Journal of Petroleum Exploration and Production Technology (2025) 15:46
https://doi.org/10.1007/s13202-025-01931-x

ORIGINAL PAPER - PRODUCTION ENGINEERING

Enhancing predictive maintenance strategies for oil and gas equipment through ensemble learning modeling

Mei Wang[1] · Xuesong Su[1,2] · Huifang Song[1,2] · Yifei Wang[1] · Xinxin Yang[1]

Received: 9 April 2024 / Accepted: 6 January 2025 / Published online: 18 February 2025

Abstract

In the field of oil and gas equipment management, frequent maintenance is conducted, resulting in unnecessary costs. Relying solely on a single artificial intelligence model often leads to low predictive accuracy and inadequate robustness because of the poor data quality. Therefore, a method based on ensemble learning modeling is proposed to accurately assess the health status of industrial equipment and predict its remaining useful life. By conducting big data analysis on the operational history data of all oil and gas equipment, various fault instances are extracted and grouped accordingly. The integration of meta-learning convolutional shrinkage neural networks (ML-CSNN), domain expert rules, and support vector machine (SVM) models forms a hybrid model aimed at constructing a robust classification model. The effectiveness of the proposed method is validated using operational data from shengli oilfield production wells in China. The proposed method achieves 0.98 in accuracy, 0.93 in precision, 0.94 in recall, and 0.93 in F1 score, which is an improvement of 9–25% compared to the popular integrated learning models, such as GBDT and XGBoost. By designing the ablation study, results demonstrate the method's ability to accurately predict potential faults of oil and gas equipment, thereby facilitating the enhancement of predictive maintenance strategies.

Keywords Decision support · Equipment predictive maintenance · Data mining · Machine learning · Industrial health assessment

✉ Xuesong Su
t-suxuesong.slyt@sinopec.com

Mei Wang
wangmei667.slyt@sinopec.com

Huifang Song
huifang.song@outlook.com

Yifei Wang
wangyifei524.slyt@sinopec.com

Xinxin Yang
xinxinyang.slyt@sinopec.com

1 Technology Inspection Center of Shengli Oilfield, SINOPEC, Dongying 257000, Shandong Province, China

2 Postdoctoral Scientific Research Workstation of Shengli Oilfield Company, Dongying 257000, Shandong Province, China

List of symbols

Dimensionless

d^i The vector of condition sample data
d^j The norms of vector
θ^* Optimal parameter
ω The parameter of the meta-model
ω^* Common parameter

Pair

D^i Support set
D^{tr} Training set
D^{trs} Support set in the training set
D^{trq} Query set
D^{ts} Testing set
T_i Possible tasks

Abbreviations

4D-TF Time-domain and frequency domain feature extraction techniques
AMD Advanced Micro Devices

CNN	Convolutional Neural Networks
FN	False Negative
FP	False Positive
GBDT	Gradient Boosting Decision Tree
GPU	Graphics processing unit
ML	Meta-learning methods
ML-CSNN	Meta-Learning Convolutional Shrinkage Neural Networks
RAM	Random Access Memory
RF	Random Forest
SNN	Shrinkage Neural Networks
SVM	Support Vector Machine
SP-SVM	Support Vector Machine model based on sensitive parameters
TN	True Negative
TP	True Positive

Introduction

In recent years, the global petroleum industry has maintained a stable development trend. There has been an increasingly stringent demand for the safety performance of oil and gas equipment in petroleum production, transportation, and management. The oil and gas equipment includes pumps and sensors such as temperature, pressure, and flow (Iraji et al. 2023b). The pumps often apply in drilling fluid circulation, crude oil extraction to transport processing, as well as water injection and environmental management. The sensors monitor and control temperature, pressure, liquid level, flow and other key parameters from drilling to production, ensuring safe and efficient oil exploration, production and environmental monitoring (Iraji et al. 2024). Simultaneously, a series of automation and intelligent technologies have been introduced, continuously improving the production scheduling and process control levels of oil and gas equipment (Jacobs 2018; Wang et al. 2012). However, it is worth noting that due to the large scale and complexity of oil and gas equipment, various safety hazards exist at different stages of production practice (Chen et al. 2022). Equipment failures and shutdowns can lead to production interruptions and losses, and in severe cases, may result in safety accidents and environmental pollution, thereby causing significant economic losses to enterprises and society (Bukhtoyarov et al. 2019; Woldesellasse and Tesfamariam 2023). Additionally, equipment failures and damage can affect the service life of equipment, increase maintenance costs, and typically require a large amount of manpower, material, and financial resources for maintenance and management (Pivovarov and Duvakina 2020).

To effectively improve the operational status of oil and gas equipment, it is essential to focus on equipment maintenance and management, aiming to minimize the failure rate of equipment and ensure its operational status tends towards stability and reliability (Johnstone and Curfew 2012). Currently, major oil and gas companies place significant emphasis on the management and maintenance of oil and gas equipment in their operational practices. However, influenced by traditional concepts, some companies still rely on manual real-time monitoring of equipment operating parameters, including temperature, pressure, vibration, flow rate, and others (Aranha et al. 2024). This approach suffers from inefficiency, subjectivity, uncertainty, and high labor costs. It often involves post-processing and emergency maintenance, making it difficult to identify potential problems and potential failures of equipment in advance, and lacking preventative maintenance measures (Nithin et al. 2021).

With the development of information technology, the field of crude oil transportation in oilfields is gradually moving towards digitization and intelligence (Choi et al. 2023). Predictive maintenance of oil and gas equipment based on machine learning can promptly identify problems existing in the operation of oil and gas equipment (Lu et al. 2020). By collecting and analyzing equipment status information, supported by data mining techniques, potential hazards of equipment can be analyzed and diagnosed, thereby achieving predictive maintenance of equipment (Xia et al. 2021).

Predictive maintenance of oil and gas equipment is regarded as a multi-class classification problem of equipment failures, with various types of equipment failures considered as distinct categories (Khalid et al. 2021). By collecting large amounts of equipment operational data and employing machine learning techniques such as support vector machines (Mohammed 2023), random forests (Xie et al. 2019), neural networks (Al-Sabaeei et al. 2023), classification models are trained to predict potential failure types based on the current equipment status (Park et al. 2021). Through this approach, appropriate maintenance measures can be taken before equipment failure occurs (Zhang and Li 2019), thereby minimizing downtime and repair costs to the greatest extent possible.

The use of machine learning algorithms has been enormously successful as a way to address the challenge of equipment predictive maintenance in various fields (Li et al. 2018). Nguyen et al. (2022) proposed a digital twin approach to system-level detection and diagnosis in thermal-hydraulic systems in nuclear power plants, but the modeling is difficult and requires the support of a large amount of physical information, which is very difficult to collect. Thus, data-driven models are considered instead of a digital twin model. Fan et al. (2020) used K-nearest neighbors and naive Bayes classifies to classify wafers as normal or abnormal. Hsu and Liu (2021) proposed a multiple time-series convolutional neural network model for fault detection in semiconductor manufacturing and the core is that key features of device sensors can be automatically learned by stacking convolutional pooling layers. Wu et al. (2023) presented a Transformer-based classifier that can effectively identify different known types and severity of fault conditions, as well as novel fault detections. But these method does not combine expert experience and actual parameters. Therefore, expert experience and oil production engineering sensitive parameters are integrated into the machine learning model to enhance its interpretability and generalization.

Moreover, machine learning methods have been widely used in the oil and gas industry, especially in the analysis of well logs, lithofacies and depositonal environments, and seismic data inversion. The K-means unsupervised classification algorithm can characterize the reservoir sequences and precisely define the reservoir rock types. Two deep learning models, ResNet and 1D CNN, were trained and evaluated for clogging porosity prediction, in which the 1D CNN model performed well. In addition, the XGBoost algorithm was used to predict the rock types with high accuracy (Iraji et al. 2023a) However, the above models do not consider the time series dynamic information and signal structure information in the data, and have poor generalization. Therefore, meta-learning techniques are introduced to enable the model to more rapidly adjust its parameters in response to the data characteristics of different wells and varying operating conditions. Then the time domain and frequency domain feature extraction technology is used to comprehensively capture the time domain characteristics and frequency domain characteristics of the oil well.

