is not easy to cause the problem of gradient disappearance, which has been widely used in regression problems. The calculation formula of the ReLU is shown in Eq. (2) (Hansson and Olsson, 2017).

$$\text{ReLU}(x) = \max(0, x) \tag{2}$$

2.2. FNN model interpretation method

When the total number of input data samples is large, to further simplify Eq. (1), **x**, **w**, and **b** are represented in vector form, respectively, as shown in Eq. (3).

$$\begin{cases} \mathbf{x} = \left[x_1, \cdots, x_{n_{\text{in}}-1}, x_{n_{\text{in}}}\right] \\ \mathbf{w}^i = \left[w_1^i, \cdots, w_{n_{\text{hidden}}-1}^i, w_{n_{\text{hidden}}}^i\right] \\ \mathbf{b}^i = \left[b_1^i, \cdots, b_{n_{\text{hidden}}-1}^i, b_{n_{\text{hidden}}}^i\right] \end{cases} \tag{3}$$

The rewriting vector operation form of Eq. (1) is:

$$\mathbf{y} = \varphi_{\text{active}}\left(\text{x}\mathbf{w}^1 + \mathbf{b}^1\right)\mathbf{w}^{\text{out}} + \mathbf{b}^{\text{out}} \tag{4}$$

The derivation of the ReLU function is shown in Eq. (5). When x is greater than zero, DR(x) is one, and when x is less than zero, DR(x) is zero. When Eq. (5) input is a vector **x**, the output can be recorded as a vector **DR**. And **x**, ReLU(**x**), and DR(**x**) satisfies the relationship in Eq. (6):

$$\text{DR}(x) = \frac{\text{dReLU}(x)}{\text{d}x} = \begin{cases} 1, x > 0 \\ 0, x \le 0 \end{cases} \tag{5}$$

$$\text{ReLU}(\mathbf{x}) = \mathbf{x}\frac{\text{dReLU}(\mathbf{x})}{\text{d}\mathbf{x}} = \mathbf{x}\text{DR}(\mathbf{x}) = \mathbf{xDR} \tag{6}$$

According to Eq. (6), the calculation process of ReLU can be converted from function operation to vector calculation. The vector **DR** is composed of numbers zero and one, which respectively represent neurons in an inactive or activated state. The vector **DR** is the activation vector of the neurons in the current hidden layer. Therefore, when the activation function of the FNN model is ReLU, Eq. (4) can be rewritten as:

$$\mathbf{y} = \left(\text{x}\mathbf{w}^1 + \mathbf{b}^1\right)\mathbf{DR}\left(\text{x}\mathbf{w}^1 + \mathbf{b}^1\right)\mathbf{w}^{\text{out}} + \mathbf{b}^{\text{out}} \tag{7}$$

According to the combination law, Eq. (7) can be simplified as:

$$\mathbf{y} = \mathbf{xw} + \mathbf{b} \tag{8}$$

In Eq. (8), **w** is $\mathbf{w}^1\mathbf{DR}(\text{x}\mathbf{w}^1 + \mathbf{b}^1)\mathbf{w}^{\text{out}}$, and **b** is $\mathbf{b}^1\mathbf{DR}(\text{x}\mathbf{w}^1 + \mathbf{b}^1)\mathbf{w}^{\text{out}} + \mathbf{b}^{\text{out}}$. After deduction, when the number of hidden layers is n, the calculation process of the FNN with ReLU as the activate function can still be simplified to the form of Eq. (8). Where **w** is $\mathbf{w}^1\text{DR}^1\mathbf{w}^2\text{DR}^2\cdots\mathbf{w}^n\text{DR}^n\mathbf{w}^{\text{out}}$, and **b** is $\mathbf{b}^1\text{DR}^1\cdots\mathbf{w}^n\text{DR}^n\mathbf{w}^{\text{out}} + \mathbf{b}^2\text{DR}^2\cdots\mathbf{w}^n\text{DR}^n\mathbf{w}^{\text{out}} + \cdots + \mathbf{b}^n\text{DR}^n\mathbf{w}^{\text{out}} + \mathbf{b}^{\text{out}}$. Where DR^n is the activation vector of the nth hidden layer, which is $\mathbf{DR}(\mathbf{y}^{n-1}\mathbf{w}^n + \mathbf{b}^n)$.

According to Eq. (8), the output result of the FNN model with ReLU as the activation function after multi-layer neuron calculation is still the linear combination of input parameters, which conflicts with the strong nonlinear fitting of the neural network (Hansson and Olsson, 2017). However, different from the conventional linear regression, the weight and bias parameters of the linear combination of the FNN model show nonlinear changes with **x**. Thus, the neural network can maintain a high fitting ability in a huge data set.

Through the analysis of **w**, **b**, and **DR**, we can understand the weight parameters of neural networks and the activation states of neurons. The activation state of neurons is active or inactive, corresponding to one or zero in the **DR**. When multiple sets of **x** are input, the **DR** vector of different input data can be obtained, and the activation state of each neuron on each hidden layer can be counted. The model neuron's activation state should be stable when the model converges. When the activation state of different input parameters is identical, these data can be considered to have some similarities. The **w** and **b** intuitively show the contribution of each feature to the ROP. When the value range difference of the feature parameters is small, the larger the weight parameter of the feature, the more important the feature of the model.

In addition, the rationality and reliability of the FNN ROP prediction model can be analyzed by observing the changing trend of the weights of different characteristics and comparing it with the clear engineering mechanism. When the two are inconsistent, it is necessary to analyze the specific data further to explore the causes of the contradiction. According to Eq. (8), when the activation state of the neural network is completely consistent, the weight parameters of these input parameters are also completely uniform, which further verifies that the input parameters with the same activation have some similarities. According to the calculation formula of **w** and **b** in Eq. (8), the weight parameters of each neuron in the FNN model are fixed after training. The change in weight parameters of different input parameters is due to the difference in the activation state. Therefore, we can approximate the FNN model with ReLU as the activation function as a multiple regression process with screening characteristics.

Fortunately, the linear representation of the FNN model with the ReLU function as the activation function can be realized through the above method, which can realize the interpretation of the intelligent model and greatly simplify the understanding of non-professionals on the intelligent drilling model. Through the linear characterization of the FNN model with the ReLU function as the activation function, on the one hand, we can intuitively understand the importance of feature parameters and remove the features with small contributions. On the other hand, the changing trend of the target parameters with the characteristic parameters can be analyzed to optimize the drilling parameters.

3. Data and modeling

3.1. Drilling data analysis

To avoid the low prediction accuracy and poor reliability of the ROP model caused by the small amount of drilling data, three vertical wells located in the same block of Tarim Oilfield were selected as the training data of the ROP prediction model. As shown in Fig. 2, the Neogene Kangcun Formation ($N_{1\text{-}2}k$) and Neogene Jidike Formation (N_1j) are encountered while drilling these three vertical wells. The thicknesses of Well 1, Well 2, and Well 3 in $N_{1\text{-}2}k$ are 1788, 1932, and 1656 m, respectively; The thicknesses in N1j are 1136, 788, and 716 m, respectively.

During the drilling process, the GT55 series PDC bit of DBS company and a similar bottom hole assembly (BHA) were used, and there was no abnormal drilling accident. In addition, the $N_{1\text{-}2}k$ formations of three wells are dominated by the conglomerate, with high lithological similarity. N_1j formation is mainly composed of mudstone and siltstone.

The drilling data used in this research can be divided into five categories: geological parameters, trajectory parameters, engineering parameters, drilling fluid parameters, and drilling bit parameters. However, since this well is straight, the influence of the trajectory on the ROP is small and can be ignored. The drilling parameters finally selected for intelligent prediction of ROP based on drilling mechanism analysis and expert consultation are shown in Table 1.

Geological parameters mainly include Gamma Ray (GR), Acoustic (AC), and D index (D). GR is used to divide lithology and reservoir, calculate formation mud content, and calculate the median of sandstone particle size. AC can be used for lithology analysis, porosity calculation, gas reservoir and fracture zone detection, and rock mechanical property analysis. And D can be used to detect formation pressure. Engineering parameters include Depth, Rotary Speed (RP), Weight on Bit (WOB), Torque (Tor), Stand Pipe Pressure (SPP), Inlet Flow (IF), and Hook Load (HL), which are used to characterize mechanical rock breaking and hydraulic impact during drilling. Drilling fluid parameters include Density (Den), Viscosity (Vis), and Plastic Viscosity (PV), which are mainly used to characterize the cuttings transportation capacity of drilling fluid. Drilling bit parameters include bit size and Bit Footage (BF). Due to the bit sizes of these three wells, all 333.38 mm, the bit size is not used as an input feature. The bit wear greatly impacts the ROP during drilling, and there is no doubt that the BF is closely related to the bit wear. Therefore, this paper uses the BF to characterize the bit wear.

3.2. Drilling data preprocess

Some drilling parameters are statistically analyzed to understand the drilling data information further. As shown in Table 2, the mean value, standard deviation (Std), maximum value (Max), and minimum value (Min) of feature parameters are calculated, respectively. The mean value of Depth is the largest, 3150.2, and D is the smallest, 0.83, with a difference of 3795 times. The Std reflects this parameter's data fluctuation, and the Depth fluctuation is the strongest. In addition, the maximum and minimum values of different parameters differ greatly. Therefore, before inputting the drilling data into the ROP intelligent prediction model, it is necessary to preprocess it.

Since the value ranges of different drilling parameters vary greatly if the original parameter values are directly used for analysis, the weight of the parameters with higher values in the regression calculation will be highlighted, and the role of the parameters with lower values will be weakened. Therefore, it is necessary to preprocess the drilling data before establishing the intelligent model to reduce its training time and improve its prediction accuracy. As shown in Eq. (9), through the normalization processing method, the drilling parameters can be subject to the normal distribution with a mean value of zero and a standard deviation of one.

