Research on Fault Diagnosis System of IOT for Oil Well Pump Based on Machine Learning

Bozhi Xiao, Yangbing Lu, Miao Wang, Haojie Li, Hengyan Zhu, Xuan Cao, Shengquan Yang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: xaitysq@163.com

***Abstract*—In order to realize automatic prediction and processing of remote fault diagnosis of oil well pumps distributed in different regions by crude oil production enterprises, a fault diagnosis system for oil well pumps based on machine learning was researched and designed. This fault diagnosis system is composed of three layers (perception layer, network layer and control application layer) Internet of Things structure. The function and characteristics of each layer are analyzed in this paper, and the hardware composition and control principle of sensor nodes and aggregation nodes of the measurement and control system are discussed and gives the node microcontroller program design flow chart and the main module content of the IoT central computer software design. This paper focuses on the principle of machine learning for fault diagnosis and prediction, and deeply explains the algorithm steps of using LSTM for fault diagnosis of oil well pumps. The enterprise application experiment results show that, compared with the traditional manual well patrol fault diagnosis method, this system has the advantages of convenient operation and maintenance, low labor intensity, high timeliness and accuracy of fault diagnosis, which can better reduce equipment maintenance costs for enterprises.**

Keywords-Machine Learning; Oil Well Pump; Internet of Things; Fault Diagnosis

I. INTRODUCTION

The oil pump is an essential and important power equipment for oil wells to export crude oil[1]. Since most oil wells are distributed in desert mountains and wilderness, oilfield enterprises basically adopt on-site instrumentation display monitoring for their management, and carry out early warning management by manual line inspection, that is, patrol The inspectors check the operation status of the oil delivery pump of the well group regularly every day, record relevant data, and obtain information such as the flow, pressure and temperature of the oil pump during the period of operation [2-3]. As the scope of oilfield production sites and pipeline networks becomes wider and wider, and the number of oil wells continues to expand, this method gradually shows shortcomings in management. It shows shortcomings such as high labor intensity and obvious lag in dealing with problems, such as: If the operating pressure of the oil pump is abnormal, if the manual inspection method finds that the treatment is not timely, it may lead to the rupture of the oil pipeline, which will cause a series of serious problems such as property loss of oilfield enterprises and environmental pollution [4-5].

The Internet of Things is the "Internet of Things Connected", which is automatically generated through information such as various sensing devices, ZigBee wireless sensor network technology, 5G network transmission technology, RFID technology, video recognition technology, infrared sensing, global positioning system, and laser scanners. Acquisition equipment, according to the agreed agreement, realizes the network connection of interconnection and intercommunication of items according to the needs, and performs information exchange and communication, so as to realize the intelligent network system of intelligent identification, positioning, tracking, monitoring and management [6].

Machine learning is one of the branches of artificial intelligence, which refers to a method that enables computer systems to automatically learn and continuously improve performance

DOI: 10.2478/ijanmc-2023-0062

through computer algorithms and models [7-8]. The functions of machine learning include but are not limited to: classification, clustering, regression, dimensionality reduction, anomaly detection, association rule mining, etc. The advantages of machine learning in fault diagnosis are many. For example, machine learning can adaptively learn a machine's diagnostic knowledge from collected data, rather than utilizing the experience and knowledge of engineers. [9-11] This method can improve the accuracy and efficiency of fault diagnosis. LSTM [12-13] is an algorithm in machine learning. It is a special cyclic neural network that can solve the problem of gradient disappearance in traditional cyclic neural networks, and can also process long sequence data. In addition, the LSTM algorithm can also discover hidden patterns and relationships in large-scale data, thereby improving the ability of fault diagnosis.

II. DESIGN OF OIL PUMP MEASUREMENT AND CONTROL SYSTEM BASED ON INTERNET OF THINGS ARCHITECTURE

A. Overall Design of IoT System

The measurement and control of oil well pump data is the foundation of digital intelligent oilfield construction. Its goal is to detect the operating status and parameters of various components of oil pump equipment in real time, and transmit the collected operating data to the Internet of Things through intelligent instrument communication nodes and dedicated 5G networks. Control center machine [14]. On the one hand, the central computer conducts analysis and processing through the upper-level intelligent control software, makes decision-making control when necessary, and then outputs it to the equipment control mechanism through a dedicated 5G network; on the other hand, the central computer can convert and standardize the collected data through algorithms and store them to the database server, and at the same time serve mobile terminal or desktop terminal users through the public 5G network or Internet network [15]. The principle composition structure of the system equipment based on the Internet of Things architecture is shown in Figure 1. The system adopts a layered design method to realize real-time data collection and monitoring of oil pipeline pressure, flow and temperature, and has low cost and strong automatic monitoring capabilities. It mainly consists of three layers: perception layer, network layer and application layer.