Despite the potential of machine learning methods in predictive maintenance of oil and gas equipment, they still face numerous challenges and shortcomings (Niyonambaza et al. 2020). In the oil and gas industry, due to the diversity of data sources and inconsistency in data quality, issues such as missing data, noise, and outliers may exist, which can affect the accuracy and robustness of machine learning models (Joseph et al. 2022). Moreover, fault samples of oil and gas equipment are often much fewer than normal samples, resulting in a severe imbalance in the training dataset. This imbalance can cause the model to tend towards predicting the majority class while neglecting the minority class, thus affecting the model's performance. Ensemble learning technology is considered to address this problem. Ensemble learning can address the shortcomings of individual models by combining the predictions of multiple models to enhance overall generalization ability. Additionally, ensemble learning can increase robustness, reduce noise interference, and improve stability (Fomin et al. 2018).

In this paper, the contributions are the following: (i) By utilizing an ensemble learning framework, a new multi-class classification model named DTSKL-Boost is proposed. The motivation for using AdaBoost is that AdaBoost demonstrates strong versatility, making it applicable to various types of classification tasks, as well as both discrete and continuous feature data. (ii) The workflow of the proposed method is to input the actual production data, and then classify the working conditions using ML-CSNN, a hybrid model based on convolutional neural network and expert rules, and SVM model based on sensitive parameters of oil production engineering. Finally, the classification results of the three methods are passed through the voting method to obtain the working condition prediction results. (iii) Expert rules and sensitive parameters of oil production engineering are incorporated into the operating condition recognition model to enhance the interpretability of the identification results. (iv) Data mining is conducted on the historical records of equipment failures to extract unique training features for different types of failures, thereby constructing a training sample library. (v) In order to evaluate the performance of DTSKL-Boost model, eight models were tested on the set of oil field production dataset. (vi) According to the results of the ablation tests, it was found that the ensemble of three methods performs better than any single method or any combination of two methods.

The paper is organized as follows: a introduction of key technologies in the field of predictive maintenance is provided in Section "The key technologies of predictive maintenance". The details of the DTSKL-Boost model are outlined in Section "Framework for the equipment predictive maintenance". The tests on the datasets among the proposed method and other techniques are presented and discussed in Section "Test results and analysis". Further discussion is in Section "Discussion". Conclusions are in Section "Conclusions".

The key technologies of predictive maintenance

Big data analytics warehouse

According to the research results of relevant scholars and experts, the so-called data warehouse refers to a collection of data, which is relatively stable and clear in subject matter, and can provide data support in management decision-making, presenting the development of data laws. The data in the data warehouse originates from the collection and integration of information sources and is stored in a unified manner in a specific environment, so that enterprises can quickly and easily query and utilize this information to support the decisions they make (Tariq et al. 2021). At present, many oil and gas companies have introduced information management technologies and methods in oilfield construction, and collected and organized data related to oil and gas production and management, and used these data to analyze the operation status of oil and gas equipment in different production processes, so as to achieve diagnosis and maintenance of equipment (Mohammadpoor and Torabi 2020; Ran et al. 2021; Abhulimen et al. 2018; Zhifeng 2019).

Data mining technology

The most representative definition of data mining is "the non-trivial process of obtaining correct, novel, potentially valuable, and ultimately understandable patterns from large amounts of data" (Fayyad et al. 1996). Data mining is divided into two main categories: descriptive and predictive. The former is to derive the inductive model of the implicit association in the data, while the predictive data mining technology is based on the existing data to analyze the current situation and then make judgments and decisions with the actual situation. The application mode of data mining technology is also divided into many kinds, such as association analysis, classification, clustering, outlier analysis, etc. (Elahifar and Hosseini 2024; Antonio et al. 2022) In the process of mode construction, it is required to collect and organize the historical data of the equipment so that it has certain reference value, however, training and learning around the data. The diagnosis knowledge base can be built based on the analysis results of equipment failure modes in the interpretation and evaluation, and the manual diagnosis and different types of data mining techniques can be combined together to diagnose the equipment failure information. The details are described as follows.

First, clustering. Diagnose and analyze the faults of the object equipment, extract its features, and classify them, and then discover the inherent common structure of the faulty equipment based on signal and system processing technology methods; second, association analysis. Through modeling to describe the state of the equipment components, is the premise of parallel use of equipment to start the comparison, according to the difference between the input and output of the equipment and the model to identify the problems that arise in the operation of the equipment, based on this to explore the correlation within the equipment module; third, classification. The equipment status is monitored and formed into data, then compared with the fault knowledge base, and the fault is diagnosed according to the monitoring value and fault type, and finally combined with fault handling methods to solve the problem; fourth, outlier analysis. Some of the data obtained by monitoring the equipment may have certain differences with its normal model parameters, at this time, it can be determined that there are problems in some aspects of the equipment, in other words, the equipment monitoring value data distribution and probability model, you can use statistical and distance metric test to detect outliers, so as to determine the normal operation status of the equipment or not (Fernandes et al. 2022).

However, in the field of predictive maintenance of oil and gas equipment, the application depth of data mining technology still has some room for improvement, and the problems exposed in the actual application are relatively more. For example, the challenges of multi-dimensional and large amount of data sets, the problem of handling redundant information and noisy data, and the invalidity of mining results (Fan et al. 2014; Mahdavi Kalatehno and Khamehchi 2024). In view of the aforementioned issues, it is necessary to adjust and optimize the application of data mining technology in the process of equipment maintenance and repair based on the findings of this research. At the same time, with the increase of automation and intelligence of the equipment itself, the demand for predictive maintenance of the equipment will become bigger and bigger (Quinlan 1986).

The application of predictive maintenance technology focuses on how to build a model, and to ensure that the monitoring of equipment operating conditions and fault diagnosis tend to be effective, it is necessary to follow the principle of applicability to build the equipment maintenance model. According to the corresponding mode to play the corresponding function, of course, considering the potential value of data mining technology is high, so in practice can be used in combination with different modes, so as to achieve predictive accuracy (Steinwart and Christmann 2008).

Framework for the equipment predictive maintenance

Oil and gas equipment has a large scale and relatively complicated types, in the production practice of each link there are different degrees of safety hazards, if equipment failure will easily lead to safety accidents, thus causing serious economic losses to enterprises and society. In view of this, in order to effectively improve the operation of oil and gas equipment, it is necessary to do a good job of maintenance and management of equipment, to maximize the control of equipment failure rate, to ensure that the operation of equipment tends to be stable and reliable. Based on data mining technology to implement predictive maintenance of oil and gas equipment, it is necessary to first diagnose equipment faults and understand the types and specific characteristics of equipment faults. In the diagnosis process, the main point of data mining technology is to collect and organize the operating parameters of the equipment unit, such as the data

Table 1 Classification of oil well operating conditions

Category	Item
Healthy production	Normal
Formation	Insufficient fluid supply, liquid percussion, gas impact, air lock, suction and spray, take time out, sand production, thick oil, waxing
Pipe	Leakage of pipe, pipe removal
Rod	Break of rod, rod clamp, oil thickens and slackens, wax slowing
Pump	Swim Vail miss, fixed vail leakage, double vane miss, piston leakage, pump body leakage, pump stuck, suction inlet block, vail hood broken
Ground	Hang on, touch down, remove the cylinder, oscillation, oil beam stuck
Data	Dead point ahead, dead end lag, period extension, cycle shortening, mutation of load, drift of load, load serration, displacement loss, disorder of displacement

of the equipment under normal operating conditions and the data of the equipment in the fault state, and at the same time, to know the fault category. In this way, based on the fault category, fault parameters and historical data, a data mining-based training sample library can be constructed. Through the application of data mining technology, the laws and features can be extracted from the huge data samples, so that the characteristics of various faults can be mined(Guo et al. 2003; Agwu et al. 2020).

Sample set construction

There are many types of oilfield equipment failures, and the troubleshooting is not targeted, mostly dealt with after the occurrence of accidents, without comprehensive equipment monitoring means for effective auxiliary management, the fundamental reason being that the weights of factors affecting equipment failures cannot be determined through manual experience. According to the common faults of oilfield equipment, the types of faults and feature sets are clarified, and a sample library is established. According to the relevant provisions of the oil and gas equipment failure standards, the sample pool is divided into seven major categories, totaling 38 items as shown in Table 1.

