Jestr

Journal of Engineering Science and Technology Review 16 (6) (2023) 177 - 184

Research Article

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

A Wax Deposition Prediction Method for Pumping Oil Wells Based on LSTM Neural Networks

Zhonghui Zhang[1], Yancen Shen[2,*], Lei Zhang[1], Shu Xiao[1] and Xin Jin[1]

[1]*Petroleum Engineering Technology Research Institute, Sinopec Shengli Oilfield Company, Dongying 257000, China*
[2]*School of Petroleum and Natural Gas Engineering, Changzhou University, Changzhou 213164, China*

Received 3 October 2023; Accepted 18 December 2023

Abstract

Wax deposition in oil wells reduces oil production efficiency, increases maintenance costs, and raises the occurrence rate of accidents. Traditional methods have difficulty accurately predicting wax deposition in oil wells, and oil fields typically adopt a passive response mode, resulting in decreased efficiency. Hence, this study proposed a wax deposition prediction method for oil wells based on the long short-term memory (LSTM) neural network to effectively extract and utilize the information content of historical dynamic monitoring data of oil wells and achieve real-time prediction and early warning for wax deposition. A predictive indicator system for predicting failures caused by wax deposition in pumping wells was established, and a wax deposition prediction dataset was prepared by integrating and normalizing collected operational data of oil wells and conducting sensitive parameter analysis. The Grey Wolf Optimization (GWO) algorithm was used to optimize the constructed LSTM prediction model, and the optimized LSTM model was trained and tested using data from the prediction dataset. The accuracy of the model was confirmed through experiments. Results demonstrate that the proposed LSTM-based wax deposition prediction method in this study achieves R-squared values of 0.8453 and 0.9439 on the test set before and after optimization, respectively, effectively improving the accuracy of wax deposition trend prediction in oil wells. This study can assist in predicting wax deposition trends in oil wells, enabling oil field workers to formulate preventive measures in advance, thereby reducing production costs and the occurrence rate of accidents.

Keywords: Wax deposition prediction; Long short-term memory neural network; Oil well production; Intelligent oilfield

1. Introduction

Wax deposition in oil wells is one of the most common issues in petroleum extraction [1-3]. Wax deposition has numerous negative impacts on the daily production of oil wells. Wax deposition can lead to blockages in the oil flow channels, increased flow resistance, reduced oil well output, and even motor shutdown and maintenance owing to wax blockage and combustion, ultimately resulting in the closure of oil wells and increased oil extraction costs. The early and accurate prediction of wax deposition in oil wells is a highly challenging research direction. On the one hand, oil wells typically reach depths of several kilometers, and many important equipment components, such as sucker rods and pumps, are difficult to monitor for wax deposition directly. On the other hand, wax deposition processes are influenced by various factors, including underground temperature, pressure, fluid properties, and flow velocity. Moreover, a clear and effective physical theory prediction model has not been established.

However, with the continuous deepening of oilfield informatization, numerous sensors installed on pumping units can collect real-time data on well temperature, pressure, and electric current, among others. These data are continuously transmitted to the oilfield data center. After years of operation, considerable real-time monitoring data on wax deposition in oil wells have been recorded and saved. Fully mining and utilizing the valuable information contained in the real-time monitoring data of wax deposition in oil wells has become a promising research direction and has surpassed the limitations of existing wax deposition prediction methods. Currently, a new generation of artificial intelligence technologies represented by deep learning has achieved a series of breakthroughs, triggering a new round of technological revolution. Deep learning excels in identifying complex relationships in large-scale multidimensional data and has achieved breakthrough applications in various fields, such as image recognition, speech processing, autonomous driving, and handwriting recognition. Owing to its reliability in data analysis and interpretation, many intelligent methods, including support vector machines (SVM) [4], genetic algorithms (GA), artificial neural networks, and fuzzy logic [5], have been widely employed in the petroleum and natural gas industry as data-driven predictive tools [6-8].

Based on the above analysis, scholars have conducted studies using big data and deep learning–related methods to solve the problem of wax deposition (waxing) prediction in oil wells [9]. However, the current study lacks the ability to continuously predict the future development trend of paraffin deposition in oil wells or suffers from insufficient prediction accuracy, which still necessitates reliance on expert experience to solve this problem. Therefore, developing a prediction method with high accuracy for forecasting wax deposition trends in oil wells is an urgent issue.

To address this issue, this study utilizes deep learning technology to solve the wax deposition prediction problem in pumping unit wells. Deep learning technology preprocesses oil well operational data, conducts sensitivity

*E-mail address: s18361672722@163.com
ISSN: 1791-2377
doi:10.25103/jestr.166.22

analysis, prepares a sample set and comprehensive evaluation index for waxing prediction, constructs a deep learning-based prediction model, optimizes the model using optimization algorithms, and aims to predict the wax deposition trend in pumping unit wells to provide references for implementing anti-wax measures in oil fields ahead of time.

