Article

Prediction of Lost Circulation in Southwest Chinese Oil Fields Applying Improved WOA-BiLSTM

Xianming Liu [1], Wen Jia [1], Zhilin Li [2], Chao Wang [1,*], Feng Guan [1], Kexu Chen [2] and Lichun Jia [2]

[1] School of Mechanical Engineering, Yangtze University, Jinzhou 434000, China; jiawen0855@163.com (W.J.)
[2] Research Institute of Drilling and Production Engineering Technology, Chuanqing Drilling Engineering Co., Ltd., Guanghan 618300, China
* Correspondence: cw_paper@163.com

Abstract: Drilling hazards can be significantly decreased by anticipating potential mud loss and then putting the right well control measures in place. Therefore, it is critical to provide early estimates of mud loss. To solve this problem, an enhanced WOA (Whale Optimization Algorithm) and a BiLSTM (Bidirectional Long Short Term Memory) optimization based prediction model of lost circulation prior to drilling has been created. In order to minimize the noise in the historical comprehensive logging data, a wavelet filtering technique was first used. Then, according to the nonlinear Spearman rank correlation coefficient between mud loss and logging parameter values from large to small, seven characteristic parameters were preferred, and the sliding window was used to extract the relevant data. Secondly, the number of neurons in the first and second hidden layers, the maximum training time, and the initial learning rate of the BiLSTM model were optimized using the enhanced WOA method. The BiLSTM network was given the acquired superparameters in order to improve the model's ability to predict occurrences. Finally, the model was trained and tested using the processed data. In comparison to the LSTM model, BiLSTM model, and WOA-BiLSTM model, respectively, the improved WOA-BiLSTM early mud loss prediction in southwest Chinese oil fields suggested in this study beat the others, receiving 22.3%, 18.7%, and 4.9% higher prediction accuracy, respectively.

Keywords: lost circulation prior to drilling; prediction model; correlation analysis; improved whale optimization algorithm; Bidirectional Long Short Term Memory

1. Introduction

Mud circulates from the annulus back to the earth as it travels through the drill pipe during the drilling operation. Mud is essential for maintaining hydrostatic pressure, wellbore stability, and bit temperature, in addition to being utilized for suspending cuttings. As a consequence, the wellbore's mud circulation does not want to cause mud loss [1]. However, deep drilling in complex geological environments is becoming more common with increased exploration and development, particularly in the process of drilling "three high" oil and gas wells, which involves drilling into the cracks of carbonate rocks or other abnormal pressure geospheres with complex and variable pressure systems [2]. It is challenging in this situation to precisely assess the ground pressure and geological conditions, and it is easy to introduce the risk of mud loss. In addition, the amount of mud loss is different, and the corresponding plugging methods are different. If the mud loss is detected in advance, it can be solved by adding plugging agents within 48 h to prevent it from further developing into more serious mud loss [3,4]. Therefore, the purpose of this study is to train neural networks using historical drilling data and establish a predictive model capable of forecasting early mud circulation loss. By predicting these early mud losses, appropriate measures can be taken in advance to prevent or, at the very least, significantly mitigate early mud loss. This approach aims to address downhole crossflow, blowout, and wellbore instability incidents in their infancy, thereby achieving safe and efficient drilling.

In recent years, several researchers both domestically and internationally have begun to apply neural network approaches for lost circulation prior prediction. Moazzenii et al. (2010) built a multilayer feed-forward network learned by backpropagation to forecast lost circulation events in the Maroun oil field [5]. Jahanbakhshi et al. (2014) created a multilayer perceptron model to estimate mud loss and demonstrate the impact of geomechanical factors [6,7]. Aljubran et al. (2017) created many ML and DL models to predict circulation loss, including RF (Random Forest), ANN (Artificial Neural Network), CNN (Convolutional Neural Network), and LSTM (Long Short-Term Memory). The CNN model was shown to be the best [8]. Sabah et al. (2019) created many smart systems to anticipate circulation loss in the Maroun oil field such as MLP (Multi-Layer Perceptron), RBF (Radial Basis Function), GA-MLP (Genetic Algorithm Multi-Layer Perceptron, DP (Decision Tree), and ANFIS (Adaptive Neuro-Fuzzy Inference System). The findings revealed that DT is the best prediction mode [9–15]. Ahmed et al. (2020) employed artificial neural network models to foresee lost circulation in both naturally occurring and artificially produced fractures [16,17]. Mardanirad et al. (2021) used a comparison between different DL (deep learning) algorithms, CNN (Convolutional Neural Network), GRU (Gated Recurrent Unit), and LSTM (Long Short-Term Memory) for the classification of mud loss intensity in the Azadegan oil field, which showed the superior accuracy of the LSTM compared to other DL algorithms [18–20]. Jafarizadeh et al. (2022) used a fusion of an optimization algorithm and a modular neural network to address the problem of mud loss. The topology, threshold, and weight of the neural network were optimized to effectively solve the shortcomings of the traditional neural network, such as improper setting of hyperparameters and easy to fall into local optimization [21]. SiamiNamini et al. (2022) carried out depth analysis using traditional RNN, LSTM, and BILSTM network algorithms in deep water drilling condition identification. The results showed that the BILSTM network has good performance [22]. Xiang et al. (2022) predicted horizontal in situ stresses by using a CNN-BiLSTM-Attention hybrid neural network. The verification showed that compared with convolutional neural networks, LSTM and BiLSTM can extract the autocorrelation characteristics of the dynamic changes of the comprehensive logging curve and can better predict [23]. Li et al. (2022) proposed a deep learning method for early mud loss prediction based on the CNN-LSTM fusion network. They verified that the prediction accuracy of the network structure fused by the optimization algorithm is better than that of the CNN or LSTM structure alone [24].

BiLSTM has a distinct advantage in dealing with the complex mapping relationship of high-dimensional nonlinear long time series and can fully account for the time effect and parameter influence of the drilling process, which has good potential in mud loss prediction, according to research of existing early mud loss prediction models. At the same time, the network's prediction accuracy varies substantially according to the effect of structural characteristics. If the network parameters are not appropriately configured, the trained model will struggle to obtain the desired result. Furthermore, at the moment, early mud loss prediction is generally used to forecast whether or not mud loss would occur, whereas there is little research on mud loss volume prediction. Therefore, it is necessary to select a suitable neural network method to realize the early prediction of mud loss during the drilling process so as to guide the drilling operation more effectively.

In order to solve the problems in the existing methods, this paper chooses a two-layer BiLSTM as the basic neural network. The improved WOA is used to optimize the number of neurons in the input layer, the number of neurons in the hidden layer, the maximum training period, and the initial learning rate in the BiLSTM structure [25]. Based on the comprehensive consideration of measurement while drilling parameters, logging parameters, and fine pressure control drilling parameters, an improved early prediction model of WOA-BiLSTM of mud loss is constructed. First, textual data are converted into numerical values. Characteristic parameters are then selected through Spearman rank correlation analysis. Subsequently, wavelet filtering is applied to mitigate the impact of noise on the data. The selected characteristic parameters are used. Finally, the

collected data are partitioned into training sets, test sets, and validation sets for training and verification purposes.

2. Relevant Theories

Data preprocessing is an important step in building a real model. A good data preprocessing process includes important steps such as data denoising, data conversion, and data dimensionality reduction [26].

2.1. Data Denoising

When we talk about real data, noise is an inevitable component, with at least 5% even under the strictest controls [27]. In this study, the term "data denoising" refers to the use of filtering to lessen the effect of noise on the data. Data conversion requires the use of data in multiple units, and the distribution of hyperparameters may be impacted by the scale disparity, homogenizing the processing. The Spearman rank correlation coefficient's correlation analysis is used to determine the order of the influencing factors during data dimensionality reduction. The main factors are then chosen from a list of parameters affecting drilling mud loss.