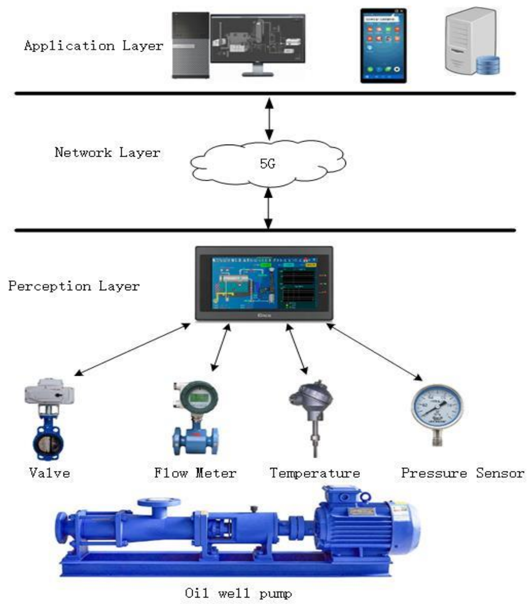

Figure 1. Organization structure of the system based on internet of things technology

B. Perceptual layer design

The sensing signals of this system for the measurement and control of oil delivery pumps in oilfield well groups mainly include: valve switches of various components, oil pump inlet and outlet pressure, oil pump inlet and outlet temperature, oil pump output flow, oil pump body temperature and other sensors. The sensing layer is composed of ZigBee wireless subnets of oil well pumps located in several areas, and the ZigBee wireless subnet is composed of sensor nodes with ZigBee communication modules and aggregation nodes (ZigBee well group control communicators, also known as gateway nodes). The convergence node in this system is the transfer station of each oil well's external communication, and its core component is the ZigBee wireless transceiver module. The aggregation node is a relatively independent processing unit. If the external network fails to contact the IoT control center, it can automatically adjust the operating status of the local oil pump according to the node's internal operation logic rule base. In addition, it can also store the local data for a period of time. The measurement and control data can be queried locally to display the real-time or historical operation data of each sensor.

C. Network layer design

The network layer mainly completes the reliable transmission of the information collected by the oil pump or the control signal of the Internet of Things center [16-17]. It is essentially based on the wireless local area network (WLAN) or the wireless mesh protocol to form a larger network of multiple aggregation nodes to monitor the distribution of multiple areas. Oil well pumps. Specifically, its function is to receive the information from the aggregation node of the perception layer, which can be encrypted and transmitted to the control application layer through appropriate algorithms as needed [18]. In addition, the information and control output of the control application layer are reversely transmitted to the aggregation node of the perception layer. The network layer mainly acts as a bridge. Since the oil wells of oilfield companies are mostly located in uninhabited mountain areas, these places often do not have public communication 5G networks, and a dedicated wireless communication network needs to be set up to serve this system.

D. Application layer design

The application layer is the highest layer in the functional structure diagram of the system, and it is the remote control center of the system. It realizes the automatic safety control of the oil pump on the basis of receiving and analyzing the sensory information of each oil well [19-21]. The control application layer is mainly composed of IoT central computer, database server, desktop terminal, mobile terminal and so on. The central computer of the Internet of Things generally has a wireless transceiver module, which can receive the equipment information transmitted by the dedicated 5G network base station of the transport layer in real time, analyze and process it according to the requirements of the oil pump operation process parameters through the software system, and store the data periodically in the database server [22-23]. Oil wells with different production volumes are equipped with different oil pump powers and different operating parameters. Various equipment parameters can be set through the man-machine interface of the control application layer. Spare. The data received by the control application layer from the network layer generally undergoes the steps of verification, unpacking, and anti-encryption conversion [24-25]. It can dynamically display data such as oil well temperature, pressure, and flow rate through a powerful host computer intelligent graphics software system, and it can monitor the abnormal operation of the equipment. It can predict in advance and make corresponding output actions quickly according to the equipment process. Mobile terminals such as mobile phones and handheld PDAs can access the web interface of the control system through the public 5G network, and can access the operating curves and data of the equipment under the permission of the authority; in addition, various departments of the oilfield enterprise can also access and browse the operating status of the system through the Internet Or remote control, which greatly facilitates the efficient management of oil pumps by oilfield companies.