2. State of the art

Currently, scholars have conducted extensive studies on prediction problems, such as well failure and oilfield production data. Nazar et al. proposed an optimized multiphase thermodynamic model for predicting the deposition of paraffin in paraffin-rich crude oil [10]. The reliability of the model was verified by simulating crystallization experiments of paraffin in oil and crude oil samples. In addition, they integrated the energy equation, paraffin deposition kinetics equation, and thermodynamic equation to handle two different flow states in pipelines, namely, laminar and turbulent flow. This integration resulted in a paraffin deposition kinetics prediction model suitable for simulating indoor annular experimental tests. Svendsen et al. specifically developed a wax deposition model for transverse oil pipelines [11]. However, the model was sensitive to changes in oil pipe height, and its accuracy error was significant in areas where the height differences change rapidly. Magnini Mirco et al. conducted a basic analysis of wax management mechanisms in oil wells and used fluid volume methods and chemical equilibrium models to locally simulate petroleum deposits [12]. They proposed a predictive method for steady-state deposition thickness. Jiang et al. proposed a relatively simple equation to predict the sedimentation of mud [13]. They described and summarized mechanisms for paraffin and asphaltene sedimentation from microscopic and macroscopic perspectives. Based on an analysis of the factors affecting wax deposition, Li established a CS-SVM wax deposition degree prediction model [14] with an average relative error of less than 5%, which could provide a practical reference for establishing a washing well operation cycle. Bhaskaran et al. used time series forecasting and data clustering methods to predict the failure rate of oil transport pipelines [15], achieving an effective rate of 91.9% on the historical dataset. Kumar et al. proposed an attention-based long short-term memory network-assisted oil field production time series prediction model [16] with a root mean square error (RMSE) of 0.0102, significantly lower than the RMSE of other models in the same study. Wu et al. developed a pumping unit fault diagnosis model based on deep learning algorithms and the DenseNet model using transfer learning methods [17]. The model was applied to classify and test various site conditions, including wax deposition in oil wells, pump leakage, and inadequate liquid supply. In evaluating the testing dataset, the model had an average accuracy rate of over 95% in identifying various conditions. Ahmadi et al. introduced a wax deposition prediction method by combining FL and GA [18], obtaining an R2 value of 0.9449, demonstrating high robustness. Manshad et al. combined neural networks with improved GA to propose a correlation model between the wax content and deposition amount of crude oil in a certain temperature range. The accuracy of the method was evaluated by predicting the wax precipitation volume of various reservoir fluids not used in the development of the prediction model [19]. Behbahani et al. applied ANN algorithms to calculate the amount of wax crystallization, achieving higher accuracy than traditional mechanical models, with an average absolute error of approximately 1% [20]. Wang et al. proposed an ensemble learning method (SCRF) for non-equilibrium data-oriented sample datasets [21], using the SMOTE method to oversample the minority class in the original dataset to increase the number of minority cases and balance the dataset, followed by the CLUSTER clustering method to generate training datasets. Finally, the random forest algorithm based on bagging technology was used for integrated learning on the training dataset to address wax deposition rate calculations, accuracy of wax deposition status determination, and other issues. The AERF algorithm was proposed by Chang et al., who combined the ADASYN and ENN algorithms to handle imbalanced datasets and then used the random forest algorithm for training and learning on balanced sample datasets [22]. For wax deposition rate calculation, Ai et al. fitted the calculation coefficients of different moisture contents based on experimental data and simplified empirical formulas to establish a wax prediction model [23]. The model's accuracy was verified by actual wax deposition data from 10 wells in the field, with an accuracy rate of 87.7%. The changes in the timing indicator diagram can reflect the degree of oil well paraffin deposition. Tan et al. used a residual convolutional neural network to extract the characteristic features of the paraffin deposition well [24], determined the paraffin deposition level using a clustering algorithm, and established a sample set by combining the extracted indicator diagram characteristics and 12 production parameters. The neural network algorithm was used to establish a paraffin deposition-level prediction model and quantitatively predict the paraffin deposition trend of the pumping unit well.

The above studies mainly focused on fault diagnosis and failure rate prediction during oil well operation, oil well production data prediction, handling of imbalanced samples, and prediction of wax deposition time and severity. However, studies on wax deposition trend prediction in oil wells are relatively limited. In this study, real-time monitoring data of wax deposition wells were collected and preprocessed to form a dataset. Feature engineering was performed to analyze the changes in various indicators before and after wax deposition in oil wells and determine the sensitivity of these indicators to wax deposition. Then, the indicators with high sensitivity were selected as inputs for wax deposition prediction. A comprehensive evaluation index for wax deposition severity was calculated by combining the highly sensitive indicators. The wax deposition prediction problem was transformed into a time series prediction problem in the field of machine learning, and a time series prediction sample set was prepared. Finally, a neural network structure based on LSTM was designed for prediction, and the Grey Wolf Optimization algorithm was used to optimize its hyperparameters to obtain the optimal network architecture f or predicting wax deposition trends in oil wells.

The rest of this study is organized as follows. Section 3 introduces data collection and preprocessing, neural network design, and optimization methods. Section 4 presents the training optimization and performance analysis of the neural network. Finally, Section 5 concludes the study and provides relevant recommendations.