2.2. Data Normalization

At present, the main methods of dimensionless data processing are standardization, averaging, and standard deviation [28]. Considering that the covariance matrix composed of the original data after averaging processing can not only reflect the difference in the degree of variation of each index in the data but also contain information on the degree of mutual influence of each index, the data in this paper chose the averaging method to normalize it to the scale range of $[-1, 1]$. The equation for the data normalization is given below:

$$X = 2\left(\frac{X - X_{min}}{X_{max} - X_{min}}\right) - 1 \tag{1}$$

where X_{max} represents the data maximum; X_{min} represents the data minimum.

2.3. Feature Selection

When using the neural network model to train the sample data, we need to consider the high dimension, which will cause the neural network model to run slowly and consume hardware. In addition, in the case of large data dimensions, there is the problem of "dimension disaster" [29]. Therefore, it is necessary to select features of the data to achieve the purpose of dimensionality reduction. The Spearman rank correlation coefficient, also called the rank correlation coefficient, is a nonparametric statistic whose value is unrelated to two groups of variables related to the specific value but only the size of the relationship between its values. Therefore, it is very suitable for studying the correlation between nonlinear relations.

2.4. LSTM Principle of Neural Network

LSTM neural network is another neural network algorithm improved on the basis of RNN neural network to solve time series problems [30,31]. The structure is shown in Figure 1. The shaded areas represent the previous and next moments, while the non-shaded area represents the current moment. By adding three control units, forgetting gate, memory gate, and output door, the network can alleviate the problem of RNN being prone to gradient explosion and gradient disappearance with a special way of storing "memory" and setting gradient range threshold. The main operation flow of LSTM is as follows:
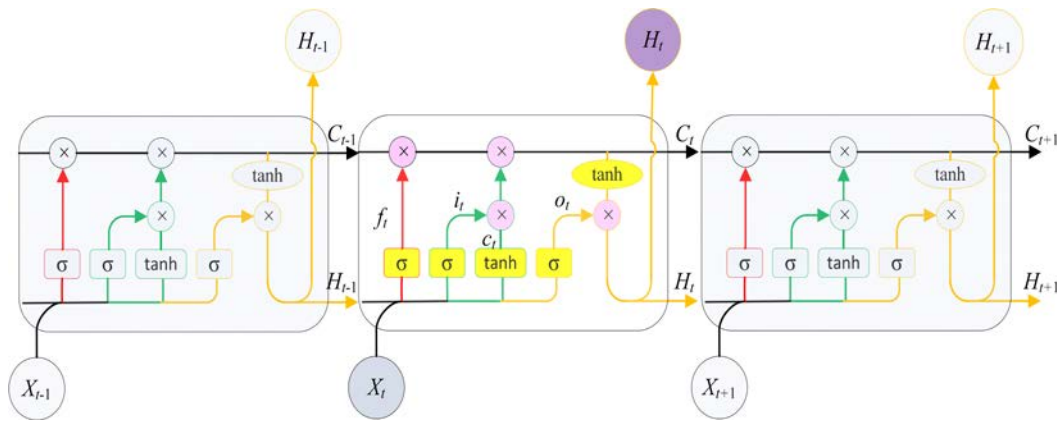
Figure 1. LSTM structure.

(1) Use the forgetting door, as shown by the red arrow in Figure 1, combined with the output of the previous moment H_{t-1}, to cellular information C_{t-1} filtering; purposefully screen out the cell information that has influence on this moment. The mathematical expression is as follows:

$$f_t = \sigma(W_f \cdot [H_{t-1}, x_t] + b_f) \tag{2}$$

where H_{t-1} represents the last moment output; f_t represents the output of the forgetting door; x_t represents the input at the current moment; σ represents the sigmoid function; W_f and b_f, respectively, represent the weight coefficient and offset of the linear relationship.

(2) The memory gate, as shown by the green arrow in Figure 1, is used to retain the effective information of the cell information C_{t-1} combined with the output H_{t-1} of the previous moment. The mathematical expression is as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$

$$c_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_c) \tag{4}$$

$$C_t = f_t \times C_{t-1} + i_t \times c_t \tag{5}$$

where i_t represents the put of the first part; c_t represents the put of the second part; W_i, b_i, and W_c represent the corresponding weight coefficient and offset, respectively; C_t represents the state of the updated cells.

(3) The output door, as shown by the orange arrow section in Figure 1, combined with the output of the previous moment H_{t-1} and current cell information C_t after the calculation, input to the neural network for operation. The mathematical expressions are as follows:

$$o_t = \sigma(W_O[h_{t-1}, x_t] + b_o) \tag{6}$$

$$h_t = o_t \times \tanh(C_t) \tag{7}$$

where o_t represents the state of the upper hidden layer; the value of h_{t-1} and x_t is computed by the sigmod function.

2.5. BiLSTM Principle of Neural Network

The basic principle of BiLSTM neural network is composed of two LSTM neural networks [32], and the training sequence can be transmitted forward and backward. It can achieve a more complete analysis of the characteristics and laws of the data. Below, Figure 2 shows the structure of a single-layer BiLSTM neural network expanded over time. The shaded areas in the diagram represent the previous and next moments, while

the non-shaded areas represent the current moment. At the same time, the blue arrows represent forward propagation, and the yellow arrows represent backward propagation, thus achieving bidirectional propagation. x is the input value of the neuron. The hidden layer of the bidirectional convolutional neural network needs to save two values, "A" participates in forward calculation, and "A^*" participates in backward calculation. The final output value y depends on both "A" and "A^*".
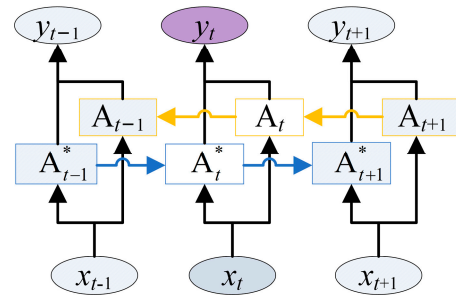

Figure 2. Single layer BiLSTM neural network.

2.6. Principle of Optimization Algorithm

2.6.1. Whale Optimization Algorithm (WOA)

WOA is a metaheuristic swarm intelligence optimization algorithm which mainly includes three steps: randomly searching for prey, encircling target prey, and preying on selected prey [33].

(1) Finding the solution to a problem is the process of finding prey by a herd of whales, choosing a prey at random, and the process can be translated into the following:

$$D = |X(t) - CX_{rand}(t)| \tag{8}$$

$$X(t+1) = X_{rand}(t) - AD \tag{9}$$

where $X_{rand}(t)$ represents randomly selected position vectors from the current population of whales; $X(t)$ represents the position vector of the individual; t represents the current iterations; A and C represent the coefficient and are calculated as follows:

$$A = 2a \cdot r_1 - a \tag{10}$$

$$C = 2r_2 \tag{11}$$

$$a = 2 - 2t/T_{max} \tag{12}$$

where r_1 and r_2 are random vectors belonging to the interval [0, 1]; a is linear reduction from 2 to 0 during iteration; T_{max} is the maximum number of iterations.

(2) The best candidate solution is the target prey or a near optimal solution. After the optimal solution is found, the other candidate positions will move closer to the target prey, surround the prey, and update its position. The mathematical model is as follows:

$$D = |X(t) - CX^*(t)| \tag{13}$$

$$X(t+1) = X^*(t) - AD \tag{14}$$

where t represents the current iterations; A represents the coefficient; $X^*(t)$ represents the current best position; $X(t)$ represents the current location.

