


Article

A Working Conditions Warning Method for Sucker Rod Wells Based on Temporal Sequence Prediction

Kai Zhang ^{1,2,*} , Chengzhe Yin ^{1,*}, Weiying Yao ³, Gaocheng Feng ³, Chen Liu ⁴, Cheng Cheng ¹ and Liming Zhang ¹¹ School of Petroleum Engineering, China University of Petroleum, Qingdao 266500, China² School of Civil Engineering, Qingdao University of Technology, Qingdao 266500, China³ CNOOC EnerTech-Drilling & Production Co., Tianjin 300452, China⁴ CNOOC Research Institute Ltd., Beijing 100028, China

* Correspondence: zhangkai@upc.edu.cn (K.Z.); b20020045@s.upc.edu.cn (C.Y.)

Abstract: The warning of the potential faults occurring in the future in a sucker rod well can help technicians adjust production strategies in time. It is of great significance for safety during well production. In this paper, the key characteristic parameters of dynamometer cards were predicted by a temporal neural network to implement the warning of different working conditions which might result in failures. First, a one-dimensional damped-wave equation was used to eliminate the dynamic loads' effect of surface dynamometer cards by converting them into down-hole dynamometer cards. Based on the down-hole dynamometer cards, the characteristic parameters were extracted, including the load change, the position of the valve opening and closing point, the dynamometer card area, and so on. The mapping relationship between the characteristic parameters and working conditions (classification model) was obtained by the Xgboost algorithm. Meanwhile, the noise in these parameters was reduced by wavelet transformation, and the rationality of the results was verified. Second, the Encoder-Decoder and multi-head attention structures were used to set up the time series prediction model. Then, the characteristic parameters were predicted in a sequence-to-sequence way by using historical characteristic parameters, date, and pumping parameters as input. At last, by inputting the predicted results into the classification model, a working conditions warning method was created. The results showed that noise reduction improved the prediction accuracy significantly. The prediction relative error of most characteristic parameters was less than 15% after noise reduction. In most working conditions, their F1 values were more than 85%. Most Recall values could be restored to over 90% of those calculated by real parameters, indicating few false negative cases. In general, the warning method proposed in this paper can predict faulty working conditions that may occur in the future in a timely manner.

Keywords: sucker rod well; dynamometer card; characteristic parameters; time series prediction; working conditions warning

MSC: 68T07



Citation: Zhang, K.; Yin, C.; Yao, W.; Feng, G.; Liu, C.; Cheng, C.; Zhang, L. A Working Conditions Warning Method for Sucker Rod Wells Based on Temporal Sequence Prediction. *Mathematics* **2024**, *12*, 2253. <https://doi.org/10.3390/math12142253>

Academic Editor: Yolanda Vidal

Received: 23 June 2024

Revised: 15 July 2024

Accepted: 17 July 2024

Published: 19 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Research Background

During the production of sucker rod wells, where equipment, such as rods and pumps, works continuously, the inflow and outflow dynamics near the wellbore will change over time. The equipment and fluid flow status is influenced by complex factors, which cause various working conditions to occur during the artificial lifting. Down-hole working conditions can be monitored in real-time by the dynamometer, which provides dynamometer cards. The dynamometer card, which consists of load and displacement, reflects the different working conditions. Based on the real-time monitoring of working conditions, technicians can quickly understand the current production status of oil wells.

However, when the technician uses real-time monitoring to identify a fault, it usually either began or occurred at a previous time in the oil well. If so, the time when technicians adjust production strategies usually lags behind when the fault occurs. Warnings about working conditions can be used to identify a potential fault occurring in the future. It can help alert technicians in advance and shorten the interval between the occurrence of the fault and the implantation of corrective measures. It is of great significance for safety and efficient well production.

During normal production, the working conditions are usually stable. When there is an abnormality, the shape of the down-hole dynamometer cards will change gradually. When the change reaches a certain degree, it will be labeled as a fault. The working condition changes of some wells are shown in Figure 1.

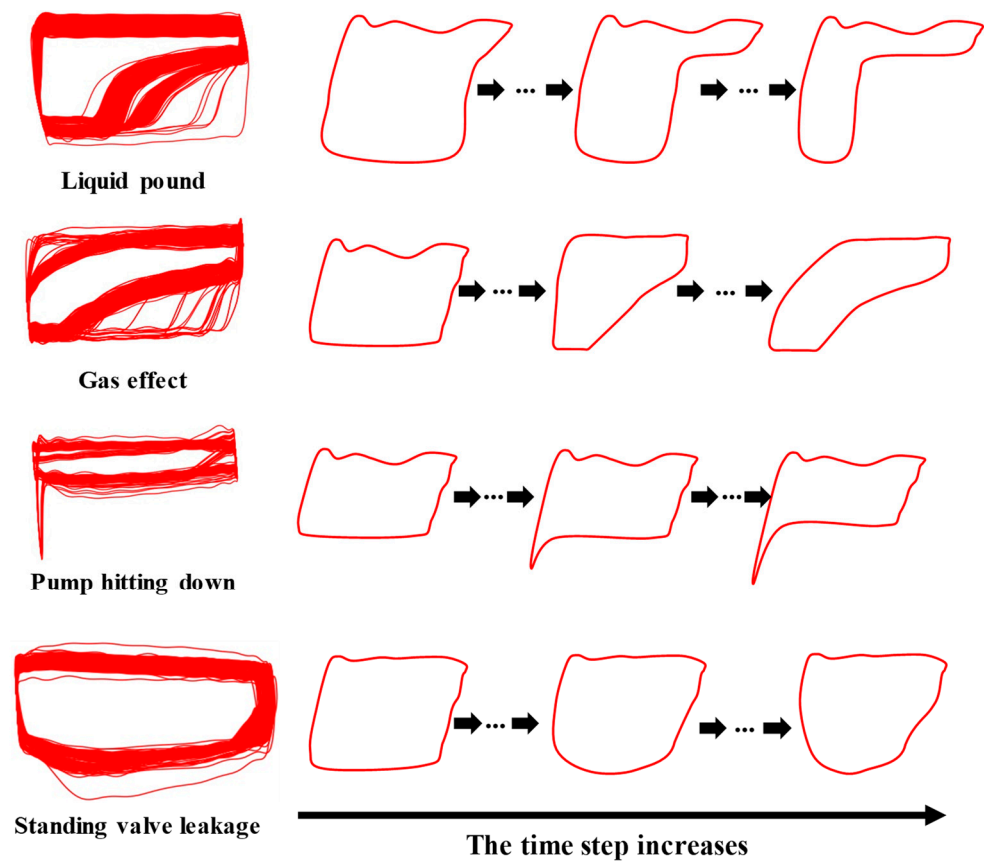


Figure 1. Change of four abnormal working conditions over time.

Figure 1 shows the progressive changes in four abnormal working conditions, including the liquid pound, gas effect, pump hitting down, and standing valve leakage. It can be seen that the faults become increasingly significant over time. In this process, the position and shape of a dynamometer card changes with the change in working conditions. It also means that the load and displacement of each point on the dynamometer card change jointly. It is not necessary to focus on all points on the dynamometer card when the working conditions change. By extracting the load, displacement, and geometric parameters (hereinafter referred to as the characteristic parameters) of the key points and areas [1,2], the dynamometer card features used for working conditions recognition can be simplified. By predicting the change in characteristic parameters, it is possible to predict the working conditions over time.

In general, future working conditions are predictable. During a working conditions warning process, there are three important tasks that need to be completed. First, set up the mapping relationship between characteristic parameters and working conditions. Second, accurately find the change rules of the characteristic parameters over a continuous time.

Third, set up the working conditions warning method based on the mapping relationship and prediction results.

1.2. Related Works

Real-time diagnosis of working conditions based on dynamometer cards is the foundation for studying working conditions warnings. Traditional methods integrated statistical analysis of expert experience to match it with working conditions, which is easy to understand but highly subjective [3,4]. By calculating key features in the dynamometer cards, such as peak and valley loads, valve opening and closing point positions, and so on, typical working conditions can also be matched with them, thereby forming recognition rules [5]. Similarly, geometric features, grayscale matrices, and Fourier descriptors can also serve as the basis for forming recognition rules [6,7]. In recent years, with the development of machine and deep learning, research on the automation and intelligence recognition of dynamometer cards has become increasingly rich. Based on interpretable key characteristic parameters of dynamometer cards, some algorithms have been used to generate structured data classification [8]. Jinze Liu et al. (2021) used an improved Fourier descriptor for feature extraction and utilized the support vector machine (SVM) as the model to classify faults [9]. Considering the occurrence probabilities of different faults, Xiaoxiao Lv (2021) proposed evolutionary SVM methods [10]. Tianqi Chen (2016) proposed the Xgboost model, which has strong generalization ability for structured data modeling [11]. Lu Chen et al. (2021) used it as an effective method for fault recognition as well [12]. Compared with traditional machine learning methods, artificial neural networks are more flexible and widely used in fault recognition [13]. Yanbin Hou et al. (2019) and Tong Xu et al. (2019) used extreme learning machines [14,15] and RBF neural networks for fault recognition in dynamometer cards. Convolutional neural networks (CNN) have strong image recognition capability [16], and it is feasible to use CNN for fault recognition when the dynamometer cards are viewed as unstructured data, such as images [17]. With further research, researchers have noticed that insufficient samples lead to a decrease in the accuracy of fault recognition. Kai Zhang et al. (2022) effectively solved the fault recognition problem with small samples by using Meta-transfer learning [18]. Xiang Wang et al. (2021) oversampled the categories with fewer samples to solve the class imbalance problem in dynamometer card recognition [19], but it was prone to overfitting. Chengzhe Yin et al. (2023) used the Conditional Generative Adversarial Neural Network (CGAN) to generate dynamometer cards and enhanced the diversity of generated samples [20]. In this way, the recognition performance of the categories with fewer samples can be improved further.

The existing research about real-time working conditions diagnosis covers various aspects, including data, features, and models. Whether it is structured data, such as key characteristic features, or unstructured data, such as images, the working conditions of oil wells can be reflected accurately. From machine learning to deep learning, various diagnostic models can solve many complex problems. These research results provide rich references for working conditions warnings.