3. Methodology

3.1 Data Collection and Data Cleaning

According to the oil well accident report of the oil field, a total of 100 records of oil well shutdowns owing to wax deposition were sorted out. Based on the well name and shutdown time, the real-time monitoring database of each wax deposition well was queried to obtain all data within 30 days before the shutdown day, and a total of 76 wax deposition wells with data records were obtained.

The original data of each wax deposition well contained 356 data indicators. Further analysis showed that there were 150 indicators with a null value rate of less than 5%, accounting for 46%, and 172 indicators with a null value rate of more than 95%, accounting for 53%. In the subsequent data analysis, only data indicators with a null value rate of less than 5% were retained. Interpolation was used to fill the remaining null values.

Among the 150 indicators retained, there were differences in the data sampling frequency. For example, the displacement and the load were 30 min/time, the current and the power were 1 min/time, and the production rate was 1 d/time. In addition, some static indicators, such as crude oil viscosity, exist. The data were standardized to 30 min/time to facilitate subsequent data mining needs.

For the conversion from low- to high-frequency data, the nearest neighbor interpolation, linear interpolation, and cubic spline interpolation algorithms were compared. Moreover, linear interpolation was selected for conversion. For the conversion from high- to low-frequency data, the downsampling method was selected.

After the above processing, the standardized data set of wax deposition wells was obtained, covering 150 indicators of 76 wells, with a total of 109,000 data records.

3.2 Feature Engineering

Not all of the 150 data indicators have an indicative significance for wax deposition prediction. If all of them are used directly for oil well wax deposition prediction, it will not only cause excessive computational load owing to a large amount of data but also introduce noise interference and affect the accuracy of the model prediction. Therefore, feature engineering studies should be conducted to identify indicators with strong sensitivity to wax deposition prediction.

A manual analysis of the data of each wax deposition well is conducted to determine the start time of wax deposition. The period before and after wax deposition is defined as the healthy and wax deposition states, respectively. If a certain indicator undergoes significant changes in the wax deposition and healthy states, it is a highly sensitive indicator of wax deposition.

For example, Figure 1 shows the variation curve of the indicator diagram area (GTMJ) of the pumping unit of a wax deposition well in the past month. The well began to experience wax deposition on March 19. The red and blue areas in the figure represent the wax deposition and healthy states, respectively. After the wax deposition began, the area of the pumping unit showed a significant upward trend.

For each time series parameter, the healthy interval and the wax deposition interval are selected. The median (value), linear regression slope (slope), and standard deviation (STD) of each indicator under the healthy state and the wax deposition state are calculated to characterize the numerical level, evolution rate, and volatility characteristics of each indicator under the healthy state and the wax deposition state.

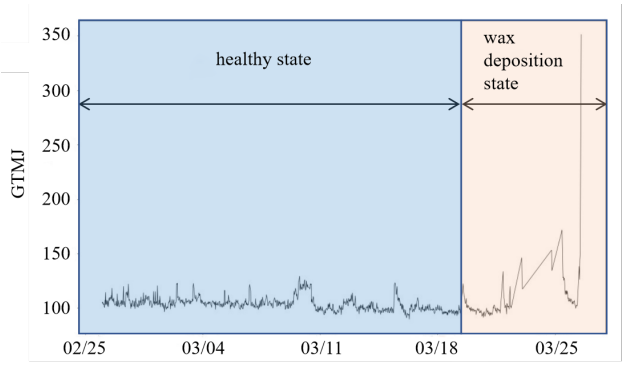

Fig. 1. Healthy state and wax deposition state for an oil well

The characteristic parameters of each index were calculated in healthy and waxing states. Figure 2 shows the results of two typical indicators, including the indicator diagram area and voltage. In this figure, the healthy state is represented by red circles, while the wax deposition state is represented by blue circles.

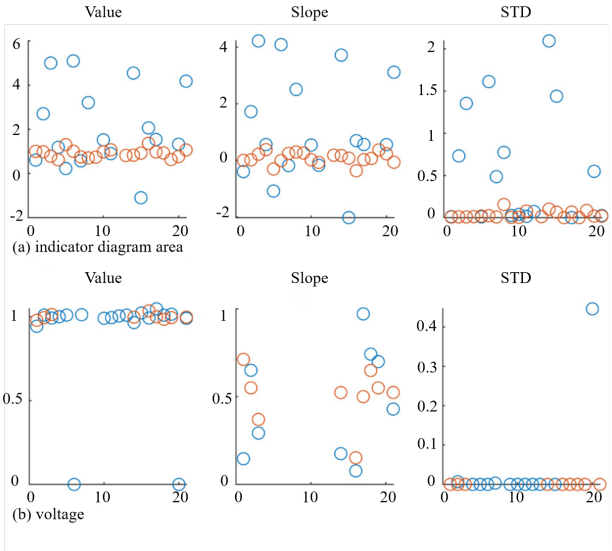

Fig. 2. Comparison of the characteristic indexes of the healthy state and the wax deposition state of oil wells

From Figure 2(a), the absolute value of the indicator diagram is generally greater than that in the healthy state when waxing occurs. When in a healthy state, the numerical value of the indicator diagram area is stable. When waxing occurs, the numerical value of the indicator diagram area generally increases positively. When in a healthy state, the numerical value of the indicator diagram area fluctuates slightly. When waxing occurs, the fluctuation of the numerical value of the indicator diagram area generally increases. Therefore, the indicator diagram area is considered an indicator with high sensitivity.