(1) Waxing

Considering the above screening results, the electrical parameters such as power, electrical energy value and current are associated with maximum load and minimum load, so finally Maximum load, Minimum load, Load difference, A current are selected as the training features for the self-learning of oil well waxing rules based on the classification method. When wax formation occurs in a pumping well, the resistance of pumping rod movement due to wax formation increases, which affects the load of the pumping well. A current, B current, and C current refer to the currents in the three different phases of the three-phase AC supply system, respectively. It is characterized by the fact that the phases of the three-phase voltage and current differ 120 degrees from each other. This means that when A current is at its peak, B current is 120 degrees behind A current and C current is 240 degrees behind A current (Table 2).

Table 2 Selection of characteristics for waxing

F-test	Correlation of discrete targets	Pearson correlation coefficient	Spearman correlation coefficient
Maximum load	Positive total active energy value	Maximum load	Maximum load/load difference
Minimum load	Load difference	Minimum load	Maximum load/minimum load
C current	Total power average over the cycle	A current/C current	Load difference
A current	Minimum load	Positive total active energy value	Cumulative power consumption of inverter
B current	Maximum load	Maximum load/load difference	B current/A current

(2) Insufficient fluid supply

Integrating the above screening results, considering those electrical parameters such as power, power consumption and current are associated with maximum load and minimum load, Sunkenness, Wellhead temperature, and daily power consumption are finally selected as the training features for the self-learning of oil well undersupply rule based on classification method. When a pumping well is undersupplied with fluid, the discharge capacity of the pump is greater than the fluid supply capacity of the formation, which leads to a low dynamic fluid level in the pumping well due to the undersupply, thus the Sunkenness of the well is low and the fluid cannot fill the pump barrel; this then leads to a decrease in the efficiency of the pumping well and an increase in daily power consumption; it also leads to a change in the flow rate of the fluid at the wellhead, which causes the wellhead temperature to change as well (Table 3).

(3) Gas impact

Combining the above screening results, Daily power consumption, Dynamic fluid level, Unloading slope, Maximum load, Minimum load, Load difference are finally selected as the training features for the self-learning of gas impact rule based on classification method.

Table 3 Selection of characteristics for insufficient fluid supply

F-test	Correlation of discrete targets	Pearson correlation coefficient	Spearman correlation coefficient
Sunkenness	Positive total active energy value	Daily power consumption	Sunkenness/daily power consumption
Wellhead temperature	Sunkenness	Minimum load	Wellhead temperature/wellhead oil pressure
Maximum of total power over the cycle	Maximum load	Wellhead oil pressure/wellhead temperature	Daily power consumption
Daily power consumption	Load difference	Positive total active energy value	Cumulative power consumption of inverter
Total power average over the cycle	Cumulative power consumption of inverter	Maximum load	Positive total active energy value

Table 4 Selection of characteristics for gas impact

F-test	Correlation of discrete targets	Pearson correlation coefficient	Spearman correlation coefficient
Inverter output voltage	Positive total active energy value	Daily power consumption	Maximum load/load difference
Daily power consumption	Load difference	Minimum load	Maximum load/minimum load
Unloading slope	Maximum load	Maximum load/load difference	Load difference
Total power average over the cycle	Total power average over the cycle	Dynamic fluid level/load difference	Cumulative power consumption of inverter
Dynamic fluid level	Minimum load	Maximum load	Positive total active energy value

When gas impact occurs in a pumping well, the maximum load, minimum load, load difference and daily power consumption of the pumping well are affected because the gas enters the pump barrel and causes the liquid column load of the pumping well to decrease; the pumping pump valve ball in gas-bearing oil wells generally opens with a lag, and only when the plunger meets the liquid level can it unload quickly, so the unloading slope can also be used as a basis for judgment; at the same time, when the dynamic well produces abnormal gas impact conditions, so the fluid level of the oil well decreases, the occurrence of at the same time, when the dynamic fluid level of the well decreases, the phenomenon of degassing at the bottom of the well produces abnormal gas impact conditions, so the dynamic fluid level can also be used to diagnose the gas impact of oil wells (Table 4).

The ensemble learning model DTSKL-Boost

The DTSKL-Boost Model is a meta-learning ensemble algorithm that integrates three sub-models: meta-learning convolutional shrinkage neural network (ML-CSNN), a fusion model based on convolutional neural network and expert rules, and a support vector machine model based on sensitive parameters of oil production engineering. Targeting different characteristic parameters of oil and gas equipment, various classification models are employed from different perspectives to predict equipment failures in advance, enhancing the reliability of the prediction results, as shown in Fig. 1. By introducing meta-learning, expert rules, and sensitive parameters of oil production engineering, the

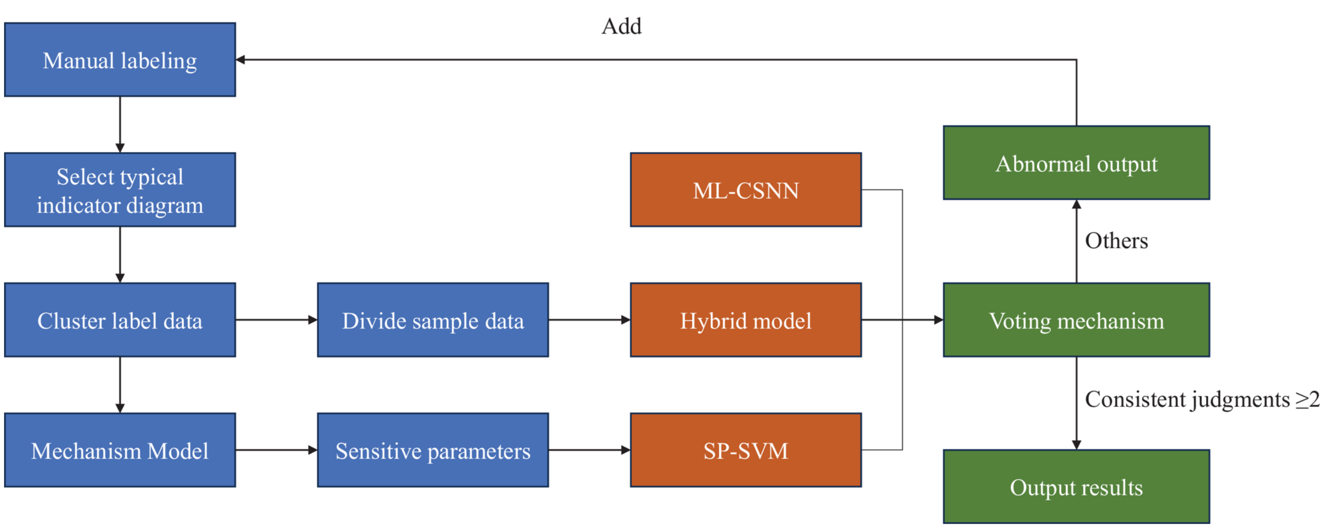

Fig. 1 The structure of DTSKL-Boost method

model exhibits stronger learning capabilities for equipment failure types with limited labeled data, and its prediction results are more interpretable. To ensure that the model can effectively adapt to different fault prediction tasks in practical applications, a voting mechanism is adopted to comprehensively evaluate the prediction results of the three sub-models, constructing an ensemble learning framework. This enables the model to achieve high prediction accuracy and good robustness.

(1) Meta-learning convolutional shrinkage neural network

To address the challenge of limited labeled data for equipment failure and the complexity and variability of operating conditions, a meta-learning convolutional shrinkage neural network model is constructed. Targeting equipment operating parameters, the neural network framework combines meta-learning modules, time-domain and frequency-domain feature extraction techniques, and embedded shrinkage neural networks, as illustrated in Fig. 2. The model consists of 1 input layer, 4 convolutional layers, 4 SNN layers, and 1 fully connected layer.