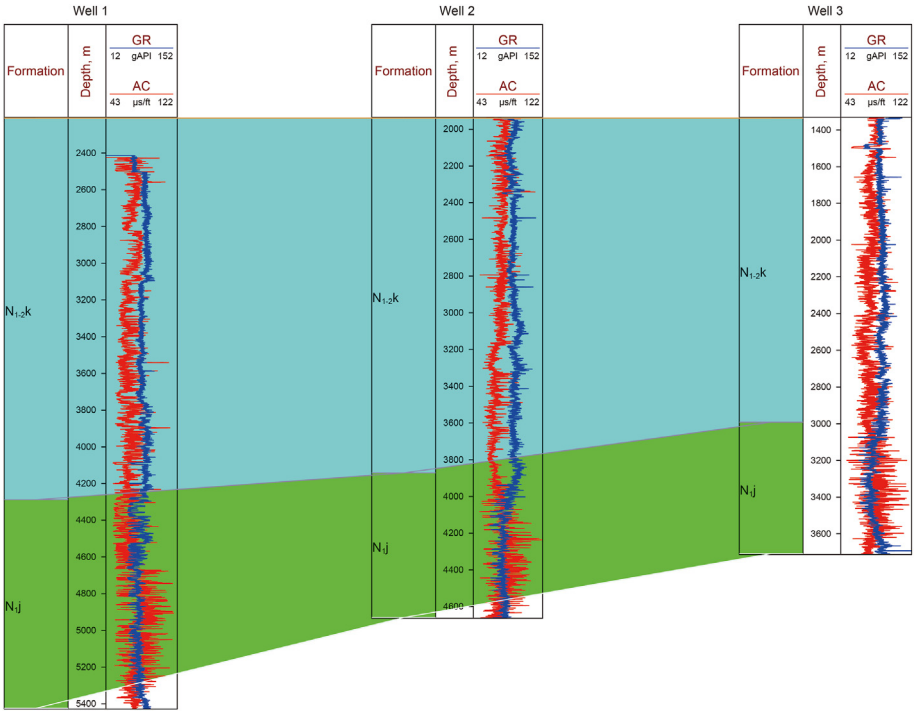

Fig. 2. Geophysical and stratigraphic column map of three wells.

Table 1
Drilling parameters.

Items	Parameter
Geological	Gamma Ray (GR), Acoustic (AC), D index (D)
Engineering	Depth, RS, WOB, Torque (Tor), Stand Pipe Pressure (SPP), Inlet Flow (IF), Hook Load (HL)
Drilling fluid	Density (Den), Viscosity (Vis), Plastic Viscosity (PV)
Drilling bit	Bit Footage (BF)

Table 2
Statistics of feature parameters.

Features	Mean	Std	Min	Max
Depth, m	3150.2	684	2000	4600
WOB, kN	78.17	26.33	15	170.4
RS, RPM	75.74	6.72	40	100
SPP, MPa	22.28	3.34	1.6	31.9
IF, L/s	126.2	318.9	48.33	3384
Tor, kN·m	12.15	3.33	3.27	24.5
HL, kN	1476	210.3	1017	1953
BF, m	173.8	122.7	0.00	574.0
GR (API)	89.02	12.46	44.18	139
AC, μs/m	74.83	8.99	49.91	117.9
D	0.83	0.13	0.5	1.21
Den, g/cm^3	1.58	0.09	1.4	1.73
Vis, s	56.79	11.44	42	109
PV, MPa·s	27.08	8.93	13	138.33
ROP, m/h	4.67	3.58	0.4	30

$$x_i' = \frac{x_i - \mu}{\sigma} \tag{9}$$

In Eq. (9), x_i' is the normalized x, μ is the average value of x, σ is the standard deviation of x.

3.3. Model establishment and evaluation

The evaluation of the FNN model should comprehensively consider the model complexity and the test set's performance. The complexity of the model refers to the number of variable parameters in the FNN model. Mean square error (MSE), mean relative error (MRE), and R^2 are used to evaluate the performance of the FNN model on the test set. In general, under the same model performance conditions, the lower the complexity of the model, the better the model.

Taking the 14 characteristic parameters shown in Table 1 as input parameters and the ROP as output parameters, an intelligent prediction model of ROP can be established based on the FNN algorithm. During the establishment of the model, the hyperparameters to be optimized mainly include the number of hidden layers, the number of neurons in each hidden layer, the activation function, the learning rate, and the optimizer.

80% of the drilling data of all formations of 3 wells are randomly selected as a training set and 20% as a test set to train and test FNN ROP prediction models under different hyperparameters. The learning rate of all the optimal ROP intelligent prediction models with different activation functions is 1×10^{-3}, and the optimizer is Adam. The network structure of neural networks can be combined in various ways. This article introduces the concept of model complexity to characterize the impact of network structure on model accuracy. The model complexity refers to the number of variables involved in the training process of a neural network model. Fig. 3 shows the change of MRE of the FNN ROP prediction model with model complexity. With the increase in model complexity, the MRE of the ROP prediction model with different activation functions first decreased rapidly and then increased slowly. This is because, in the beginning, as the complexity of the model increases, the fitting ability of the FNN model rises rapidly, resulting in a rapid reduction of MRE. All the ROP prediction models with different activation functions fully exploit the hidden information of drilling data and achieve the best fitting performance when the complexity is 10,337. Later, with the increase of complexity, the model is gradually overfitting, increasing the MRE of the model.

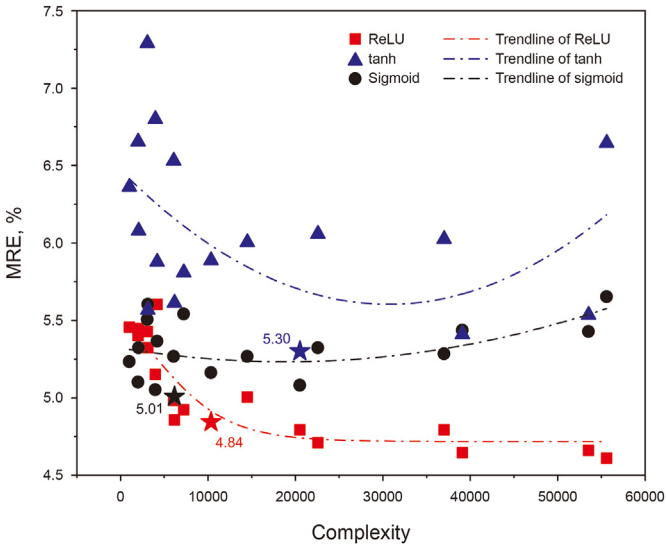

Fig. 3. Scatter diagram of model complexity and MRE with different activation functions.

The five-pointed star in Fig. 3 is the inflection point of the MRE change trend line of different activation functions. It is also the optimal model of different activation functions under the comprehensive consideration of model complexity and MRE conditions. The performance of the optimal ROP prediction model with different activation functions is shown in Fig. 4. The structure of the hidden layer of the optimal FNN ROP prediction model under the three activation functions is [64, 64, 64, 16], and the complexity is 10,337. The MSE, MRE, and R^2 of the FNN model with ReLU as the activation function are 0.38, 8.52%, and 0.97, respectively. The MSE, MRE, and R^2 of the FNN model with tanh as the activation function are 0.49, 10.22%, and 0.91, respectively. The MSE, MRE, and R^2 of the FNN model with sigmoid as the activation function are 0.42, 10.17%, and 0.94, respectively. According to Figs. 3 and 4(a), the FNN penetration rate prediction model with ReLU as the activation function has low complexity, good prediction performance, and strong stability and is the optimal activation function.

In addition, this paper also established the ROP prediction model based on the Bourgoyne and Young (B&Y) equation (Bourgoyne and Young, 1974) and the RF algorithm and compared it with the FNN model. As shown in Fig. 4(b), the performance of the FNN model is significantly better than other models.

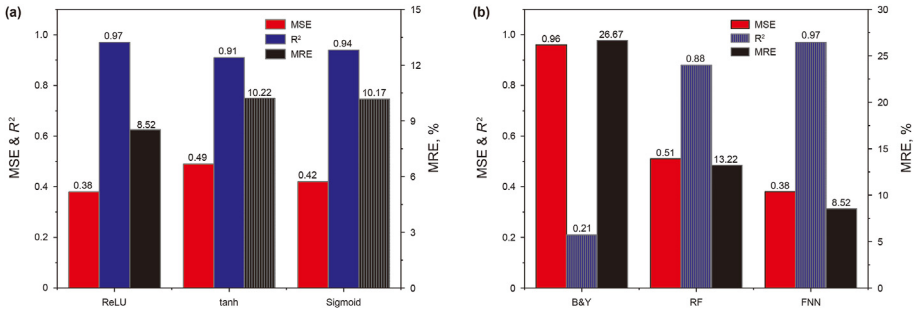

Fig. 4. Comparison of performance of different ROP prediction models. (**a**) FNN model with different activation functions, (**b**) different ROP prediction models.

3.4. Model error analysis

The FNN ROP prediction model with ReLU as the activation function was established using the optimal hyperparameters, and the drilling data of Well 1 and Well 2 were used as the training set to predict the ROP of Well 3. According to Fig. 5, the MRE of Well 3 is 7.95%, which is lower than that of the three wells. In addition, the MRE of the Kangcun Formation is 8.63%, and the MRE of the Jidik Formation is 6.63%. The MRE of the Jidike Formation is lower than the average level, while the MRE of the Kangcun Formation is higher than the average of the whole well. Nevertheless, the MREs of the established ROP prediction model of different formations and different wells are not very different.

Fig. 6 shows the error distribution of the ROP prediction model in Well 3. The error distribution of ROP prediction in different formations tends to be normal, and 90% of the error distribution is in the range of −0.58 to 0.26. The error distribution conforms to the performance of random error, which is also evidence that the model has fully mined the hidden information of drilling data, indicating that the prediction effect of the FNN model is good.