From Figure 2(b), the absolute value, evolution rate, and fluctuation of the voltage are not significantly distinguished when waxing occurs, indicating that the sensitivity of the voltage to waxing prediction is low.

Based on this, the comprehensive significance of each index in wax deposition prediction was calculated, and Figure 3 shows the top-ranking indicators. In this diagram, the variables corresponding to each abbreviation are as follows: GTMJ: area of the indicator diagram; YGGL: active power; GGGL: light rod power; ZXZH: minimum load; ZDZH: maximum load; HDL: power consumption; ZBX: overall pump efficiency; CC1: frequency of stroke; HS: water cut; GLYS: power factor; RCYL1: daily liquid

production; XXDL: downstream current; JK_WD: wellhead temperature. Based on this figure, six indicators with a comprehensive significance greater than 20% were selected for wax deposition prediction, including the area of the indicator diagram, active power, light rod power, minimum load, maximum load, and power consumption.

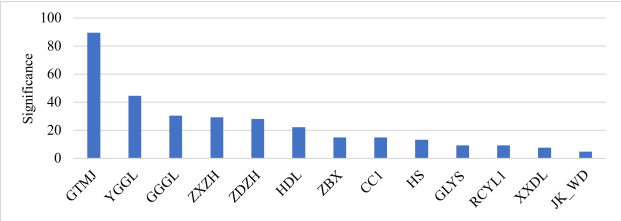

Fig. 3. Top-ranking indicators with high comprehensive significance in wax deposition prediction

3.3 Comprehensive Evaluation Index for Wax Deposition Degree

For each indicator related to wax deposition failure prediction, the indicator value of each well in the healthy state was set to 0, and the value of each well from wax deposition to well shutdown was set to 100. The time series of each indicator was transformed into a value between 0 and 100 through function transformation. Considering each sensitive indicator, a weighted average was performed to construct a comprehensive evaluation index of wax deposition degree. Table 1 shows the weights of each indicator in the comprehensive evaluation index of wax deposition degree.

Table 1. Weights of each indicator in the comprehensive evaluation index of the wax deposition

Indicator	Significance	Weight
GTMJ	89.547	0.367
YGGL	44.566	0.183
GGGL	30.370	0.124
ZXZH	29.273	0.120
ZDZH	28.093	0.115
HDL	22.223	0.091

The constructed comprehensive evaluation index of wax deposition degree has a value of 0 in the healthy state and a value of 100 in the most severe wax deposition that leads to well shutdown.

3.4 Introduction to the Principle of LSTM

The LSTM network is currently one of the most widely used and effective methods for addressing time series problems [25]. This network improves upon the gradient vanishing issue and can deal with long-term dependencies, making it highly efficient when processing time series data [26-27]. Given that data on wax deposition in oil wells also exhibit temporal characteristics, this study opts to utilize the LSTM neural network to establish a predictive model to yield superior results.

At its core, the LSTM introduces a "cell state" to store historical information. It leverages a "forget gate" to control the extent of forgetting this historical data, an "input gate" to manage the incorporation of new information, and an "output gate" to regulate data output. The specific formulas are as follows:

The forget gate is responsible for determining which information in the cell state needs to be forgotten based on the current input and the previous hidden state. Its calculation can be represented using Eq. (1):

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \tag{1}$$

where w_f denotes the weight parameter, b_f denotes the bias term, and σ denotes the activation function.

The input gate, which determines the degree of influence of the new input data based on the current input and the previous hidden state, plays a role in the process. Its calculation can be represented using Eq. (2):

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \tag{2}$$

where w_i denotes the weight parameter and b_i denotes the bias term.

The cell state update, which updates the cell state based on the calculations of the input gate and the forget gate, serves the purpose, which can be represented using Eq. (3):

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t \tag{3}$$

where $\tilde{c}_t$ denotes the new candidate value, and the formula for its computation is as follows:

$$\tilde{c}_t = \tanh(w_c[h_{t-1}, x_t] + b_c) \tag{4}$$

where c_t denotes the weight term, b_c denotes the bias term, and tanh denotes the hyperbolic tangent function.

The output gate, which determines which information needs to be output from the cell state based on the current input and the previous hidden state, serves the purpose. Its calculation can be represented using Eq. (5):

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \tag{5}$$

where w_o denotes the weight parameter and b_o denotes the bias term.