Warnings about working conditions are based on real-time diagnosis. It can predict the future production status of sucker rod wells. However, the features used to describe future working conditions are usually unknown. Therefore, the recognition of future working conditions is more difficult. Shaoqiang Bing (2019) proposed a comprehensive index related to wax deposition and used LSTM to predict the degree of wax deposition quantitatively [21]. Lin Xia et al. (2020) used LSTM and CNN to predict the trend in the dynamometer card shape over time [22]. Hui Tian et al. (2021) extracted the dynamometer card features by CNN to predict the occurrence time of severe paraffin problems [17]. Chaodong Tan et al. (2022) used the dynamometer card image and electrical features to predict the wax accumulation level through LSTM [23]. In other industrial fields, there are also some fault warnings research results worth learning from. Mathis Riber Skydt et al. (2021) defined three severity levels of power grid states and classified time series by data augmentation and LSTM [24]. Yulei Yang et al. proposed a novel method based

on SAE-LSTM to excavate features highly related to time series data, which can warn of defects 85 h in advance compared with the traditional threshold warning method [25]. Congzhi Huang et al. (2024) proposed a dual warning method considering univariate abnormal detection and multivariate coupling thresholds to learn about coal mill faults in time [26]. Kuan-Cheng Lin et al. (2024) used LSTM to analyze historical pre-failure information to predict the future status of wind turbine health [27]. Yunxiao Chen et al. (2024) analyzed the reasons for wind power prediction model failure and used variance to assist Bi-LSTM in 1-h-ahead early warning to solve the problems [28]. Hongqian Zhao et al. (2024) combined the gated recurrent unit neural network and multi-step ahead prediction scheme to accurately predict the battery voltage 1 min in advance [29]. In addition, the noise inside the data will affect the model performance during prediction. Na Qu et al. (2020) used discrete wavelet transformation for noise reduction and to decompose electrical signals, which improved the prediction accuracy of fault monitoring on the ENET public dataset [30]. Jun Ling et al. (2020) used multi-resolution wavelet transformation for noise reduction and fault warning for nuclear power machinery through RNN and Bayesian statistical inference [31].

The existing research on fault warning is generally based on time series prediction models. In the petroleum field, the warning research of sucker rod wells has mainly focused on a single working condition, and there are few studies on collaborative warnings for multiple working conditions. In other industrial fields, researchers generally predict faults from key signal data related to the specific field. When the noise in the data has an obvious influence on the prediction performance, it is necessary to reduce the noise and then reconstruct the data. During fault warning, many strategies exist, including category sequence prediction, threshold delineation, and binary classification, based on predicting results in advance. Although these studies focused less on multiple working conditions warnings, they also provided us with many references in terms of data processing and algorithms.

With the help of existing research results, in this paper, we extracted key characteristic parameters from dynamometer cards and used deep learning to predict them. Based on the prediction results, a working conditions warning method was implemented. The main content and innovations are as follows:

First, based on the sequence-to-sequence prediction mode, the time series of multiple characteristic parameters of the dynamometer cards could be predicted.

Second, the influence of noise on the characteristic parameters prediction results was considered in this paper.

Finally, multiple working conditions warning methods of the sucker rod well were proposed based on characteristic parameters prediction.

2. Characteristic Parameters Extraction

There may be deviations when extracting characteristic parameters from surface dynamometer cards due to the influence of dynamic loads. Based on the research of Gibbs S. G. and Qi Zhang et al. [32,33], to extract the characteristic parameters more accurately, in this paper, the surface dynamometer cards were converted into down-hole dynamometer cards by solving the one-dimensional damped-wave equation with a Fourier series and taking surface load and displacement as the boundary conditions. In this way, the calculation deviations caused by dynamic loads, such as vibration, inertia, and friction, can be eliminated to some degree.

In this paper, the characteristic parameters were extracted to describe the difference among several typical working conditions based on the down-hole dynamometer cards, as shown in Figure 2.

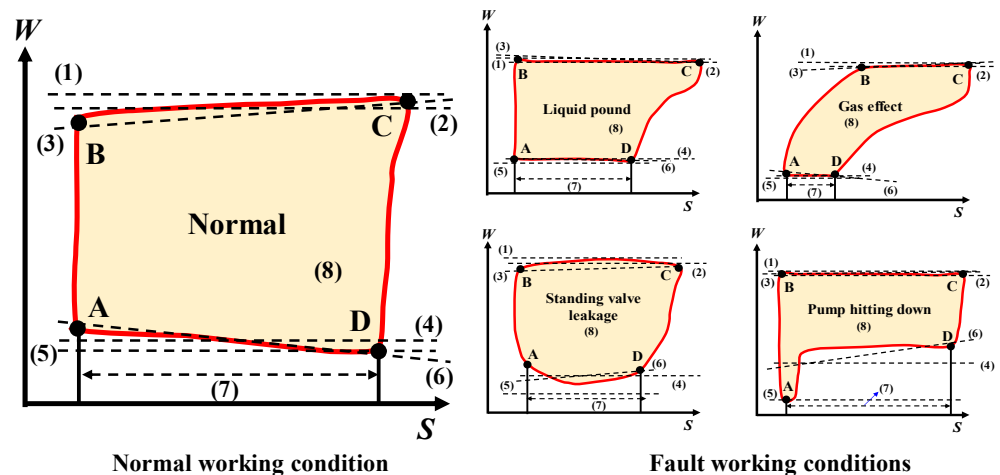


Figure 2. Characteristic parameters extraction results of the down-hole dynamometer cards.

In Figure 2, S represents the displacement, and W represents the load. In addition, A, B, C, and D represent the traveling valve closed point, standing valve open point, standing valve closed point, and traveling valve open point, respectively. It can be seen that the basic shape of the down-hole dynamometer cards was determined by the load and displacement change of the valve opening and closing points. Based on these, other key characteristic parameters can be obtained to describe different working conditions further. The detailed names are as follows, and the serial numbers are consistent with those in Figure 2.

(1) Maximum load

Maximum load always appears in the upstroke. It can be obtained by selecting the maximum value of the load data from B to C;

(2) Average load of the upstroke

The average load of the upstroke reflects the average level of the load change in the upstroke. It can be obtained by calculating the arithmetic mean of the load data from B to C;

(3) Slope of the upstroke

The slope of the upstroke is used to describe the uncertainty of the load change quantitatively. In this paper, the slope can be calculated by fitting the linear relationship between load and displacement from B to C;

(4) Average load of the downstroke

Similar to that of the upstroke, it can be obtained by calculating the arithmetic mean of the load data from D to A,

(5) Minimum load

The maximum load always appears in the downstroke. It can be obtained by selecting the minimum value of the load data from D to A;

(6) Slope of the downstroke

Similar to that of the upstroke, the slope can be calculated by fitting the linear relationship between load and displacement from D to A;

(7) Effective stroke

The effective stroke, also known as the piston stroke, usually reflects the production efficiency of the pump. It can be obtained by calculating the displacement difference between D and A;

(8) Area of the dynamometer card

The area of the dynamometer card indicates how much work the pump has done. It can be obtained by the numerical integration of the closed curve.

Different working conditions can be distinguished further by the above calculated parameters. Moreover, if the working conditions are described only by a specific characteristic parameter, there will be significant recognition deviation when multiple working conditions coexist. Therefore, they need to be determined by multiple characteristic parameters together.

In addition, it can be seen from Figure 1 that A, B, C, and D should be located at positions with significant curvature changes. However, because down-hole dynamometer cards eliminate the influence of vibration loads, the load change with displacement becomes smoother in the upstroke and downstroke. Therefore, it is more difficult to identify curvature changes, which leads to greater uncertainty of A, B, C, and D. When extracting the load and displacement of A, B, C, and D, there will be significant fluctuations. At the same time, the other extraction results also depend on the positions of A, B, C, and D, so there is noise in the time series of each characteristic parameter, which will have an influence on the prediction performance.

3. Methodology

3.1. Model of Working Conditions Classification

When the characteristic parameters extracted are in different intervals, they correspond to different working conditions. The mapping relationship between the working conditions and the parameter values can be set up through conditional judgment. However, it relies on expert experience. A trainable and self-learning knowledge tree (also named the classification model) can help construct a mapping relationship that is similar to the rules established between characteristic parameters and working conditions.

Xgboost is the improved version of tree models such as decision trees, Random forests, and GBDT. It establishes K regression trees (CART trees) to make the predicted values of the tree clusters as close to the true values as possible with the greatest generalization ability at the same time. The objective function is shown as Equation (1).

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (1)$$

where i represents the i -th sample. The objective function consists of two parts: loss function l and regularization term Ω . $l(\hat{y}_i, y_i)$ is the prediction error of the i -th sample, and $\Omega(f_k)$ is the function that reflects the complexity of a tree. The less $L(\phi)$ is, the lower the error and the stronger the generalization ability.

Through Taylor expansion and split search algorithms, the model can be trained by optimizing the objective function values. For detailed information, please refer to Part 3 of Tianqi Chen et al.'s article [9].

3.2. Noise Reduction Based on Wavelet Transformation

According to the analysis in Part 2, parameters to be predicted can be considered as one-dimensional signal data with noise. They are non-stationary signals composed of regular parts that change-over-time and noise is generated during extraction. Therefore, wavelet transformation is suitable for noise reduction. The process is shown in Figure 3.

First, after inputting a signal with noise, a wavelet basis function (db8) is used to perform multi-scale wavelet decomposition of the time sequence. The actual sequence length of each well determines the number of decomposed layers. Then, based on the setting threshold, the decomposed wavelet coefficients are selected. If the amplitude of the wavelet coefficients is lower than the threshold, it would be caused by noise, and the corresponding decomposed coefficients should be removed. The threshold is calculated by the hard threshold function in this paper to reserve more information on working conditions. Finally, by reconstructing the remaining decomposed coefficients, the time sequence after noise reduction can be obtained [34,35].

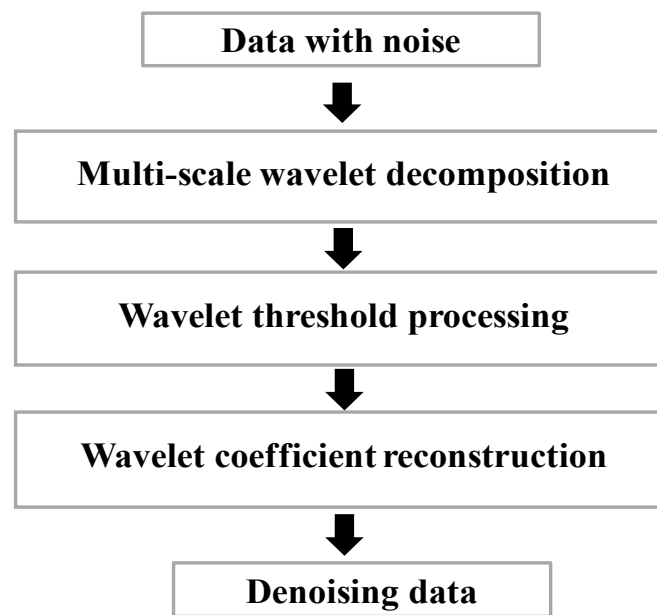


Figure 3. Noise reduction process by wavelet transformation.

The selection of thresholds is a challenging task. If the threshold is too large, the sequence will be smoother, and prediction accuracy will be higher. However, at the same time, it removes more valuable information, which makes the sequence unable to reflect the true rules. If the threshold is too small, the sequence will be closer to the real data and reserve more information that is valuable. But it also contains more noise, which results in poorer prediction performance. Therefore, the noise reduction performance under different thresholds has been shown in the Experiment Results analysis in order to select a reasonable threshold.

3.3. Model of Characteristic Parameters Prediction

The characteristic parameters of the future M -days can be predicted according to that of the past N -days. The prediction mode includes many-to-one rolling prediction and many-to-many sequence prediction.

