ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

ROP Prediction Method Based on PCA−Informer Modeling

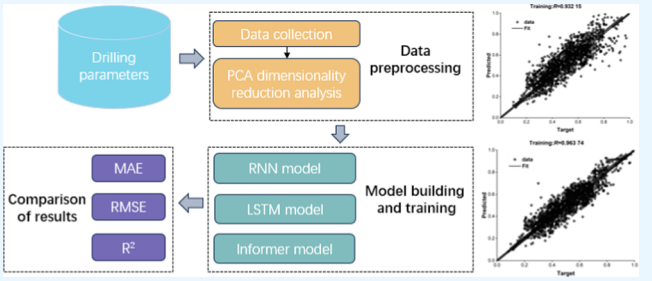
Yefeng Wang, Yishan Lou, Yang Lin, Qiaoling Cai, and Liang Zhu*

Cite This: *ACS Omega* 2024, 9, 23822−23831

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Increasing the rate of penetration (ROP) is an effective means to improve the drilling efficiency. At present, the efficiency and accuracy of intelligent prediction methods for the rate of penetration still need to be improved. To improve the efficiency and accuracy of rate of penetration prediction, this paper proposes a ROP prediction model based on Informer optimized by principal component analysis (PCA). We take the Taipei Basin block oilfield as an example. First, we use principal component analysis to extract data features, transforming the original data into low-dimensional feature data. Second, we use the PCA-optimized data to build an Informer model for predicting ROP. Finally, combined with actual data and using the recurrent neural network (RNN) and long short-term memory (LSTM) as baselines, we perform algorithm performance comparative analysis using root-mean-square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). The results show that the average MAE, RMSE, and R^2 of the PCA−Informer model are 9.402, 0.172, and 0.858, respectively. Compared with other methods, it has a larger R^2 and smaller RMSE and MAPE, indicating that this method significantly outperforms existing methods and provides a new solution to improve the rate of penetration in actual drilling operations.



1. INTRODUCTION

The rate of penetration is one of the most effective evaluation indicators in drilling engineering, directly related to drilling costs and efficiency.[1,2] Currently, rate of penetration (ROP) prediction mainly relies on the professional knowledge of field engineers and postdrilling data analysis, and the results are often quite subjective, lacking reliable analytical basis.[3] Accurate prediction of the rate of penetration can better plan drilling operations and shorten the drilling cycle. Therefore, it is an important means for engineers to increase the drilling speed.[4,6] The current research on the rate of penetration in drilling can be divided into the following three stages.[5]

From the 1950s to the 1990s, the main approach was to obtain ROP equations by establishing physical models. For example, the rate of penetration model proposed by Bourgoyne[7] laid the foundation for subsequent research. Walker and Guo et al.[8,9] considered rock mechanics and drilling parameters and used statistical regression methods to derive rate of penetration equations. Ju et al.[10] derived the mathematical relationship between the rate of penetration and rock drillability through linear regression analysis of field data. Anemeangely et al.[11] utilized evolutionary algorithms to determine the constant coefficients of the rate of penetration model. Moraveji et al.[12] introduced 6 quantities including mechanical parameters, hydraulic parameters, and rheological parameters to optimize the rate of penetration equation based on the traditional rate of penetration equation. Aarsnes et al.[13] modified the rate of penetration prediction model by introducing rock mechanical parameters, pressure parameters, and hydraulic parameters. Bilim et al.[14] established a mathematical model relating the uniaxial compressive strength, density, hardness, and porosity of the rock to the mechanical rate of penetration based on the drillability of the rock. The aforementioned regression analyses only covering limited field drilling parameters are restricted to certain formation types and have low accuracy. At the same time, reliance on field drilling data severely limits the improvement of drilling efficiency.

With the development of deep learning, intelligent algorithms have gradually been applied to the field of ROP prediction. For example, Yan and Amer et al.[15,16] proposed a new method for predicting the rate of penetration based on artificial neural network models. Amer et al.[17] considered formation changes, drilling parameters, and bit data and proposed a new method using an artificial neural network model to predict ROP. Song et al.[18] designed an intelligent ROP prediction based on support vector machine (SVM) regression. Yu et al.[19] used the gradient boosting decision tree (GBDT) algorithm for predictive

Received: February 25, 2024
Revised: May 5, 2024
Accepted: May 15, 2024
Published: May 22, 2024

https://doi.org/10.1021/acsomega.3c10339
ACS Omega 2024, 9, 23822−23831

analysis. Abbas et al.[20−22] studied the feasibility of intelligent algorithm ROP prediction using the support vector regression (SVR), extreme learning machine (ELM), gradient-boosted decision tree (GNDT), and other algorithms. Elkatatny et al.[23] conducted real-time prediction through an artificial neural network model. The results showed that the introduced real-time drilling fluid rheological property prediction model had high accuracy and can be used for real-time prediction. Elkatatny and Mahmoud et al.[24] utilized multiple measured data points to train and construct network models. The results showed that intelligent models not only can be used to predict the rate of penetration but also have high prediction accuracy.

In addition, Anemeangely and Bajolvand et al.[25,26] conducted rate of penetration prediction using a multilayer perceptron (MLP) neural network combined with the particle swarm optimization (PSO) algorithm. Hui et al.[27] designed a new method based on particle swarm optimization of long short-term memory (LSTM) neural networks to participate in model prediction. These methods indicate that applying artificial neural network models for ROP prediction is a feasible and useful approach. Zhang et al.[28] used a principal component analysis-LSTM (PCA-LSTM) neural network for downhole tool stick−slip vibration prediction, which effectively suppressed the lag effect caused by overfitting. Tang et al.[29] proposed a new rate of penetration prediction model based on PCA optimization of the BP neural network, successfully improving the performance of drilling operations. These optimization examples illustrate that parameter optimization is an important means to improve model performance and also verify the accuracy of PCA in model parameter optimization. The aforementioned rate of penetration prediction models mostly only perform simple correlation analysis on the input parameters, leading to long model training time, low prediction accuracy, and the problem of overfitting during the prediction process. In deep learning,[30−34] Informer is a model based on the improved Transformer for time series prediction. It can handle multitime scale and irregular time interval data and has good performance in many prediction domains. It can effectively overcome the problems of overfitting and long training times in previous methods.

Previous research has proven the efficiency of PCA in optimizing intelligent algorithms.[35−38] PCA is based on multivariate statistical analysis and provides powerful data feature extraction capabilities; therefore, it is often used to solve various optimization problems. In this paper, we use PCA to extract the principal components of the dataset, effectively reducing the dimensionality of the data, saving a significant amount of computation time, and improving the efficiency of dataset utilization.

To address the issues with the existing rate of penetration prediction methods, we combine the prediction advantages of the PCA and Informer models. In this method, we use PCA to obtain the relevant feature indicators from the original data, input them into the Informer model for ROP prediction, and perform a comparative analysis of the prediction results before and after optimization. It is verified that the Informer model after PCA optimization has higher prediction accuracy and shorter training time. This method helps drilling engineers achieve ultralong ROP prediction, providing support for optimizing drilling operations and shortening the drilling cycle.