III. OIL PUMP IOT MEASUREMENT AND CONTROL NODE HARDWARE

A. Sensor node

The main function of the sensor node is to collect current and voltage data, perform A/D conversion on the data and undergo a digital filtering process. These sensors are all equipped with wireless ZigBee communication modules. They can upload data to the sink node and accept commands from the sink node at the same time. Make the necessary adjustments to the output. A typical oil pump sensor node is usually composed of a sensor module, an analog filter module, an A/D conversion module, a D/A output module, an I/O control module, a node processor CPU, a wireless RFID module, a node storage module, and a power management module, etc [26]. Composition, the block diagram of its hardware composition is shown in Figure 2.

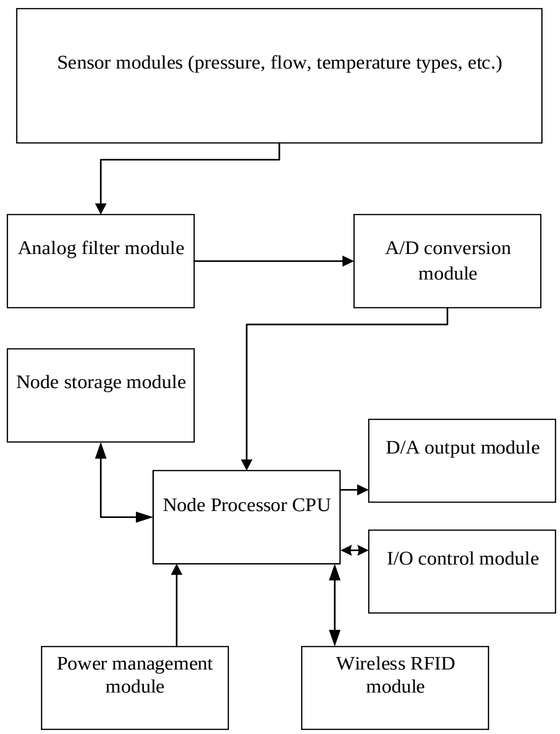

Figure 2. Hardware components theory block diagram of the fuel pump sensor instrument node

The pressure sensor of this system adopts Rosemount sensor with a measurement accuracy of 0.1%, the measurement range is 0-10 Mpa, the DC power supply voltage is DC24V, the output sensing signal is 4-20mA, and the medium temperature is -20-90℃. The flow sensor adopts FD-M100AT sensor, the detection distance is 4-150 mm, the switching frequency is 1000 Hz, the output mode is NPN, and the response time is less than 0.5 ms. It can realize the detection of liquid flow in harsh environments, and is very suitable for oil pipelines in oil field enterprises, field arrangement environment. The model used for the temperature sensor is DS18B20, and its range is -55 to 125℃. It can sample and quantify the temperature data, and the resolution can generally reach 0.0625. The working voltage is 3.0 to 5.5V. The format only needs a maximum of 740ms, which is suitable for the low power consumption of sensor nodes.

B. Sink node

Convergence node (ZigBee well group control communicator, also known as gateway node) is responsible for collecting local sensor signals and interacting data with the IoT central computer through the network layer [27]. When the central computer of the Internet of Things sends a control command through data calculation, the sink node immediately sends it to the sensor node through wireless communication in real time after receiving it. After the sensor node unpacks the communication command, it specifically performs D/A output or I/O control. The aggregation node is mainly composed of the lower wireless communication module, the upper wireless communication module, the node processor CPU, the node LCD display module, the node key control module, the node storage module and the power management module [28-29]. The block diagram of its hardware composition is shown in the figure 3. The sink node itself is a relatively independent local oil pump control instrument. It has two modes: Remote and Local. Normally, it is in the remote mode. Once the communication delay or communication failure occurs, the sink node will automatically switch to the local mode according to the instrument backup rule base Automatic regulation of the oil pump.