Fig. 7 shows the change of MRE with well depth, and the red curve is the curve after fast Fourier transform (FFT) smoothing. Fig. 8 compares the predicted and raw values of ROP. At some well depths, such as 2268 and 2671 m, the MRE of ROP increases sharply, decreases to the normal level, and fluctuates normally. According to Figs. 7 and 8, a targeted analysis of the mutation point at 2268 m shows that the actual value of ROP at the mutation point is 2.61 m/h, the predicted value is 3.66 m/h, the error is 1.05, and the relative error is 47.9%. Although the prediction error of ROP has increased to some extent comparable to the adjacent well point, the change of MRE is too big. This is mainly because the actual value of ROP at this point is small, and under the same error condition, the MRE is larger, which leads to a sharp mutation of the MRE curve. This can also explain why the MRE of the Kangcun Formation is greater than that of the Jidik Formation.

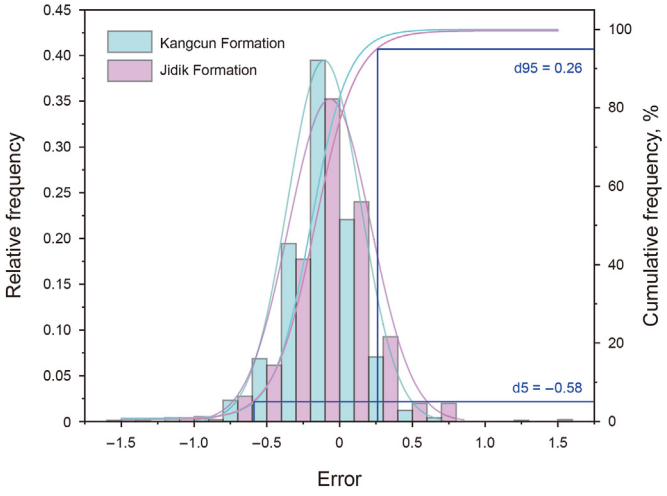

Fig. 6. Error distribution of the FNN model.

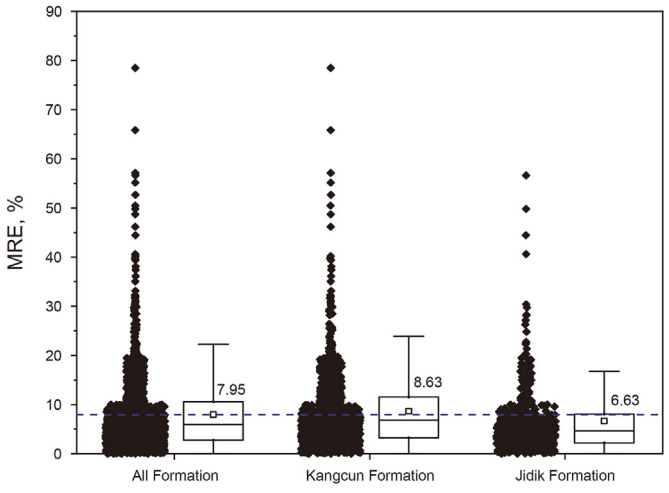

Fig. 5. MRE distribution of the FNN model.

4. Model interpretation and characterization

The training process of the FNN model is the process of error backpropagation. With continuous training, the model's parameters will gradually converge, and the model will slowly transition from the under-fitting state to the perfect-fitting state. However, when we improve the tolerance of the FNN model to errors, the model error and MSE can be further reduced. At this time, the model is often in a state of over-fitting. To analyze the activation state of the FNN model in different model performances, the ROP prediction models with MRE of 5%, 10%, 15%, and 20% were established, respectively. The performance of ROP prediction models of different MREs on Well 3 is shown in Table 4. The maximum differences of MSE, MRE, and R^2 in different formations under different performances are 0.14, 3.61%, and 0.28, respectively. Based on the error analysis in Section 3.4.2, we can conclude that the prediction effects of ROP in different formations are excellent, and the performances of the model in different layers are small.

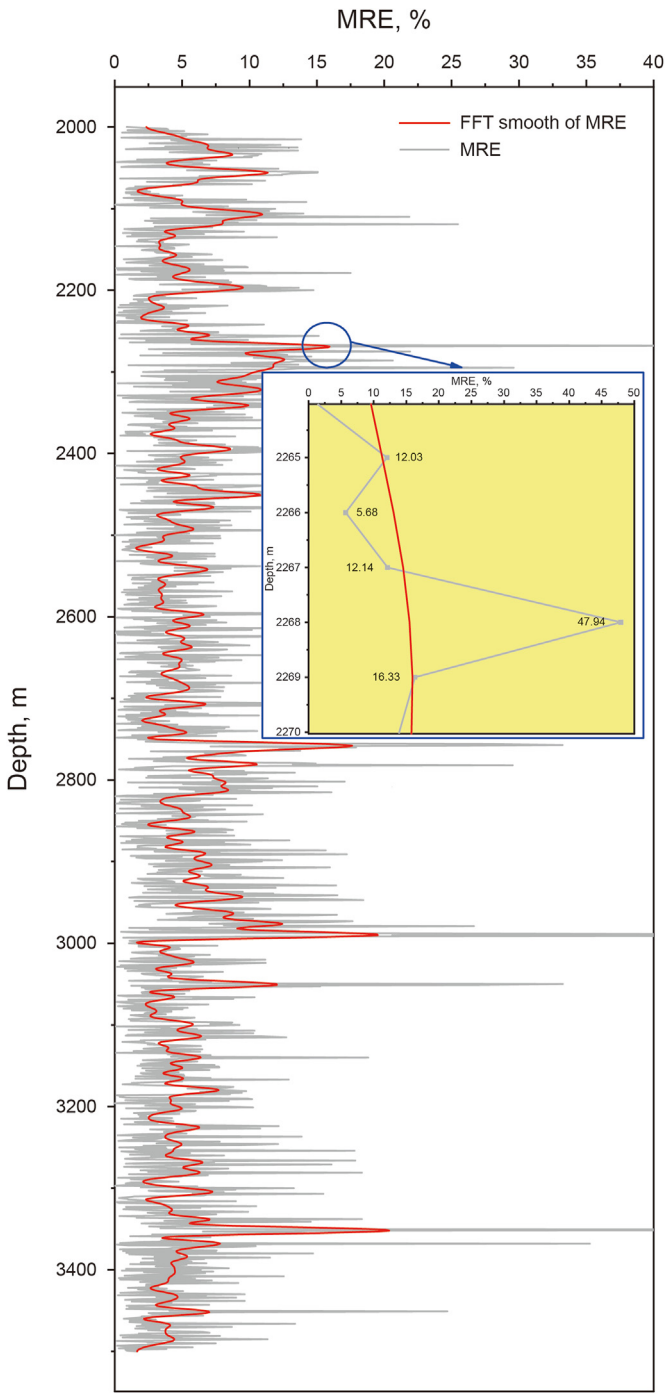

Fig. 7. Change of MRE with well depth.

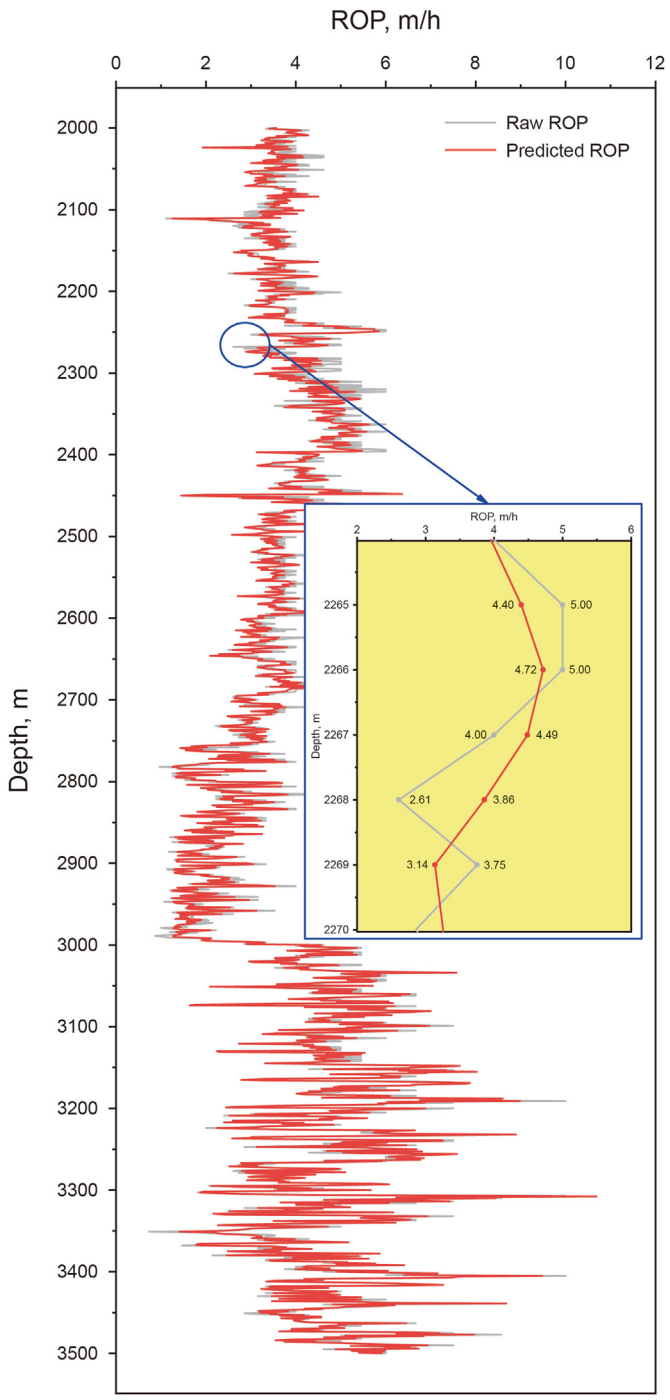

Fig. 8. Comparison between predicted ROP and raw ROP.