The main contribution of this article is that it combines the advantages of PCA in feature analysis and the strengths of the Informer model. We propose a ROP prediction method based on PCA combined with the Informer model. Compared with other models and the prediction results of the Informer model before and after optimization, the optimized PCA−Informer model has advantages in both efficiency and accuracy. This introduces a new method for predicting oil engineering-related parameters.

Figure 1 shows the workflow of our proposed method. It includes four main stages: data processing, model construction, model training, and result evaluation.

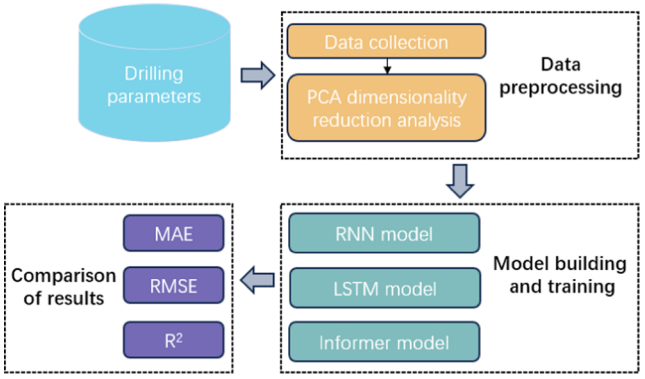


Figure 1. Forecasting workflow diagram.

2. DATA PROCESSING

2.1. Data Selection. The factors affecting the rate of penetration are mainly divided into uncontrollable factors and controllable factors. Uncontrollable factors are determined by the natural geological environment, such as rock strength, drillability, and sanding volume.[39−41] Controllable factors mainly include three categories: mechanical parameters, drilling fluid parameters, and hydraulic parameters. This paper uses drilling data from a well section in the Taipei Basin block, with a total of 2365 datasets. The while-drilling (Figure 2) parameters include the well depth (Depth), acoustic time difference (DT), γ (GR), formation density (ZDEN), pore pressure (PP), weight on bit (WOB), bit speed (BRS), well diameter (CAL), pump displacement (PD), pump pressure (Pumpp), drilling fluid density (DFD), and rate of penetration (ROP). Table 1 summarizes the basic information on these parameters, including the unit, minimum value, maximum value, and average value of each parameter.

In the process of drilling in deep and complex formations, the drilling rate exhibits an increasingly sensitive trend due to the variation in attribute values. Therefore, it is essential to fully utilize drilling data to construct an efficient drilling rate prediction model applicable to complex formations. Hence, we chose to employ PCA to perform dimensionality reduction on the original data, selecting parameters with strong correlations for training purposes.

2.2. Dimensionality Reduction and Correlation Analysis. During the process of deep and complex formation drilling, due to the variation of property values, the mechanical rotational speed exhibits an increasingly sensitive trend, so it is necessary to perform relevant processing of the drilling data. In the previous methods, the selection of relevant parameters for ROP prediction was often arbitrary or correlation analysis was simply conducted to select some parameters to participate in model training, which affected the integrity of the original information.

https://doi.org/10.1021/acsomega.3c10339
ACS Omega 2024, 9, 23822−23831

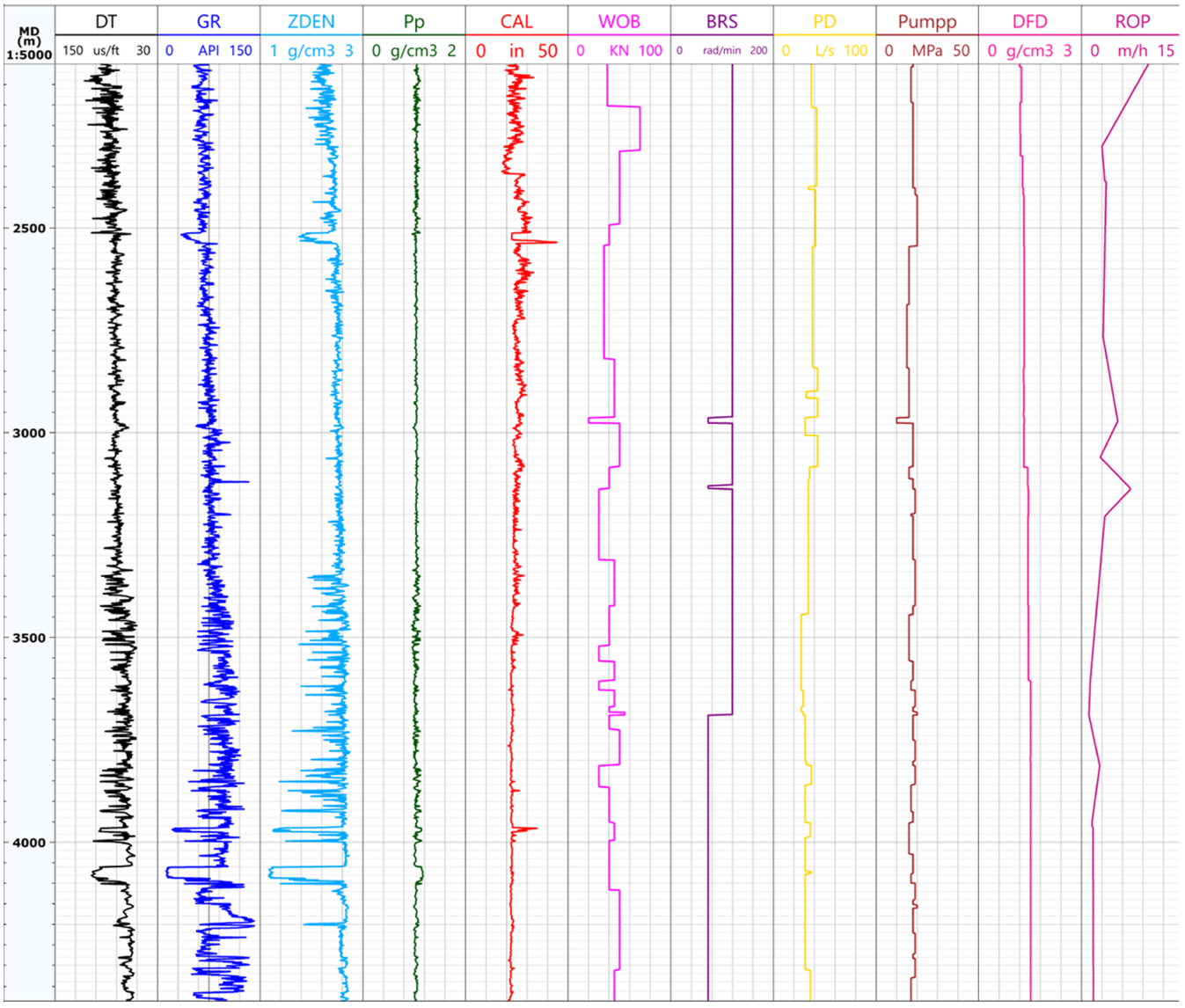


Figure 2. Raw collection data.