Many-to-one rolling prediction refers to predicting the parameters for the next day based on those from the past N days. Then, the parameters of the second next day are predicted based on the above prediction results, and so on, until the M -th day. In this way, the sequential relationship between future time series can be considered. However, due to errors in each prediction step, error propagation and accumulation can occur. Many-to-many sequence prediction refers to directly predicting the parameters for the next M days based on those from the past N days without error propagation and accumulation. However, the model cannot learn the sequential internal relationships of the predicted time series by sequence-to-sequence prediction alone. In summary, it is necessary to choose a prediction method that can both avoid error propagation and learn the sequential inner relationships.

In addition, the pumping parameters, including pump diameter, pump depth, stroke, and stroke frequency, can describe the production difference in different wells. They can contribute to characteristic parameter predictions as supplementary features. In general, they change little in a short period and are known in prediction time. During the prediction, they are used in the prediction interval (M -days). There is a one-to-one relationship between them and the predicted parameters. However, the historical features are different because they are used in the history interval (N -days), and there is a sequence-to-sequence (M - N) relationship between them and the predicted parameters. Therefore, it is necessary to choose a model that can fuse the features within different time intervals (M -days or N -days) for prediction.

Based on the above analysis, the model should have the following two functions. First, it can achieve sequence-to-sequence prediction and learn the sequential internal relationships. Second, it can effectively fuse historical characteristic parameters and supplementary features. Bryan Lim et al. (2020), from Google, proposed the Temporary Fusion Transformers (TFT) [36], which combine the advantageous strategies of existing time series prediction research and can meet the need for parameter prediction in this paper. The main structure of the model is shown in Figure 4.

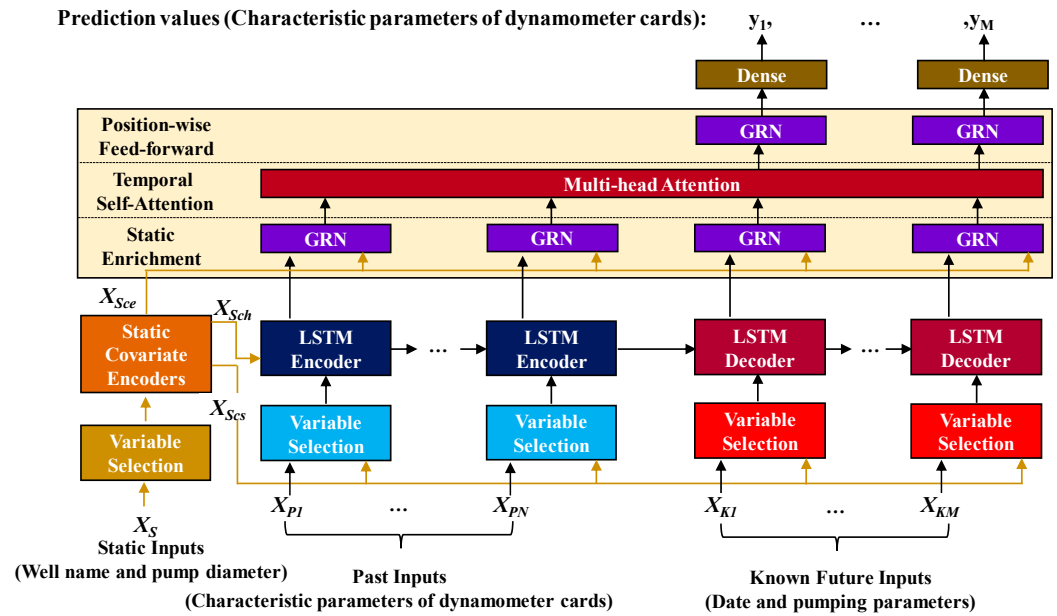


Figure 4. The model structure of Temporal Fusion Transformers (TFT).

It can be seen that TFT can solve the problem of sequence-to-sequence prediction by Encoder and Decoder. It divides features into static inputs $X_S = \{X_{S_1}, X_{S_2}, \dots, X_{S_{N+M}}\}$, past inputs $X_P = \{X_{P_1}, X_{P_2}, \dots, X_{P_N}\}$, and known future inputs $X_K = \{X_{K_1}, X_{K_2}, \dots, X_{K_M}\}$. Based on Encoder and Decoder, those features are fused effectively. The main principle and function of the model are explained as follows:

Encoder–Decoder is popularly used in Seq2Seq semantic models. The Encoder and Decoder are composed of Long-Short-Term Memory (LSTM) units, which enable the flexible lengths of the input and output of sequences, as shown in Figure 5.

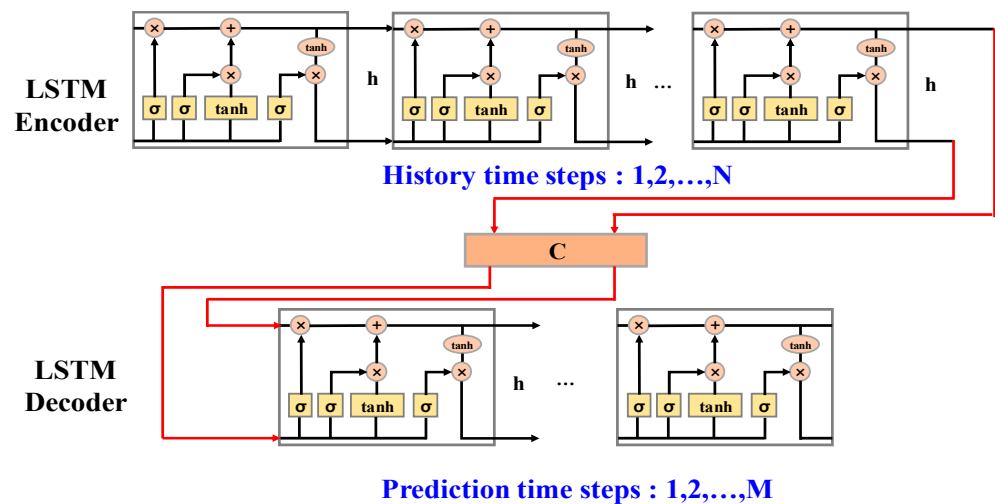


Figure 5. The structure of Encoder–Decoder based on LSTM units.

It can be seen that because the Encoder is composed of N LSTM units, it is possible to learn the sequence internal relationship of the N historical time steps of data as input. The output vector of the Encoder is C, which is also the input of the Decoder. The Decoder consists of M LSTM units corresponding with prediction time steps, which enables learning the internal relationship of predicted sequences. This method meets the demand for many-to-many time series prediction in this paper.

The supplementary features consist of the static and known future inputs, as shown in Figure 4. Static inputs are concatenated with historical and prediction time steps. Known future inputs correspond to predicted time steps together with past inputs by Encoder–Decoder. The simplified details of the correspondence between the above features (without considering hidden layers such as Gate and Variable Selection) are as follows:

First, there are three output components $X_{S_{ce}}, X_{S_{ch}}, X_{S_{cs}}$ after inputting X_s into Static Covariate Encoders. $X_{S_{ce}}$ enters into the static enrichment layer, $X_{S_{ch}}$ and X_P enter into LSTM Encoder and $X_{S_{cs}}$ is directly connected with the input layer. The input layer can be simplified as Equation (2).

$$\begin{aligned} & \{X_{Input_i} | i \in \{1, 2, \dots, N + M\}\} \\ & = \{(X_{P_1}, X_{S_{cs1}}), \dots, (X_{P_N}, X_{S_{csN}}), (X_{K_1}, X_{S_{csN+1}}), \dots, (X_{K_M}, X_{S_{csN+M}})\} \end{aligned} \tag{2}$$

After inputting $X_P, X_{S_{ch}}$, and $\{X_{Input_i} | i \in \{1, 2, \dots, N\}\}$ into the LSTM Encoder, the intermediate output in each LSTM unit in the Decoder is shown as Equation (3).

$$\widetilde{X_{PS_{ch}}} = \{\widetilde{X_{PS_{ch1}}}, \widetilde{X_{PS_{ch2}}}, \dots, \widetilde{X_{PS_{chM}}}\} \tag{3}$$

Therefore, after the Encoder–Decoder, the output of the static enrichment layer can be simplified as Equation (4).

$$\begin{aligned} X_{Output} = & \{Encoder((X_{P_1}, X_{S_{cs1}}), \dots, (X_{P_N}, X_{S_{csN}}))\} \\ & \cup \{Decoder((\widetilde{X_{PS_{ch1}}}, X_{K_1}, X_{S_{csN+1}}), \dots, (\widetilde{X_{PS_{chM}}}, X_{K_M}, X_{S_{csN+M}}))\} \end{aligned} \tag{4}$$

Through the above calculation, the model can incorporate autoregressive historical features and other supplementary features, such as pump diameter, pump depth, stroke, and stroke frequency.

In addition, in the feedforward of the neural network structure, a gated residual network (GRN) is also applied to TFT in order to control the degree of nonlinear transformation. Variables a and optional context vector c (which can be ignored in this article) are the inputs of GRN, which can be shown in Equation (5) to Equation (7).

$$GRN(a, c) = LayerNorm(a + GLU(\eta_1)) \tag{5}$$

$$\eta_1 = W_1\eta_2 + b_1 \tag{6}$$

$$\eta_2 = ELU(W_2a + W_3c + b_2) \tag{7}$$

where W is the weight, b is the bias, and ELU is the activation function [32]. The expression for GLU is shown as Equation (8).

$$GLU(\gamma) = \sigma(W_4\gamma + b_4) \odot (W_5\gamma + b_5) \tag{8}$$

where σ is the sigmoid activation function and \odot is the Hadamard product. When the data volume of some wells is small, complex nonlinear transformation does not need to be performed during training. After using GLU, its output is a vector close to 0, which is almost equivalent to no nonlinear transformation.

Finally, the TFT adopts a multi-head attention mechanism to solve the problem of information loss caused by excessive information in the Encoder–Decoder process.

3.4. Working Conditions Warning Technical Route

By inputting the predicted characteristic parameters into the above working condition classification model, the future working conditions can be predicted as well, which provides intuitive working conditions warning results for the site. The research technical route is shown in Figure 6.