4.1. Neuronal activation interpretation

4.1.1. Neuronal activation of different MREs

Input the drilling data of Well 3 into the FNN ROP prediction model of different MREs on the whole drilling data, and the activation status of neurons in the hidden layer corresponding to each group of drilling data of Well 3 can be obtained. Although we cannot clearly explain the meaning of each neuron, we can perform statistical analysis on the activation status of neurons. As shown in Fig. 9, the activation times of each neuron in the hidden layer are counted, and the relative activation frequency of each neuron is calculated. The activation curve of each hidden layer of the FNN ROP prediction model for different MREs can be obtained by ascending the order of relative activation frequency.

According to Fig. 9, with the continuous reduction of the MRE of the model, the relative activation frequency curve of each hidden layer gradually approximates the activation curve when the MSE is 10%, and the relative activation frequency curves of each hidden layer slowly tend to a stable and smooth state. Continue to increase the MRE of the model, and the activation curves have changed significantly and violently. And the activation frequency curve when MRE is 5% is substantially different from that when MRE is

Table 4
FNN ROP prediction model performance of some hyperparameters.

MRE on three wells	Well 3			Kangcun Formation of Well 3			Jidik Formation of Well 3		
	MSE	MRE, %	R^2	MSE	MRE, %	R^2	MSE	MRE, %	R^2
5%	0.21	5.50	0.96	0.20	5.90	0.93	0.21	4.73	0.96
10%	0.38	10.36	0.87	0.33	10.12	0.82	0.47	10.83	0.83
15%	0.59	15.62	0.69	0.58	16.52	0.47	0.61	13.85	0.72
20%	0.83	21.82	0.43	0.81	23.04	0.24	0.85	19.43	0.52

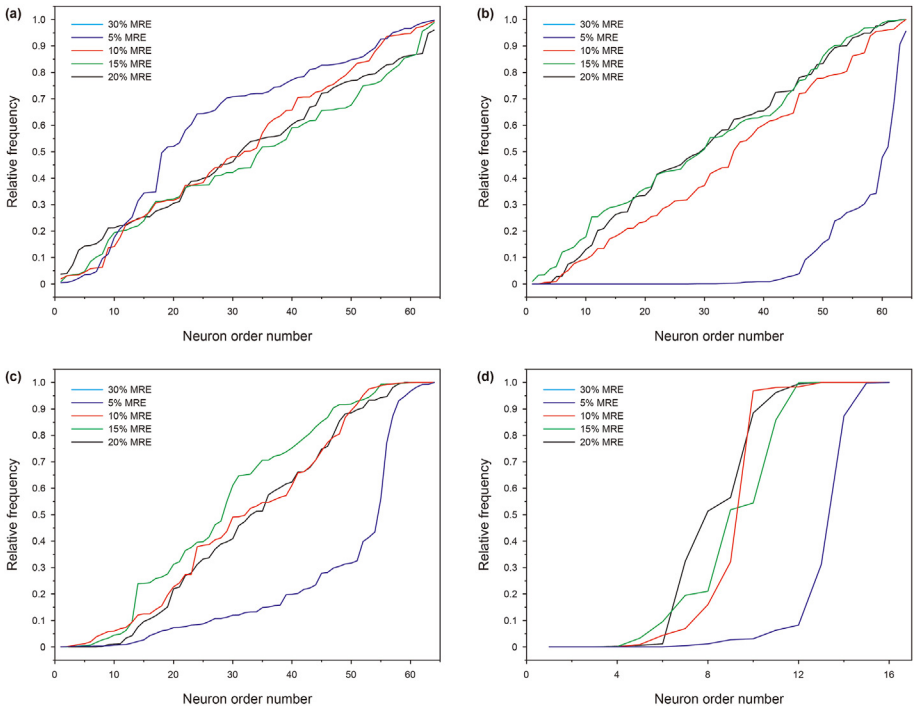

Fig. 9. The relative neuron activation frequency of the four hidden layers. (**a**) The relative neuron activation frequency of hidden layer 1, (**b**) the relative neuron activation frequency of hidden layer 2, (**c**) the relative neuron activation frequency of hidden layer 3, (**d**) the relative neuron activation frequency of hidden layer 4.

10%, 15%, and 20%. In addition, it is also found that when the MRE is higher than 10%, the relative activation frequency curves of hidden layers 1, 2, and 3 are linear. However, when the MRE is 5%, the relative activation frequency curves of hidden layers 1, 2, and 3 are S-shaped. For hidden layer 4, different MREs' relative activation frequency curves are all S-shaped.

Fig. 10 shows neurons' relative activation frequency distribution in each hidden layer when MRE is 5%. For hidden layer 1, the relative activation frequency of 46% of neurons was greater than 0.7; For hidden layer 2, the relative activation frequency of 89% of neurons was less than 0.3; For hidden layer 2, the relative activation frequency of 82% of neurons was less than 0.35; For hidden layer 3, the relative activation frequency of 89% of neurons is less than 0.3.

Fig. 11 shows neurons' relative activation frequency distribution in each hidden layer when MRE is 10%. The distribution of the relative activation frequency of neurons in hidden layers 1, 2, and 3 is relatively uniform. When the cumulative frequency is 50%, the relative activation frequency of hidden layers 1, 2, and 3 is 0.48, 0.4, and 0.4, respectively. However, for hidden layer 4, 73% of neurons, the relative activation frequency is less than 0.35. The distribution of neuron relative activation frequency in hidden layer 4 is significantly different from the other three hidden layers, possibly due to the small number of neurons in hidden layer four and a certain statistical error.

Based on Figs. 10 and 11, the relative activation frequency curve of S-type represents that the distribution of relative activation frequency of neurons is centralized, and the relative activation frequency of most neurons is in a small relative frequency range. The linear relative activation frequency curve represents that the distribution of the relative activation frequency of neurons is uniform, and the number of neurons with different relative activation frequencies is equal. The activation of each neuron in the hidden layer in the FNN ROP model is a process of screening the drilling data. The neuron with low relative activation frequency means that the neuron has only activated a few drilling data, and only a few drilling data meet the screening conditions of this neuron. On the contrary, the neuron with high activation frequency means that the neuron has activated most of the drilling data, and most of the drilling data meets the screening conditions of this neuron, which also means that the neuron is screening a common feature.

Based on the above analysis, combined with the study of the weight parameters of different MREs in Section 4.2.1, when the MRE is 5%, the relative activation frequency of most neurons is low, and the number of neurons used for mining common features is small. Although the accuracy of the FNN ROP prediction model at this time is higher, the model is overfitting, and the stability and mobility of the model are low. When the MRE of the model is 10%, the number of neurons used to capture different features is evenly

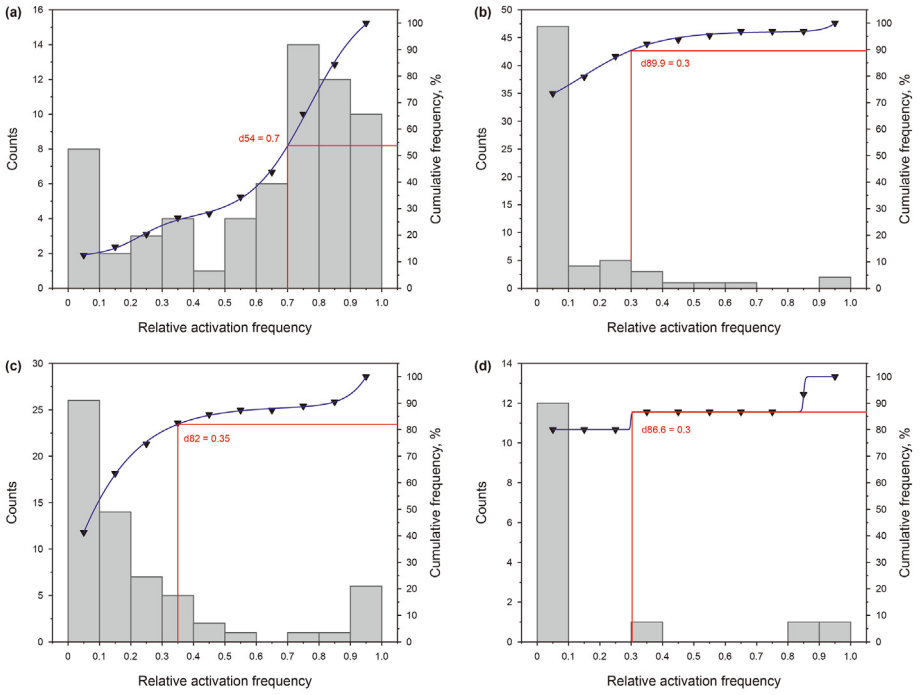

Fig. 10. Relative activation frequency distribution of neurons in each hidden layer when MRE is 5%. (**a**) Relative activation frequency distribution of neurons in hidden layer 1, (**b**) relative activation frequency distribution of neurons in hidden layer 2, (**c**) relative activation frequency distribution of neurons in hidden layer 3, (**d**) relative activation frequency distribution of neurons in hidden layer 4.