Table 1. Partial Drilling Data

parameter	unit	minimum	maximum	average
depth	m	2100	4379	3239.5
DT	μs/ft	66.13	90.02	78.07
GR	API	61.53	100.48	81.00
ZDEN	g/cm^3	2.19	2.67	2.43
Pp	g/cm^3	1.04	1.05	1.05
CAL	in	20.87	24.70	22.79
WOB	KN	38	45	41.5
BRS	r/min	73	120	96.5
PD	L/s	36	37	36.5
Pumpp	MPa	17	18	17.5
DFD	g/cm^3	1.2	1.52	1.36
ROP	m/h	1.76	9.80	5.78

Based on the advantages of principal component analysis (PCA) in high-dimensional data processing, it can be used to perform dimensionality reduction on the original data, representing the high-dimensional drilling parameters with a low-dimensional matrix linear expression. While the principal components that best represent the original data are determined, the loss of original data information can be minimized as much as possible.

Using PCA to screen the principal components affecting the rate of penetration involves four steps.[42,43] The first step is standardization, which can ensure that the difference in the magnitudes of different property data will not affect the cumulative variance contribution rate. The standardization expression is defined as

$$x = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} \\ x_{21} & x_{22} & \cdots & x_{2j} \\ \vdots & \vdots & \vdots & \\ x_{i1} & x_{i2} & \cdots & x_{ij} \end{bmatrix} \tag{1}$$

$$X_{ij}' = \frac{X_{ij} - \overline{X_j}}{\sqrt{\sum (X_{ij} - \overline{X_j})^2}}, \overline{X_j} = \sum_{j=1}^{n} X_{ij} (i = 1, 2, \cdots, m; j = 1, 2, \cdots, n) \tag{2}$$

https://doi.org/10.1021/acsomega.3c10339
ACS Omega 2024, 9, 23822−23831

Table 2. Correlation Coefficient Matrix eigenvectors

feature vectors	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	⋯	λ_{11}
depth(X_1)	0.05	0.15	0.07	0.07	0.74	0.65	⋯	0.05
GR(X_2)	0.33	−0.44	0.34	−0.13	−0.57	0.23	⋯	−0.54
WOB(X_3)	0.43	−0.33	0.20	−0.21	−0.34	−0.35	⋯	0.27
BRS(X_4)	0.49	0.24	0.15	−0.44	0.18	−0.21	⋯	−0.04
Pp(X_5)	−0.06	0.58	−0.48	−0.21	0.15	0.07	⋯	0.56
DFD(X_6)	0.26	−0.03	0.01	0.67	−0.14	−0.42	⋯	−0.47
PD(X_7)	−0.40	0.10	0.84	0.05	−0.28	−0.79	⋯	0.01
CAL(X_8)	0.15	0.21	0.41	−0.71	0.71	0.52	⋯	0.36
ZDEN(X_9)	0.04	0.14	−0.03	0.27	−0.35	−0.15	⋯	−0.51
DT(X_{10})	0.26	0.24	0.07	0.11	0.14	0.06	⋯	0.07
Pumpp(X_{11})	−0.12	0.14	−0.25	0.16	0.28	−0.24	⋯	0.14
eigenvalues	3.66	1.51	1.12	0.78	0.55	0.31	⋯	0.17
contribution rate/%	50.3	16.38	11.22	7.24	5.11	3.23	⋯	2.21
cumulative contribution rate/%	50.3	66.68	77.9	85.14	90.25	93.48	⋯	100

where X'_{ij} is the processed data in the ith row and jth column, X_{ij} is the sample value in the ith row and jth column, $\overline{X_j}$ denotes the mean of the jth column, m represents the number of original data, and n is the dimensionality of the original data.

The second step is to solve the covariance matrix $R_{n\times n}$ of the original data matrix and calculate the corresponding eigenvalues and eigenvectors.[44] The calculation process is

$$R_{n\times n} = \mathrm{Cov}(x) = \frac{1}{n-1}\sum_{k=1}^{n}(X_{ki} - \overline{X_i})(X_{ki} - \overline{X_j}) \tag{3}$$

The calculation of the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ and the corresponding eigenvectors of the covariance matrix $R_{n\times n}$ is as follows

$$a_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix}, a_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{bmatrix}, \cdots, a_n = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{nn} \end{bmatrix} \tag{4}$$

The third step is to calculate the contribution rate R and cumulative contribution rate RS using the following formulas. The calculated results of the eigenvalues, variance contribution rate, and cumulative variance contribution rate of the original data are shown in Table 2.

The fourth step is to select the principal components and analyze their significance. In practical applications, the principle for selecting the principal components is when the cumulative variance contribution rate reaches 80%.[45] The ith principal component is

$$F_i = a_{1i}X_1 + a_{2i}X_2 + \cdots + a_{ji}X_j (i, j = 1, 2, \cdots, n) \tag{5}$$

From Table 2, we can see that the cumulative contribution rate of the first 5 principal components reaches 93.84%. it is considered reasonable to only take the first 5 principal components to represent the original data characteristics, according to formula 5.

The first principal component F1 has moderate positive loadings on WOB and BRS and a moderate negative loading on PD, while the loadings on the remaining variables are relatively small. Therefore, F1 can be termed the mechanical parameter component. The second principal component F2 has a relatively large positive loading on Pp, while the loadings on the remaining variables are relatively small. Therefore, F2 can be termed the pressure factor component. The third principal component F3 has a relatively large positive loading on PD and a moderate negative loading on Pp, while the loadings on the remaining variables are relatively small. Therefore, F3 can be termed the hydraulic parameter component. The fourth principal component F4 has a relatively large positive loading on DFD and a relatively large negative loading on CAL, and the loadings on the remaining variables are relatively small. Therefore, F4 can be termed the drilling fluid parameter component. The fifth principal component F5 has relatively large positive loadings on Depth and CAL, a moderate negative loading on GR, and a large negative loading on the remaining variables, relatively small. Therefore, F5 can be termed as the formation factor component.

The calculation results of the original dataset are shown in Table 3.

Table 3. Results after Dimensionality Reduction of the Original data

	principal component				
example	F1	F3	F2	F4	F5
1	0.297	1.274	3.912	1.354	2.312
2	0.284	1.411	2.902	1.235	2.641
3	0.198	1.340	2.870	1.479	2.511
4	0.354	1.212	2.783	1.315	2.931
5	0.157	1.089	2.744	1.011	2.130
⋮	⋮	⋮	⋮	⋮	⋮
2365	2.521	0.534	1.068	0.475	1.301

3. PREDICTION MODEL

3.1. Informer Model. The Informer model was proposed by Zhou in 2020.[46] Its main feature is the ability to adaptively capture the long-term dependencies and multiscale patterns in time series data. The Informer model adopts a multilevel self-attention mechanism, which allows the model to consider the input data with weighted attention based on the different features and contextual information on the data.