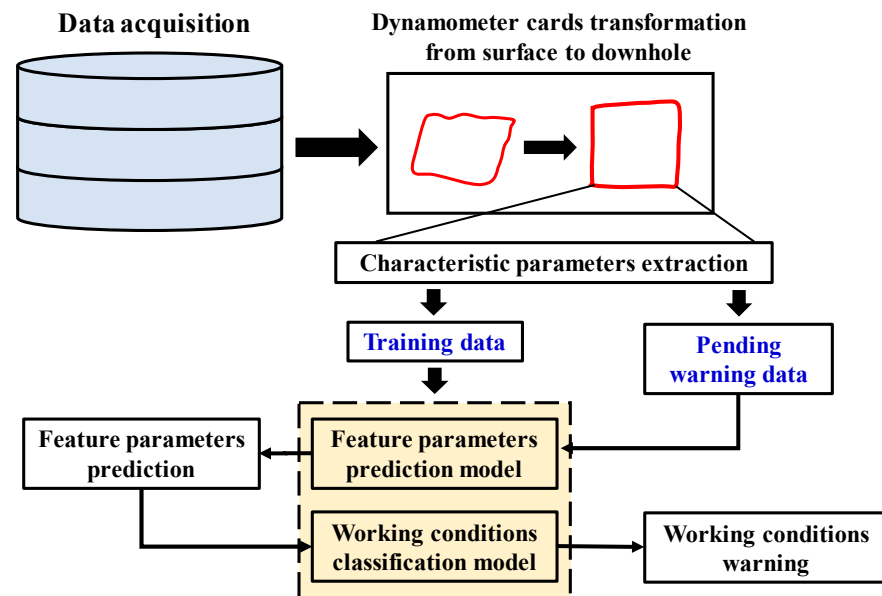


Figure 6. The research technical route of working conditions warning.

4. Experiment Results Analysis

4.1. Real-World Dataset from D Oilfield

In the D oilfield, the dynamometer card was to have been collected about once every 20 min. However, due to facilities and real-site condition differences, the collection frequency of dynamometer cards for each well was different. Furthermore, the dynamometer card would not change obviously within 20 min. Therefore, to unify the data frequency, we diluted the dynamometer cards to one sample per day, which can also meet the needs of actual oilfield applications. Then, by converting the surface dynamometer cards into down-hole dynamometer cards and the characteristic parameters extraction (Part 2), the real-world dataset could be obtained.

In the dataset, there were 182414 samples from 565 wells in the D oilfield, covering a total of 664 days in the past two years. The samples from the first 295 days account for 70% of all the samples, which were used as the training set. The remaining samples (from the 296th to 664th days) which account for 30% of all the samples were used as the validation set. The data size of different working conditions in the training and validation set is shown in Table 1.

Table 1. Data size of different working conditions.

Working Conditions	Data Size of Training Set	Data Size of Validation Set
Normal	71,174	31,336
Liquid pound	48,622	22,248
Gas effect	3402	687
Standing valve leakage	2808	36
Pump hitting down	2396	392
Total	128,402	54,012

It can be seen that normal and liquid pound samples account for over 90%. Due to the temporal characteristics, the date of the validation set must be after that of the training set. Therefore, stratified sampling cannot be used to split the data.

4.2. Working Conditions Classification Results

Based on the extraction results of the characteristic parameters, the Xgboost classification model was constructed to obtain the mapping relationship between parameters and working conditions. Based on the data volume, number of features, and experience, the values of the main hyperparameters are set out in Table 2.

Table 2. Values setting of the Xgboost model hyperparameters.

Hyper Parameters Name	Value	Meaning
Learning rate	0.06	It is used to control the iteration rate of trees.
Numbers of estimators	800	The total number of iterations, i.e., the number of decision trees.
Max depth	8	The max depth of trees represents the complexity of the tree model.
Min child weight	10	The minimum number of samples specified on the child nodes. The larger the value, the easier it is to cause under-fitting.
Gamma	1	The penalty coefficient represents the minimum loss function descent required for specified node splitting.
Subsample	0.8	The proportion of data used in training each tree to the entire training set.
Colsample btree	1	The proportion of features used in training each tree to all features.

To eliminate the influence of dimensionality and accelerate model convergence, the characteristic parameters were standardized by Equation (9) before model training.

$$X_{std} = \frac{X - \mu}{\sigma} \tag{9}$$

where X_{std} and X are the real and standardized values of some characteristic parameters, respectively, μ and σ are the mean value and standard deviation of the corresponding parameter, respectively.

The binary classification model performance is usually evaluated by the confusion matrix in machine learning, as shown in Table 3.

Table 3. Confusion matrix of the classification results.

Actual Results	Prediction Results	
	Positive Category	Negative Category
Positive category	TP (True positive)	FN (False negative)
Negative category	FP (False positive)	TN (True negative)

Based on Table 3, Precision, Recall, and F1 were used to evaluate the model’s performance. The detailed calculation method is shown in Equation (10) to Equation (12).

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{12}$$

Recall refers to the proportion of positive samples that are correctly classified among all positive samples. It reflects whether the model has underreporting for the specific working condition. Precision refers to the proportion of samples that are truly positive among all samples classified as positive. It reflects whether the model has misreporting. The F1 value is the harmonic mean of Recall and Precision, which reflects the general performance of the model classification.

When evaluating the performance of multiple classifications, the specified category is usually viewed as the positive category, and the other categories are viewed as the negative category. Therefore, the model of multiple working conditions classification can also be evaluated using the above method. The results are shown in Table 4.

Table 4. Working conditions classification results of the training and validation set.

Working Conditions	Training Set			Validation Set		
	Precision	Recall	F1	Precision	Recall	F1
Normal	1.00	1.00	1.00	1.00	1.00	1.00
Liquid pound	1.00	1.00	1.00	1.00	1.00	1.00
Gas effect	0.99	0.98	0.99	0.96	0.94	0.95
Pump hitting down	1.00	1.00	1.00	0.99	1.00	0.99
Standing valve leakage	1.00	0.98	0.99	0.89	0.86	0.87

It can be seen that the F1 values of most working conditions were over 0.95 on the validation set. Recall values for the standing valve leakage were relatively lower than the other working conditions. Furthermore, the confusion matrix of the classification results on the validation set is shown in Figure 7.

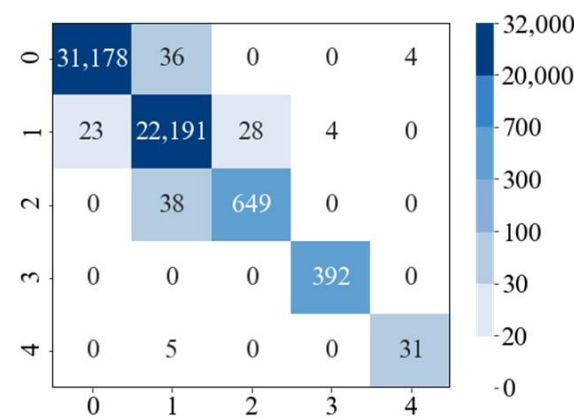


Figure 7. The confusion matrix of the classification model on the validation set. Explanation of the working conditions number: Normal: 0, Liquid pound: 1, Gas effect: 2, Pump hitting down: 3, Standing valve leakage: 4.

It can be seen that some samples of the standing valve leakage were classified as liquid pound, resulting in Recall decrease. The detailed reasons will be discussed further in the working conditions warning results.

4.3. Noise Reduction and Soundness Verification

The noise of characteristic parameters was reduced by wavelet transformation with the db8 as the basis function. The high-frequency part in the parameters may be either noise or an indication of some faulty working conditions. It is important to reserve the high-frequency part, which indicates the fault as much as possible. Therefore, the hard threshold function was chosen for noise reduction. In this paper, the reciprocal of the standard deviation of parameters (reflecting the degree of noise reduction) and the average

Recall of working conditions after noise reduction under different thresholds have been applied to help obtain an appropriate threshold value. The results are shown in Figure 8.

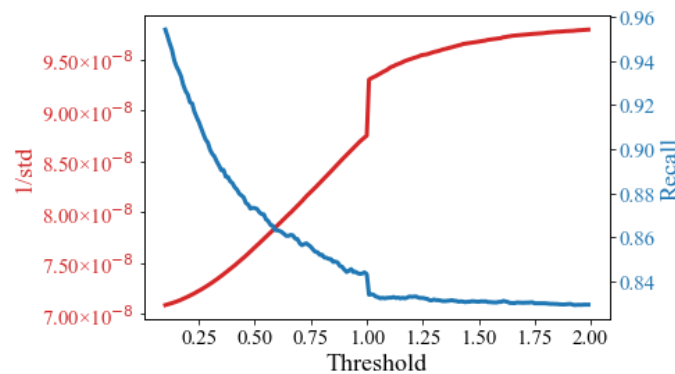


Figure 8. Noise reduction performance under different thresholds.

It can be seen that as the threshold increased, the reciprocal of the standard deviation of the parameters increased, as well, which indicates that more high-frequency parts were discarded and the parameter changed more smoothly over time. Meanwhile, the decrease in Recall indicates that the noise reduction discarded some high-frequency parts indicating faults. When the threshold was greater than 1.0, the reciprocal and Recall appeared to change suddenly. Therefore, 1.0 can be the threshold for wavelet noise reduction.

When the threshold was 1.0, the average Recall was 0.84, which decreased by 0.14 compared with 0.98 of the original data in the training set (threshold = 0). In order to reserve more fault information on the noise reduction data, we restored the characteristic parameters after noise reduction to the original values, except those of normal and liquid pound, so that the prediction model learned fault information as much as possible. The classification model performance with the noise reduction data and original data on the validation set is shown in Table 5.

Table 5. Classification results of the validation set before and after noise reduction.

Working Conditions	Before Noise Reduction			After Noise Reduction		
	Precision	Recall	F1	Precision	Recall	F1
Normal	1.00	1.00	1.00	1.00	1.00	1.00
Liquid pound	1.00	1.00	1.00	1.00	1.00	1.00
Gas effect	0.96	0.94	0.95	0.97	0.94	0.96
Pump hitting down	0.99	1.00	0.99	1.00	1.00	1.00
Standing valve leakage	0.89	0.86	0.87	0.78	0.86	0.82

Similarly, the confusion matrix before and after parameters noise reduction is shown in Figure 9.

It can be seen that after noise reduction, the Precision, Recall, and F1 values were nearly equal to those before noise reduction. Due to the change in characteristic parameters after noise reduction, more normal samples were wrongly classified into standing valve leakage. Therefore, the Precision of the standing valve leakage decreased compared to before noise reduction.

In this paper, we also proposed the Restored rate to evaluate the performance after the data process or prediction. It can be defined as Equation (13).

$$Restored\ rate = \frac{AR}{R} \tag{13}$$

where R is the value of original data, such as the Recall before noise reduction or prediction. AR is the value of processed data, such as the Recall after noise reduction or prediction. The Restored rate of Precision, Recall, and F1 values in Table 5 are shown in Table 6.

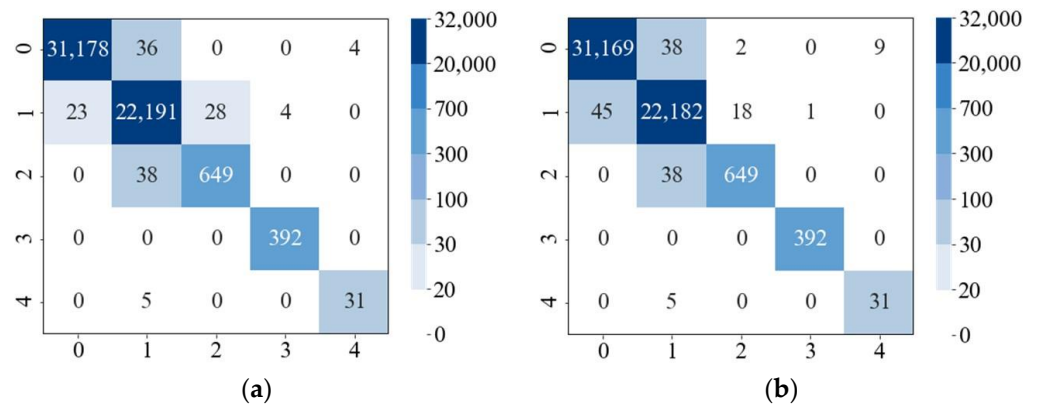


Figure 9. The confusion matrix of the classification model on the validation set before and after noise reduction. (a) Before noise reduction, (b) After noise reduction.