(3) Humpback whales update their positions by spiraling up to hunt selected prey. The mathematical model is as follows:

$$X(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) \tag{15}$$

$$D' = \left| X^* - X(t) \right| \tag{16}$$

where b represents the constant; l represents a random number in the interval $[-1, 1]$. In the Formula (16), D' represents the distance between the best whale individual in the t iteration and the current whale.

In order to achieve this simultaneous behavior, it is assumed that there is a 50% chance to choose whether to shrink the encirclement mechanism or the spiral model to update the position of the whale in the optimization process. The mathematical model is as follows:

$$X(t+1) = \left\{ \begin{array}{ll} X^*(t) - A \cdot D & p < 0.5 \\ X^*(t) + D' \cdot e^{bl} \cdot \cos(2\pi l) & p \geq 0.5 \end{array} \right. \tag{17}$$

where p represents a random number in the interval $[0, 1]$.

2.6.2. Improved Whale Optimization Algorithm (WOA)

The original WOA still has some disadvantages similar to other swarm intelligence optimization algorithms [34], such as low solution accuracy, slow convergence speed, and easy to fall into local optimization. In order to overcome these shortcomings, this paper will improve WOA from two aspects: location update strategy and prevention of falling into local optimization [35].

(1) Nonlinear convergence factor

As the WOA Formula (9) knows, the global and local exploration abilities of the algorithm mainly depend on the parameters A as setting a larger A in the early stages of the iteration can speed up the algorithm. The algorithm's ability to perform local searches is enhanced by lowering parameter A in later iterations. By Formulas (10) and (12), it is known that the value of the parameter A mainly depends on the convergence factor a. In this research, a nonlinear convergence factor a is proposed because the linear variation of convergence factor a cannot demonstrate its searching ability. The mathematical model is as follows:

$$a = 2 - 2\sin(\mu \frac{t}{T_{max}} \pi + \varphi) \tag{18}$$

where t is current iterations; T_{max} is maximum iterations.

(2) Adaptive weight strategy and random difference variation strategy

In order to keep the diversity of the population and jump out of the local optimization in time, Yao Ning proposes an adaptive weight strategy and a random difference mutation strategy [31]. The mathematical expression of the adaptive weight strategy is as follows:

$$w(t)_i = \left\{ \begin{array}{ll} w_1 - \frac{w_2 - w_1}{T_{max}} \cdot \frac{f(t)_i - f(t)_{min}}{f(t)_{max} - f(t)_{min}} & f(t)_i < f(t)_{avg} \\ w_1 - \frac{w_2 - w_1}{T_{max}} \cdot \frac{f(t)_i - f(t)_{min}}{f(t)_{max} - f(t)_{min}} & f(t)_i > f(t)_{avg} \end{array} \right. \tag{19}$$

where $w(t)_i$ represents the weight of the search i in the iteration t; T_{max} represents the maximum number of iterations; w_1 represents the initial minimum weight; w_2 represents the initial maximum weight; $f(t)_{avg}$ represents the average adaptation value of the population after the current iteration of t times; $f(t)_{min}$ and $f(t)_{max}$ represent the minimum and maximum fitness values after the iteration of t times, respectively. The position update strategy expression occurs when the improved prey is formed by substituting the Formula (19) into (17) is as follows:

$$X(t+1) = \left\{ \begin{array}{ll} w(t)gX^*(t) - AgD & p < 0.5 \\ w(t)gX^*(t) + D'ge^{bl}g\cos(2\pi l) & p \geq 0.5 \end{array} \right. \tag{20}$$

The mathematical expression of the random difference variation strategy is as follows:

$$X(t+1) = r_1 \times (X^*(t) - X(t)) + r_2 \times (X'(t) - X(t)) \tag{21}$$

where r_1 and r_2 are random numbers within the range of [0, 1]; $X'(t)$ is randomly selected individuals from a population.

2.7. Improved WOA-BiLSTM Prediction Model of Early Mud Loss

The selection of structural parameters of the BiLSTM prediction model has a great influence on the final prediction ability of the model. In order to find the optimal superparameters of the early mud loss prediction model, the WOA is used to optimize the number of units in the hidden layer (L_1, L_2), the maximum cycle period (T), and initial learning rate of the cycle (I_r). Taking these four key hyperparameters as the characteristics of optimization, the WOA algorithm is used to adjust and optimize the LSTM model to make the network structure model more compatible with the characteristics of comprehensive logging data. The main implementation steps are shown in Figure 3. Firstly, the yellow flowchart represents the computational process of BiLSTM. The initial steps involve inputting historical well data and conducting correlation analysis on the data. Subsequently, training is performed using a predictive model with conventional parameter settings to select highly correlated predictive parameters. Then, the selected parameters are subjected to wavelet filtering for noise reduction. After identifying the four hyperparameters that need optimization, the blue flowchart illustrates the process of optimizing these parameters using the training error of the BiLSTM model as the fitness value. Finally, an ideal predictive model is obtained through training.
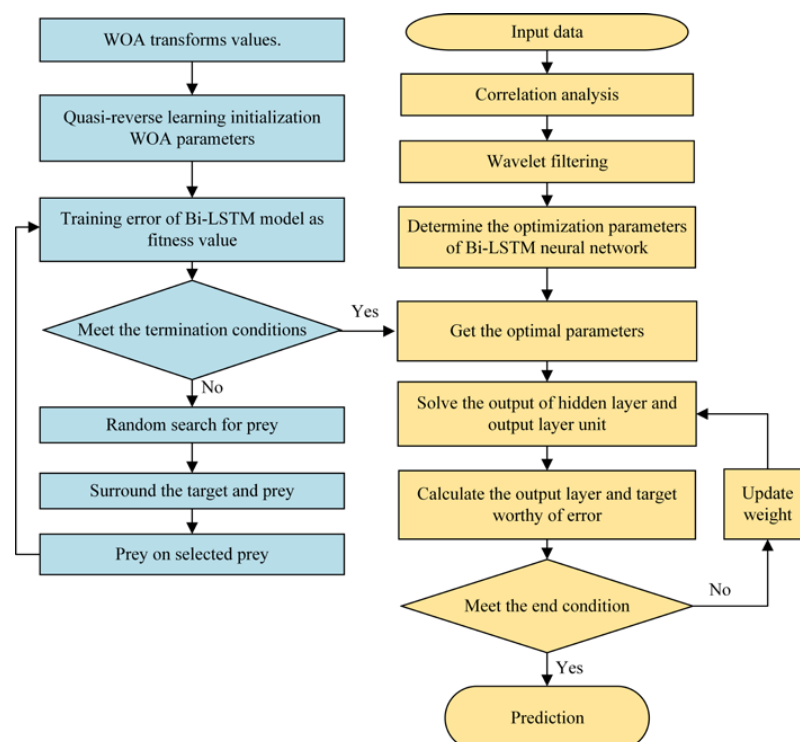

Figure 3. WOA-BiLSTM predictive model flow chart.