The overall architecture of the Informer model (Figure 3) consists of an encoder and a decoder. The encoder is used to capture the features of the original input sequence, while the decoder is used to generate the sequence prediction results. After the input parameters undergo masked multihead probabilistic sparse self-attention processing, they undergo multihead self-attention operations with the feature mapping

https://doi.org/10.1021/acsomega.3c10339
ACS Omega 2024, 9, 23822−23831

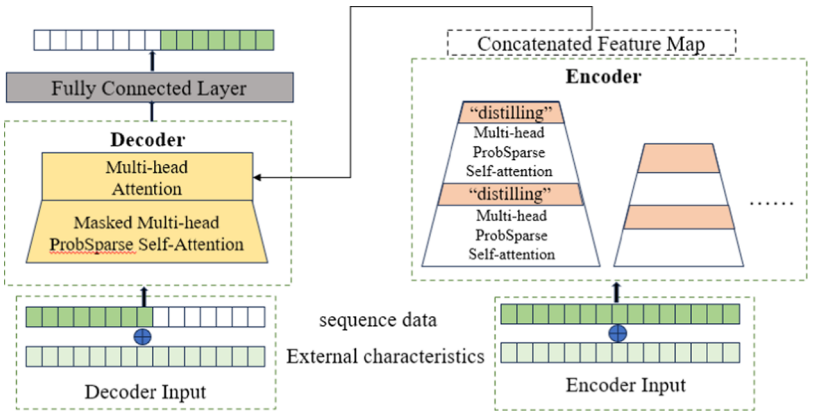


Figure 3. Informer model structure.

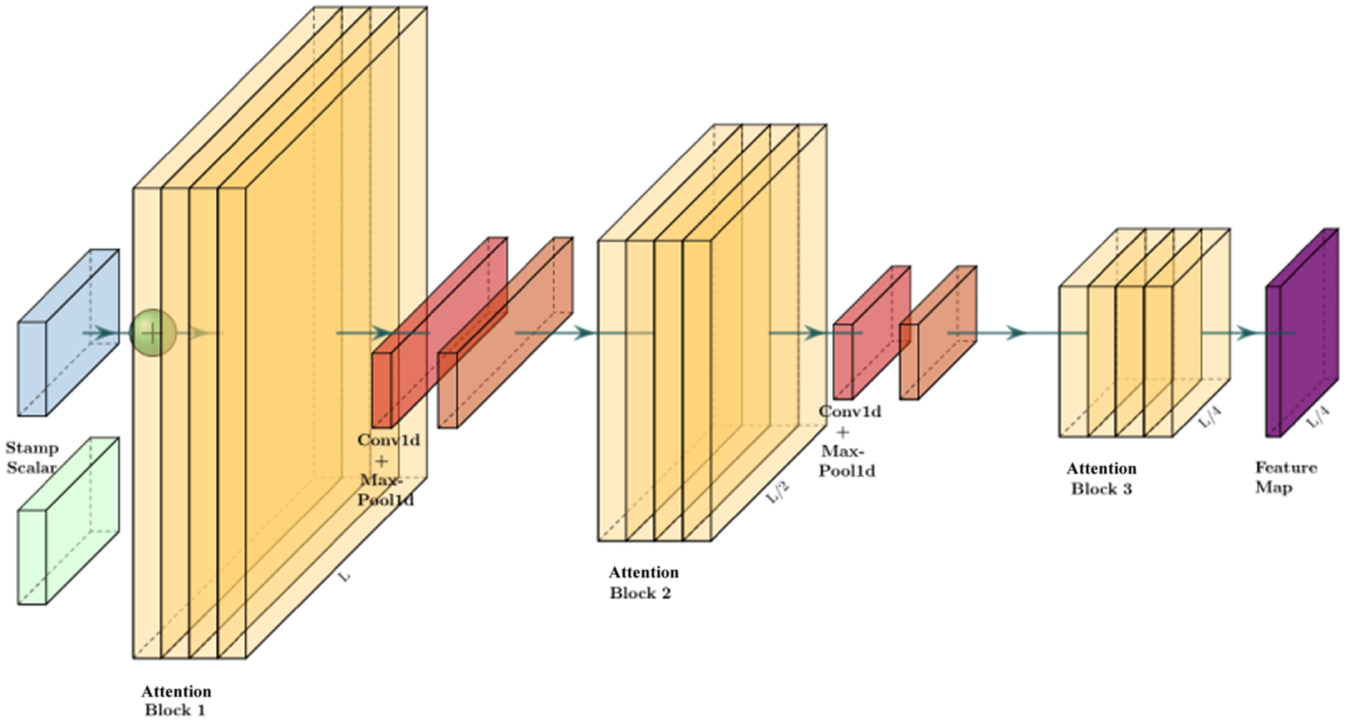


Figure 4. Informer's encoder.

output from the encoder and finally obtain the output through a fully connected layer.

3.1.1. Encoder. The Informer encoder (Figure 4) can process longer sequential inputs under memory usage constraints. It consists of multiple levels, and each level includes a multihead ProbSparse self-attention module and a "Distilling" module.

The ProbSparse self-attention module in the Informer encoder is used to select more important queries (Q) from the input, improving the computation efficiency. The calculation formula is as follows[47,48]

$$\text{attention}(Q, K, V) = \text{softmax}\left(\frac{\bar{Q}K^T}{\sqrt{d}}\right)V \tag{6}$$

Q (query), K (key), and V (value) are matrices of the same size obtained by linear transformation with weight matrices of input features; Q is obtained after probabilistic sparsification; softmax is the activation function; and d represents the dimension.

The "Distilling" module is a generalization of one-dimensional convolution and max-pooling operations. As a result of the multihead probabilistic sparse self-attention, the feature mapping of the encoder contains redundant combinations of the V values. To reduce the network size, the "distillation" operation is used to privilege the dominant high-level features.The "Distilling" operation formula from layer j to layer $j + 1$ is shown as formula 7[49,50]

$$X_{j+1} = \text{MaxPool}(\text{ELU}(\text{Convld}([X_j^t]_{\text{AB}}))) \tag{7}$$

$[\bullet]_{\text{AB}}$ is an attention block containing multiple probabilistic self-attention and elementary operations; Conv1d(*) represents one-dimensional convolution along the time dimension; ELU is the activation function; and MaxPool is the maximum pooling operation.

3.1.2. Decoder. The Informer model adopts a standard decoder structure, which is composed of two stacked identical multihead attention layers internally. At the same time, in order to alleviate the rapid drop in speed in long sequence prediction

https://doi.org/10.1021/acsomega.3c10339
ACS Omega 2024, 9, 23822−23831

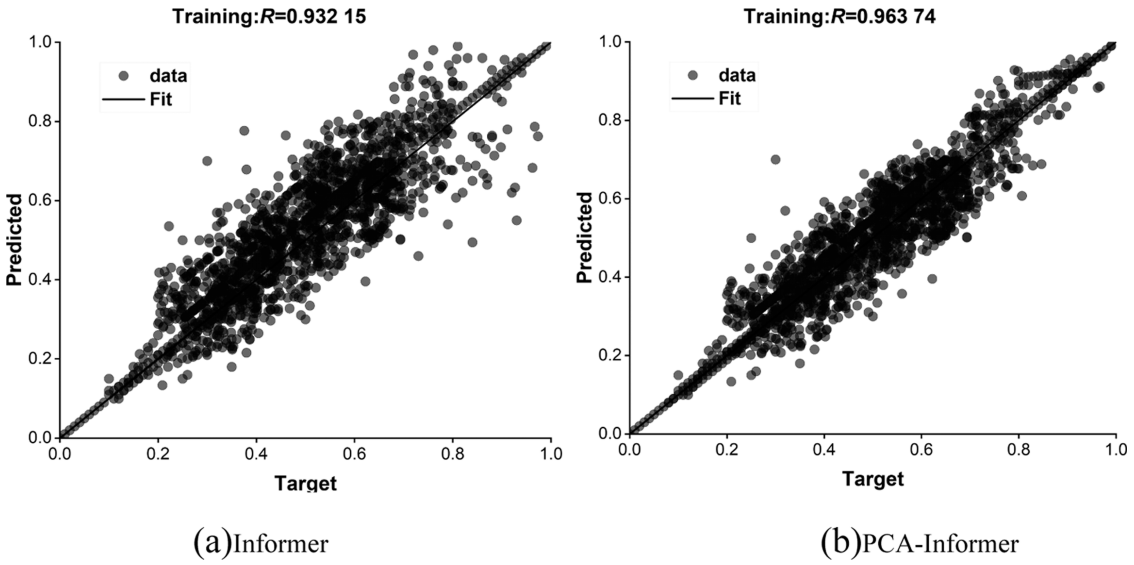


Figure 5. Accuracy of the model on the training set before and after PCA optimization.