The hidden state, obtained based on the calculations of the output gate and the current cell state, represents the hidden state at the current time step. Its calculation can be represented using Eq. (6):

$$h_t = o_t \tanh(c_f) \tag{6}$$

3.5 Design of the Wax Deposition Prediction Model Based on LSTM

The detailed model architecture is as follows: First is an LSTM layer, which is based on feature engineering from Section 3.2. The input for this study includes six features, with each feature having a dimension of 1. Therefore, the input shape of this LSTM layer is (number of samples, timesteps, 6). To prevent overfitting, a dropout layer is added after the first LSTM layer, followed by a second LSTM layer to further capture the correlation of the time series, thereby enhancing the performance of the model. Subsequently, two dense layers are introduced: the first dense layer employs a Relu activation function, which assists in greatly fitting the model. The second dense layer contains only a single neuron, used for outputting a solitary prediction value; therefore, the output shape of the model is (number of samples, 1). Figure 4 shows the precise model structure.

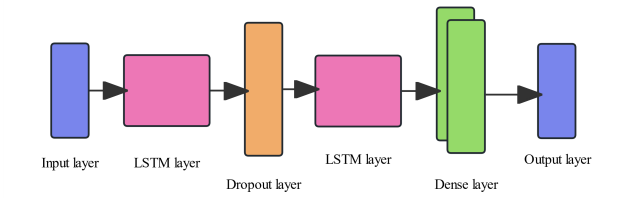

Fig. 4. LSTM model architecture designed for wax deposition prediction

The model employs mean squared error (MSE) as the loss function for compilation and utilizes the Adam optimizer for training. This study evaluates the prediction accuracy of the model using MSE and R squared as the evaluation metrics. The expression for MSE can be represented as Eq. (7):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{7}$$

Where Y_i denotes the true value associated with sample i, whereas $\hat{Y}_i$ denotes the respective predicted value.

The specific expression for R squared can be found in Eq. (8):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{8}$$

where $\bar{y}$ denotes the sample mean, $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ denotes the discrepancies between predicted and actual values, and $\sum_{i=1}^{n}(y_i - \bar{y})^2$ denotes the errors attributed to the mean value.

3.6 Methodology for Hyperparameter Optimization

This study employs the Grey Wolf Optimization algorithm to fine-tune the hyperparameters of the model. Specifically, the algorithm is leveraged to optimize the number of neurons in the LSTM layer, the dropout rate in the dropout layer, and the quantity of neurons in the dense layer. The designated search window spans from (128,0.1,32) to (256,0.5,128).

The principle of the GWO algorithm is as follows:

Within the GWO framework, the wolves' ensemble (comprising various hyperparameter permutations) evolves its position according to the optimal coordinates of the three lead wolves (i.e., alpha, beta, and delta). The positional shift of the wolves, represented as D, from iteration t to t + 1, can be articulated using Eqs. (9) and (10):

$$D = CX_P(t) - X(t) \tag{9}$$

where X denotes the grey wolf's coordinates and D denotes the spatial separation between the grey wolf and its target prey.

$$X(t+1) = X_P(t) - AD \tag{10}$$

where X_P denotes the position of the prey, equivalently, the coordinates of the alpha, beta, and delta wolves. The calculation methods for A and C in the above equation are represented by Eqs. (11) and (12):

$$A = 2ar_1 - a \tag{11}$$

$$C = 2r_2 \tag{12}$$

where r_1 and r_2 denote random numbers within the interval from 0 to 1, respectively. Concurrently, a denotes a linear decrement from an initial value of 2 to a final value of 0 proportional to the increasing count of iterations.

We independently determine the values of Dalpha, Dbeta, Ddelta, $X(t+1)_a$, $X(t+1)_b$, and $X(t+1)_d$. Consequently, the position, denoted by $X(t+1)$, which requires adjustment by each wolf individual, is as follows:

$$X(t+1) = \frac{X(t+1)_a + X(t+1)_b + X(t+1)_d}{3} \tag{13}$$

The procedure for enhancing the model is elaborated as follows:

Initially, based on the LSTM prediction model delineated in Section 3.5, we establish an evaluation function catering to model performance. This function incorporates the number of neurons within the LSTM layer, the rate of dropout within dropout layers, and the neural count within the dense layer as input parameters. As an output, this function bestows the model error as a performance score.

Subsequently, we construct the GWO algorithm that stipulates the dimensions concerning the wolves and the iterative count, along with defining the range of exploration for the hyperparameters. Initialization of the population, the values represented by the hyperparameter combinations, including objective function values, occurs. An ongoing update facilitated by the algorithm influences the values of the hyperparameter combinations and the objective function until it reaches the pre-set limit of iterations. Eventually, the result comprises the optimal combination of hyperparameters and ends with a model score.

In the final phase, we implement the acquired optimally suited hyperparameter combination into the LSTM model, facilitating the evolution of an enhanced predictive model.

4. Result Analysis and Discussion

4.1 Parameters and Environments for Training

During training, we randomly apportioned 10% of the total samples as a validation dataset, with the remaining 90% forming the training dataset. Seventy epochs of training were completed, wherein each epoch incorporated 300 samples for its neural network processing. The training environment was facilitated through Keras 2.13.1.

Concerning the workstation architecture, the hardware framework comprised a 13th Gen Intel® Core™ i9-13900HX 2.20 GHz processor coupled with 32GB RAM. Regarding the software infrastructure, we utilized the 64-bit Windows 11 operating system along with PyCharm Community Edition 2023.1.3.