The lower layer wireless communication module selects the CC1100 communication module, which can directly receive the information collected by the perception layer. The upper wireless communication module selects the SRWF508A wireless module, and the transmission distance between the node and the network layer base station can reach 3.5km. It is easy to realize long-distance transmission of collected information by building a wireless network. The node storage module uses the AT 24C256 chip, and the storage capacity can be easily expanded through multi-chip interconnection. The node LCD display module can display the oil pump operation collection information and equipment status transmitted by the current sensor node, which is convenient for users to directly view on site. The Key button control module is convenient to use direct local manual input to run the control rule library, set the working mode of the node, or directly issue control commands to start and stop the oil pump equipment, etc., and it uses the matrix scanning method to work.

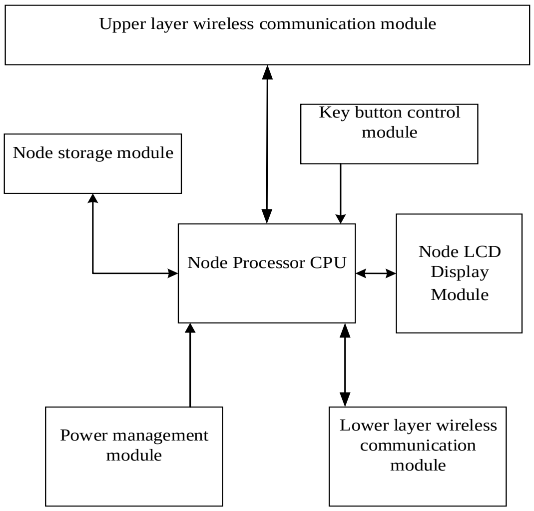

Figure 3. Hardware components theory block diagram of the IOT aggregation instrument node

IV. SOFTWARE DESIGN OF INTERNET OF THINGS SYSTEM FOR OIL WELL OIL PUMP

The software design of the intelligent oil well pump wireless network monitoring system is based on the basic principles of reliability, safety, and control robustness, and is designed under the guidance of modularization and scalable upgrades [30-31]. The software design mainly includes the sensor node program design, the convergence node program design and the software design of the intelligent monitoring system of the central computer of the Internet of Things.

A. Node programming

The node program is mainly based on the design of the single-chip microcomputer program. In this system, the sensor node and the sink node have many similarities in their working principles, and their programming methods are similar. Due to space limitations, only the workflow of the sink node single-chip microcomputer program is given below, as shown in the figure As shown in figure 4, the node program uses a modularized subroutine design mode, which is composed of multiple relatively independent subroutines called by the main control loop program. Each round of cycle through these subroutines mainly completes the lower sensor data communication acquisition, data storage and processing, Key button processing, upper-layer wireless communication data submission, node data display and other tasks, set the corresponding maximum running cycle time for each subroutine during the main loop running, so that if a node occurs during the running of a loop subroutine For exceptions, the system uses the stack to record the exception numbers in sequence, and unify to the last step to focus on popping the stack to execute the exception handling subroutine processing, which can improve the program operation efficiency and the timeliness of peripheral wireless communication response.