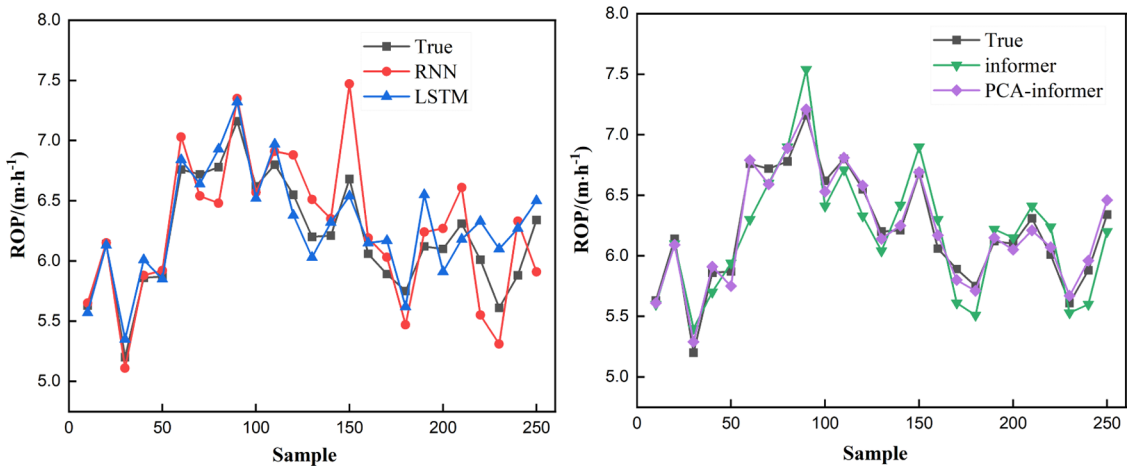


Figure 6. Prediction results of all methods on the test set.

problems, Informer adopts a generative inference structure with masking, that is, the input vector contains part of the original sequence data and placeholder (which can be initially set to 0) for the target prediction sequence.

3.2. PCA−Informer ROP Prediction Model. Traditionally, most ROP prediction methods have relatively simple data preprocessing, leading to long training times and low prediction accuracy. PCA is an effective method for extracting features from complex data with good performance and has been used to solve complex optimization problems in various fields.

After research, we set (explanation in Section 4.2) the Informer to include a 4-layer stack and a 1-layer stack in the encoder as well as a 2-layer decoder. The Adam optimizer is used for optimization, with the learning rate starting from 0.001 and halving every epoch.

3.3. Model Predictive Performance Evaluation. In order to more reasonably evaluate the model performance, this paper adopts root-mean-square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) as evaluation metrics. The calculation steps are given by eqs 8−10. RMSE reflects the precision of the prediction, with a smaller value indicating that the predicted value is closer to the true value. MAE reflects the average difference between the predicted and actual values. R^2 is used to measure the goodness of fit of the model, with a larger value (closer to 1) indicating higher prediction accuracy.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - x_i^*)^2} \tag{8}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|x_i - x_i^*| \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - x_i^*)^2}{\sum_{i=1}^{n}(x_i^* - \overline{x_i^*})^2} \tag{10}$$

where x_i is the predicted value; x_i^* is the original value; $\overline{x_i^*}$ is the average of the original values; and n is the total number of test samples.

4. RESULTS AND DISCUSSION

4.1. Result Discussion and Analysis. To verify the performance of the PCA−Informer model, we take the Taipei Basin block oilfield as an example and use different models to

https://doi.org/10.1021/acsomega.3c10339
ACS Omega 2024, 9, 23822−23831

predict the ROP of this well. The logged data from the drilled section are used as the training set to predict the ROP for the next 200 m depth interval.

In Figure 5, we compare the accuracy of the Informer training set before and after optimization. It can be observed that the prediction accuracy of 0.963 for the optimized Informer model is greater than 0.932 before optimization, indicating that PCA achieved good optimization results in model parameter optimization. Figure 6 shows the prediction results of different models on the test set. The black curve represents the original data, and the other colors represent the prediction results of the test set using different models. It can be seen that the difference between the predicted values and the true values for the PCA-optimized Informer model is significantly smaller than that of other models. Figure 7 shows the distribution of predicted values

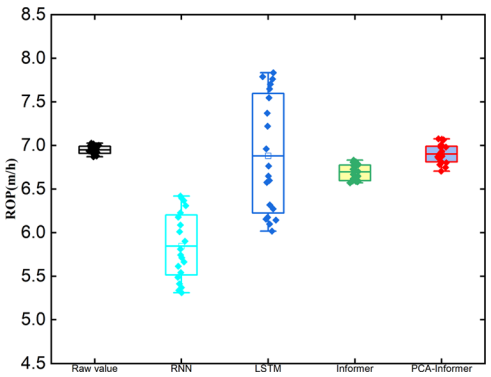


Figure 7. Distribution of predicted values for the models.

on the test set. It can be seen that the predicted values of PCA−Informer are generally closer to the original values, with the predictions being more concentrated and stable.

Table 4 summarizes the evaluation results of all methods on the dataset, with the best results shown in boldface. We can observe the following: (1) the proposed PCA−Informer model significantly improves the inference performance (number of accurate predictions) on the dataset, and as the prediction range continues to increase, the prediction error rises steadily, indicating that the PCA−Informer model is successful in predicting the rate of penetration. (2) The performance of the PCA−Informer model is significantly better than that of RNN and LSTM, with an average RMSE reduction of 35 and 29.9%, respectively, compared to these methods. (3) Compared to the single Informer model, the RMSE of our method is reduced by 11.8%. This indicates that PCA can better optimize the input parameters and improve the prediction accuracy of the model.

4.2. Parameter Sensitivity Analysis. We propose analyzing the parameters of the PCA−Informer model. In Figure 8(a), it can be observed that when the prediction sequence is 48, increasing the input length of the encoder/decoder will decrease the model performance, but further increasing it will lead to a decrease in the RMSE, as it brings repetitive short-term ROP patterns. However, when the prediction sequence is 96, the longer the input, the lower the RMSE because a longer encoder input contains more dependency information, while a longer decoder has richer local information.

At the same time, we studied the effect of the number of encoder stacks on the model in Figure 8(b). We found that longer stacks are more sensitive to the input. We chose to connect a 4-layer stack and a 1-layer stack as the most robust strategy.