4.2 Analysis of Training Outcomes

Employing the training conditions delineated in Section 4.1, well operational data were fed into the model for training purposes. On average, the entire process required approximately 8 seconds per training epoch. The resulting predictions were visualized and scrutinized, with key observations outlined below:

An analysis of the model's error rates with respect to variations in training epochs was conducted (Figure 5). During the preliminary epochs, substantial discrepancies existed between the validation and training losses, thereby signaling inadequate model training. Nonetheless, commencing from the 20th epoch, no considerable distinction was observed between the two, thereby indicating the absence of overfitting or underfitting issues and, subsequently, effective training outcomes. Additionally, a steady decline in training and validation losses was observed as the number of training epochs escalated, ultimately converging to a significantly low error rate, which further attested to the model's effectiveness.

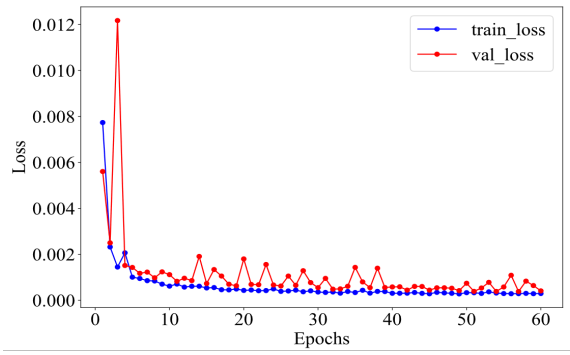

Fig. 5. Loss values of the pre-optimized model on the validation set and the training set

Concurrently, the predictive performance of the model prior to optimization was visually represented. Figure 6 provides an illustration of the model's performance based on the predictions made on the training dataset, whereas Figure 7 delineates the predictive outcomes on the test dataset.

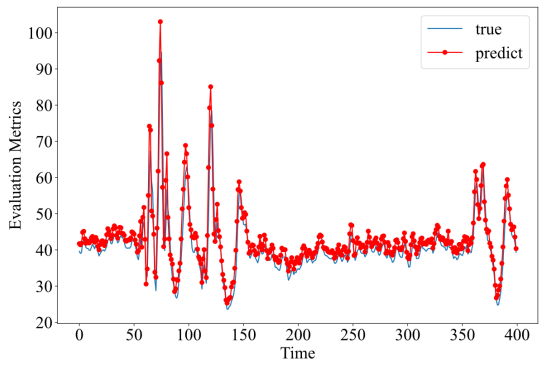

Fig. 6. Prediction performance of the pre-optimized model on the training set

The refined model was applied to the same data used for pre-optimization predictions to compare the predictive performance of the model before and after optimization. Subsequently, Figure 8 shows the loss values pertaining to the validation and training datasets. Figures 9 and 10 present the enhanced model's predictive performance on the training and test datasets, respectively.