3. Experimental Study

3.1. Input Data

The data used for the comprehensive recording of wells came from three oil and gas wells located in the Sichuan Basin in the southwestern region of China, named respectively as A, B, and C. A full set of drilling data was collected and integrated in the drilling

conditions of the three wells within 5 h prior to the loss of sludge, including PWD (Pressure While Drilling) underground pressure and surface microflow monitoring data during drilling. The extracted data were stacked together to form a sample data set. The interval between each set of data collected was 20 s, the data within 10 min were selected as a time sequence and the length of the sequence was 30 time steps [34]. The integrated data consisted of 144,000 sets of data. The 115,200 (80%) groups of data were modeled as data A and the remaining 28,800 (20%) groups were modeled for data B validation. Some of the primary historical data collected include drilling parameters as shown in Figure 4. Figure 4A–H offers an in-depth depiction of the drilling process, showcasing the fluctuations in crucial parameters such as well depth, casing pressure, hook weight, torsion moment, standpipe pressure, drilling pressure, inlet flow, and total pool volume. The substantial variations in the hook weight's load and torque are particularly striking, underscoring the dynamic nature of the drilling operation.
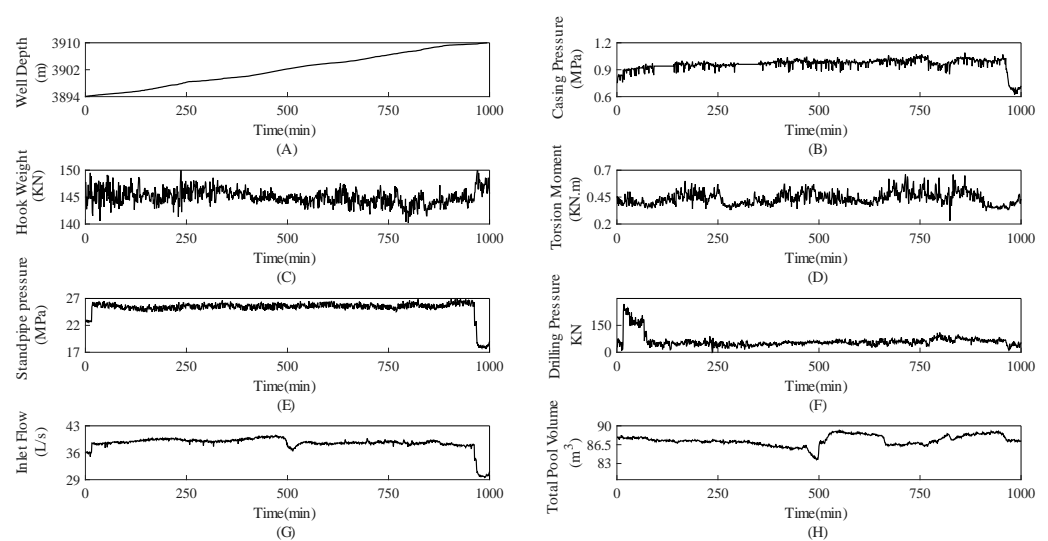
Figure 4. Partial logging data.

From Figure 4A,H, it can be clearly seen that at approximately 3902 m, a significant turning point is observed in the total volume of the mud pool. This decrease suggests a possible mud loss incident prior to this depth. This critical observation underscores the importance of continuous monitoring and analysis in identifying potential issues during drilling operations. In addition to these findings, it's also crucial to consider other factors that may influence these parameters. Therefore, we conducted a correlation analysis on multiple drilling parameters to identify the most sensitive key parameters, in order to better establish a predictive model.

3.2. Correlation Analysis

The occurrence of mud loss will be demonstrated by comprehensive drilling recording parameters. The characteristic parameters for extracting the state of mud loss from large amounts of data are a key step in predicting well loss using the BiLSTM neural algorithm. The actual drilling process at the site is mostly used to observe the changes in the total pool volume as a judgment criterion for the mud loss. Therefore, in this article, the amount of change in the total pool volume is used as a "reference value", by analyzing the Spearman rank correlation coefficient to evaluate the depth of the well, the steering pressure, the measurement of the pipe pressure and set the piping pressure, the position of the flow valve, suspension, drilling pressure, pumping, input flow, input density, meter drilling time, working current, height, speed, torque, and the degree of association of mud loss. This article provides an analysis based on relevance. From Figure 5 it can be seen that the surface pressure pump correlation is the highest at 0.865, followed by torque correlations of

0.657. According to the correlation values of the feature parameters, the feature parameters are sorted from high to low. The sorted parameters are shown in Table 1.
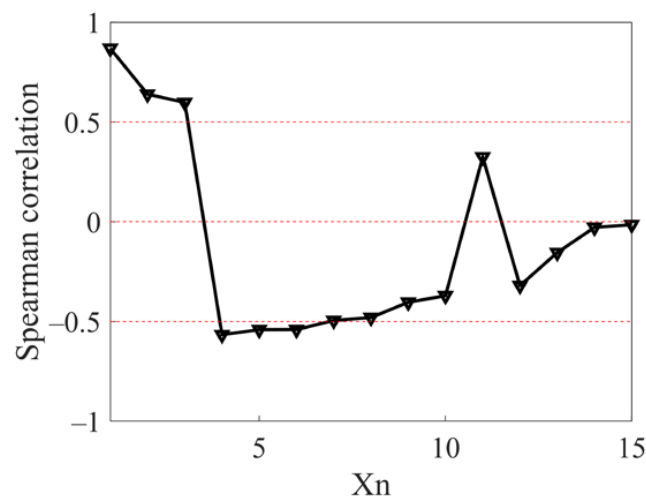

Figure 5. Grey relational degree comprehensive evaluation.

Table 1. Values of the gathered parameters drilled in Southwest Chinese Oil fields.

X_n	Parameter	Unit	Min	Max	Mean
X_1	Surface Pressure Pump	MPa	10.02	29.98	20.00
X_2	Torque	KN.m	0	1.73	0.86
X_3	Casing Pressure	MPa	0	8.77	4.39
X_4	Hook Weight	KN	0	774.4	387.2
X_5	Inlet Flow	L/min	0	49.03	24.51
X_6	Inlet Density	g/cm^3	0	2.52	1.26
X_7	Measured Depth	m	810.00	5540.00	3175.00
X_8	Overpull	KN	210.80	1000.00	605.40
X_9	Tripping in	h	0.89	2.98	1.93
X_{10}	Choke position	m	13.23	95.03	54.13
X_{11}	Rotational speed	rpm	12.00	70.00	41.00
X_{12}	Traveling block height	m	10.00	30.00	20.00
X_{13}	Pore pressure	MPa	8.65	72.45	40.55
X_{14}	Fracture pressure	MPa	13.58	93.83	53.71
X_{15}	Drilling time	min/m	0.94	10.00	5.47

Based on extracted feature parameters, it is observed that within the same block, mud loss events in oil and gas wells typically occur within a specific range of measurement depths. This phenomenon arises due to variations in reservoir pressure, differing rock properties at various depths, and discrepancies in inlet flow rates and densities. These factors reflect distinct drilling conditions, thereby influencing the incidence of mud loss.

The recorded changes in casing pressure reflect fluctuations in annular pressure, while variations in mud density affect the pressure differential between the wellbore and the formation. During mud loss incidents, drilling pressure, annular casing pressure, and torque also undergo changes in response to subsurface conditions. When mud loss occurs, drilling pressure decreases, annular casing pressure diminishes, and concurrently, drill bit speed increases, resulting in an upsurge in inlet flow rates. In contrast, during wellbore overflow events, mud density decreases, and with the overall increase in mud pit volume, annular casing pressure reduction occurs gradually. This, in turn, leads to a decrease in the impact of buoyancy on downhole fluids as mud density decreases, ultimately resulting in an increase in hook load.

Considering the on-site testing accuracy of engineering parameters and the computational efficiency of the prediction model, we determined the order of sorted drilling

parameters based on correlation analysis to increase the number of input neurons in the BiLSTM model, denoted as 'n'. The criterion for selecting the number of input parameters is based on the prediction model's error. From Figure 6, it is evident that when the number of input neurons reaches 7, the error reaches its minimum. With 8 neurons, there is a slight increase in error, after which the error remains relatively stable. Therefore, to conserve computational resources and optimize the efficiency of the prediction model, we chose 7 drilling parameters as input neurons.
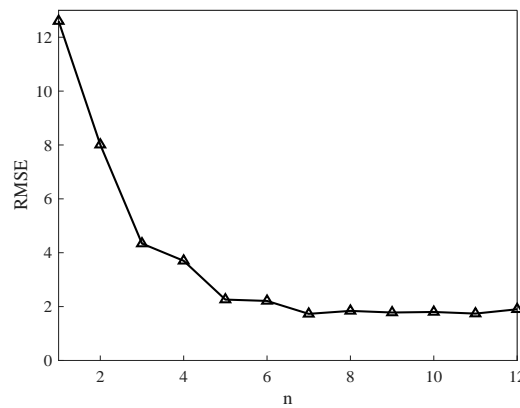
Figure 6. Feature parameter extraction.