Table 6. Restored rate of the validation set.

Working Conditions	Restored Rate		
	Precision	Recall	F1
Normal	1.00	1.00	1.00
Liquid pound	1.00	1.00	1.00
Gas effect	1.01	1.00	1.01
Pump hitting down	1.01	1.00	1.01
Standing valve leakage	0.88	1.00	0.94

It can be seen that after noise reduction, the Precision, Recall, and F1 values of most working conditions were restored to over 95%, except the Precision of standing valve leakage, which was restored to 88%. It indicates that the majority of fault characteristics were reserved, and the degree of noise reduction was controlled reasonably. Moreover, in terms of the gas effect and pump hitting down, the Restored rates of Precision and F1 values were over 1.0, indicating that the noise reduction can contribute to the classification performance. Generally, in this way, we can remove the invalid high-frequency noise and reserve the effective high-frequency fault information at the same time, which can provide higher-quality training data for parameters prediction.

4.4. Characteristic Parameters Prediction Results

According to the model structure shown in Figure 4, the feature names corresponding to different feature types are shown in Table 7.

Table 7. Features name of different feature types inside the TFT model.

Feature Name	Feature Type
Prediction characteristic parameters	Past inputs
Date (month, day)	Known inputs
Pump depth	Known inputs
Stroke	Known inputs
Stroke frequency	Known inputs
Pump diameter	Static inputs
Well name	Static inputs

Historical characteristic parameters, such as past inputs, are the most important features used for autoregressive prediction. Date, pump depth, stroke, and stroke frequency are usually known in the operation process, so they can be viewed as known inputs. If the pump is not replaced, the pump diameter usually remains unchanged. During prediction, the most recent pump diameters can be used as static inputs along with the well name.

Based on the data volume, prediction length and experience, the values of main TFT hyperparameters were set as shown in Table 8. The values of the other hyperparameters can be set according to Bryan Lim's research paper [36].

Table 8. Values setting of TFT model hyperparameters.

Hyper Parameters Name	Value	Meaning
Learning rate	4×10^{-4}	It is used for controlling the speed of gradient descent
Dropout rate	0.1	It is used for preventing overfitting. During the training process, 10% of neurons will be randomly discarded to improve the model's generalization ability
Batch size	64	It represents the amount of training data used for updating network weights each time.
Hidden layer size	160	Number of hidden layer neurons of Variable Selection, LSTM Encoder, LSTM Decoder, GRN, Dense layers, and so on.
Max gradient norm	0.01	It is used to set the degree of gradient clipping so that the 2-norm of gradient values cannot exceed this value, which can help prevent gradient explosion or disappearance.
Number of heads	4	Number of heads in multi-head attention mechanism.
Activation	ELU	Name of the activation function.
Early stopping patience	1	The training epoch that stops early. If the validation loss does not improve before the number of this epoch, the training process will be terminated.
Number of encoder steps	28	Historical time steps.
Total time steps	35	The sum of historical time steps and prediction time steps.

The prediction models were trained using the original and noise reduction data. In this article, the data from the first four weeks (28 days) were used to predict the characteristic parameters of the following one week (7 days). Before model training, the input features and output labels were also standardized by Equation (1). The mean absolute error was used for Loss calculation. In the training process, when the loss of the validation set increased, the model training should be terminated. The loss changes of each characteristic parameter before and after noise reduction on the training and validation sets are shown in Figure 10.

It can be seen that the loss of each parameter on the validation set was lower than that without noise reduction. Before noise reduction, due to invalid high-frequency data, the model overfitting was severe, resulting in poor prediction performance on the validation set. In addition, after noise reduction, the number of epochs was larger before the loss of the validation set increased, which made the model training more complete.

The normalized absolute error (relative error) of the validation set was used to evaluate the prediction model performance. It can be shown as Equation (14).

$$relative\ error = \frac{100\%}{7N} \sum_{t=1}^7 \sum_{i=1}^N \left| \frac{y_{true}(i,t) - y_{pred}(i,t)}{y_{true}(i,t)} \right| \quad (14)$$

The relative errors of different characteristic parameters before and after noise reduction are shown in Table 9.

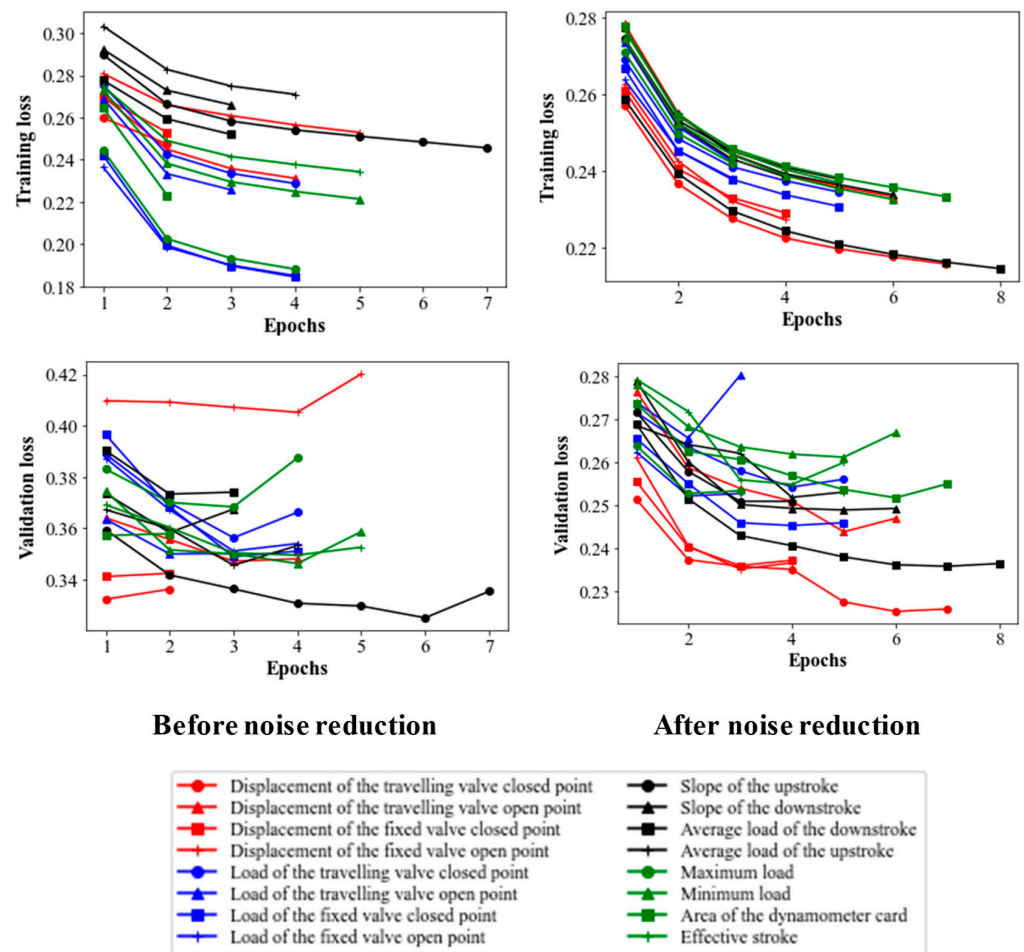


Figure 10. Loss changes before and after noise reduction on the training and validation set.

Table 9. Prediction relative error of characteristic parameters.

Predicted Parameters	Before Noise Reduction	After Noise Reduction
Displacement of the traveling valve closed point	33.9%	13.61%
Displacement of the traveling valve open point	4.58%	2.00%
Displacement of the fixed valve closed point	1.93%	0.77%
Displacement of the fixed valve open point	28.99%	11.68%
Load of the traveling valve closed point	11.65%	5.25%
Load of the traveling valve open point	11.08%	5.75%
Load of the fixed valve closed point	6.2%	2.58%
Load of the fixed valve open point	6.2%	2.82%
Slope of the upstroke	33.7%	14.4%
Slope of the downstroke	41.58%	21.71%
Average load of the downstroke	21.08%	10.07%
Average load of the upstroke	1.00%	0.35%
Maximum load	2.39%	0.92%
Minimum load	3.5%	1.48%
Area of the dynamometer card	6.87%	2.49%
Effective stroke	5.27%	2.53%

It can be seen that the relative errors were generally lower than without noise reduction, especially for the Displacement of the traveling valve closed point and the fixed valve open point, Slope of the upstroke and downstroke, and Average load of the downstroke. The relative prediction errors of most parameters after noise reduction were less than 15%, which can meet the requirements of engineering applications.

Due to the space limitation, taking two significantly improved parameters (Slope of down stroke and Displacement of the traveling valve open point) as examples, the comparison before and after noise reduction is shown in Figures 11 and 12.

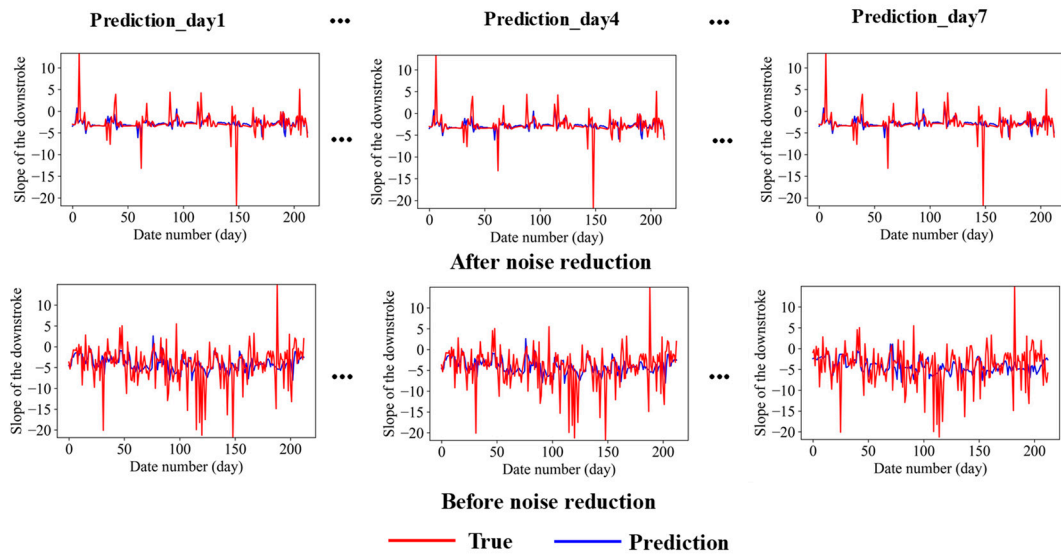


Figure 11. Slope of the downstroke prediction before and after noise reduction.