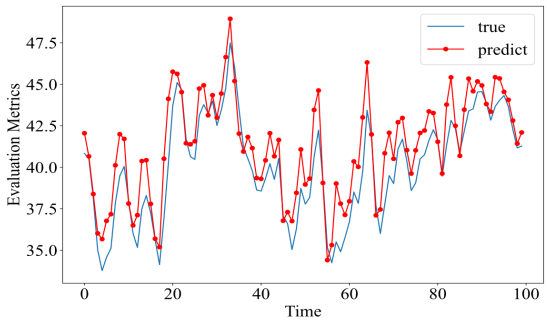

Fig. 7. Prediction performance of the pre-optimized model on the test set

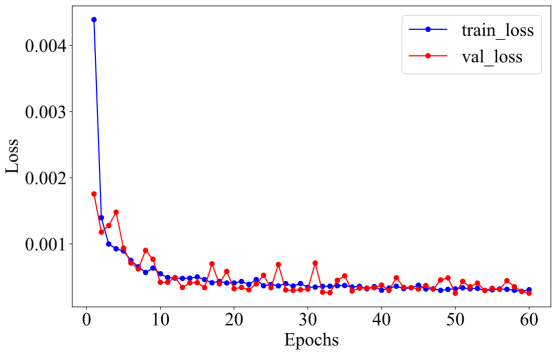

Fig. 8. Loss values of the optimized model on the validation and training sets

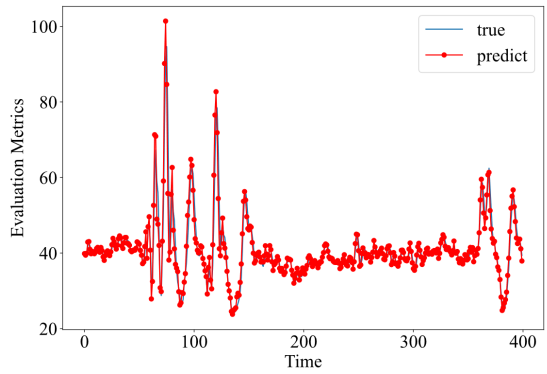

Fig. 9. Prediction performance of the optimized model on the training set

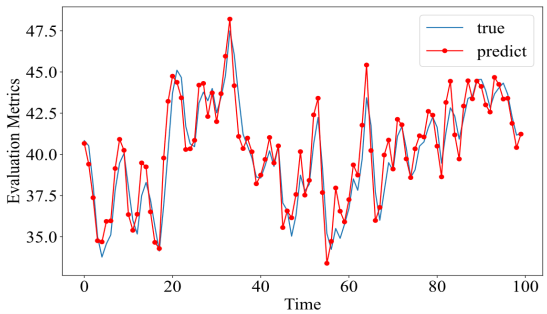

Fig. 10. Prediction performance of the optimized model on the test set

From the visualizations drawn after optimization, the refined model manifests an improved predictive performance on the training and test datasets, outperforming its initial state. With an expedited rate of convergence, the optimization has indeed enhanced the model's performance.

Table 2 catalogs the variations in the model's hyperparameters before and after the optimization process.

Notably, the hyperparameters prior to optimization were empirically established through manual configuration.

Table 2. Variations in the model's hyperparameters pre- and post-optimization

Hyperparameters	Pre-optimization	Post-optimization
LSTM neurons	128	256
Dropout rate	0.1	0.2
Dense neurons	128	64

This study employs R-squared as an illustrative measure of the predictive efficacy of the model, pre- and post-optimization, with the nuanced findings explicated in Table 3. An in-depth examination of the results presented in Table 3 elucidates that the post-optimization model yields high performance across the training and testing sets. This result corroborates the premise of the optimization methodology posited in this manuscript and underscores its capacity to improve prediction precision.

Table 3. Efficacy of the model's predictions prior to and subsequent to the process of optimization

Data set	Pre-optimization	Post-optimization
R-squared (training set)	0.8769	0.9570
R-squared (test set)	0.8453	0.9439

5. Conclusions

This study first preprocessed the operational data of oil wells, performed feature analysis, and constructed a comprehensive evaluation index of wax deposition severity using highly sensitive features. The objectives are to effectively mine and utilize the information contained in historical dynamic monitoring data of oil wells and achieve real-time prediction and advance warning of wax deposition. Then, a predictive algorithm based on the LSTM network was developed and applied to the dataset. Finally, the GWO algorithm was used to optimize the hyperparameters of the predictive algorithm, and the prediction performance of the model was validated on the test set. The following conclusions could be drawn:

(1) The comprehensive significance of each indicator in wax deposition prediction was calculated, and six indicators with a comprehensive significance greater than 20% were determined: pump card area, active power, polished rod power, minimum load, maximum load, and power consumption.

(2) The proposed LSTM-based wax deposition prediction method in this study can predict the wax deposition trend in oil wells using the dataset used in this study, and it has good prediction performance.

(3) The optimization algorithm used in this study can further improve the performance of the prediction model.

This study proposes a feasible prediction method by applying neural network methods to wax deposition prediction in oil wells. The developed algorithm exhibits high accuracy in prediction, providing a reference for oilfield managers in formulating anti-wax measures. This study only uses the LSTM neural network to construct the predictive algorithm, and LSTM networks have certain limitations. Therefore, ensemble learning is being considered in future studies to improve the predictive model and further enhance its performance.

Acknowledgements
The authors acknowledge fundings from Sinopec Science and Technology Research Project (No. 20071), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. SJCX23_1574).