In the process of oil drilling, the movement of the previous period of time affects the next movement. Therefore, in the tag construction, the 8 drilling parameters selected by the feature over the length of the time series are used as the multidimensional variable X. The change in total pool volume after the time series length is used as a regression prediction of mud loss (label Y). In this paper, a time series matrix with time series length of 30 is constructed by using the time window sample structure method. Select the time series matrix using the form of window slide, and the time step of each slide is 1. The drilling data within the first 10 min are used to predict the mud loss at the next moment, achieving the purpose of predicting 10 min in advance, as shown in Figure 7. The red box represents the matrix of the first sample data input, while the green and blue boxes represent the input of sample data for the next and the following moments, respectively, proceeding in an orderly manner.
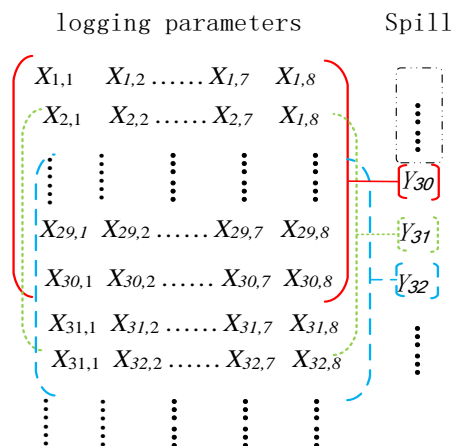

Figure 7. Time window matrix extraction.

3.3. Wavelet Filtering

Data processing involves hard thresholding, soft thresholding, and fixed thresholding. Common indicators for evaluating the effect of wavelet threshold denoising include Signal-to-Noise Ratio (SNR), Root Mean Square Error (RMSE), smoothness, and correlation

coefficient. In this study, we use SNR and RMSE to evaluate the effect of wavelet threshold denoising. SNR is defined as the ratio of the energy of the original signal to the noise signal. The higher the SNR, the better the denoising effect. RMSE is the square root of the variance between the original signal and the denoised signal. The smaller the RMSE value, the better the denoising effect.

As shown in Figure 8, the green line, blue line, and red line represent denoising methods using hard thresholding, soft thresholding, and fixed thresholding respectively. From Figure 8A, it can be seen that when using hard thresholding and soft thresholding, the SNR values are similar, but overall lower than the noise reduction effect of using a fixed threshold. This indicates that filtering with a fixed threshold is better. Similarly, from Figure 8B, it can be seen that when using a fixed threshold, the RMSE values are overall lower than those obtained using hard thresholding and soft thresholding methods. This again proves the superiority of filtering with a fixed threshold. Therefore, in our research, we chose to filter with a fixed threshold.
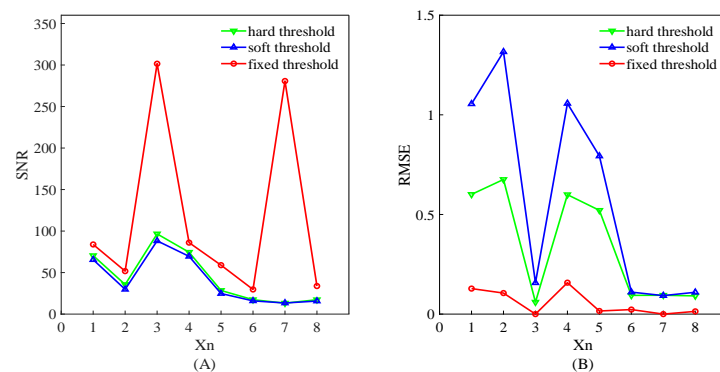

Figure 8. Wavelet noise reduction ratio and square error.

The 8 data after feature selection are displayed by wavelet filtering. Figure 9A–H represent measured depth (MD), surface pressure pump (SPP), torque (TQ), casing pressure (CP), weight on bit (WOB), inlet flow rate, outlet flow rate, and total pool volume (TPV) respectively. the black line represents the original data before filtering, and the red line represents the data after wavelet filter denoising.
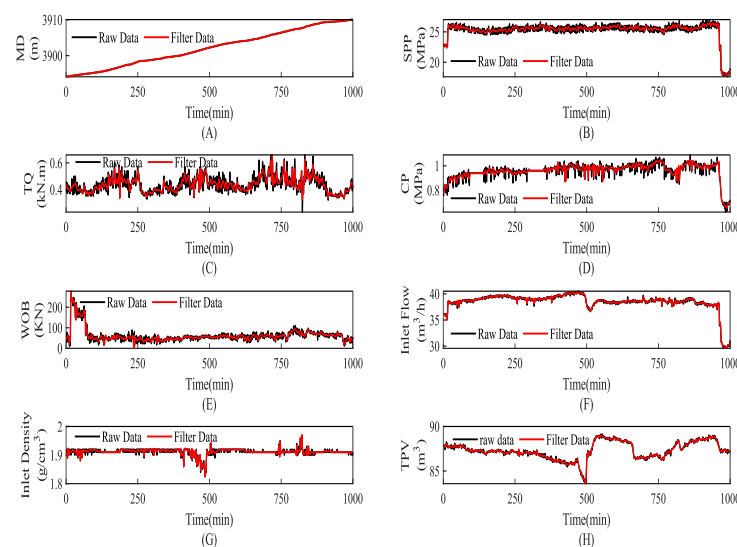

Figure 9. Wavelet Filtering.

Figure 9H illustrates two significant events in the change of total pit volume. At approximately 500 min, there is a decrease in total pit volume, indicating mud loss occurring

around a drilling depth of approximately 3902 m. Around 890 min, the total pit volume first increases and then decreases, signifying an occurrence of mud overflow followed by mud loss around a drilling depth of approximately 3910 m. We can observe noticeable fluctuations in inlet density, inlet flow rate, torque, drilling pressure, and casing pressure before and after these events, indicating their correlation with changes in total pit volume. This suggests that these parameter variables can be used to establish a predictive model.

3.4. Improved WOA Optimization Algorithm Parameter Design

Selecting the appropriate number of hidden layer neurons and the right initial learning rate is a crucial decision in neural network design. Typically, this requires experimentation and validation to determine the optimal configuration that meets the specific task requirements, while ensuring the network possesses excellent performance and generalization ability. Increasing the number of hidden layer neurons can expand the network's capacity and generalization capabilities, enabling it to better fit complex data patterns and functional relationships. However, if there are too many neurons, it may lead to overfitting the training data, thereby reducing generalization performance.

The choice of learning rate is equally critical. Smaller learning rates usually demand more training epochs because each weight update has a smaller magnitude, while larger learning rates can result in rapid model changes, requiring fewer training epochs. Inappropriate learning rate settings can lead to the model converging quickly to suboptimal solutions. Due to the lack of mature theoretical guidance, this study relied on existing research results to determine these key parameters [36,37].