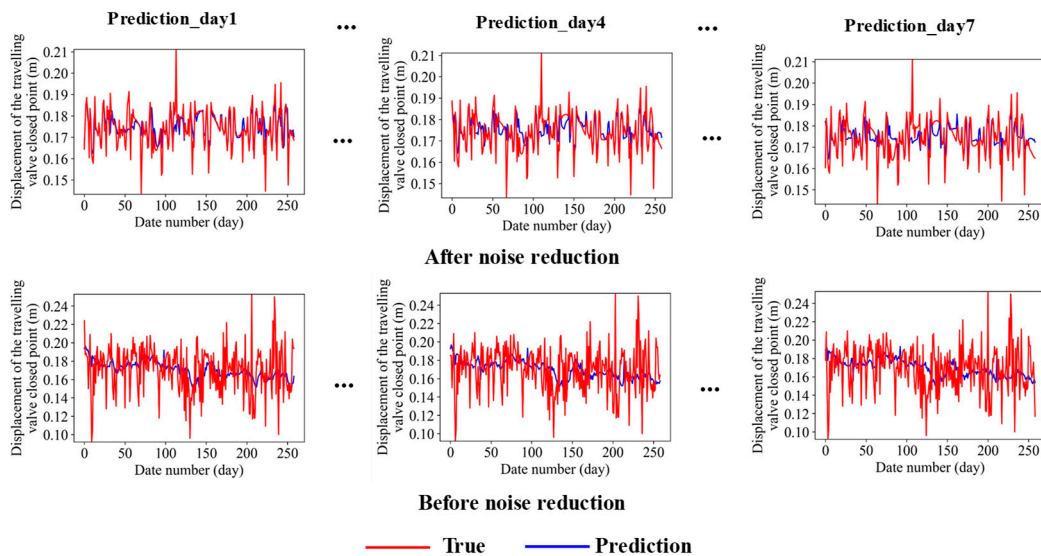


Figure 12. Displacement of the traveling valve open point before and after noise reduction.

It can be seen that the high-frequency noise decreased after noise reduction, and the fluctuation of parameters decreased as well, which can help obtain better prediction performance. In fact, deep learning methods are better at handling time series with internal correlation during prediction. If the fluctuation is too large, it will greatly increase the difficulty of model prediction and even cause overfitting. Therefore, a fluctuation decrease is beneficial for improving the prediction performance on the validation set.

As the prediction time step increased, the gap between the prediction and real values increased, which indicates that the effective information provided by the past 28 time steps for the future time steps decayed continuously. It is normal for time series predictions that the prediction results rely on the information that was closer in time to them. In addition, during some time intervals, there is a time lag in the prediction results (i.e., compared to the real value at time $t + 1$, the predicted result at time $t + 1$ is closer to the true value at time

t). The result is that the internal correlation of the time sequence over time is not obvious, which means that the time sequence is closer to a random walk sequence. Moreover, this phenomenon also results from the lack of effective features that change over time, which is also a common problem in time series prediction.

4.5. Working Conditions Warning Results

The working conditions of the next seven days for each well could be predicted by inputting the characteristic parameters into the classification model. In this way, a short-term working conditions warning method can be achieved. Within 7 days, the comparisons of working conditions warning results, including Precision, Recall, and F1 value, based on real and predicted parameters is shown in Tables 10–12, respectively.

Table 10. Precision comparison between real and predicted parameters.

Working Conditions Number		0	1	2	3	4
Day1	Real	1	1	0.97	1	0.71
	Prediction	1	0.99	0.87	0.93	0.51
Day2	Real	1	1	0.97	1	0.7
	Prediction	1	0.99	0.87	0.93	0.45
Day3	Real	1	1	0.97	1	0.68
	Prediction	1	0.99	0.87	0.93	0.43
Day4	Real	1	1	0.97	1	0.69
	Prediction	1	0.99	0.86	0.93	0.4
Day5	Real	1	1	0.97	1	0.7
	Prediction	1	0.99	0.86	0.93	0.4
Day6	Real	1	1	0.97	1	0.7
	Prediction	1	0.99	0.86	0.93	0.38
Day7	Real	1	1	0.97	1	0.72
	Prediction	1	0.99	0.86	0.93	0.34

Table 11. Recall comparison between real and predicted parameters.

Working Conditions Number		0	1	2	3	4
Day1	Real	1	1	0.94	1	0.81
	Prediction	0.99	0.99	0.96	0.82	0.85
Day2	Real	1	1	0.95	1	0.81
	Prediction	0.99	0.99	0.96	0.81	0.85
Day3	Real	1	1	0.94	1	0.79
	Prediction	0.99	0.99	0.95	0.81	0.73
Day4	Real	1	1	0.94	1	0.8
	Prediction	0.99	0.99	0.95	0.8	0.76
Day5	Real	1	1	0.94	1	0.81
	Prediction	0.99	0.99	0.94	0.81	0.73
Day6	Real	1	1	0.94	1	0.81
	Prediction	0.99	0.99	0.94	0.8	0.69
Day7	Real	0.99	1	0.94	1	0.81
	Prediction	0.99	0.99	0.94	0.8	0.62

Table 12. F1 score comparison between real and predicted parameters.

Working Conditions Number		0	1	2	3	4
Day1	Real	1	1	0.96	1	0.76
	Prediction	0.99	0.99	0.92	0.87	0.64
Day2	Real	1	1	0.96	1	0.75
	Prediction	0.99	0.99	0.91	0.86	0.59
Day3	Real	1	1	0.96	1	0.73
	Prediction	0.99	0.99	0.91	0.87	0.56
Day4	Real	1	1	0.96	1	0.74
	Prediction	0.99	0.99	0.9	0.86	0.52
Day5	Real	1	1	0.96	1	0.75
	Prediction	0.99	0.99	0.9	0.86	0.51
Day6	Real	1	1	0.96	1	0.75
	Prediction	0.99	0.99	0.9	0.86	0.49
Day7	Real	1	1	0.96	1	0.76
	Prediction	0.99	0.99	0.9	0.86	0.44

It can be seen that compared with real parameters, Precision, Recall, and F1 score based on predicted parameters decreased during a working conditions warning, especially in Precision and F1 score of standing valve leakage. In order to make the results more vivid, the working conditions warning results of change-over-time are shown in Figures 13 and 14 based on Tables 10–12.

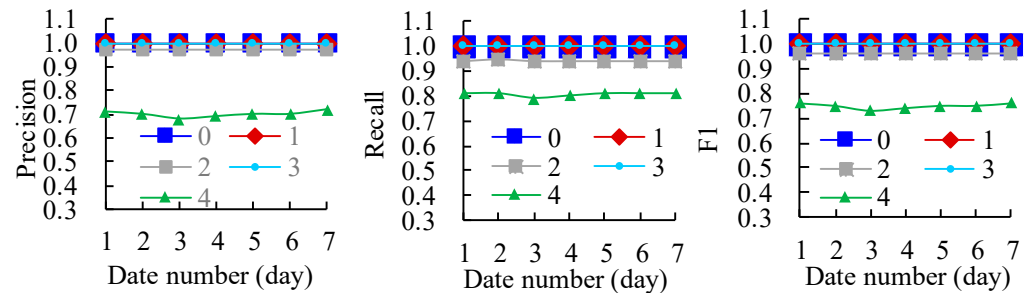


Figure 13. Warning results of change-over-time based on real parameters.

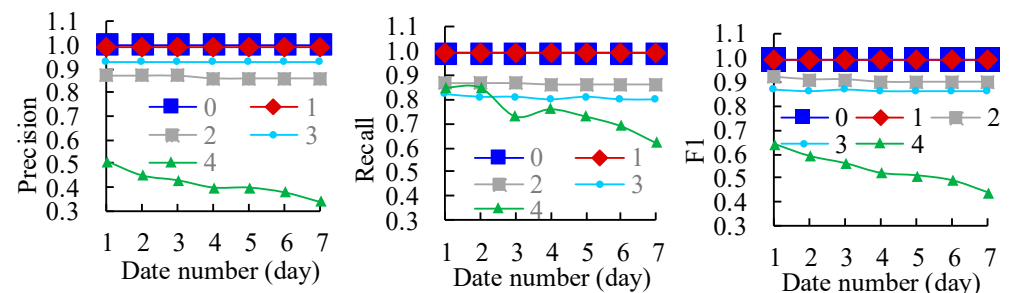


Figure 14. Warning results of change-over-time based on predicted parameters.

In order to compare the differences between warning and real classification results for change-over-time quantitatively, based on Figures 13 and 14 and the definition of Restored rate, the Restored rates of different working conditions are shown in Figure 15.

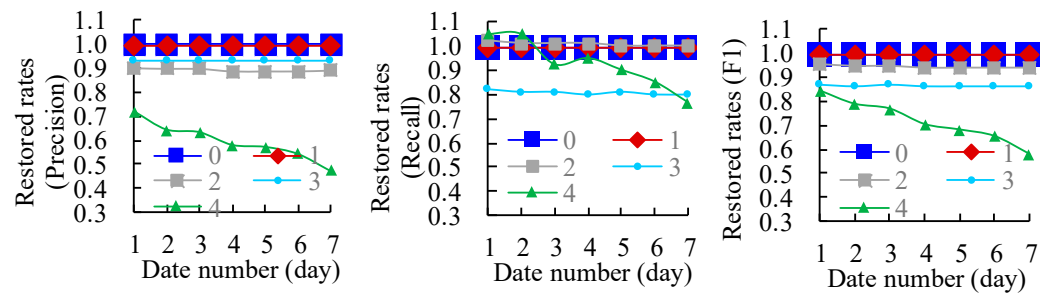


Figure 15. Restored rates of change-over-time of different working conditions.

It can be seen that as the prediction time went by, the classification performance with real parameters changed little. However, the warning performance with predicted parameters became worse over time, especially for standing valve leakage. Within a 7-day prediction interval, the Restored rates of F1 values under most working conditions reached over 85%, and those of Recall values reached over 80%. Moreover, the Recall values of the liquid pound, normal, and gas effect were restored to over 95%, indicating that there were few fault-underreporting samples.

In addition, the Precision values and the Restored rates of the standing valve leakage were generally low within the prediction interval. The classification results are shown in detail in the confusion matrixes (Figure 16) within a 7-day prediction.