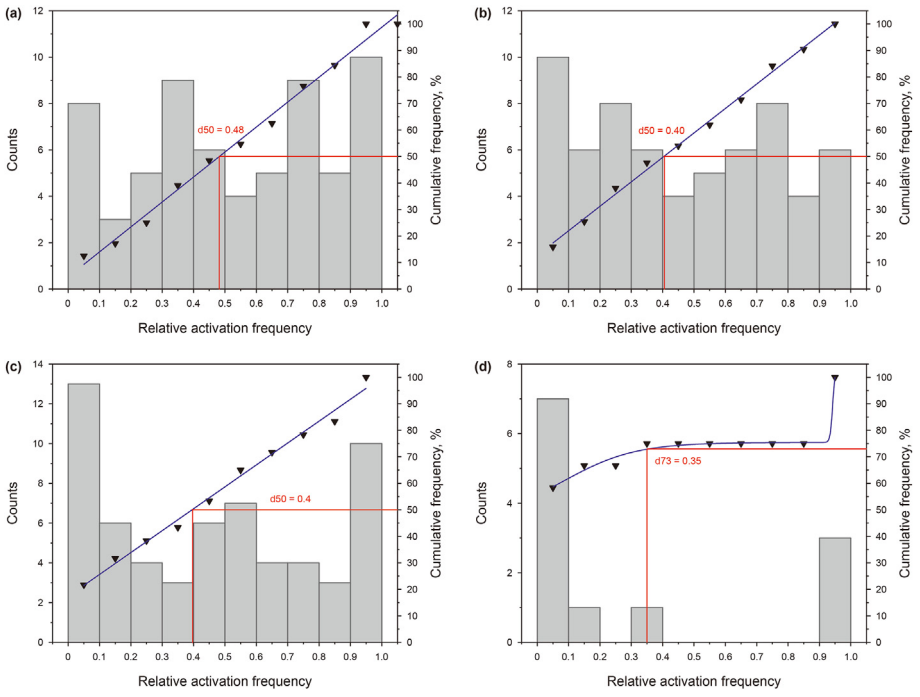

Fig. 11. Relative activation frequency distribution of neurons in each hidden layer when MRE is 10%. (**a**) Relative activation frequency distribution of neurons in hidden layer 1, (**b**) relative activation frequency distribution of neurons in hidden layer 2, (**c**) relative activation frequency distribution of neurons in hidden layer 3, (**d**) relative activation frequency distribution of neurons in hidden layer 4.

distributed, and the screening characteristics of neurons are fully utilized. At this time, the model is in a convergent state, which can improve the stability and mobility of the model under the condition that the accuracy meets the application requirements.

4.1.2. *Neuronal activation of different formations*

According to Fig. 12, when the MRE of the FNN ROP prediction model is 10%, the relative activation frequency curves of four hidden layers are S-shaped, and the distribution of neuron relative activation frequency is centralized. Compared with Fig. 9, The

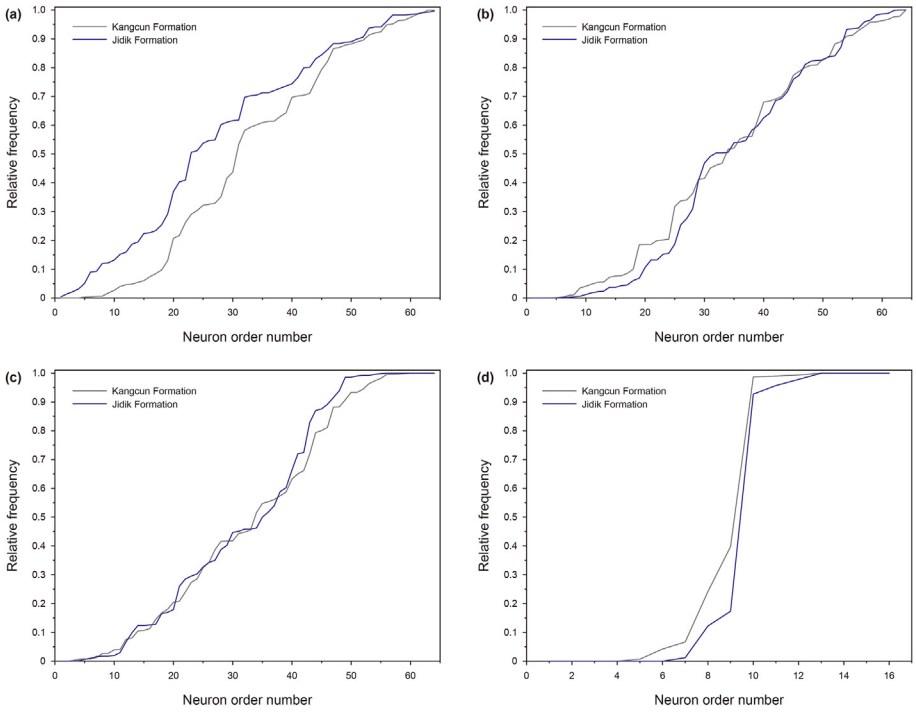

Fig. 12. The relative neuron activation frequency in four hidden layers of Kangcun Formation and Jidik Formation. (**a**) The relative neuron activation frequency of hidden layer 1, (**b**) the relative neuron activation frequency of hidden layer 2, (**c**) the relative neuron activation frequency of hidden layer 3, (**d**) the relative neuron activation frequency of hidden layer 4.

distribution of relative neuron activation frequency in hidden layers 1, 2, and 3 of all formations is uniform, which means that the same neuron in the hidden layer has an obvious relative activation frequency difference on the drilling data of different formations. The FNN ROP prediction model established has good distinction and fits the drilling data of different formations.

4.1.3. Similar neuronal activation

The relative activation frequency curve of each hidden layer on the drilling data with well depth from 3255 to 3266 m is shown in Fig. 13. Compared with the S-shape relative activation frequency curve of Jidik Formation in Fig. 12, this section's relative activation frequency curve is Z-shape. The activation state of neurons with relative activation frequency greater than 0 and less than 1 in the drilling data of this well section is not completely consistent. The number of neurons with relative activation frequency greater than 0 and less than 1 in each hidden layer is the projection length of the transition curve of relative activation frequency from 0 to 1 on the *x*-axis, which we can call the transition segment length. When the activation state of neurons in a group of data is completely consistent, the relative activation curve of each hidden layer is also Z-shape, and the length of the transition segment is 0. Therefore, the smaller the length of the transition segment, the smaller the number of neurons with differences, and the more similar the activation state of neurons corresponding to this group of data.

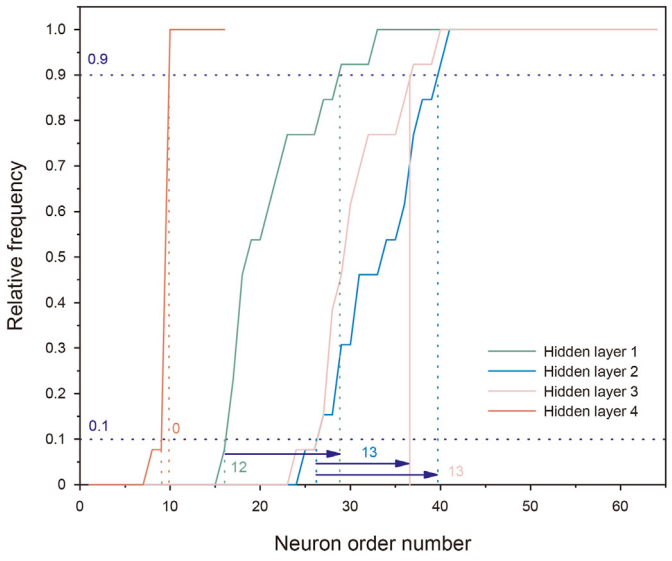

Fig. 13. The neuron activation frequency of some drilling data with similar neuron activation states.

Considering the measurement error of drilling data, when the relative activation frequency of neurons is greater than 0.9 or less than 0.1, it is assumed that the difference in the activation state of neurons in this group of drilling data can be ignored, and the length of transition section is not included. In the well section from 3255 to 3266 m, the length of transition segments of hidden layers 1, 2, 3, and 4 are respectively 12, 13, 13, and 0, accounting for 18.72%, 20.31%, 20.31%, and 0% of the total number of neurons in each hidden layer. In general, the activation difference of drilling data in this section is small, and the activation status has a certain similarity.

The statistics of the drilling data in this section are shown in Table 5. The distribution of drilling features can be known from the mean value and standard deviation (Std). This well section's most volatile drilling feature is Tor, with the mean and variance of 11.81 and 0.86, respectively, and the Std accounting for only 7.3% of the mean value. In addition, according to cuttings logging records, the lithology of the formation in this section is brown argillaceous siltstone, and the lithology also has high similarity. We can also

Table 5
Drilling data with similar activation.

No.	Depth, m	WOB, kN	RP, RPM	SPP, MPa	IF, L/s	Tor, kN·m	HL, kN	BF, m	GR, API	AC, μs/m	D	Den, g/cm^3	Vis, s	PV, MPa·s
1	3255	116	74	21.1	49.0	11.9	1436.9	204.7	72.5	73.8	0.8	1.7	69.4	25.9
2	3256	113	74	20.3	49.1	13.2	1435.4	205.7	74.4	79.7	0.8	1.7	69.5	26.0
3	3257	115	74	20.2	49.2	13.4	1437.5	206.7	79.6	80.6	0.8	1.7	69.6	26.0
4	3258	114	74	20.3	49.3	12.2	1436.2	207.7	76.3	84.1	0.8	1.7	69.7	26.0
5	3259	115	74	20.3	49.2	11.0	1435.1	208.7	80.4	89.7	0.8	1.7	69.8	26.0
6	3260	116	74	20.3	49.1	10.8	1436.6	209.7	80.9	86.3	0.8	1.7	69.9	26.0
7	3261	112	74	20.3	49.2	12.3	1439.3	210.7	80.8	84.8	0.8	1.7	70.0	26.0
8	3262	111	74	20.3	49.1	12.1	1440.3	211.7	83.8	81.1	0.8	1.7	69.9	26.0
9	3263	112	74	20.6	49.4	12.0	1438.7	212.7	87.5	82.1	0.8	1.7	69.8	26.0
10	3264	112	74	20.7	49.6	11.6	1436.8	213.7	84.5	78.8	0.8	1.7	69.7	26.0
11	3265	114	74	20.7	49.6	11.2	1437.7	214.7	89.3	84.4	0.8	1.7	69.6	25.9
12	3266	112	74	20.6	49.6	11.2	1440.4	215.7	78.0	82.7	0.8	1.7	69.6	25.9
13	3267	112	74	20.4	49.5	10.6	1440.2	216.7	76.8	75.3	0.9	1.7	69.5	25.9
Mean	3261	113.4	74	20.47	49.3	11.8	1437.8	210.7	80.4	81.8	0.8	1.7	69.7	25.9
Std	3.89	1.71	0	0.26	0.23	0.86	1.84	3.91	4.95	4.35	0.03	0	0.19	0.03

infer that the similarity of neuron activation state is fundamentally derived from the resemblance of drilling parameters.