The parameter settings for the WOA optimization algorithm are as follows: a population size of 100, with a maximum weight of 0.9 and a minimum weight of 0.2. Additionally, the whale algorithm is configured with a population size of 50 and a maximum iteration count of 30. Considering the need to optimize four parameters, each corresponding to a dimension, constraints were applied within a limited parameter search space, as outlined in Table 2. Simultaneously, the number of neurons in the first and second layers of the neural network was set within a range of 10 to 50, the initial learning rate was within the range of 0.001 to 0.01, and the maximum cycle count was within the range of 50 to 200.

Table 2. Bi-LSTM optimizing parameter range table.

Parameter	L_1	L_2	T	I_r
scope	10~50	10~50	0.001~0.01	50~200
step size	2	2	0.0005	5

According to the parameters and constraints set above, the genetic algorithm, the particle swarm optimization algorithm, and the improved whale optimization setting parameters are shown in Table 2 (WOA maximum weight, 0.9; minimum weight, 0.5; population size, 50; the maximum number of iterations, 100).

The weight of WOA-BiLSTM tends to be stable at about 63 generations and the best fitness value is 0.022. Compared with the genetic algorithm and the particle swarm optimization algorithm, the convergence speed is faster and the error value is the smallest. Select the population optimal solution when the number of iterations is 100, $[L_1, L_2, T, I_r]$ = [9, 12, 16, 0.03, 100], as the combined value of the parameters to be optimized for the Bi-LSTM structure.

4. Model Evaluation

Figure 10A,B illustrate the comparison between the predicted results of four models and the actual values from randomly selected test data sets. The black line represents our prediction target, which is the total pool volume change (i.e., mud circulation loss). The green line shows the prediction results obtained using the LSTM model, the dark blue line shows the prediction results obtained using the BiLSTM model, the purple line shows the

prediction results obtained using the WOA-BiLSTM model, and the red line shows the prediction results obtained using the improved WOA-BiLSTM model.
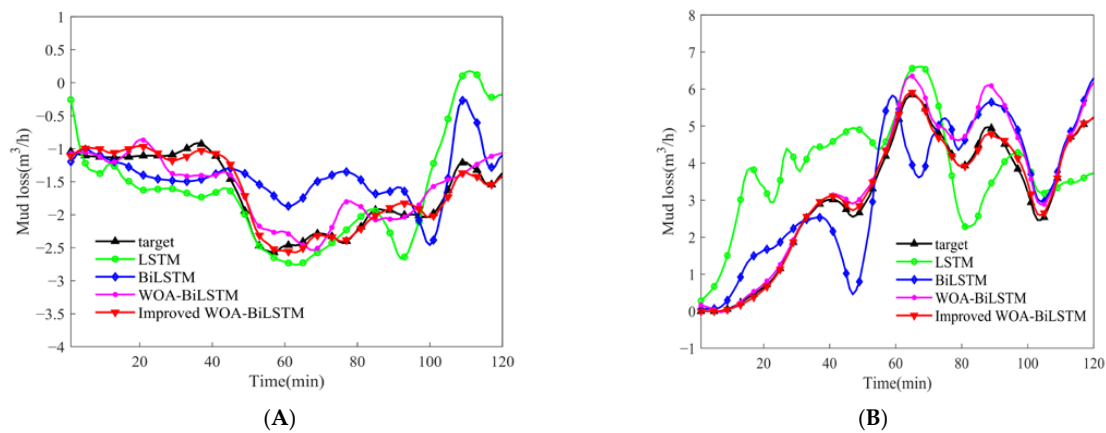

Figure 10. Regression graph.

In Figure 10A, real drilling data from a measurement depth range of 3900 m to 3901.5 m are used, while in Figure 10B, data from a measurement depth range of 4710 m to 4712 m are employed. The objective is to predict the change in total pool volume 10 min in advance to assess the occurrence of mud circulation loss, as indicated in Figure 10A for potential mud circulation loss and in Figure 10B for potential drilling fluid overflow, in the respective depth ranges.

Through comparative analysis, it was observed that LSTM and Bi-LSTM models with random parameters exhibited significant disparities between their training and testing sets. These models displayed fluctuations and deviations from the target values throughout the training process, indicating overfitting and resulting in suboptimal test results. Conversely, the WOA-BiLSTM and the improved WOA-BiLSTM models did not exhibit overfitting and demonstrated a more accurate performance. The research findings underscore the substantial impact of hyperparameter configurations on BiLSTM neural network models.

The statistic that measures goodness of fit is the coefficient of determination, also known as the coefficient of certainty (R^2), with a maximum value of 1. The closer the value of R^2 is to 1, the better the regression line fits the observed value. In addition, RMSE is also an evaluation index to measure the fitting performance with the target value, and the smaller the RMSE, the smaller the error. After 4320 groups of data verification, the comparison results are shown in Table 3. It can be found that the BiLSTM neural network prediction model shows better prediction performance than LSTM. At the same time, the realization of LSTM and BiLSTM in the training set and test set is very different, and the overfitting phenomenon appears. The prediction model optimized for BiLSTM neural network parameters by the WOA optimization algorithm shows relatively stable regression fitting performance, which also indicates that for different regression problems, the superparameter setting of LSTM and BiLSTM has a great impact on the performance of the neural network. Improper setting makes it easy to fall into the local optimal solution and has poor generalization ability. The swarm intelligent optimization algorithm can solve the problem of improper setting of superparameters. At the same time, prediction accuracy can be improved by improving the WOA optimization algorithm.

Table 3. R^2 and RMSE values.

		LSTM	BiLSTM	WOA-BiLSTM	Improved-WOA-BiLSTM
Training	RMSE	0.365	0.315	0.254	0.206
Training	R^2	0.821	0.908	0.942	0.996
Test	RMSE	0.448	0.412	0.271	0.225
Test	R^2	0.783	0.823	0.932	0.984

Furthermore, the improved WOA-BiLSTM model achieved the best performance in predicting mud loss on the test dataset, with an RMSE (Root Mean Square Error) of 0.225 and an R^2 (Coefficient of Determination) of 0.984. When compared to the three models mentioned above, this model demonstrated significantly closer alignment with actual values in both trends and accuracy. As shown in Figure 11, the graph illustrates the absolute error between the predicted and actual values. Figure 11A presents the error when predicting mud loss, with the absolute error primarily concentrated around ±0.2. When the absolute error of predicting mud loss is less than 0.2, the prediction accuracy can reach up to 90.8%. On the other hand, Figure 11B displays the error when predicting mud overflow, where the prediction accuracy can reach 88.3% if the absolute error of predicting mud overflow is less than 0.2. These results indicate that this model performs excellently in terms of fitting accuracy, stability, and predictive performance, and can effectively predict mud circulation loss. The output during the training phase shows that in the early stages of mud loss prediction for all models, the predicted values lag behind the actual values, which may be due to the time window function we used. At the same time, we found that the accuracy of predicting mud circulation loss is higher than that of predicting mud overflow. The reason for this may be that there are more sample data sets where mud loss occurs in the training sample data, so the training effect is better.
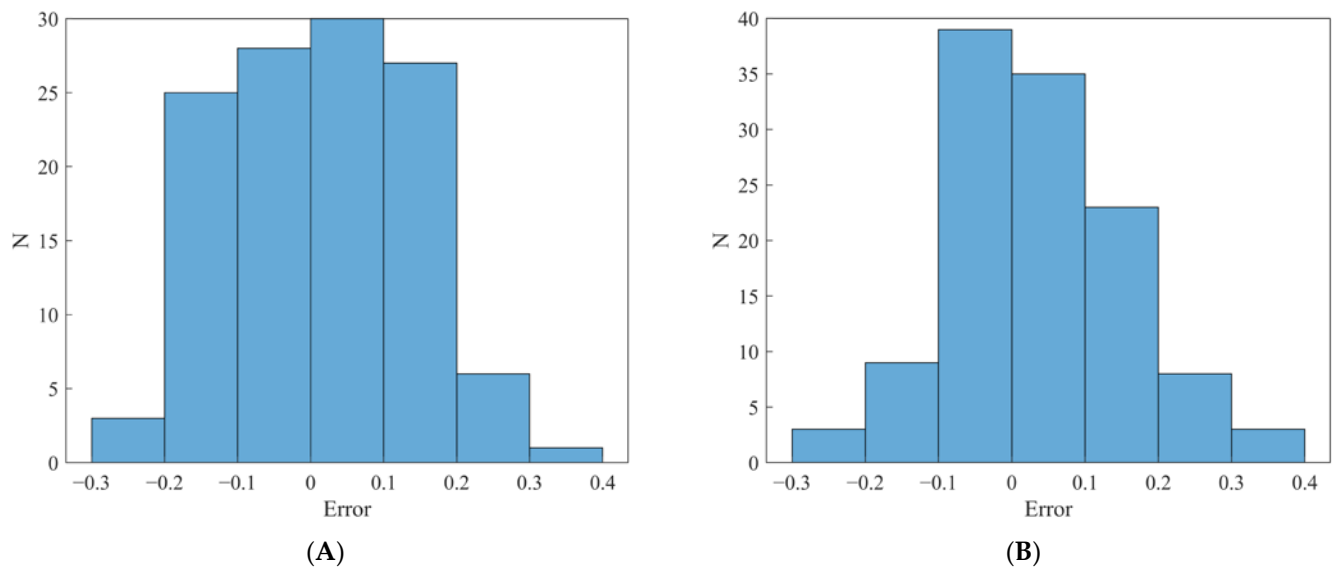