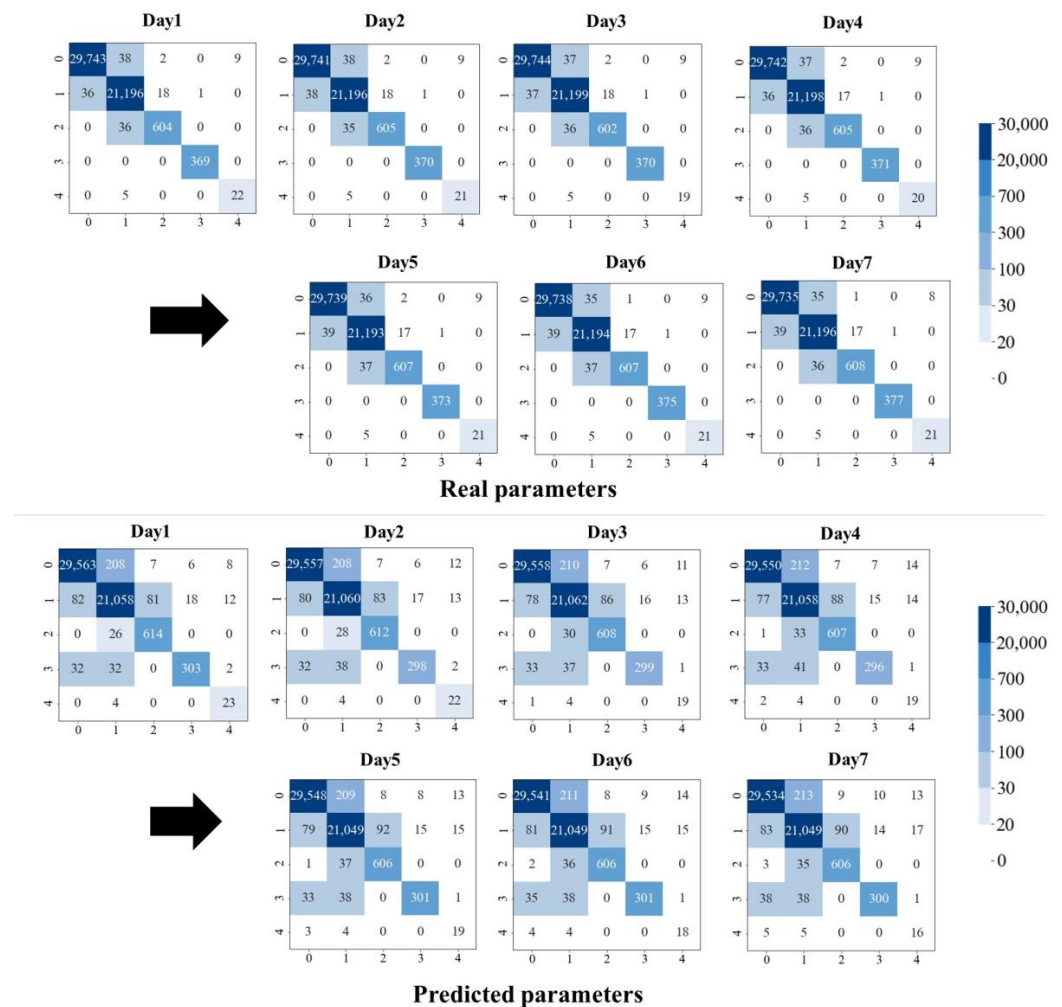


Figure 16. Confusion matrixes of change-over-time based on real and predicted parameters.

It can be seen that there were more samples wrongly classified during warning by predicted parameters than true parameters. For normal and liquid pound, samples were wrongly classified into the standing valve leakage and pump hitting down. Due to the larger number of them than in other working conditions, their down-hole dynamometer cards were more diverse as well. Therefore, some samples were not identical but similar to other working conditions that existed in them, which is shown in Figure 17.

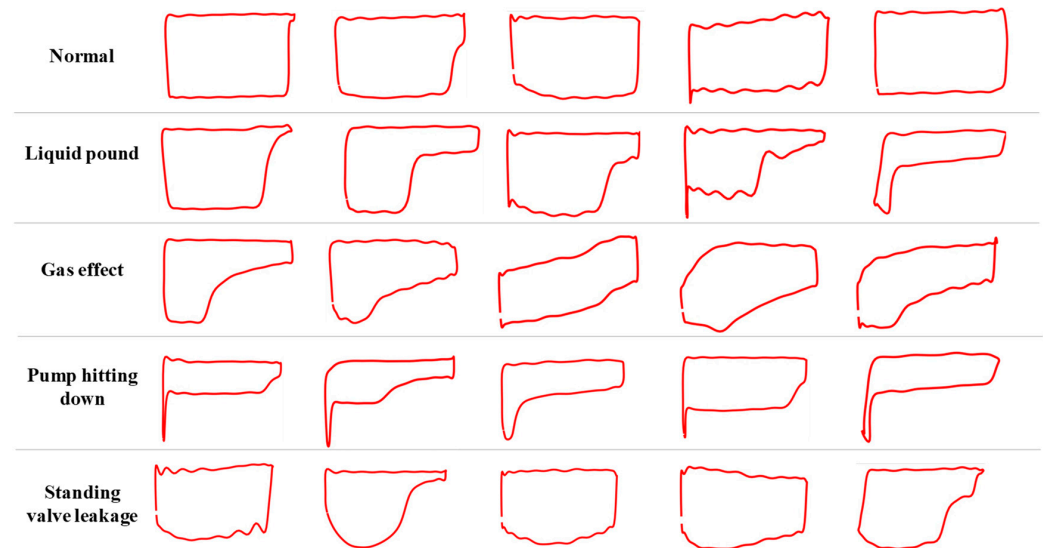


Figure 17. Examples that are easily confused between different working conditions.

From Figure 17, some conclusions can be summarized as follows. First, in the samples of normal condition, there were some standing valve leakage and pump hitting down samples with weak severity. Second, in the samples of liquid pound, not only were there composite working conditions between liquid pound and standing valve leakage but there were also composite working conditions between liquid pound and pump hitting down. It may cause the liquid pound to be prone to misclassification as standing valve leakage and pump hitting down. Third, in the samples of pump hitting down, there were some samples that were similar to severe liquid pound. It may cause the pump hitting down to be prone to misclassification as liquid pound. Finally, in the samples of standing valve leakage, there were composite working conditions between liquid pound and standing valve leakage. It may cause the standing valve leakage to be prone to misclassification as liquid pound as well.

In summary, there were similar or composite shapes in different working conditions. When there were errors during parameter prediction, these samples were more likely to be misclassified. These samples only account for a small proportion of the normal condition and liquid pound due to their large data size. However, in the validation set, the data size of standing valve leakage was much smaller than the other working conditions. Therefore, these easily misclassified samples account for a large proportion of standing valve leakage, which resulted in lower precision during working conditions warning.

In order to show the warning results in detail, representative time points of some wells were selected. In Figure 18, the warning process (28–34) with historical data (0–27) over time can be observed visually.

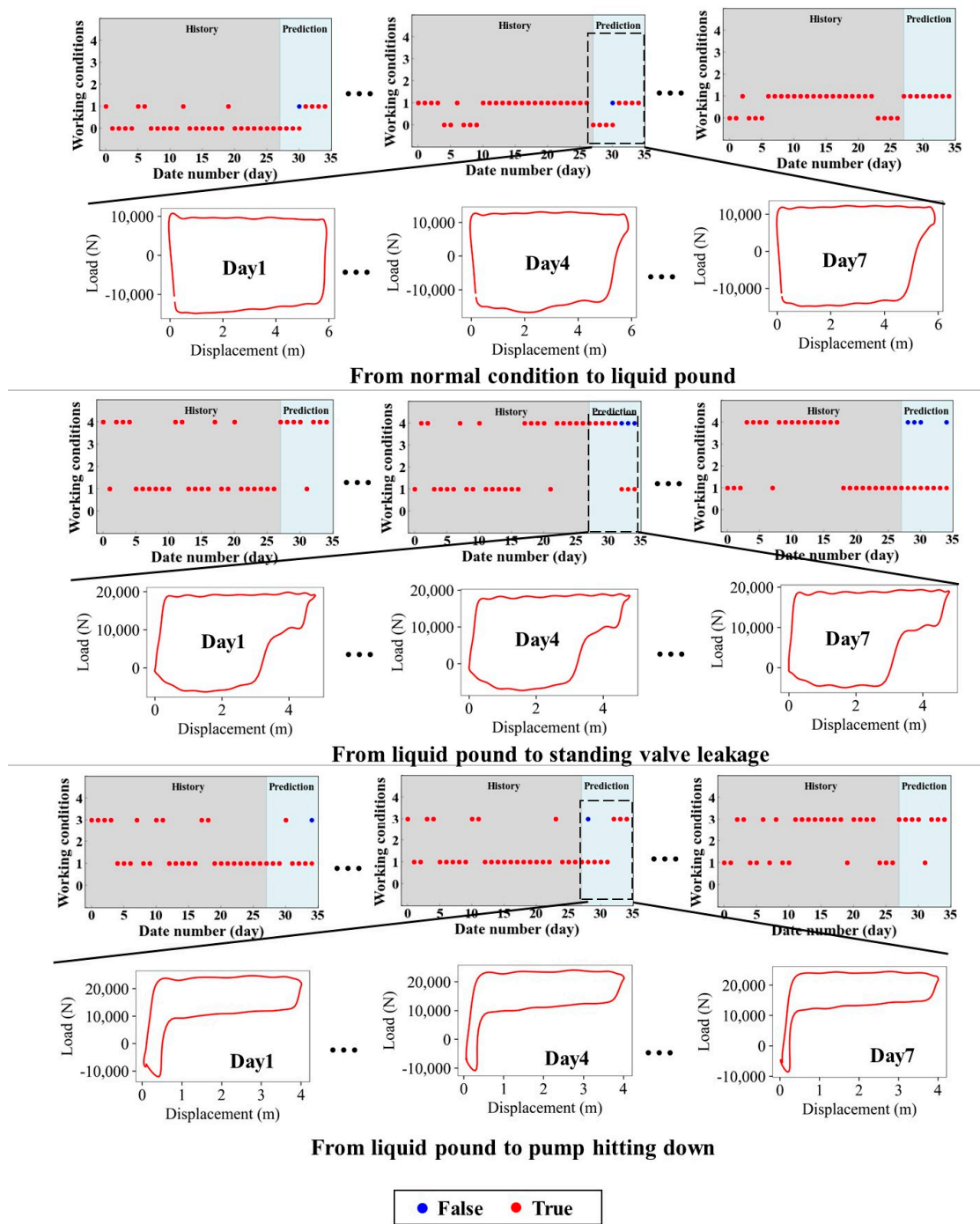


Figure 18. The process of fault types changing over time.

It can be seen that the unknown working conditions of the future 7 days were predicted by the samples of the past 28 days, which restored the warning process in the oilfield site. In the working conditions warning interval (28–34), most samples were predicted correctly with the method proposed in this paper when the working conditions began to change. Because the predicted results were based on the information of the past time steps, there was also a time lag in the warning process. Nonetheless, the warning results can be corrected after several time steps (usually 1–3 steps).

5. Conclusions

In this paper, the characteristic parameters of down-hole dynamometer cards were extracted to describe the temporal change of different working conditions. Based on them and the classification model, the mapping relationship between the parameters and working conditions were obtained.