4.3. Computational Efficiency. In order to select the best implementation from all models, we performed a strict runtime comparison in Figure 9. In the training stage, the PCA−Informer method achieved the best training efficiency. In the testing stage, our method was much faster than the other methods in generative decoding. This shows that the Informer model optimized by PCA has a shorter training time and a higher runtime efficiency.

5. CONCLUSIONS

This paper proposes a ROP prediction method based on PCA and the Informer neural network. When constructing the model, we use the PCA algorithm to optimize the input parameters of Informer. The following are our conclusions:

1. For the oilfield in the Taipei Sag Basin, by adopting the principal component analysis method, 5 principal components were selected, which can cover 90.25% of the original data characteristics. This can effectively reduce the dimensionality of the data, save a lot of computation time, and improve the utilization efficiency of the dataset.
2. Based on the PCA−Informer neural network, a mechanical drilling rate prediction model was established, which compared to RNN, LSTM, and Informer reduced the root-mean-square error by 34.8, 29.7, and 11.7%, respectively, reduced the mean absolute percentage error by 31.8, 22.1, and 6.9%, respectively, and improved the coefficient of determination by 20, 11.7, and 7.6%, respectively, indicating that the PCA−Informer model has higher prediction accuracy and more stable error changes.
3. The four models compared show that the Informer model optimized with the PCA algorithm has the shortest running time in both the training and testing stages, indicating that the method we proposed is highly efficient with fast iteration speed.

The PCA−Informer model we proposed can help drilling engineers improve drilling operations and shorten the drilling cycle. In future work, the authors will further explore the

Table 4. Comparison of Model Evaluation Metrics

methods	RNN			LSTM			informer			PCA−informer		
metric	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
24	10.451	0.154	0.785	10.123	0.153	0.812	**8.104**	**0.101**	**0.898**	9.054	0.121	0.813
48	11.301	0.262	0.793	9.524	0.241	0.841	10.121	0.147	0.836	**9.326**	**0.125**	**0.864**
72	16.848	0.263	0.648	14.313	0.231	0.705	10.620	0.254	0.733	**9.546**	**0.213**	**0.871**
96	16.546	0.376	0.634	14.322	0.354	0.712	11.581	0.276	0.721	**9.681**	**0.227**	**0.885**
average	13.787	0.264	0.715	12.071	0.245	0.768	10.107	0.195	0.797	**9.402**	**0.172**	**0.858**

https://doi.org/10.1021/acsomega.3c10339
ACS Omega 2024, 9, 23822−23831

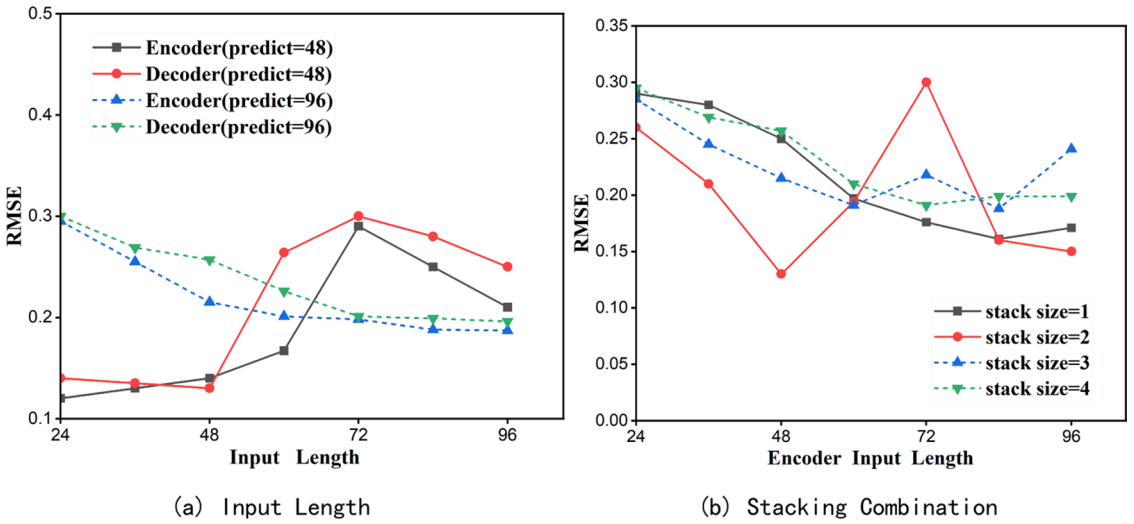


Figure 8. Parameter Sensitivity of Components in Informer.

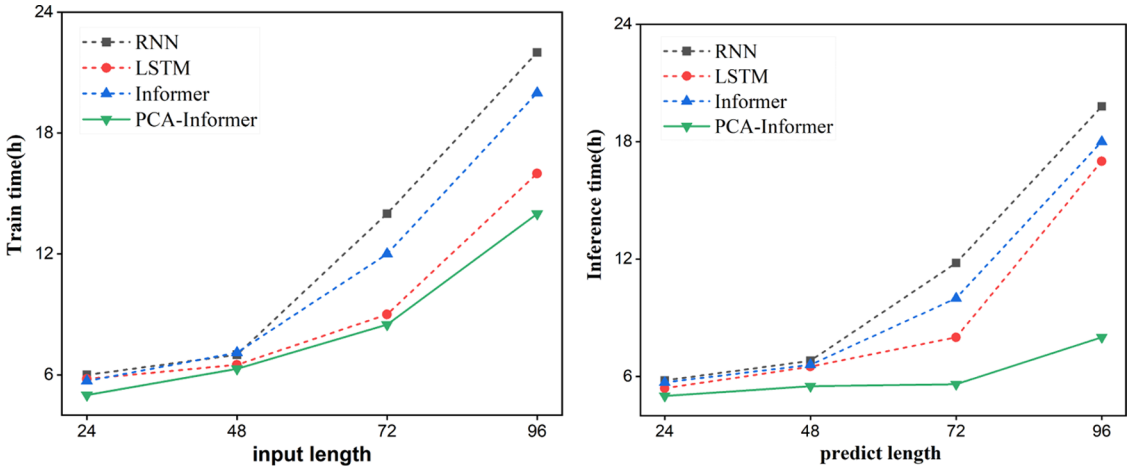


Figure 9. Total runtime in the training/testing stage.

application of the model in other petroleum-related fields and attempt to apply other prediction methods to the prediction of drilling parameters. In addition, the authors will also consider the combination of the model with real-time data to achieve real-time optimization of the drilling process.

AUTHOR INFORMATION

Corresponding Author

Liang Zhu − *School of Petroleum Engineering, Changjiang University, Wuhan 430100, China; Hubei Key Laboratory of Oil and Gas Drilling and Production Engineering, Wuhan 430100, China*; Email: zhuliang123@yangtzeu.edu.cn

Authors

Yefeng Wang − *School of Petroleum Engineering, Changjiang University, Wuhan 430100, China; Hubei Key Laboratory of Oil and Gas Drilling and Production Engineering, Wuhan 430100, China*; orcid.org/0009-0009-5143-9391

Yishan Lou − *School of Petroleum Engineering, Changjiang University, Wuhan 430100, China; Hubei Key Laboratory of Oil and Gas Drilling and Production Engineering, Wuhan 430100, China*

Yang Lin − *School of Petroleum Engineering, Changjiang University, Wuhan 430100, China; Hubei Key Laboratory of Oil and Gas Drilling and Production Engineering, Wuhan 430100, China*

Qiaoling Cai − *School of Petroleum Engineering, Changjiang University, Wuhan 430100, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c10339

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Y.W., L.Z., and Y.L. thank the National Engineering Research Center for Oil & Gas Drilling and Completion Technology for research funding under project NERCDCT202319.