(**A**)

(**B**)

Figure 11. Error distribution.

5. Discussion

The primary innovation of this study lies in the amalgamation of historical drilling data with an improved WOA-BiLSTM. This fusion resulted in the development of a mud loss circulation prediction model. This model, by utilizing a time window matrix, is capable of forecasting changes in the total pool volume ten minutes in advance, thus replacing the conventional manual recording of the total pool volume. The application of this predictive model indirectly accomplishes the prediction of mud circulation loss, with the potential to assist in early risk identification and the implementation of corresponding well control measures, ultimately mitigating the risk of blowouts at an early stage. Key aspects highlighted by our research include:

(1) The research underscores the critical role of the time window matrix in data processing. Through the incorporation of the time window matrix during the training process, we have successfully achieved early predictions of total pool volume changes. This method provides a new avenue for applying artificial neural networks to predict other drilling data and highlights the critical role of the time window matrix in data processing. It enables early predictions of total pool volume changes, opening possibilities for applying neural networks to other drilling data.

(2) Using the improved WOA for hyperparameter selection enhances predictive accuracy, offering valuable guidance for handling complex drilling datasets.

(3) While our study has made significant strides, limitations remain. It focuses on specific block wells, necessitating validation across different blocks for mud circulation loss prediction efficacy. Limited dataset size and quality may constrain model performance. Future research should consider expanding data diversity and quantity to enhance capabilities.

6. Conclusions

The upgraded WOA-BiLSTM neural network is the foundation for the early mud loss prediction model used in this study. The model primarily makes use of the upgraded WOA to overcome the challenge of configuring the parameters of the conventional BiLSTM neural network and raise the model's prediction accuracy. The following are the main findings:

(1) According to the size of the linked coefficients, extracted characteristic parameters are sorted using Spearman rank-related coefficients. The outcomes of WOA optimization reveal that seven criteria can be used to obtain extremely good accuracy. Therefore, these seven traits were tested in the modeling of the BiLSTM neural network algorithm: total pool volume (TPV), inlet flow rate, inlet iensity rate, weight on bit (WOB), surface pressure pump (SPP), torque (TQ), casing pressure (CP), and measured depth (MD).

(2) In this study, the maximum cycle, the initial learning rate, and the number of units in hidden layers 1 and 2 of the Bi-LSTM neural network structure are all optimized using the enhanced WOA. Based on this, the three prediction models LSTM, BiLSTM, and WOA-BiLSTM are compared with the early prediction model of WOA-BiLSTM of mud loss. There has been an increase in prediction accuracy of 22.3%, 18.7%, and 4.9%, respectively. The findings demonstrate that the enhanced WOA-BiLSTM model is more accurate in estimating early mud loss.

(3) The model can estimate changes in the total pit volume 10 min in advance, thereby predicting loss circulation with a high degree of accuracy. This precise forecasting contributes significantly to taking timely countermeasures, reducing the adverse effects of mud loss on drilling operations. For on site operators, this functionality is crucial as it allows for better work planning and management, resulting in increased production efficiency and reduced environmental risks.