Due to the position shifts during parameters extraction, there was noise inside the extraction results, which had a negative influence on the prediction results. Therefore, we proposed the selective noise reduction method based on wavelet transformation and verified the rationality of the noise reduction results by the Restored rates. The results showed that noise reduction can remove the invalid high-frequency noise and reserve the effective high-frequency fault information at the same time.

In order to excavate the temporal relationships better, the Temporal Fusion Transformers (TFT) was used as the prediction model. It mainly consists of a multi-scale feature input structure, Encoder–Decoder structure and self-attention structure. The experimental results showed that before noise reduction, the prediction relative errors of some parameters were over 20%. After noise reduction, the errors of nearly all the parameters were less than 20%.

By inputting the prediction results into the classification model, warnings can be flagged for future working conditions. The results showed that Recall values of most working conditions within the next 7 days were high, which indicates there were few underreporting cases for future faults. However, due to prediction bias and similarity between different working conditions, Precision values of the standing valve leakage were lower, which can be improved in future research.

In addition, only five common working conditions were used for warning research in this paper. Therefore, the research in this paper is only applicable to these common working conditions. In addition, there is still room for improvement in the warning accuracy of time-sequence conversion between different working conditions. In the future, more working conditions and their conversion relationship should be considered to extend the existing research content.

Author Contributions: Conceptualization, K.Z. and C.Y.; methodology, C.Y. and C.C.; software, C.Y. and C.C.; validation, C.Y., W.Y., G.F., C.L. and C.C.; formal analysis, K.Z., C.Y. and L.Z.; investigation, K.Z., W.Y., G.F. and L.Z.; resources, W.Y., G.F. and C.L.; data curation, K.Z. and W.Y.; writing—original draft preparation, C.Y. and C.C.; writing—review and editing, C.Y.; visualization, W.Y., G.F., C.L. and C.C. funding acquisition, K.Z., W.Y., G.F., C.L. and L.Z. project administration, K.Z. and L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Natural Science Foundation of China under Grant 52325402, 52274057, 52074340 and 51874335, the National Key R&D Program of China under Grant 2023YFB4104200, the Major Scientific and Technological Projects of CNOOC under Grant CCL2022RCPS0397RSN, 111 Project under Grant B08028.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to data secrecy requirements of the oilfield which is the data source of this paper.

Conflicts of Interest: Authors Weiyang Yao and Gaocheng Feng was employed by the company CNOOC EnerTech-Drilling & Production Co. Author Chen Liu was employed by the company CNOOC Research Institute Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Zhang, R.; Yin, Y.; Xiao, L.; Chen, D. A real-time diagnosis method of reservoir-wellbore-surface conditions in sucker-rod pump wells based on multidata combination analysis. *J. Pet. Sci. Eng.* **2021**, *198*, 108254. [[CrossRef](#)]
2. Zhang, R.; Yin, Y.; Xiao, L.; Chen, D. Calculation Method for Inflow Performance Relationship in Sucker Rod Pump Wells Based on Real-Time Monitoring Dynamometer Card. *Geofluids* **2020**, *2020*, 8884988. [[CrossRef](#)]

3. Schirmer, P.; Gay, J.C.; Toutain, P. Use of advanced pattern-recognition and knowledge-based system in analyzing dynamometer cards. *SPE Comp. App.* **1991**, *3*, 21–24. [[CrossRef](#)]
4. Foley, W.L.; Svinos, J.G. Expert Adviser Program for Rod Pumping (includes associated paper 19367). *J. Pet. Technol.* **1989**, *41*, 394–400. [[CrossRef](#)]
5. Wang, K. Fault diagnosis of rod-pumping unit based on production rules system. *Pet. Explor. Dev.* **2010**, *37*, 116–120.
6. Dickinson, R.R.; Jennings, J.W. Use of Pattern-Recognition Techniques in Analyzing Downhole Dynamometer Cards. *SPE Prod. Eng.* **1990**, *5*, 187–192. [[CrossRef](#)]
7. Wang, D.; Liu, H.; Ren, H. Fault diagnosis of sucker rod pumping system based on matching method of shape context. *J. China Univ. Pet.* **2019**, *43*, 159–165.
8. Nascimento, J.; Maitelli, A.; Maitelli, C.; Cavalcanti, A. Diagnostic of Operation Conditions and Sensor Faults Using Machine Learning in Sucker-Rod Pumping Wells. *Sensors* **2021**, *21*, 4546. [[CrossRef](#)]
9. Liu, J.; Feng, J.; Liu, S. Fault Diagnosis of Rod Pump Oil Well Based on Support Vector Machine Using Preprocessed Dynamometer card. In Proceedings of the 2021 IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS), Suzhou, China, 25 June 2021. [[CrossRef](#)]
10. Lv, X.; Wang, H.; Zhang, X. An evolutionary SVM method based on incremental algorithm and simulated dynamometer cards for fault diagnosis in sucker rod pumping systems. *J. Pet. Sci. Eng.* **2021**, *203*, 108806. [[CrossRef](#)]
11. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016. [[CrossRef](#)]
12. Chen, L.; Gao, X.; Li, X. Using the motor power and XGBoost to diagnose working conditions of a sucker rod pump. *J. Pet. Sci. Eng.* **2021**, *199*, 108329. [[CrossRef](#)]
13. Ashenayi, K.; Nazi, G.A.; Lea, J.F.; Kemp, F. Application of Artificial Neural Network to Pump Card Diagnosis. *SPE Comp. App.* **1994**, *6*, 9–14. [[CrossRef](#)]
14. Hou, Y.; Chen, B.; Gao, X. Fault Diagnosis of Sucker Rod Pumping Wells Based on GM-ELM. *J. Northeast. Univ.* **2019**, *40*, 1673–1678.
15. Xu, T. *Research on Oilfield Intelligent Fault Diagnosis System Based on Optimized RBF Neural Network*; Xi'an Shiyou University: Xi'an, China, 2019.
16. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*. [[CrossRef](#)]
17. Tian, H.; Deng, S.; Wang, C.; Ni, X.; Wang, H.; Liu, Y.; Ma, M.; Wei, Y.; Li, X. A novel method for prediction of paraffin deposit in sucker rod pumping system based on CNN dynamometer card feature deep learning. *J. Pet. Sci. Eng.* **2021**, *206*, 108986. [[CrossRef](#)]
18. Zhang, K.; Wang, Q.; Wang, L. Fault diagnosis method for sucker rod well with few shots based on meta-transfer learning. *J. Pet. Sci. Eng.* **2022**, *212*, 110295. [[CrossRef](#)]
19. Wang, X.; He, Y.; Li, F.; Dou, X.; Wang, Z.; Xu, H.; Fu, L. A Working Condition Diagnosis Model of sucker rod pumping wells based on deep learning. *SPE Prod. Oper.* **2021**, *36*, 317–326. [[CrossRef](#)]
20. Yin, C.; Zhang, K.; Zhang, L.; Wang, Z.; Liu, P.; Zhang, H.; Yang, Y.; Yao, J. Imbalanced Working States Recognition of Sucker Rod Well Dynamometer Cards Based on Data Generation and Diversity Augmentation. *SPE J.* **2023**, *28*, 1925–1944. [[CrossRef](#)]
21. Bing, S. An Early Warning Method Based on Artificial Intelligence for Wax Deposition in Rod Pumping Wells. *Pet. Drill. Tech.* **2019**, *47*, 97–103.
22. Lin, X.; Wu, B.; Mi, L.; Liu, L. Trend prediction of pumping well conditions using temporal dynamometer cards. In Proceedings of the 2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), Taiyuan, China, 23–25 October 2020.
23. Tan, C.; Chen, P.; Yang, Y.; Yu, Y.; Song, J.; Feng, G.; Sun, X. Prediction of paraffin deposition and evaluation of paraffin removal effect for pumping wells driven by timing indicator diagram. *Oil Drill. Prod. Technol.* **2022**, *44*, 123–130.
24. Skydt, M.R.; Bang, M.; Shaker, H.R. A probabilistic sequence classification approach for early fault prediction in distribution grids using long short-term memory neural networks. *Measurement* **2021**, *170*, 108691. [[CrossRef](#)]
25. Yang, Y.; Zhang, S.; Su, K.; Fang, R. Early warning of stator winding overheating fault of water-cooled turbogenerator based on SAE-LSTM and sliding window method. *Energy Rep.* **2023**, *9*, 199–207. [[CrossRef](#)]
26. Huang, C.; Qu, S.; Ke, Z.; Zheng, W. Dual fault warning method for coal mill based on Autoformer WaveBound. *Reliab. Eng. Syst. Saf.* **2024**, *245*, 110030. [[CrossRef](#)]
27. Lin, K.C.; Hsu, J.Y.; Wang, H.W.; Chen, M.Y. Early fault prediction for wind turbines based on deep learning. *Sustain. Energy Technol. Assess.* **2024**, *64*, 103684. [[CrossRef](#)]
28. Chen, Y.; Lin, C.; Zhang, Y.; Liu, J.; Yu, D. Proactive failure warning for wind power forecast models based on volatility indicators analysis. *Energy* **2024**, *305*, 132310. [[CrossRef](#)]
29. Zhao, H.; Chen, Z.; Shu, X.; Shen, J.; Liu, Y.; Zhang, Y. Multi-step ahead voltage prediction and voltage fault diagnosis based on gated recurrent unit neural network and incremental training. *Energy* **2023**, *266*, 126496. [[CrossRef](#)]
30. Qu, N.; Li, Z.; Zuo, J.; Chen, J. Fault Detection on Insulated Overhead Conductors Based on DWT-LSTM and Partial Discharge. *IEEE Access.* **2020**, *8*, 87060–87070. [[CrossRef](#)]

31. Ling, J.; Liu, G.; Li, J.; Shen, X.; You, D. Fault prediction method for nuclear power machinery based on Bayesian PPCA recurrent neural network model. *Nucl. Sci. Tech.* **2020**, *31*, 75. [[CrossRef](#)]
32. Gibbs, S.G. Predicting the behavior of sucker-rod pumping systems. *J. Pet. Technol.* **1963**, *15*, 769–778. [[CrossRef](#)]
33. Zhang, Q.; Wu, X. Pumping well diagnostic technique and its application. *J. East China Pet. Inst.* **1984**, *2*, 145–159.
34. Guo, T.; Zhang, T.; Lim, E.; Lopez-Benitez, M.; Ma, F.; Yu, L. A Review of Wavelet Analysis and Its Applications: Challenges and Opportunities. *IEEE Access* **2022**, *10*, 58869–58903. [[CrossRef](#)]
35. Wodarski, P.; Jurkojć, J.; Gzik, M. Wavelet Decomposition in Analysis of Impact of Virtual Reality Head Mounted Display Systems on Postural Stability. *Sensors* **2020**, *20*, 7138. [[CrossRef](#)] [[PubMed](#)]
36. Lim, B.; Arik, S.Ö.; Loeff, N.; Pfister, T. Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.