NOMENCLATURE

depth	drilling depth, m
DT	formation acoustic transit time, μs/ft
GR	γ, API
ZDEN	formation density, g/cm^3
Pp	formation pore pressure, g/cm^3
WOB	weight on bit, KN
BRS	bit speed, r/min

https://doi.org/10.1021/acsomega.3c10339
ACS Omega 2024, 9, 23822−23831

CAL	well diameter, in
PD	pump displacement, L/s
Pumpp	pump pressure, MPa
DFD	drilling fluid density, g/cm^3
ROP	rate of penetration, m/h
`ProbSparse`	method to reduce the number of parameters while maintaining model performance
self-attention	capturing sequence relationships through relevance analysis
`Conv1d`	operation that extracts features or performs filtering on a 1D data sequence
`MaxPool`	operation that extracts the maximum values from the input feature map
`Q (query)`	model needs to focus on, dimensionless
`K (key)`	model needs to correlate, dimensionless
`V (value)`	model needs to output, dimensionless
`Softmax/ELU`	neural network activation function

REFERENCES

(1) Hegde, C.; Daigle, H.; Gray, K. E. Performance comparison of algorithms for real-time rate-of-penetration optimization in drilling using data-driven models. *SPE J.* **2018**, *23*, 1706−1722.

(2) Oyedere, M.; Gray, K. ROP and TOB optimization using machine learning classification algorithms. *J. Nat. Gas Sci. Eng.* **2020**, *77*, No. 103230.

(3) Ashrafi, S. B.; Anemangely, M.; Sabah, M.; Ameri, M. J. Application of hybrid artificial neural networks for predicting rate of penetration (ROP): A case study from Marun oil field. *J. Pet. Sci. Eng.* **2019**, *175*, 604−623.

(4) Brenjkar, E.; Delijani, E. B. Computational prediction of the drilling rate of penetration (ROP): A comparison of various machine learning approaches and traditional models. *J. Pet. Sci. Eng.* **2022**, *210*, No. 110033.

(5) Mazen, A. Z.; Rahmanian, N.; Mujtaba, I.; Hassanpour, A. Prediction of Penetration Rate for PDC Bits Using Indices of Rock Drillability, Cuttings Removal, and Bit Wear. *SPE Drill. Completion* **2021**, *36*, 320−337.

(6) Liu, S.; Sun, J.; Gao, X.; Wang, M. Analysis and establishment of drilling speed prediction model for drilling machinery based on artificial neural networks. *Comput. Sci.* **2019**, *46*, 605−608.

(7) Bourgoyne, A. T.; Young, F. S. A Multiple Regression Approach to Optimal Drilling and Abnormal Pressure Detection. *Soc. Pet. Eng. J.* **1974**, *14* (04), 371−384.

(8) Walker, B. H.; Black, A. D.; Klauber, W. P.; Little, T.; Khodaverdian, M. In *Roller-bit Penetration Rate Response as a Function of Rock Properties and Well Depth*, SPE Annual Technical Conference and Exhibition; OnePetro, 1986.

(9) Guo, Y. Prediction of optimum footage and penetration rate of bit by regression analysis. *Pet. Drill. Prod. Technol.* **1994**, *16*, 24−26.

(10) Ju, M. Rock Drillability and the Application of Linear Regression During Pure Drilling with a Drill Bit. *Pet. Drill. Tech.* **1992**, No. 01, 19−24.

(11) Anemangely, M.; Ramezanzadeh, A.; Tokhmechi, B. Determination of constant coefficients of Bourgoyne and Young drillingrate model using a novel evolutionary algorithm. *J. Min. Environ* **2017**, *8*, 693−702, DOI: 10.22044/jme.2017.842.

(12) Moraveji, M. K.; Naderi, M. Drilling rate of penetration prediction and optimization using response surface methodology and bat algorithm. *J. Nat. Gas Sci. Eng.* **2016**, *31*, 829−841, DOI: 10.1016/j.jngse.2016.03.057.

(13) Aarsnes, U. J. F.; Aamo, O. M.; Krstic, M. In *Extremum Seeking for Real-time Optimal Drilling Control*, 2019 American Control Conference (ACC); IEEE: Philadelphia, PA, USA, 2019; pp 5222−5227.

(14) Bilim, N.; Karakaya, E. Penetration rate prediction models for core drilling. *Min. Metall. Explor.* **2021**, *38* (1), 359−366.

(15) Tie, Y. A. N.; Xueliang, B. I.; Chuntian, L. I. U.; et al. An artificial neural network method for predicting deep well drilling speed. *Pet. Drill. Technol.* **2001**, *29* (6), 9−10.

(16) Tie, Y. A. N.; Rui, X. U.; Weikai, L. I. U.; et al. Research and development of intelligent drilling technology in china. *J. Northeast Pet. Univ.* **2020**, *44* (4), 15−21.

(17) Amer, M. M.; Dahab, A. S.; El-Sayed, A.-A. H. In *An ROP Predictive Model in Nile Delta Area Using Artificial Neural Networks*, SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition; OnePetro, 2017.

(18) Song, X.; Pei, Z.; Wang, P.; et al. Intelligent Prediction of Mechanical Drilling Speed Based on Support Vector Machine Regression. *Xinjiang Pet. Nat. Gas* **2022**, *18* (01), 14−20.

(19) Yang, Yu.; Huang, K.; Hui, Li. Research on Mechanical Drilling Speed Prediction Method Based on Machine Learning and Multi-source Data Preprocessing Techniques. *China Pet. Chem. Stand. Qual.* **2021**, *41* (20), 133−136.

(20) Abbas, A. K.; Rushdi, S.; Alsaba, M. In *Modeling Rate of Penetration for Deviated Wells Using Artificial Neural Network*, Abu Dhabi International Petroleum Exhibition & Conference; OnePetro: Abu Dhabi, UAE, 2018.

(21) Najjarpour, M.; Jalalifar, H.; Norouzi-Apourvari, S. The effect of formation thickness on the performance of deterministic and machine learning models for rate of penetration management in inclined and horizontal wells. *J. Pet. Sci. Eng.* **2020**, *191*, No. 107160, DOI: 10.1016/j.petrol.2020.107160.

(22) framework: for the prediction of drilling rate of peneln-tion [C]/APCEC17 International Conference on Advances in Petroleum Chemical & Energy Challenges. 2017, pp 1−11.

(23) Elkatatny, S.; Tariq, Z.; Mahmoud, M. Real time prediction of drilling fluid rheological properties using Artificial Neural Networks visible mathematical model (white box). *J. Pet. Sci. Eng.* **2016**, *146*, 1202−1210.

(24) Elkatatny, S.; Tariq, Z.; Mahmoud, M.; Abdulraheem, A.; Mohamed, I. An integrated approach for estimating static Young modulus using artificial intelligence tools. *Neural Comput. Appl.* **2019**, *31*, 4123−4135.