(4) The research addresses a critical research gap in the field of petroleum drilling. To date, there have been relatively limited methods for predicting and managing mud loss events, and the model offers an efficient approach to addressing this issue. This is of great significance in ensuring the smooth progress of drilling operations and minimizing unnecessary downtime. Additionally, it underscores the potential of machine learning and deep learning in the petroleum engineering field y introducing advanced computational methods into traditional drilling processes. This will contribute to accelerating the digital transformation of petroleum engineering, improving industry efficiency, and sustainability.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Peng, Q.; Fan, H.; Xu, S.; Zhou, H.; Lai, M.; Ma, G.; Fu, S. *A Real-Time Warning System for Identifying Drilling Accidents*; SPE: Astana, Kazakhstan, 2014; p. SPE-172303-MS. [CrossRef]
2. Yi, H.; Ming, L.; Wentuo, L.; Wenbiao, D.; Chao, Z.; Zhenxiang, Z.; Tianwei, Z.; Xin, Z.; Yuqun, H.; Can, Z.; et al. *Application of Formation Pressure While Drilling Prediction Technology in Offshore HTHP Wells*; OnePetro: Richardson, TX, USA, 2019.
3. Arshad, U.; Jain, B.; Ramzan, M.; Alward, W.; Diaz, L.; Hasan, I.; Aliyev, A.; Riji, C. *Engineered Solution to Reduce the Impact of Lost Circulation During Drilling and Cementing in Rumaila Field, Iraq*; IPTC: Doha, Qatar, 2015; p. IPTC-18245-MS. [CrossRef]
4. Liu, Z.; Luo, R.; Yang, Z.; Wang, L.; Wang, L. Research and Practice of Risk Early Warning Technology for Lost Circulation with Drilling under the Conditions of Geological Engineering Information Fusion: The Example of the Yuanba Area. *Processes* **2022**, *10*, 2516. [CrossRef]
5. Moazzeni, A.R.; Nabaei, M.; Shahbazi, K.; Shadravan, A. *Mechanical Earth Modeling Improves Drilling Efficiency and Reduces Non-Productive Time (NPT)*; SPE: Manama, Bahrain, 2010; p. SPE-131718-MS. [CrossRef]
6. Alkinani, H.H.; Al-Hameedi, A.T.T.; Dunn-Norman, S. Artificial neural network models to predict lost circulation in natural and induced fractures. *SN Appl. Sci.* **2020**, *2*, 1980. [CrossRef]
7. Carpenter, C. Liner-Drilling Technology Mitigates Lost Circulation Offshore Mexico. *J. Pet. Technol.* **2014**, *66*, 104–107. [CrossRef]
8. Aljubran, M.; Ramasamy, J.; Albassam, M.; Magana-Mora, A. Deep Learning and Time-Series Analysis for the Early Detection of Lost Circulation Incidents During Drilling Operations. *IEEE Access* **2021**, *9*, 76833–76846. [CrossRef]
9. Sabah, M.; Talebkeikhah, M.; Agin, F.; Talebkeikhah, F.; Hasheminasab, E. Application of decision tree, artificial neural networks, and adaptive neuro-fuzzy inference system on predicting lost circulation: A case study from Marun oil field. *J. Pet. Sci. Eng.* **2019**, *177*, 236–249. [CrossRef]
10. Al-Hameedi, A.T.T.; Alkinani, H.H.; Dunn-Norman, S.; Flori, R.E.; Hilgedick, S.A.; Amer, A.S.; Alsaba, M. Mud loss estimation using machine learning approach. *J. Pet. Explor. Prod. Technol.* **2018**, *9*, 1339–1354. [CrossRef]
11. Tang, H.; Zhang, S.; Zhang, F.; Venugopal, S. Time series data analysis for automatic flow influx detection during drilling. *J. Pet. Sci. Eng.* **2019**, *172*, 1103–1111. [CrossRef]
12. Alkinani, H.H.; Al-Hameedi, A.T.T.; Dunn-Norman, S.; Flori, R.E.; Alsaba, M.T.; Amer, A.S.; Hilgedick, S.A. Using data mining to stop or mitigate lost circulation. *J. Pet. Sci. Eng.* **2019**, *173*, 1097–1108. [CrossRef]
13. Geng, Z.; Wang, H.; Fan, M.; Lu, Y.; Nie, Z.; Ding, Y.; Chen, M. Predicting seismic-based risk of lost circulation using machine learning. *J. Pet. Sci. Eng.* **2019**, *176*, 679–688. [CrossRef]
14. Abbas, A.K.; Bashikh, A.A.; Abbas, H.; Mohammed, H.Q. Intelligent decisions to stop or mitigate lost circulation based on machine learning. *Energy* **2019**, *183*, 1104–1113. [CrossRef]
15. Abbas, A.K.; Al-Haideri, N.A.; Bashikh, A.A. Implementing artificial neural networks and support vector machines to predict lost circulation. *Egypt J. Pet.* **2019**, *28*, 339–347. [CrossRef]
16. Ahmed, A.; Elkatatny, S.; Abdulraheem, A.; Abughaban, M. *Prediction of Lost Circulation Zones Using Support Vector Machine and Radial Basis Function*; IPTC: Dhahran, Saudi Arabia, 2020; p. IPTC-19628-MS. [CrossRef]
17. Hou, X.; Yang, J.; Yin, Q.; Liu, H.; Chen, H.; Zheng, J.; Wang, J.; Cao, B.; Zhao, X.; Hao, M.; et al. *Lost Circulation Prediction in South China Sea Using Machine Learning and Big Data Technology*; OTC: Houston, Texas, USA, 2020; p. OTC-30653-MS. [CrossRef]
18. Mardanirad, S.; Wood, D.A.; Zakeri, H. The application of deep learning algorithms to classify subsurface drilling lost circulation severity in large oil field datasets. *SN Appl. Sci.* **2021**, *3*, 785. [CrossRef]
19. Sabah, M.; Mehrad, M.; Ashrafi, S.B.; Wood, D.A.; Fathi, S. Hybrid machine learning algorithms to enhance lost-circulation prediction and management in the Marun oil field. *J. Pet. Sci. Eng.* **2021**, *198*, 108125. [CrossRef]
20. Yingzhuo, X.; Liupeng, W.; Yang, Z.; MingZhen, L. Real-time monitoring and early warning of well leakage based on big data analysis. *J. Phys. Conf. Ser.* **2021**, *1894*, 012084. [CrossRef]
21. Jafarizadeh, F.; Larki, B.; Kazemi, B.; Mehrad, M.; Rashidi, S.; Neycharan, J.G.; Gandomgoun, M.; Gandomgoun, M.H. A new robust predictive model for lost circulation rate using convolutional neural network: A case study from Marun Oilfield. *Petroleum* **2022**, 468–485. [CrossRef]
22. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The Performance of LSTM and BiLSTM in Forecasting Time Series. In Proceedings of the IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 3285–3292. [CrossRef]
23. Ma, T.; Xiang, G.; Shi, Y.; Liu, Y. Horizontal in situ stresses prediction using a CNN-BiLSTM-attention hybrid neural network. *Géoméch. Geophys. Geo-Energy Geo-Resour.* **2022**, *8*, 152. [CrossRef]
24. Li, P.; Zhang, J.; Krebs, P. Prediction of Flow Based on a CNN-LSTM Combined Deep Learning Approach. *Water* **2022**, *14*, 993. [CrossRef]
25. Shen, Y.; Zhang, C.; Gharehchopogh, F.S.; Mirjalili, S. An improved whale optimization algorithm based on multi-population evolution for global optimization and engineering design problems. *Expert Syst. Appl.* **2023**, *215*, 119269. [CrossRef]
26. Fukuda, S.; Uchida, H.; Fujii, H.; Yamada, T. Short-term prediction of traffic flow under incident conditions using graph convolutional recurrent neural network and traffic simulation. *IET Intell. Transp. Syst.* **2020**, *14*, 936–946. [CrossRef]

27. Ulman, H.; Gütter, J.; Niebling, J. Uncertainty is not sufficient for identifying noisy labels in training data for binary segmentation of building footprints. *Front. Remote Sens.* **2023**, *3*, 1100012. [CrossRef]

28. Liu, Z.; Ma, Q.; Cai, B.; Liu, Y.; Zheng, C. Risk assessment on deepwater drilling well control based on dynamic Bayesian network. *Process Saf. Environ. Prot.* **2021**, *149*, 643–654. [CrossRef]

29. Gooneratne, C.P.; Gomez Gonzalez, E.S.; Al-Musa, A.S.; Osorio, H.F. *Thirsty Reservoirs—Challenges in Drilling Through Severe Lost Circulation Zones*; SPE: Abu Dhabi, United Arab Emirates, 2017; p. SPE-188461-MS. [CrossRef]

30. Chaweewanchon, A.; Chaysiri, R. Markowitz Mean-Variance Portfolio Optimization with Predictive Stock Selection Using Machine Learning. *Int. J. Financ. Stud.* **2022**, *10*, 64. [CrossRef]

31. Chatterjee, S.; Mahapatra, S.S.; Bharadwaj, V.; Upadhyay, B.N.; Bindra, K.S. Prediction of quality characteristics of laser drilled holes using artificial intelligence techniques. *Eng. Comput.* **2021**, *37*, 1181–1204. [CrossRef]

32. Liu, P.-L.; Du, Z.-C.; Li, H.-M.; Deng, M.; Feng, X.-B.; Yang, J.-G. Thermal error modeling based on BiLSTM deep learning for CNC machine tool. *Adv. Manuf.* **2021**, *9*, 235–249. [CrossRef]

33. Deng, H.; Liu, L.; Fang, J.; Qu, B.; Huang, Q. A novel improved whale optimization algorithm for optimization problems with multi-strategy and hybrid algorithm. *Math. Comput. Simul.* **2023**, *205*, 794–817. [CrossRef]

34. Mirjalili, S.; Lewis, A. The Whale Optimization Algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [CrossRef]

35. Aljarah, I.; Faris, H.; Mirjalili, S. Optimizing connection weights in neural networks using the whale optimization algorithm. *Soft Comput.* **2018**, *22*, 1–15. [CrossRef]

36. Wood, D.A.; Mardanirad, S.; Zakeri, H. Effective prediction of lost circulation from multiple drilling variables: A class imbalance problem for machine and deep learning algorithms. *J. Pet. Explor. Prod. Technol.* **2022**, *12*, 83–98. [CrossRef]

37. Sirisha, B.; Goud, K.K.C.; Rohit, B.T.V.S. A Deep Stacked Bidirectional LSTM (SBiLSTM) Model for Petroleum Production Forecasting. *Procedia Comput. Sci.* **2023**, *218*, 2767–2775. [CrossRef]