The condition sample data is divided into training set D^{tr} and testing set D^{ts}. On both the training and testing sets, a similar matching method is employed to separately partition them into support sets and query sets. The similarity calculation formula is represented as Eq. (1).

$$\text{similarity}(d_i, d_j) = \frac{d_i \cdot d_j}{\|d_i\| \cdot \|d_j\|} \tag{1}$$

where d_i represents the vectors of condition sample data and d_j represents the norms of vectors. Samples for which the similarity calculation result exceeds a certain threshold are considered similar samples. These similar samples are then designated as the support set, while the remaining samples are designated as the query set. The support set is utilized for inner-loop training of the model, facilitating the updating of model parameters and adaptation to new tasks. Consequently, the model can perform accurate predictions on the query set. The query set is employed to evaluate the model's generalization ability after adapting to new tasks, thereby enhancing the model's ability for few-shot learning. For the support set D^{trs} in the training set, the initial parameters θ of the ML-CSNN model are trained. This is achieved by summing up all losses in the query set D^{trq} and updating the optimal parameters θ^*. In the testing set, the optimal parameters θ^* are fine-tuned through the support set, and the ML-CSNN model is utilized to test the parameters in the query set, forming the process of "model training-model updating-model testing". Through meta-learning, the model can find common parameters ω^* suitable for different oil wells and different dataset tasks. The formula for calculating the common parameters is shown in Eq. (2).

$$\omega^* = \arg\min_{\omega} \sum_{T_i \sim p(T), D_i \sim T_i} L_i(D_i, \omega) \tag{2}$$

where ω^* represents the optimal parameter values obtained through meta-learning, T_i denotes all possible tasks, D_i is the support set sampled from task T_i, $p(T)$ denotes the task distribution, and $L_i(D_i, \omega)$ represents the loss incurred by training on support set D_i for task T_i. ω is the parameter of the meta-model. Equation (2) aims to find the parameter ω^* that minimizes the sum of losses incurred from training on the support sets of all tasks.

(2) Hybrid model based on convolutional neural networks and expert rules

By utilizing convolutional neural networks, the model can efficiently and accurately process massive and diverse dynamometer card data, thereby ensuring good real-time performance. Moreover, to address the issue of low accuracy in recognizing dynamometer cards with similar but distinct patterns, this paper introduces expert

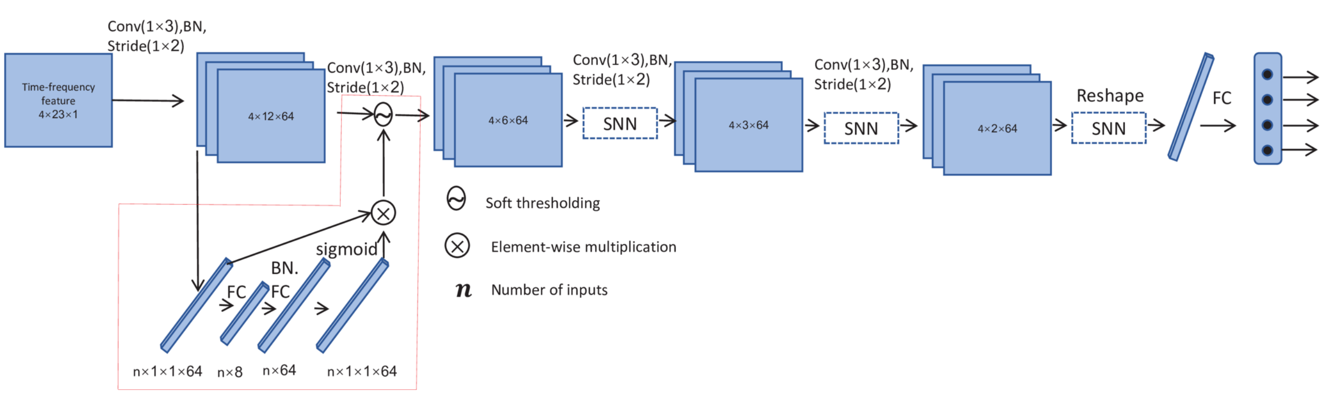

Fig. 2 The structure of the ML-CSNN model

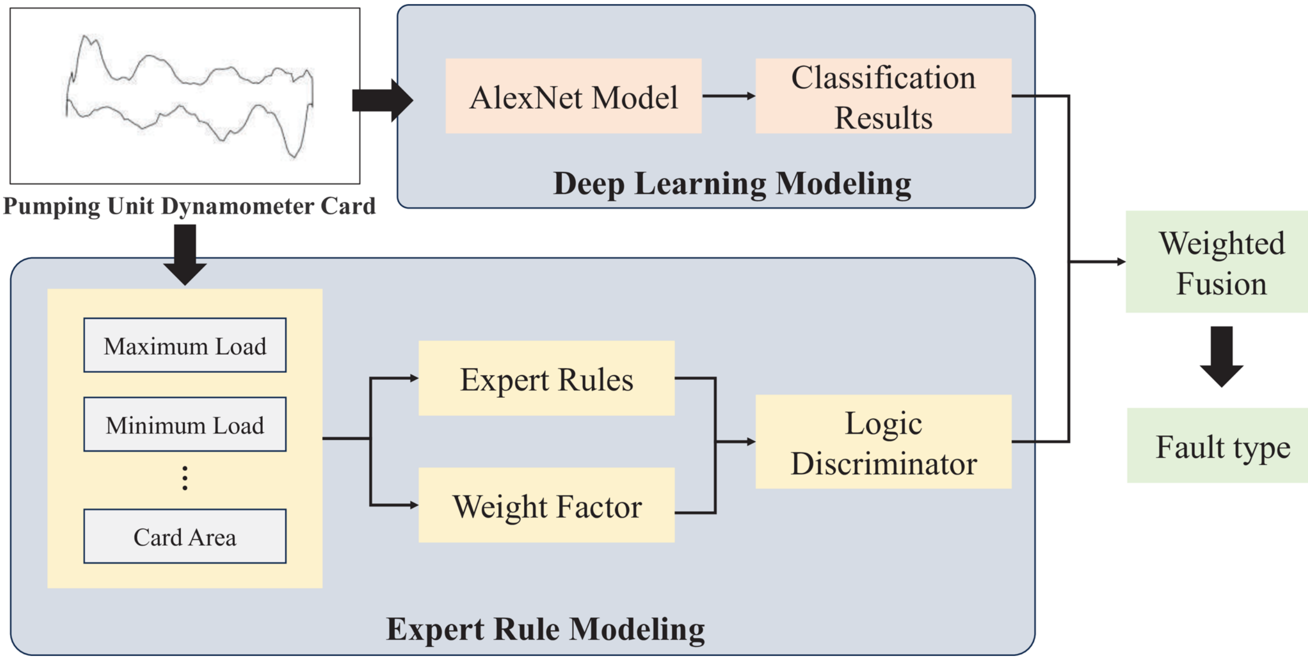

Fig. 3 The flowchart of hybrid model

Table 5 The structure of ALEXNET model

Layer type	Kernel size	Stride	Number	Weight
Input	–	–	–	–
Conv	11 × 11	4	32	11 × 11×1 × 32
Max pooling	3 × 3	2	32	–
Conv	5 × 5	2	64	5 × 5 × 32 × 64
Max pooling	3 × 3	2	64	–
Conv	3 × 3	1	128	3 × 3 × 64 × 128
Conv	3 × 3	1	128	3 × 3 × 64 × 128
Conv	3 × 3	1	128	3 × 3 × 64 × 128
Max pooling	3 × 3	2	128	–
Flatten	–	–	–	4096
Fully connected	–	–	–	1024
Dropout	–	–	–	1024
Fully connected	–	–	–	512
Dropout	–	–	–	512
Fully connected	–	–	–	128

Table 6 The structure of ALEXNET model

Fault type	Condition	Weight factor
Leaking valve flap	Maximum load reduction > 30%	0.6
	Minimum load increase > 2.5%	0.1
	Card area reduction > 35%	0.3
Fixed valve leakage	Maximum load reduction > 10%	0.1
	Minimum load increase > 30%	0.6
	Card area reduction > 35%	0.3
Pump rod breakage	Minimum load/Minimum load under normal operating conditions ≤ 95%	1.0
Continuous jet with pump	Maximum load − Minimum load/Maximum load ≤ 8%	1.0
Pumping rod slowing down	Minimum load = 0	0.7
	Card area reduction > 20%	0.3
Leakage of fixed valve and free valve	Maximum load reduction > 30%	0.3
	Minimum load increase > 20%	0.3
	Card area reduction > 40%	0.4

knowledge to construct a logic discriminator based on expert rules. Weight factors are incorporated to enhance the accuracy and reliability of condition identification (Fig. 3).