4.2. Weight parameters interpretation

In the process of establishing the FNN ROP prediction, the input parameters of the model are normalized. Therefore, the weight parameters of the model need to be modified to obtain the actual weight of the drilling parameters. The correction method is shown in Eqs. (10) and (11).

$$w_i = \frac{w_i'}{\sigma_i} \tag{10}$$

$$b = b' + \sum_{i=1}^{n} \frac{-w_i'\mu_i}{\sigma_i} \tag{11}$$

In Eqs. (10) and (11), w_i' is the weight parameter of the FNN model, b' is of the FNN model, σ_i is the standard deviation of x_i , μ_i is the average value of x_i, w_i is the corrected weight parameter, and b is the fixed bias.

4.2.1. Weight parameters of different MREs and formations

Fig. 14 shows the change of weight parameters of WOB and RS with well depth after FFT smoothing. Like the rule of relative activation frequency, with the reduction of model MRE, the changing trend of feature weight parameters tends to be smooth and stable at 10% MRE and continues to reduce MRE, and feature weight parameters change dramatically.

For WOB, generally speaking, WOB and ROP are positively correlated. When MRE exceeds 10%, only a few data weight parameters are less than 0. In comparison, when MRE is 5%, many drilling data are negatively correlated to ROP, which is inconsistent with existing physical principles. For RS, the weight parameters of different MREs are the most distinct in the depth interval of 2700 and 2800 m. However, the RS in this drilling interval has not changed, so the weight parameters should not be greatly reduced under normal fluctuation, which is not in line with reality. Based on the above analysis, we can be sure that the established FNN ROP prediction model achieves a good fitting state when the MRE is 10%, and the model is over-fitting when the MRE is 5%.

Further, analyze the FNN ROP prediction model with MRE of 5% and 10%, and the statistical information of the correction weight parameters of the FNN ROP prediction model is shown in Table 6. The MRE of the model and the classification of the formation have a great influence on the weighting parameters of the drilling characteristics. When the MRE of the FNN ROP prediction model is 5%, on the drilling data of all formations, the input features with the largest fluctuation of the weight parameters are GF, BF, and AC. Their mean values are 1.93×10^{-4}, 4.25×10^{-3}, and 2.95×10^{-4}. The standard deviation is 10.78×10^{-4}, $17.55 \times 10^{-3,}$ and 10.34×10^{-4}, respectively, and the standard deviation is 5.59, 4.13, and 3.51 times the mean value. When the MRE of the FNN ROP prediction model is 10%, on the drilling data of all formations, the input features with the largest fluctuation of the weight parameters are GF, BF, and AC. Their mean values are -2.11×10^{-4}, 0.69×10^{-3}, 4.13×10^{-4}, and the standard deviation is 21.34×10^{-4}, 29.29×10^{-3} and 19.54, respectively. The standard deviation is 10.09, 42.35, and 19.54 times the mean value.

When the weight parameters of input features are counted according to the classification of formations, the variance of the weight parameters of most features is reduced, and the variance of only a few feature parameters is increased. Fig. 15 compares the mean weight parameters of different MRE FNN ROP prediction models on different formations. In the Kangcun Formation, the weight parameters of HL, Den, and Vis are quite different. When the MRE of the model is 5% and 10%, the correlation between Den and ROP is negative and positive, and between Vis and ROP is positive and negative. The correlation between these two parameters of the model under different MRE and ROP is opposite. In the Jidikd Formation, under different MRE conditions, the weight parameters of IF, Tor, Den, and Vis vary greatly, and the correlation between all features and ROP is consistent. In Kangcun Formation, the number of features with a large difference in weight parameters of the ROP prediction model with 5% and 10% MRE is small, but the correlation between a few features and ROP has changed. In the Jidik Formation, the correlation between all features and ROP is consistent.

4.2.2. Weight parameters of similar neuronal activation

When the activation state of neurons in a group of data is similar, according to Eq. (8), the weight parameters of these data also have a greater similarity. Table 7 shows the statistical information of weighting parameters of drilling data with similar activation status. The maximum fluctuation of weighting parameters is BF, and the minimum is SPP. The standard deviation accounts for 77.04% and 10.66% of the mean value, respectively. Most drilling characteristics' standard deviation accounts for less than 50% of the average value. The fluctuation of weighting parameters is small, and the similarity is high.

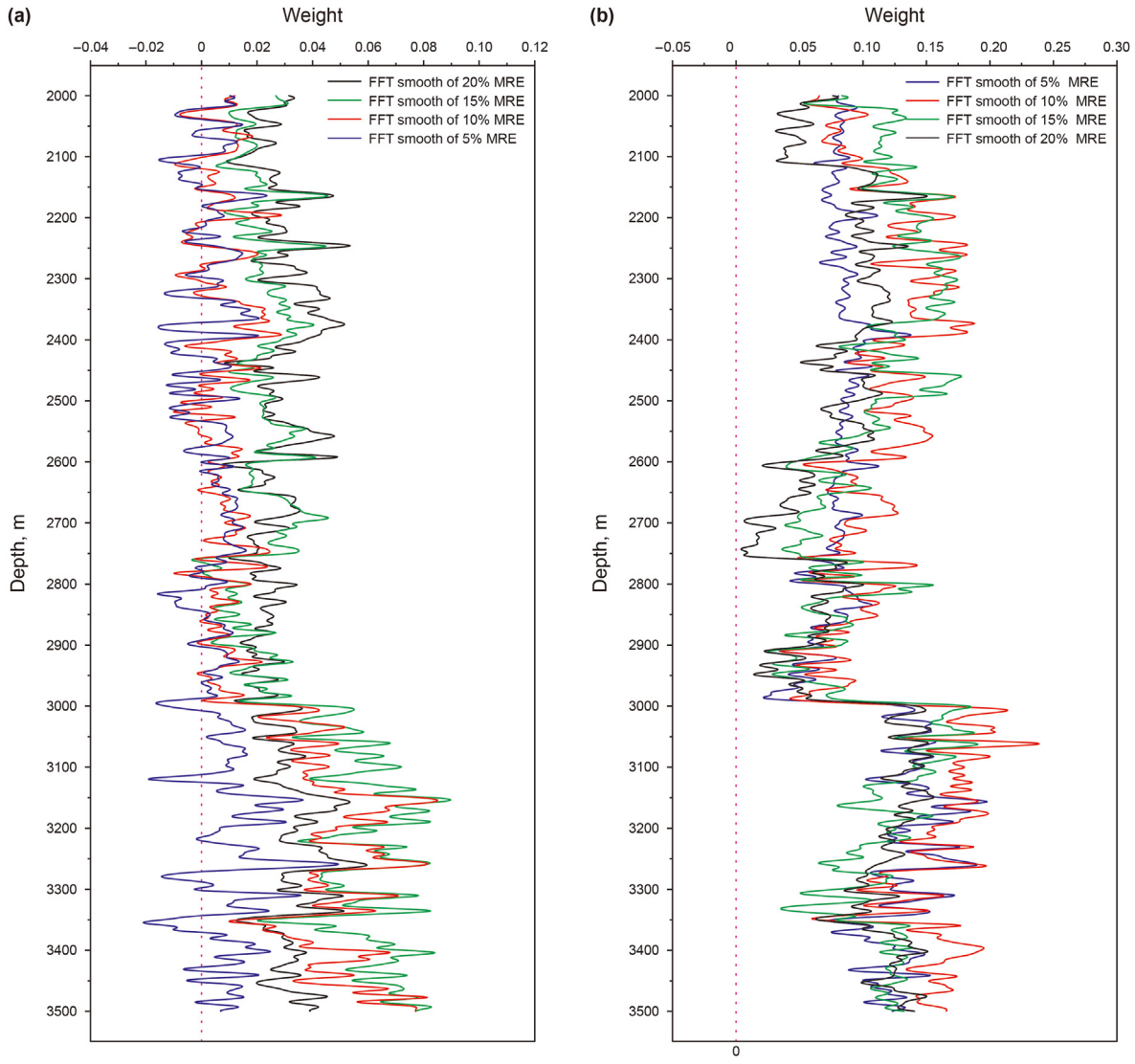

Fig. 14. The change of weight parameters with well depth after FFT smoothing. (**a**) WOB weight parameters, (**b**) RS weight parameters.

Table 6
Statistics of weight parameters of FNN ROP prediction model.

Items	5% MRE						10% MRE					
	All formations		Kangcun		Jidik		All formations		Kangcun		Jidik	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Depth ($\times 10^{-4}$)	−7.80	7.57	−3.04	1.04	−17.10	6.03	−8.68	3.99	−7.70	3.16	−10.58	4.70
WOB ($\times 10^{-3}$)	5.29	14.88	3.21	11.6	9.37	19.13	20.73	25.23	6.87	12.39	47.82	21.71
RP ($\times 10^{-2}$)	9.85	3.95	8.03	2.27	13.39	4.13	12.73	4.97	11.04	4.29	16.04	4.54
SPP ($\times 10^{-2}$)	−9.97	4.83	−8.73	3.05	−12.38	6.47	−20.90	8.09	−17.88	6.81	−26.81	7.07
IF ($\times 10^{-4}$)	−50.19	17.09	−41.87	13.5	−66.44	10.48	−43.04	11.38	−43.17	12.43	−42.78	8.98
Tor ($\times 10^{-2}$)	7.06	7.82	9.70	7.55	1.91	5.39	−5.78	12.67	1.76	6.25	−20.52	8.31
HL ($\times 10^{-4}$)	−28.17	11.74	−27.04	9.19	−30.38	15.34	−43.98	17.81	−50.14	16.76	−31.96	13.04
BF ($\times 10^{-4}$)	4.25	17.55	13.34	11.4	−13.52	13.42	0.69	29.29	13.87	17.22	−25.07	30.87
GR ($\times 10^{-3}$)	1.93	10.78	6.11	8.28	−6.24	10.42	−2.11	21.34	6.97	17.55	−19.86	16.33
AC ($\times 10^{-3}$)	2.95	10.34	1.44	7.66	5.91	13.73	4.13	19.54	5.82	17.35	0.81	22.89
D ($\times 10^{0}$)	−20.25	10.99	−14.64	3.70	−31.22	12.17	−21.53	13.34	−14.68	6.47	−34.92	13.15
Den ($\times 10^{-1}$)	−10.74	32.13	−21.47	22.9	10.24	36.78	43.56	36.81	35.34	33.59	59.61	37.55
Vis ($\times 10^{-3}$)	38.20	22.92	25.85	13.3	62.34	17.98	−25.46	46.14	−55.37	17.98	33.01	22.07
PV ($\times 10^{-2}$)	−18.85	5.22	−17.35	5.12	−21.78	4.04	−20.38	6.32	−18.84	6.18	−23.38	5.47
Bias ($\times 10^{0}$)	24.34	11.47	20.40	6.27	32.03	14.92	22.01	10.32	19.11	6.73	27.67	13.33