References

[1] R. Dalirsefat and F. Feyzi, "A thermodynamic model for wax deposition phenomena," *Fuel*, vol. 86, no. 10-11, pp. 1402-1408, Aug. 2007.
[2] J. Y. Zuo, D. D. Zhang, and H. J. Ng, "An improved thermodynamic model for wax precipitation from petroleum fluids," *Chem. Eng. Sci.*, vol. *56*, no. 24, pp. 6941-6947, Dec. 2001.
[3] H. P. Rønningsen, "Production of waxy oils on the Norwegian continental shelf: experiences, challenges, and practices," *Energy Fuels*, vol. 26, no. 7, pp. 4124-4136, May. 2012.
[4] W. S. Noble, "What is a support vector machine?," *Nat. Biotechnol.*, vol. 24, no. 12, pp. 1565-1567, Dec. 2006.
[5] S. A. Kalogirou and M. Bojic, "Artificial neural networks for the prediction of the energy consumption of a passive solar building," *Energy*, vol. 25, no. 5, pp. 479-491, May. 2000.
[6] M. A. Ahmadi, M. Zahedzadeh, S. R. Shadizadeh, and R. Abbassi, "Connectionist model for predicting minimum gas miscibility pressure: application to gas injection process," *Fuel*, vol. 148, pp. 202-211, May 2015.
[7] S. R. Moosavi, D. A. Wood, M. A. Ahmadi, and A. Choubineh, "ANN-based prediction of laboratory-scale performance of CO 2-foam flooding for improving oil recovery," *Nat. Resour. Res.*, vol. *28*, pp. 1619-1637, Jan. 2019.
[8] M. Ramezanizadeh, M. A. Ahmadi, M. H. Ahmadi, and M. A. Nazari, "Rigorous smart model for predicting dynamic viscosity of Al_2O_3/water nanofluid," *J. Therm. Anal. Calorim.*, vol. 137, pp. 307-316, 2019.
[9] S. Bing, "An Early Warning Method Based on Artificial Intelligence for Wax Deposition in Rod Pumping Wells," *Pet. Drill. Tech.*, vol. 47, no. 4, pp. 97-103, Jul. 2019, doi: 10.11911/syztjs.2019093.
[10] A. S. Nazar, B. Dabir, and M. R. Islam, "A multi-solid phase thermodynamic model for prediction wax precipitation in petroleum mixtures," *Energ. Source*, vol. 27, no. 1-2, pp. 173-184, Jan. 2005.
[11] J. A. Svendsen, "Mathematical modeling of wax deposition in oil pipeline systems," *Aiche J.*, vol. 39, no. 8, pp. 1377-1388, Aug. 1993.
[12] M. Magnini and O. K. Matar, "Fundamental study of wax deposition in crude oil flows in a pipeline via interface-resolved numerical simulations," *Ind. Eng. Chem. Res.*, vol. 58, no. 47, pp.21797-21816, Oct. 2019.
[13] H. Jiang, Y. Wang, C. Nie, F. Yan, X. Ouyang, and J. Gong, "Oil Sludge Deposition in Storage Tanks: A Case Study for Russian Crude Oil in Mo-he Station," *Appl. Sci.*, vol.11, Dec. 2020, Art. no. 321.
[14] Z. Li, "Cause analysis of wellbore waxing formation and degree prediction in a block of North China," *World Pet. Ind.*, vol. 29, no. 4, Aug. 2022, Art. no. 1006-0030(2022)04-0061-006.
[15] P. E. Bhaskaran, M. Chennippan, T. Subramaniam, "Future prediction & estimation of faults occurrences in oil pipelines by using data clustering with time series forecasting," *J. Loss Prev. Process Ind.*, vol. 66, Jul. 2020, Art. no. 104203.
[16] I. Kumar, B. K. Tripathi, and A. Singh, "Attention-based LSTM network-assisted time series forecasting models for petroleum production," *Eng. Appl. Artif. Intell.*, vol. 123, Aug. 2023, Art. no.

106440.

[17] Y. Wu, Z. Feng, J. Liang, Q. Liu, and D. Sun, "Fault Diagnosis Algorithm of Beam Pumping Unit Based on Transfer Learning and DenseNet Model" *Appl. Sci.*, vol. 12, no. 21, Nov. 2022, Art. no. 11091.

[18] M. Ahmadi, "Data-driven approaches for predicting wax deposition," *Energy*, vol. 265, Feb. 2023, Art. no. 126296.

[19] A. Khaksar Manshad, S. Ashoori, M.Khaksar Manshad, and P. Omidvar, "The prediction of wax precipitation by neural network and genetic algorithm and comparison with a multisolid model in crude oil systems," *Pet. Sci. Technol.*, vol. 30, no.13, pp. 1369-1378, Apr. 2012.

[20] T. J. Behbahani, A. A. M. Beigi, Z. Taheri, and B. Ghanbari, "Investigation of wax precipitation in crude oil: Experimental and modeling," *Petroleum*, vol. 1, no. 3, pp. 223-230, Sep. 2015.

[21] L. Wang, Z. Zhi, L. Jia, and W. Li, "Study on Optimized Method for Predicting Paraffin Deposition of Pumping Wells Based on SCRF," *Comput. Sci.*, vol. 46, no. 11A, pp. 599-603, Nov.2019.

[22] Y. Chang, Q. Li, and K. Li, "Prediction of Oil Well Wax Deposition Based on AERF Model," *Comput. Syst. Appl.*, vol. 30, no. 9, pp. 138-144, 2021, doi: 10.15888/j.cnki.csa.008060.

[23] X. Ai, Y. He, L. Qiang, Y. Xin, and K. He, "Prediction model optimization of wellbore wax deposition in high water cut wells in Yanchang oilfield," *Chem. Eng. Oil Gas*, vol. 50, no. 3, pp. 85-89, 2021, doi: 10.3969/j.issn.1007-3426.2021.03.014.

[24] C. Tan *et al.*, "Prediction of paraffin deposition and evaluation of paraffin removal effect for pumping wells driven by timing indicator diagram," *Oil Drill. Prod. Technol.*, vol. 44, no. 1, pp. 123-130, Jan. 2022, doi: 10.13639/j.odpt.2022.01.019.

[25] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235-1270, Jul. 2019.

[26] D. Chi, "Research on electricity consumption forecasting model based on wavelet transform and multi-layer LSTM model," *Energy Rep.*, vol. 8, no. 4, pp. 220-228, Jul. 2022.

[27] R. Hassanian, Á. Helgadóttir, and M. Riedel, "Deep learning forecasts a strained turbulent flow velocity field in temporal Lagrangian framework: Comparison of LSTM and GRU," *Fluids*, vol. 7, Nov. 2022.