The convolutional neural network model is constructed based on the AlexNet model, which exhibits powerful capabilities in handling large-scale image data with multiple classification categories. Additionally, it includes mechanisms to prevent overfitting of data, such as dropout. The structure of the neural network model is detailed in Table 5.

The data of oil well indicator diagram is processed and analyzed in a period of 7 days. The expert rules and weight factors are formed by combining expert experience. Expert rules corresponding to oil well faults and their weight factors are shown in Table 6.

The comparison between the condition recognition results of the AlexNet model and the inference results obtained through expert rules is conducted as follows: if the results of both methods are consistent, the judgment result is directly outputted. However, if there is a discrepancy between the two results, the judgment result with higher confidence is outputted.

(3) Support vector machine model based on sensitive parameters of oil production engineering

Support Vector Machine model based on sensitive parameters (SP-SVM) combines with practical production experience in oil fields. It extracts key points, key segments, and areas of the dynamometer card through data mining methods which enriches the technical means of dynamometer card data mining and recognition. Thereby it can achieve the prediction of equipment. The model architecture is shown in Fig. 4.
For the input real-time production data, the data is normalized spatially, and the derivative of the normalized data is calculated. The points where the derivative is zero represent the points of maximum change, which are used to determine the inflection points of the dynamometer card. By calculating the positions and loads of points A, B, C, and D, as well as the average load of segment BC, the average load of segment DA, and the area of the dynamometer card, sensitive parameter data related to oil production engineering are extracted. These sensitive parameters are then fed into the support vector machine model for fault classification, resulting in the results of this method.

Test results and analysis

To validate the feasibility and effectiveness of the proposed model, tests were conducted using pump-jack wells in an oilfield located in eastern China. The model takes real-time production data from the wells, such as oil pressure, temperature, etc., as inputs, and outputs the types of faults. The following contents provides details for testing configurations and results.

Datasets

One hundred wells from the Shengli oil field in China were selected for theoretical validation, and historical operating parameters of each well were collected through sensors from 2019 to 2022. The parameters included wellhead oil pressure, wellhead temperature, daily power consumption, maximum load, minimum load, submergence, dynamic fluid level, and current, among others. The data collection interval is one record per minute so the dataset comprises approximately 200 million records. Table 7 shows only some of the sample data.

Model training

The model in this paper is obtained using python programming in an environment of AMD Ryzen 5 4600U with Radeon Graphics, 2.10 GHz, Tesla V100s GPU and 16 GB of RAM. There are many hyperparameters in the DTSKL-Boost model, and these parameters can affect the method classification effect to a certain extent, so the parameter setting is of great significance. Different from the cluster validity in the traditional K-means clustering which use within-cluster sum of squares or silhouette coefficient, a 10-fold cross-validation approach is used to evaluate model performance. Meanwhile, automated tuning techniques, such as grid search, are utilized to search for the optimal parameter combinations, thereby enhancing the efficiency

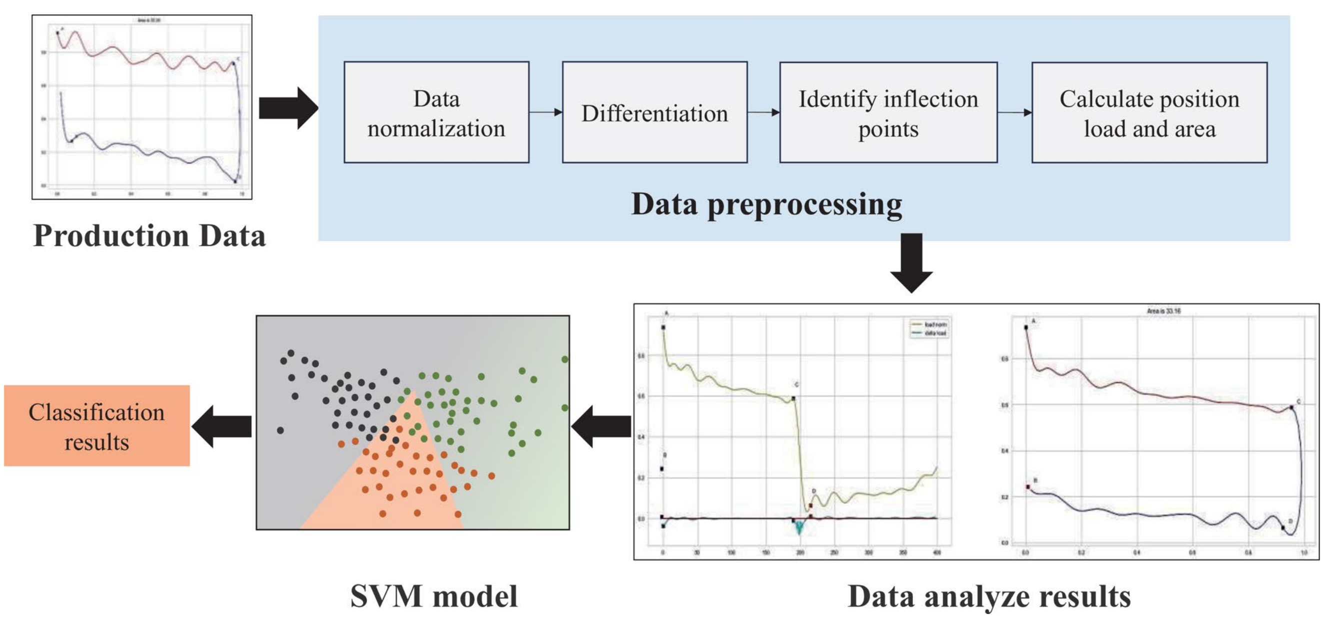

Fig. 4 The structure of SVM model

Table 7 Sample data

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
Wellhead oil pressure	0.18	0.63	0.45	0.98	0.25	0.46
Wellhead temperature	4.35	3.65	4.32	3.98	4.81	3.73
Daily power consumption	64.3	69.5	61.2	57.3	53.9	51.2
Maximum load	51.2	53.5	52.9	48.3	45.7	59.5
Minimum load	42.1	46.4	44.6	42.7	39.4	46.4
Submergence	4.67	3.76	6.32	9.23	4.12	7.92
Dynamic fluid level	1265	1209	1187	1098	1208	1361
A current	32.1	35.4	31.9	37.4	29.8	20.5
B current	31.5	34.9	33.1	36.4	31.2	23.9
C current	33.5	35.6	32.8	35.9	30.4	22.9
Fault type	Wax	Thick oil	Insufficient fluid supply	Piston leakage	Gas impact	Hang on

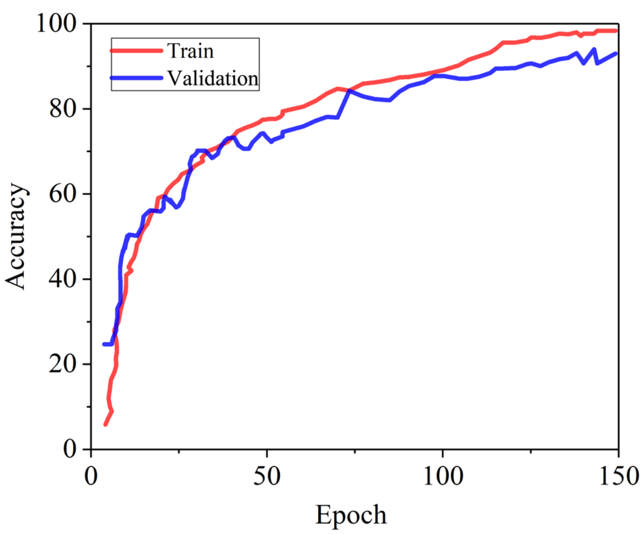

Fig. 5 The accuracy of train and validation process

Table 8 The results of ablation test

Ground truth	Predict class	
	Positive class	Negative class
Positive class	True Positive	False Negative
Negative class	False Positive	True Negative

of parameter tuning during the training process. To prevent overfitting in the AdaBoost algorithm, an early stopping strategy is introduced. This strategy involves monitoring the model's performance on a validation set and stopping the iteration when the performance no longer improves. Finally, after 150 iterations, the model achieved an accuracy of 0.98. Moreover, the accuracy of validation is 0.92 (Fig. 5).