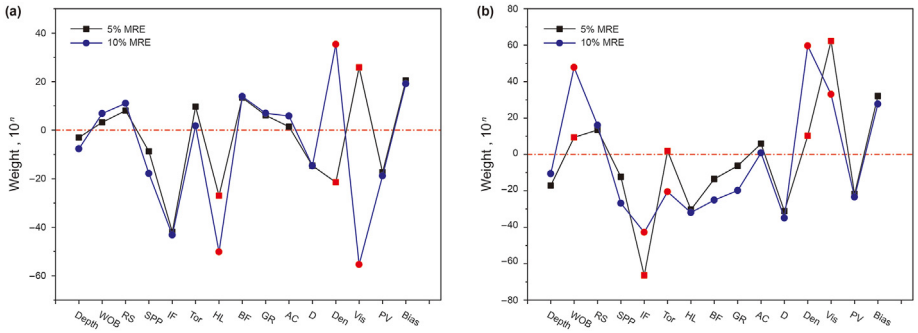

Fig. 15. Comparison of mean weight parameters under different MREs. (**a**) Mean weight parameters of Kangcun Formation, (**b**) mean weight parameters Jidik Formation.

Table 7
Statistics table of weights of drilling features with similar activation.

Features	Mean	Std	100 Std/Mean
Depth ($\times 10^{-4}$)	−4.62	1.96	42.51
WOB ($\times 10^{-3}$)	7.69	1.60	20.76
RP ($\times 10^{-2}$)	18.69	2.19	11.71
SPP ($\times 10^{-2}$)	−27.02	2.88	10.66
IF ($\times 10^{-4}$)	−39.96	6.90	17.28
Tor ($\times 10^{-2}$)	−7.81	6.17	79.06
HL ($\times 10^{-4}$)	−20.60	7.35	35.71
BF ($\times 10^{-4}$)	15.74	12.13	77.04
GR ($\times 10^{-3}$)	−2.11	1.27	60.29
AC ($\times 10^{-3}$)	−1.39	1.12	80.91
D ($\times 10^{0}$)	−53.92	11.16	20.69
Den ($\times 10^{-1}$)	15.95	4.84	30.34
Vis ($\times 10^{-3}$)	34.93	18.43	52.76
PV ($\times 10^{-2}$)	−21.96	2.98	13.59
Bias ($\times 10^{0}$)	41.00	15.28	37.26

4.3. Model characterization

4.3.1. Characterization of similar neuronal activation

According to the mean value of the weight parameters of the drilling features in Tables 7 and in the drilling data of similar neuronal activation of the section from 3255 to 3266 m, since the variance of the weighting parameters of these drilling data is small and the distribution of weights is relatively stable, the mean weight value of each characteristic weighting parameter can be used to measure the impact of the features on the ROP. The ROP calculation equation based on the FNN ROP prediction model in this section is established as follows:

$$\begin{aligned}\mathrm{ROP} = &-0.0005\mathrm{Depth} + 0.0769\mathrm{WOB} + 0.1869\mathrm{RS} - 0.2702\mathrm{SPP}\\ &- 0.004\mathrm{IF} - 0.0781\mathrm{Tor} - 0.0021\mathrm{HL} + 0.0016\mathrm{BF}\\ &- 0.0211\mathrm{GR} - 0.0139\mathrm{AC} - 53.92\mathrm{D} - 1.595\mathrm{Den}\\ &+ 0.035\mathrm{Vis} - 0.2196\mathrm{PV} + 41\end{aligned} \tag{12}$$

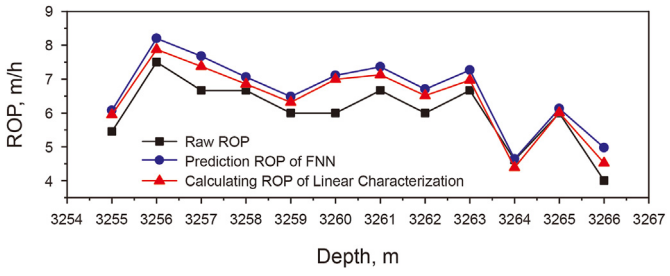

Fig. 16. Comparison of ROP prediction of similar neuron activation states drilling data.

Table 8
Comparison of ROP prediction performance of similar neuron activation states drilling data.

Item	MSE	MRE, %	R^2
FNN	0.45	10.53	0.68
Linear characterization	0.24	7.28	0.83

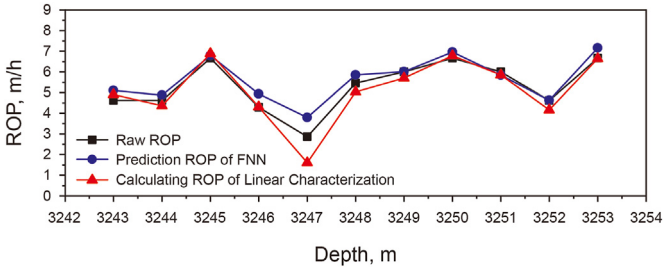

Fig. 17. Comparison of ROP prediction of the verification data.

Equation (12) is this section's linear characterization of the FNN ROP prediction. Fig. 16 is a comparison diagram of the raw ROP, the predicted ROP of the FNN model, and the calculating ROP of the linear characterization equation. Through linear characterization, the ROP prediction results in this section are still highly consistent with the raw ROP change trend.

Table 8 shows the performance of the FNN model and linear characterization equation in this section. Compared with the FNN model, the MSE and MRE of the linear characterization calculation equation in this well section decreased by 0.21, 3.25%, and R^2 increased by 0.15, which shows that the established linear characterization equation has the characteristics of strong interpretability and high accuracy.

4.3.2. Characterization application and verification

According to the linear characterization equation, the main influencing factors of ROP can be analyzed according to the weight parameters of the input features. However, because the value ranges of different drilling parameters are different, conducting a comprehensive analysis in combination with the value ranges and weight parameters of the drilling features is necessary. In this

Table 9
Comparison of ROP prediction performance of the verification data.

Item	MSE	MRE, %	R^2
FNN	0.19	7.97	0.86
Linear characterization	0.21	7.86	0.85

Table 10
Comparison of ROP prediction performance of the other six sections.

No.	Characterization equation	Depth interval	Lithology	MSE	MRE, %
1	Establishment	2020–2029 m	Argillaceous siltstones	0.28	12.37
	Verification	2030–2039 m		0.56	17.67
	ROP increasing equation	$\Delta ROP = -0.0016\Delta WOB + 0.0922\Delta RS - 0.1469\Delta SPP - 0.0032\Delta IF$			
2	Establishment	2320–2329 m	Glutenite	0.50	10.76
	Verification	2330–2339 m		0.52	12.28
	ROP increasing equation	$\Delta ROP = -0.0019\Delta WOB + 0.1561\Delta RS - 0.1825\Delta SPP - 0.0052\Delta IF$			
3	Establishment	2350–2359 m	Glutenite	0.14	5.22
	Verification	2360–2369 m		0.08	4.92
	ROP increasing equation	$\Delta ROP = 0.0197\Delta WOB + 0.1345\Delta RS - 0.209\Delta SPP - 0.0061\Delta IF$			
4	Establishment	3033–3042 m	Argillaceous siltstones	0.30	8.57
	Verification	3053–3062 m		0.39	12.25
	ROP increasing equation	$\Delta ROP = 0.05\Delta WOB + 0.2095\Delta RS - 0.1955\Delta SPP - 0.0044\Delta IF$			
5	Establishment	3153–3162 m	Siltstone	0.32	7.51
	Verification	3163–3172 m		0.64	13.73
	ROP increasing equation	$\Delta ROP = 0.084\Delta WOB + 0.1913\Delta RS - 0.2451\Delta SPP - 0.0048\Delta IF$			
6	Establishment	3273–3282 m	Argillaceous siltstones	0.29	11.49
	Verification	3283–3292 m		0.03	4.12
	ROP increasing equation	$\Delta ROP = 0.0435\Delta WOB + 0.1228\Delta RS - 0.2596\Delta SPP - 0.0034\Delta IF$			

paper, the weight parameter is multiplied by the mean value of the feature parameter to determine the importance of the feature parameter. According to Tables 5 and 7, The importance order of features is RP, Den, Vis, WOB, BF, AC, GR, IF, Tor, Depth, HL, SPP, and Den.

According to the cuttings logging data, the lithology of the formation with the well depth from 3243 to 3253 m and the lithology of the formation with the well depth from 3255 to 3266 m are brown argillaceous siltstones. To further verify the linear characterization equation's performance, taking the formation's drilling data with the well depth from 3243 to 3253 m as the verification data, the ROP is calculated using Eq. (12) and compared with the FNN ROP prediction. Fig. 17 compares the ROP of different models on the verification data, and Table 9 shows the performance of different models on the verification data.

On the verification data, the calculation ROP result of the linear characterization equation is still consistent with the changing trend of raw ROP. Compared with the FNN model, the MSE of the linear characterization calculation equation on the verification data increased by 0.02, and the MRE and R^2 decreased by 0.11% and 0.15, which proves that the established linear characterization equation has high accuracy, strong stability, and strong mobility.