(25) Anemangely, M.; Ramezanzadeh, A.; Tokhmechi, B.; Molaghab, A.; Mohammadian, A. Drilling rate prediction from petrophysical logs and mud logging data using an optimized multilayer perceptron neural network. *J. Geophys. Eng.* **2018**, *15*, 1146−1159.

(26) Matinkia, M.; Sheykhinasab, A.; Shojaei, S.; Kand, A. V. T.; Elmi, A.; et al. Developing a New Model for Drilling Rate of Penetration Prediction Using Convolutional Neural Network. *Arabian J. Sci. Eng.* **2022**, *47*, 11953−11985, DOI: 10.1007/s13369-022-06765-x.

(27) Ji, H.; Lou, Y.; Cheng, S.; et al. An Advanced Long Short-Term Memory (LSTM) Neural Network Method for Predicting Rate of Penetration (ROP). *ACS Omega* **2023**, *8* (1), 934−945, DOI: 10.1021/acsomega.2c06308.

(28) Xin, Z.; Tao, Z.; Yumei, L., et al. Research on Stick-Slip Vibration Level Assessment Method Based on PCA-LSTM *China Petroleum Machinery* **2023**; Vol. *51* 02, pp 18−25 DOI: 10.16082/j.cnki.issn.1001-4578.2023.02.003.

(29) Tang, M.; Wang, H.; He, S.; et al. Research on Prediction of Mechanical Rate of Penetration Based on PCA-BP Algorithm. *Pet. Mach.* **2023**, *51* (10), 23−31.

(30) He, Z.; He, Z.; Li, S.; et al. A ship navigation risk online prediction model based on informer network using multi-source data. *Ocean Eng.* **2024**, *298*, No. 117007, DOI: 10.1016/j.oceaneng.2024.117007.

(31) Zhuang, W.; Li, Z.; Wang, Y.; et al. GCN−Informer: A Novel Framework for Mid-Term Photovoltaic Power Forecasting. *Appl. Sci.* **2024**, *14* (5), No. 2181, DOI: 10.3390/app14052181.

(32) Zhang, X.; Yang, K.; Zheng, L. Transformer Fault Diagnosis Method Based on TimesNet and Informer. *Actuators* **2024**, *13* (2), No. 74, DOI: 10.3390/act13020074.

(33) Shi, Z.; Li, J.; Jiang, Z.; et al. WGformer: A Weibull-Gaussian Informer based model for wind speed prediction. *Eng. Appl. Artif. Intell.* **2024**, *131*, No. 131107891, DOI: 10.1016/j.engappai.2024.107891.

https://doi.org/10.1021/acsomega.3c10339
ACS Omega 2024, 9, 23822−23831

(34) Wang, S.; Chang, L.; Liu, H.; et al. Short-term prediction of wind power based on temporal convolutional network and the informer model. *IET Gener. Transm. Distrib.* **2023**, *18* (5), 941−951, DOI: 10.1049/gtd2.13064.
(35) Kumar, T.; Kumar, S. N.; Rao, G. S. Automatic lithology modelling of coal beds using the joint interpretation of principal component analysis (PCA) and continuous wavelet transform (CWT). *J. Earth Syst. Sci.* **2023**, *132* (1), No. 10, DOI: 10.1007/s12040-022-02018-5.
(36) Chen, F.; Jiang, X.; Wu, C.; et al. Evaluation of coalbed methane resources in Xinjing Baoan block based on PCA, TOPSIS, & MLFM. *Energy Explor. Exploit.* **2022**, *40* (5), 1457−1481, DOI: 10.1177/01445987221096750.
(37) Hu, T.; Li, X.; You, J. The weighted fusion prediction algorithm of acoustic interval optimized by PCA-PSO-BP and MLRM. *J. Phys.: Conf. Ser.* **2022**, *2258* (1), No. 012004, DOI: 10.1088/1742-6596/2258/1/012004.
(38) Ji, B.; Zhang, H.; Liu, S.; et al. A new recognition method for oil pipeline leakage using PCA and SOM neural networks. *IOP Conf. Ser.: Earth Environ. Sci.* **2021**, *783* (1), No. 012167, DOI: 10.1088/1755-1315/783/1/012167.
(39) Liu, Y.; Wei, S.; Guan, Z.; et al. Experimental Study on Rock Breaking and Drilling Characteristics of Hard Rock in Rotary Percussion Drilling Method. *Exp. Technol. Manage.* **2022**, *39* (05), 44−59.
(40) Wang, W.; Liu, X.; Dou, D.; et al. Deep Mechanical Drilling Speed Prediction Method Based on Neural Networks. *Pet. Drill. Tech.* **2018**, *40* (S1), 121−124.
(41) Su, W.; Wu, Y.; Ha, F. Theoretical and Technical Research of Drilling Fluids to Improve Mechanical Drilling Speed. *China Pet. Chem. Stand. Qual.* **2020**, *40* (19), 124−126.
(42) Wang, X. Doubts on the Comprehensive Score Method in Principal Component Analysis. *Stat. Decis.* **2007**, No. 08, 31−32.
(43) Wang, X. Noteworthy Issues in the Application of Principal Component Analysis and Factor Analysis. *Stat. Deci.* **2007**, No. 11, 142−143.
(44) Zhou, S.; Bai, Y. Convergence analysis of Oja's iteration for solving online PCA with nonzero-mean samples. *Sci. China Math.* **2021**, *64* (04), 849−868.
(45) Cui, Y.-b.; Liu, C.-s. Blended Coal's Property Prediction Model Based on PCA and SVM. *J. Cent. South Univ. Technol.* **2008**, *15* (S2), 331−335, DOI: 10.1007/s11771-008-0482-0.
(46) Zhou, H.; Zhang, S.; Peng, J.; et al. Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35* (12), 11106−11115.
(47) Lin, D.; Lin, S.; Feng, O. Prediction of PM_2.5 Concentration Prediction Based on Informer. *Environ. Eng.* **2022**, *40* (06), 48−54, DOI: 10.13205/j.hjgc.202206006.
(48) He, Z.; He, Z.; Li, S.; et al. A ship navigation risk online prediction model based on informer network using multi-source data. *Ocean Eng.* **2024**, *298*, No. 117007, DOI: 10.1016/j.oceaneng.2024.117007.
(49) Zhuang, W.; Li, Z.; Wang, Y.; et al. GCN−Informer: A Novel Framework for Mid-Term Photovoltaic Power Forecasting. *Appl. Sci.* **2024**, *14* (5), No. 2181, DOI: 10.3390/app14052181.
(50) Zili, W.; Yujun, Y.; Shuyou, Z.; et al. A multi-state fusion informer integrating transfer learning for metal tube bending early wrinkling prediction. *Appl. Soft Comput.* **2024**, *151*, No. 110991, DOI: 10.1016/j.asoc.2023.110991. (1) Bourgoyne, A. T.; Young, F. S. A Multiple Regression Approach to Optimal Drilling and Abnormal Pressure Detection. *Soc. Pet. Eng. J.* **1974**, *14* (04), 371−384.

https://doi.org/10.1021/acsomega.3c10339
ACS Omega 2024, 9, 23822−23831