Evaluation indicators

In the classification task, the performance of the classification method is evaluated comprehensively by using three evaluation metrics: accuracy, precision, recall and F1 score. The confusion matrix is usually used to calculate the above four evaluation metrics, and the confusion matrix is shown in Table 8.

Where True Positive is the positive samples that are predicted to be positive by the model; False Positive is the negative samples that are predicted to be positive by the model; False Negative is the positive samples that are predicted to be negative by the model; True Negative is the negative samples that are predicted to be negative by the model. The three evaluation indexes of accuracy, precision and recall are calculated as shown in (3)–(6).

$$accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{3}$$

$$precision = \frac{TP}{TP + FP} \tag{4}$$

$$recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{6}$$

Test results and analysis

Daily production data from various wells for the period from January to July 2023 were collected to validate the model.

(1) Ablation study on ensemble learning

To validate the effectiveness of the three methods in ensemble learning, separate ablation tests were conducted for each model. Firstly, classification experiments were performed for each of the four models individually. Then, based on the AdaBoost ensemble framework, two of these models were combined to conduct six sets of tests. Furthermore, three of the models

Table 9 The results of ablation study on ensemble learning

Model			Accuracy	Precision	Recall	F1 score
ML-CSNN	Hybrid model	SP-SVM				
✓			0.91	0.88	0.89	0.94
	✓		0.87	0.82	0.87	0.89
		✓	0.85	0.84	0.87	0.89
✓	✓		0.93	0.86	0.86	0.92
✓		✓	0.89	0.88	0.90	0.93
	✓	✓	0.85	0.88	0.83	0.85
✓	✓	✓	0.98	0.93	0.94	0.93

Table 10 The results of compared test on ensemble learning

Model	Accuracy	Precision	Recall	F1 score
GBDT	0.81	0.84	0.79	0.88
XGBoost	0.92	0.89	0.91	0.90
CatBoost	0.94	0.91	0.89	0.85
LightGBM	0.89	0.85	0.90	0.87
RF	0.83	0.84	0.81	0.82
Stacking	0.89	0.85	0.90	0.87
ours	0.98	0.93	0.94	0.93

were combined to obtain four different sets of ablation experiments. The results are shown in Table 9.

From the data in Table 9, it can be observed that the test results of single models and those integrating only two models exhibit relatively small differences across the four evaluation metrics, with none yielding satisfactory results. The performance of integrating three models surpasses that of integrating two models but still falls short of DTSKL-Boost. The analysis reason may be that the three models identify the type of working condition from different aspects, which makes the accuracy reach the highest.

(2) Compared test on ensemble learning

To further verify the accuracy of the DTSKL-Boost model, the test was designed to compare the proposed model with several established algorithms, including Gradient Boosting Decision Tree (GBDT), XGBoost, CatBoost, LightGBM, Random Forest (RF), and StackingClassifier. These models have approximately equal number of parameters. The XGBoost, CatBoost, LightGBM and StackingClassifier models also integrate the same model to maintain the consistent design approach. Given the inherent randomness inherent to algorithms, it is not uncommon for some methods to occasionally achieve the best score due to chance occurrences. To ensure a robust and comprehensive assessment, the methods are executed multiple times, and the average score across these runs is utilized to mitigate any reliance on chance outcomes. The test results are shown in Table 10.

Table 11 The results of ablation study on ML-CSNN

Model	Accuracy	Precision	Recall	F1 score
ML	0.78	0.73	0.70	0.81
4D-TF+CSNN	0.84	0.86	0.81	0.93
CSNN	0.75	0.74	0.69	0.84
ML-CSNN	0.91	0.88	0.89	0.94

Table 12 The results of ablation study on hybrid model

Model	Accuracy	Precision	Recall	F1 score
AlexNet	0.67	0.61	0.59	0.72
AlexNet + Expert rules	0.87	0.82	0.87	0.89

From Table 10, it can observe that within the same ensemble learning framework, the proposed method demonstrates the best performance across all four evaluation metrics. DTSKL-Boost exhibited enhancements ranging from 0.04−0.17 on accuracy, 0.02−0.09 on precision, 0.03−0.15 on recall, 0.03−0.11 on f1 score, compared to the other six methods. The analysis reason may be that the GBDT and RF method integrate multiple decision trees, and the method cannot accurately extract abnormal information of working conditions. The method use AdaBoost because it demonstrates strong versatility, making it applicable to various types of classification tasks, as well as both discrete and continuous feature data.

(3) Ablation study on ML-CSNN

To demonstrate the effectiveness of incorporating time-domain and frequency-domain feature extraction techniques (4D-TF) and meta-learning methods (ML) into the shrinkage neural network, four sets of comparative experiments were conducted. The experimental results are shown in Table 11. It can be observed that the performance of ML-CSNN is better than others. The reason for the analysis may be that ML can quickly adapt to the data characteristics of different Wells and different working conditions, and 4D-TF can comprehensively capture the trend and periodic pattern of data over time.

Table 13 The results of ablation study on SP-SVM

Model	Accuracy	Precision	Recall	F1 score
SVM	0.58	0.53	0.55	0.62
SP-SVM	0.85	0.84	0.87	0.89

(4) Ablation study on hybrid model

To demonstrate the effectiveness of the hybrid model based on convolutional neural networks and expert rules, with AlexNet model as the baseline, two sets of comparative experiments are designed. The experimental results are shown in Table 12. From the results, it can be observed that using expert rules can significantly improve the accuracy of predictions. The reason for this improvement may be that the quality of the sample dataset is poor, leading to the inability of pure deep learning methods to accurately extract features from the data. The reason may be that the hybrid model can integrate the flexibility of empirical data and the reliability of physical principles, and improve the prediction accuracy and generalization ability of the model.

(5) Ablation Study on SP-SVM

To demonstrate the effectiveness of extracting oil production engineering sensitive parameters before training the SVM model, two sets of comparative experiments are designed. The experimental results are shown in Table 13. It can be seen that extracting oil production engineering sensitive parameters from production data and then importing them into SVM for further feature extraction can significantly improve the accuracy of equipment failure prediction.

Discussion

The DTSKL-Boost method proposed in this paper integrates multiple models of different types, which increases the complexity and computational cost of the models. Particularly in real-time applications, it may face high computational resource requirements. Combining different types of models may encounter challenges in consistency and stability, as there may be significant performance differences among different models. Therefore, engineering processing of different types of features is needed, and the input and output formats of different models need to be unified, which may increase the complexity of data integration and processing.

In the future, techniques for adaptively selecting the most appropriate parameters for the model will be investigated to improve overall performance. Techniques such as Bayesian optimization, genetic algorithms, or reinforcement learning can dynamically adjust model hyperparameters based on continuous evaluation metrics, ensuring that the model operates at maximum efficiency. On the other hand, meticulous feature engineering to deepen the understanding of the underlying data is crucial. Utilizing techniques such as mutual information, feature importance ranking, and even automatic feature selection algorithms can help identify the most discriminative features of each model and thus improve their predictive power.

In engineering practice, the integrated model is deployed in the actual equipment predictive maintenance system to monitor and predict various data from oil fields in real-time. This supports maintenance decision-making and continuously optimizes and improves the model and system based on feedback information, further enhancing the model's predictive performance and stability. Adopting a DevOps approach with continuous integration/continuous deployment (CI/CD) pipelines ensures rapid iteration and improvement of both models and the supporting systems. Feedback from operational use, including maintenance outcomes, should inform ongoing model refinements and system optimizations.

Conclusions

This paper proposes a deep learning-based predictive maintenance method for oil and gas equipment. The main conclusions of this study are as follows.

1. By analyzing the common faults of oil field equipment, seven categories and 38 fault types are determined. Each fault type is correlated with electrical parameters such as power, current, and load to screen the training features, so as to establish correlation links and form a feature set. The fault type and the corresponding feature set are integrated to establish a sample set for deep learning model training.
2. The AdaBoost algorithm is employed to integrate three models, thereby forming a robust classification model that significantly enhances the predictive capability for equipment failure.
3. Meta-learning is incorporated into the Convolutional Neural Network (CSNN) to enable the model to rapidly adjust its parameters using a small amount of new data, thus achieving more effective transfer learning. The combination of time-domain and frequency-domain feature extraction techniques effectively captures the dynamic changes and periodic characteristics of the equipment's operating state.
4. Hybrid modeling approach is employed to integrate expert knowledge with AlexNet. Domain-specific rules

and logic are directly integrated into the neural network, providing prior knowledge during the model training process to guide decision-making in alignment with domain expertise.