In the drilling process, the features that can be controlled in real time are WOB, RS, SPP, and IF. According to Eq. (12), we can also establish the ROP increasing equation shown in Eq. (13) to provide a basis for optimizing drilling parameters. According to Eq. (13), in the process of drilling from 3242 to 3266 m, when the values of WOB, RS, SPP, and IF increase by 1, the ROP increases by 0.0769, 0.1869, −0.2702, −0.0004. Therefore, we should properly increase the WOB and RS and reduce the SPP and IF to improve the ROP, shorten the drilling cycle, and reduce the drilling cost.

$$\Delta ROP = 0.0769\Delta WOB + 0.1869\Delta RS - 0.2702\Delta SPP - 0.004\Delta IF \quad (13)$$

Table 10 shows the results of linear characterization and verification of the FNN ROP prediction model on the drilling data of the other six sections. The results show that the linear characterization method proposed in this paper can realize the interpretation of the FNN ROP model based on ensuring the performance of ROP prediction, and the established linear characterization equation has strong stability and mobility.

5. Conclusions

An interpretation method of the FNN model with ReLU as the activation function is proposed and applied to the intelligent prediction of ROP.

(1) The FNN ROP prediction model with ReLU as the activation function performs best.
(2) With the reduction of the MRE of the FNN model, the neuronal activation states and the weight parameters of input features gradually stabilize at first and then change sharply, which is due to the gradual change of the model from underfitting to overfitting.
(3) When the lithology and drilling data of a well section are similar, the FNN ROP prediction model also has high similarity in the activation state of neurons and the weight parameters of features in this section.
(4) In the well section with similar drilling data, the linear characterization equation of the FNN ROP prediction model has high prediction accuracy, strong mobility, stability, and explainability, which can be regarded as an important reference for drilling parameter optimization.

CRediT authorship contribution statement

Zhi-Jun Pei: Data curation, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Xian-Zhi Song:** Data curation, Formal analysis, Funding acquisition, Investigation, Resources, Writing – review & editing. **Hai-Tao Wang:** Conceptualization, Data curation, Methodology. **Yi-Qi Shi:** Data curation, Formal analysis, Resources. **Shou-Ceng Tian:** Investigation, Resources, Validation. **Gen-Sheng Li:** Data curation, Resources, Supervision.

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

The authors greatly thanked the financial support from the National Key Research and Development Program of China (funded by National Natural Science Foundation of China, No. 2019YFA0708300), the Strategic Cooperation Technology Projects of CNPC and CUPB (funded by China National Petroleum Corporation, No. ZLZX2020-03), the National Science Fund for Distinguished Young Scholars (funded by National Natural Science Foundation of China, No. 52125401) and Science Foundation of China University of Petroleum, Beijing (funded by China University of petroleum, Beijing, No. 2462022SZBH002).

References

Abdulmalek, A.S., Salaheldin, E., Abdulazeez, A., et al., 2018. Prediction of rate of penetration of deep and tight formation using support vector machine. SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition. Dammam, Saudi Arabia.

Amadi, K., Iyalla, I., Prabhu, R., 2021. Modeling and predicting performance of autonomous rotary drilling system using machine learning techniques. SPE Nigeria Annual Int. Conf. Exhib. Lagos, Nigeria. https://doi.org/10.2118/208450-MS.

Bardhan, A., Kardani, N., GuhaRay, A., et al., 2021. Hybrid ensemble soft computing approach for predicting penetration rate of tunnel boring machine in a rock environment. J. Rock Mech. Geotech. Eng. 13 (6), 1398–1412. https://doi.org/10.1016/j.jrmge.2021.06.015.

Bingham, M., 1964. How to interpret drilling in the performance region. Oil Gas J. 62, 173–176.

Bourgoyne, A.T., Young, 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.

Bourgoyne, A.T., Millheim, K.K., Chenevert, M.E., et al., 1986. Applied Drilling Engineering, 2. Society of Petroleum Engineers Richardson.

Brito, L.C., Susto, G.A., Brito, J.N., et al., 2022. An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery. Mech. Syst. Signal Process. 163. https://doi.org/10.1016/j.ymssp.2021.108105.

Elmgerbi, A.M., Ettinger, C.P., Tekum, P.M., et al., 2021. Application of Machine Learning Techniques for Real Time Rate of Penetration Optimization. SPE/IADC Middle East Drilling Technology Conference and Exhibition. Abu Dhabi, UAE.

Encinas, M.A., Tunkiel, A.T., Sui, D., 2022. Downhole data correction for data-driven rate of penetration prediction modeling. J. Petrol. Sci. Eng. 210. https://doi.org/10.1016/j.petrol.2021.109904.

Ersoy, A., Waller, M., 1995. Wear characteristics of PDC pin and hybrid core bits in rock drilling. Wear 188 (1–2), 150–165.

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.

Garcia-Gavito, D., Azar, J., 1994. Proper nozzle location, bit profile, and cutter arrangement affect PDC-bit performance significantly. SPE Drill. Complet. 9 (3), 167–175.

Gers, F.A., Schmidhuber, J., Cummins, F., 2000. Learning to forget: continual prediction with LSTM. Neural Comput. 12 (10), 2451–2471. https://doi.org/10.1162/089976600300015015.

Gupta, I., Tran, N., Devegowda, D., et al., 2020. Looking ahead of the bit using surface drilling and petrophysical data: machine-learning-based real-time geosteering in volve field. SPE J. 25 (2), 990–1006. https://doi.org/10.2118/199882-pa.

Hansson, M., Olsson, C., 2017. Feedforward Neural Networks with ReLU Activation Functions Are Linear Splines. Bachelor's Theses in. Mathematical Sciences. Lund University.

Hazbeh, O., Aghdam, S.K.Y., Ghorbani, H., et al., 2021. Comparison of accuracy and computational performance between the machine learning algorithms for rate of penetration in directional drilling well. Petrol. Res. 6 (3), 271–282. https://doi.org/10.1016/j.ptlrs.2021.02.004.

Hegde, C., Wallace, S., Gray, K., 2015. Using Trees, Bagging, and Random Forests to Predict Rate of Penetration during Drilling. SPE Middle East Intelligent Oil and Gas Conference and Exhibition. Abu Dhabi, UAE. https://doi.org/10.2118/176792-MS.

Lawal, A.I., Kwon, S., Onifade, M., 2021. Prediction of rock penetration rate using a novel antlion optimized ANN and statistical modelling. J. Afr. Earth Sci. 182. https://doi.org/10.1016/j.jafrearsci.2021.104287.

Li, Y., Samuel, R., 2019. Prediction of Penetration Rate Ahead of the Bit through Real-Time Updated Machine Learning Models. In: SPE/IADC International Drilling Conference and Exhibition. The Hague, the Netherlands.

Mahmoodzadeh, A., Nejati, H.R., Mohammadi, M., et al., 2022. Forecasting tunnel boring machine penetration rate using LSTM deep neural network optimized by grey wolf optimization algorithm. Expert Syst. Appl. 209. https://doi.org/10.1016/j.eswa.2022.118303.

Mazen, A.Z., Rahmanian, N., Mujtaba, I., et al., 2021. Prediction of penetration rate for PDC bits using indices of rock drillability, cuttings removal, and bit wear. SPE Drill. Complet. 36 (2), 320–337. https://doi.org/10.2118/204231-pa.

Nasiri, H., Homafar, A., Chelgani, S.C., 2021. Prediction of uniaxial compressive strength and modulus of elasticity for Travertine samples using an explainable artificial intelligence. Results in Geophysical Sciences 8. https://doi.org/10.1016/j.ringps.2021.100034.

Negara, A., Saad, B., 2020. Combining Insight from Physics-Based Models into Data-Driven Model for Predicting Drilling Rate of Penetration. In: International Petroleum Technology Conference. Dammam, Saudi Arabia.

Pei, Z., Song, X., Ji, Y., et al., 2023. Wide and deep cross network for the rate of penetration prediction. Geoenergy Sci. Eng. 212066. https://doi.org/10.1016/j.geoen.2023.212066.

Pei, Z., Song, X., Wang, P., et al., 2022. Intelligent prediction for rate of penetration based on support vector machine regression. Xinjiang Oil&Gas 18 (1), 14–20. https://doi.org/10.12388/j.issn.1673-2677.2022.01.002.

Soares, C., Armenta, M., Panchal, N., 2020. Enhancing reamer drilling performance in deepwater Gulf of Mexico Wells. SPE Drill. Complet. 35 (3), 329–356.

Tsoka, T., Ye, X., Chen, Y., et al., 2022. Explainable artificial intelligence for building energy performance certificate labelling classification. J. Clean. Prod. 355. https://doi.org/10.1016/j.jclepro.2022.131626.

Van Lent, M., Fisher, W., Mancuso, M., 2004. An explainable artificial intelligence system for small-unit tactical behavior. In: Proceedings of the national conference on artificial intelligence. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press.

Warren, T., 1987. Penetration-rate performance of roller-cone bits. SPE Drill. Eng. 2 (1), 9–18.

Xiong, C., Huang, Z., Yang, R., et al., 2020. Comparative analysis cutting characteristics of stinger PDC cutter and conventional PDC cutter. J. Petrol. Sci. Eng. 189, 106792. https://doi.org/10.1016/j.petrol.2019.106792.

Young, F., 1969. Computerized drilling control. J. Petrol. Technol. 21 (4), 483–496.

Zhang, C., Cho, S., Vasarhelyi, M., 2022. Explainable artificial intelligence (XAI) in auditing. Int. J. Account. Inf. Syst. https://doi.org/10.1016/j.accinf.2022.100572.

Zhang, H., Lu, B., Liao, L. et al., 2021. Combining Machine Learning and Classic Drilling Theories to Improve Rate of Penetration Prediction. SPE/IADC Middle East Drilling Technology Conference and Exhibition. Abu Dhabi, UAE. https://doi.org/10.2118/202202-MS.

Zhou, F., Fan, H., Liu, Y., et al., 2022. Application of Xgboost Algorithm in Rate of Penetration Prediction with Accuracy. International Petroleum Technology Conference. Saudi Arabia, Riyadh.