5. Support vector machine model is constructed based on sensitive parameters from oil production engineering. Based on oil production experience, sensitive parameters are extracted from the indicator diagram as input features for the model, which helps reduce the interference of irrelevant features and enhances the accuracy of condition predictions.
6. The comparative experiment is conducted using real production data. The experimental results demonstrate that the DTSKL-Boost model improves accuracy, precision, recall rate, and F1 score by 9–25% compared to other models in predicting oil well equipment faults. Ablation experiments demonstrate that each module in the DTSKL-Boost method contributes to the accuracy of equipment fault prediction. The model can be deployed in actual production settings to optimize production efficiency.

Appendix A. The causes and phenomena of different fault types

1. Insufficient fluid supply

   The causes of insufficient fluid supply are as follows: (1) The fluid supply pressure at the wellhead is low, which may be due to issues with the fluid supply system, such as insufficient pump power of the lift pump or aging of the fluid supply pipelines. (2) Blockage in the fluid supply pipeline prevents the fluid from smoothly entering the oil pumping system. (3) Blockage in the crude oil transportation pipeline, which may be caused by the deposition of impurities in the oil–water mixture or corrosion inside the pipeline. (4) Damage to internal components of the lift pump, such as valve failure or pipeline rupture, affects the flow rate and pressure, leading to insufficient fluid supply. (5) Oversized equipment results in the "over-equipment for underperforming wells" phenomenon in low-yield wells, reducing equipment utilization efficiency. (6) Excessive stroke frequency increases system energy consumption and reduces system efficiency. (7) Deep burial of oil layers and small deep pumping force limit the full utilization of the well's production capacity.
   The phenomena is as follows: (1) Insufficient fluid supply directly affects the oil production of the well, leading to a significant drop in output. (2) The oil pump cannot operate fully, resulting in reduced pump efficiency. (3) To maintain or increase production, the operation time or frequency of the oil pumping unit may need to be increased, thus increasing energy consumption. (4) It may cause the oil pumping unit and other related equipment to operate under abnormal conditions, accelerating equipment wear and tear. (5) It may lead to instability in the oil pumping system, increasing the risk of safety accidents such as equipment overheating and pipeline rupture.

2. Liquid vibration

   The causes of vibration mainly include plunger-barrel collision and unstable liquid flow. Due to pump body damage or displacement, the plunger's movement within the barrel may become unstable, resulting in collisions with the barrel during the upstroke and downstroke. These collisions lead to instantaneous increases in hydraulic pressure within the pump, thereby generating vibration phenomena. In addition, pump body damage or displacement can also cause unstable liquid flow within the pump. When the liquid flow is unstable, it is prone to causing phenomena such as vortices and bubbles, which exacerbate the vibration phenomena.
   Prolonged vibration accelerates the wear and aging of the oil pumping unit and its related equipment, shortening their service life. Vibration can also trigger safety accidents such as equipment detachment and pipeline rupture.

3. Sucker rod fracture

   During the oil production process, gas impact may occur due to changes in formation pressure or gas accumulation during pumping. This impact generates intense vibration and shock forces on the sucker rod, leading to fractures in stress-concentration areas such as the arc transition zone and heat-affected zone. Under the effect of gas impact, the surface of the sucker rod may be more susceptible to corrosion, forming stress concentration points, which accelerate fatigue fractures.
   Macroscopically, the fracture surface of the sucker rod is usually relatively flat, with no obvious signs of eccentric wear, bending, or diameter reduction, demonstrating typical characteristics of fatigue fracture. Fractures often occur in the arc transition zone of the sucker rod joint and its adjacent areas, which is closely related to stress concentration and the impact of gas.

4. Oil well clogging

Under gas impact conditions, gas from the formation may be sucked into the pump, forming gas accumulation. These gases occupy space in the wellbore, reducing the flow channels for liquids, leading to oil well clogging. Gas impact may also bring impurities and particles from the formation into the wellbore. These impurities deposit in the wellbore, further exacerbating oil well

The components of the clogged material are diverse, including gas, impurities, particles, etc. The clogging locations are mostly located in the near-well area, which is related to the influence range of gas impact and impurity deposition.

5. Airlock

The causes of airlock are as follows: (1) When the liquid level drops too rapidly, the suction end of the oil well pump may inhale air, leading to the accumulation of gas within the pump. (2) After a pipeline leak occurs, external air can be sucked into the pipeline, subsequently causing an airlock phenomenon within the oil well pump. (3) During maintenance, component replacement, or pump restart, if the air within the pipeline is not properly expelled, it can also lead to airlock. (4) When there are foreign objects or obstacles in the oil pipe, eccentric distances may occur, resulting in the accumulation of gas in the pipe and the formation of an airlock. (5) As the well temperature changes, the volume of oil vapor also changes, affecting the oil flow rate and oil pressure, which may also lead to an airlock phenomenon.

Due to the obstruction of liquid flow caused by gas accumulation, the oil well pump is unable to effectively extract the oil–water mixture, leading to a sharp decline in oil well production. When operating in an airlock state, the oil well pump requires more energy to overcome the gas resistance, thereby increasing energy consumption. Operating under an airlock condition for an extended period may accelerate the wear and aging of the oil well pump and related equipment, shortening their service life.

6. Pumping unit blockage

The interior of the oil well pump can become clogged due to the accumulation of impurities, waxy substances from crude oil, sand particles, and other materials, resulting in malfunction of the pump. Additionally, aging or damaged seals can lead to oil leakage or gas ingress within the pump, further impacting its efficiency. Over time, oil pipes or suction lines may become clogged due to the deposition of impurities, wax solidification, or other factors, hindering the normal flow of oil. Improper pipeline design or installation can also contribute to blockage issues.

7. Sucker rod vibration and wobble

When the oil well pump is blocked, the pressure distribution within the pump changes, potentially generating lateral or impact forces. These forces are transmitted to the sucker rod, causing it to vibrate and wobble. In particular, the formation of a low-pressure gas pocket within the pump and its subsequent impact with the liquid surface during the downstroke of the plunger can result in a severe hydraulic hammer effect, further intensifying the vibration and wobble of the sucker rod. Changes in pressure within the pump due to blockage can also affect the overall mechanical balance of the pumping unit. When mechanical components are subjected to uneven forces, additional vibration and wobble are generated. Blockage can impede the smooth flow of oil, affecting the lubrication of bearings and other moving parts. Inadequate lubrication increases friction and wear between components, ultimately triggering vibration and wobble.

8. Wellhead eccentric wear in pumping unit

Wellhead eccentric wear in a pumping unit refers to the abnormal friction and abrasion between the polished rod and the wellhead sealing device (such as the packing box) due to poor alignment during operation. The primary causes include installation errors of pumping unit, well deviation and tree inclination, and unreasonable design of polished rod and wellhead sealing device. Eccentric wear accelerates the wear of the polished rod and wellhead sealing device, shortening their service life. Severe eccentric wear can lead to oil leakage, shutdowns, and other failures, disrupting normal oil production operations. Frequent repairs and replacements of equipment components escalate production costs.

9. Wellhead leakage

Wellhead leakage refers to the phenomenon where crude oil or oil-gas mixtures escape into the environment due to inadequate sealing at the pumping unit wellhead. The primary causes include poor sealing between polished

rod and wellhead, failure of sealing devices, and excessive backpressure.
The impacts include environmental pollution, safety hazards, and economic losses. Leaked crude oil and oil-gas mixtures contaminate soil, water sources, and other environments. Leaked oil-gas mixtures can easily trigger fires, explosions, and other safety incidents. Leakage results in crude oil losses, hiking production costs and environmental remediation expenses.

Funding This work was supported by China Petroleum & Chemical Corporation Science and Technology Special Project "Research on Intelligent Perception Technology of Oil and Gas Station Safety and Environmental Protection Risks" (323016) and Postdoctoral Innovation Program of Shandong Province (SDCX-ZG-202302035).

Data availability According to the oilfield data confidentiality requirements, the data cannot be shared publicly because the data of this study contain sensitive oilfield production information.

Materials availability No material was used for the research described in the article.

Declarations

Conflict of interest The authors declared that they have no conflict of interest in this work.

Ethical statements This study did not involve human or animal subjects, and thus, no ethical approval was required. The study protocol adhered to the guidelines established by the journal.

Consent for publication Written informed consent for publication was obtained from all participants.

References

Abhulimen KE, Fashanu S, Idialu P (2018) Modeling fracturing pressure parameters in predicting injector performance and permeability damage in subsea well completion multi-reservoir system. J Petrol Expl Prod Technol 8:813–838

Agwu OE, Akpabio JU, Dosunmu A (2020) Artificial neural network model for predicting the density of oil-based muds in high-temperature, high-pressure wells. J Petrol Expl Prod Technol 10:1081–1095

Al-Sabaeei AM, Alhussian H, Abdulkadir SJ et al (2023) Prediction of oil and gas pipeline failures through machine learning approaches: A systematic review. Energy Rep 10:1313–1338

Antonio N, de Almeida A, Nunes L (2022) Data mining and predictive analytics for e-tourism. In: Handbook of e-Tourism. Springer, New York, pp 531–555

Aranha PE, Policarpo NA, Sampaio MA (2024) Unsupervised machine learning model for predicting anomalies in subsurface safety valves and application in offshore wells during oil production. J Petrol Expl Prod Technol 14(2):567–581

Bukhtoyarov V, Tynchenko V, Petrovskiy E et al (2019) Intelligently informed control over the process variables of oil and gas equipment maintenance. Int Rev Autom Control 12(2):59–66

Chen Z, Selere OO, Seng NLC (2022) Equipment failure analysis for oil and gas industry with an ensemble predictive model. In: International conference on computational science and technology, Springer, New York, pp 569–581

Choi JE, Seol DH, Kim CY, et al. (2023) Generative adversarial network-based fault detection in semiconductor equipment with class-imbalanced data. Sensors 23(4): 1889

Elahifar B, Hosseini E (2024) A new approach for real-time prediction of stick-slip vibrations enhancement using model agnostic and supervised machine learning: a case study of norwegian continental shelf. J Petrol Expl Prod Technol 14(1):175–201

Fan J, Han F, Liu H (2014) Challenges of big data analysis. Natl Sci Rev 1(2):293–314

Fan SKS, Hsu CY, Tsai DM et al (2020) Data-driven approach for fault detection and diagnostic in semiconductor manufacturing. IEEE Trans Autom Sci Eng 17(4):1925–1936

Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. AI Mag 17(3):37–37

Fernandes M, Corchado JM, Marreiros G (2022) Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: a systematic literature review. Appl Intell 52(12):14246–14280

Fomin DA, Antipov AE, Lapin DG (2018) Protection from corrosive destruction of deep-pump equipment in oil and gas well. Key Eng Mater 769:290–295

Guo G, Wang H, Bell D, et al. (2003) Knn model-based approach in classification. In: On the move to meaningful internet systems 2003: CoopIS, DOA, and ODBASE: OTM confederated international conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3–7, 2003. Proceedings. Springer, pp 986–996

Hsu CY, Liu WC (2021) Multiple time-series convolutional neural network for fault detection and diagnosis and empirical study in semiconductor manufacturing. J Intell Manuf 32(3):823–836

Iraji S, Soltanmohammadi R, Matheus GF et al (2023a) Application of unsupervised learning and deep learning for rock type prediction and petrophysical characterization using multi-scale data. Geoenergy Sci Eng 230:212241

Iraji S, Soltanmohammadi R, Munoz ER et al (2023b) Core scale investigation of fluid flow in the heterogeneous porous media based on x-ray computed tomography images: Upscaling and history matching approaches. Geoenergy Sci Eng 225:211716

Iraji S, De Almeida TR, Munoz ER et al (2024) The impact of heterogeneity and pore network characteristics on single and multi-phase fluid propagation in complex porous media: An x-ray computed tomography study. Pet Sci 21(3):1719–1738

Jacobs T (2018) Ar headsets give oil and gas sector the quicker fix. J Petrol Technol 70(07):32–34

Johnstone J, Curfew J (2012) Twelve steps to engineering safe onshore oil and gas facilities. Oil Gas Facilities 1(04):38–46

Joseph D, Gallege T, Bekar ET et al (2022) A predictive maintenance application for a robot cell using lstm model. IFAC Pap Online 55(19):115–120

Khalid S, Hwang H, Kim HS (2021) Real-world data-driven machine-learning-based optimal sensor selection approach for equipment fault detection in a thermal power plant. Mathematics 9(21):2814

Li T, Liu H, Ding D (2018) Predictive energy management of fuel cell supercapacitor hybrid construction equipment. Energy 149:718–729

Lu Y, Ding EJ, Du J et al (2020) Safety detection approach in industrial equipment based on rssd with adaptive parameter optimization algorithm. Saf Sci 125:1–15

Mahdavi Kalatehno J, Khamehchi E (2024) Development of a novel packer fluid for high-temperature and high-pressure oil and gas wells with using design of experiments and artificial intelligence. J Petrol Expl Prod Technol 14:1–25

Mohammadpoor M, Torabi F (2020) Big data analytics in oil and gas industry: an emerging trend. Petroleum 6(4):321–328

Mohammed A (2023) Data driven-based model for predicting pump failures in the oil and gas industry. Eng Fail Anal 145:107019

Nguyen TN, Ponciroli R, Bruck P et al (2022) A digital twin approach to system-level fault detection and diagnosis for improved equipment health monitoring. Ann Nucl Energy 170:109002

Nithin A, Sriramula S, Ebinum T (2021) Reliability-centered maintenance and cost optimization for offshore oil and gas components. J Phys Conf Ser 1730:012007

Niyonambaza I, Zennaro M, Uwitonze A (2020) Predictive maintenance (pdm) structure using internet of things (iot) for mechanical equipment used into hospitals in Rwanda. Future Internet 12(12):224

Park YS, Yoo DY, Lee JW (2021) Programmable motion-fault detection for a collaborative robot. IEEE Access 9:133123–133142

Pivovarov VY, Duvakina I (2020) Development of damage evaluation methods in uneasily accessible units of oil and gas equipment. IOP Conf Ser Earth Environ Sci 459:042014

Quinlan JR (1986) Induction of decision trees. Mach Learn 1:81–106

Ran S, Jiang F, Tao X et al (2021) Research on information management system of oil and gas field surface engineering system. In: 2021 International conference on computer, internet of things and control engineering (CITCE). IEEE, pp 130–133

Steinwart I, Christmann A (2008) Support vector machines. Springer Science Business Media, Heidelberg

Tariq Z, Aljawad MS, Hasan A et al (2021) A systematic review of data science and machine learning applications to the oil and gas industry. J Petrol Expl Prod Technol 11:4339–4374

Wang Y, Tian M, Wang D et al (2012) Study on the hse management at construction site of oil and gas processing area. Proc Eng 45:231–234

Woldesellasse H, Tesfamariam S (2023) Failure assessment of oil and gas transmission pipelines using an integrated Bayesian belief network and gis model. Int J Press Vessels Pip 205:104984

Wu H, Triebe MJ, Sutherland JW (2023) A transformer-based approach for novel fault detection and fault classification/diagnosis in manufacturing: a rotary system application. J Manuf Syst 67:439–452

Xie L, Håbrekke S, Liu Y et al (2019) Operational data-driven prediction for failure rates of equipment in safety instrumented systems: a case study from the oil and gas industry. J Loss Prev Process Ind 60:96–105

Xia Z, Ye F, Dai M et al (2021) Real-time fault detection and process control based on multi-channel sensor data fusion. Int J Adv Manuf Technol 115(3):795–806

Zhang X, Li H (2019) Research on transformer fault diagnosis method and calculation model by using fuzzy data fusion in multi-sensor detection system. Optik 176:716–723

Zhifeng Y (2019) Cloud computing and big data for oil and gas industry application, china. J Comput 14:268–282