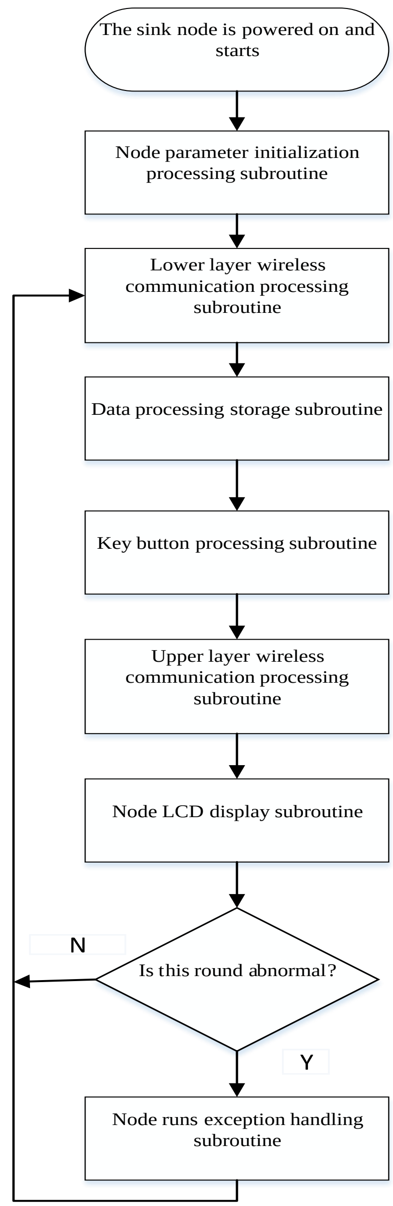

Figure 4. MCU control program flow chart of the IOT aggregation instrument node

The single-chip microcomputer program of the aggregation node is written in an MCU development environment compatible with C language syntax, and the code of the main loop program is as follows:

```
void main() /*main loop function definition*/
{
 GatherGaugeLoad();/*Gathering node power-on parameter initialization*/
  while (1) /*Infinite loop until the node is powered off*/
  {
 DownWirelessComm();/*lower layer communication, collect oil pump data*/
   DataProcessSave();/*data normalization processing and saving*/
   KeyInService();/*User Key button processing*/
   UpWirelessComm();/*upper layer communication, execute remote command*/
   KeyInService();/*Node instrument LCD displays required data*/
   if (RunError()!=0) /*There is an exception in this round*/
    Exceptionhand();/*Unified processing cycle exception*/
}
}
```

The wireless network communication between the sensor node and the aggregation node of this system is based on the traditional MAC protocol. It mainly realizes the establishment and maintenance of wireless data links between the devices of each layer of communication. The MAC data frame adopts the time slot CSMA/CA mechanism. It uses 3 the three parameters are the back-off index NB, the collision window CW and the back-off index BE to achieve reliable data transmission [32].

B. Central Computer Monitoring System Software Design

According to the actual needs of oil field enterprises for remote monitoring of well group oil pumps, the IoT central computer monitoring system software is mainly composed of 7 modules, as shown in Figure 5, including the system user management module, network communication processing module, data record processing module, process screen A display module, a dynamic curve display module, a sensor node parameter module, and a convergence node parameter module.

The system user management module is used to control the user's authority to use the software, including remote desktop terminal and mobile terminal users, and can display or maintain basic information records of logged-in users. The network communication processing module is the key core module of the system. On the one hand, it performs real-time data exchange with multiple convergence nodes of the system through the dedicated network layer network, and on the other hand, it provides data services for enterprise remote mobile terminals and desktop terminals through the public 5G network and the Internet. . The system control center adopts the C/S design mode and adopts the TCP/IP asynchronous communication mode. It handles network-related work through proxy callback functions, so that there is no need to block or suspend threads when performing network operations. The system provides a standard Modbus TCP RTU-based interface for remote desktop or mobile terminal web-based page access.

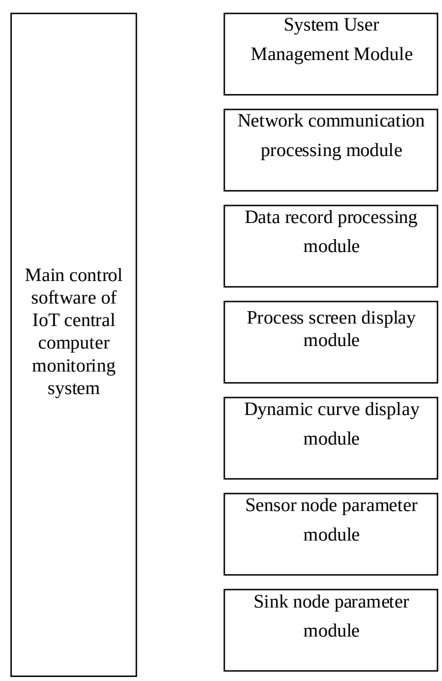

Figure 5. Module structure diagram of the IOT center computer monitor system software

The main function of the data record processing module is to normalize the data of all oil pumps in the oil station monitored by the system, and then periodically (the time can be set, usually 5 minutes) record it to the database server [33-34]. The process screen display module is the most important man-machine interface of the system. It uses virtual instrument technology to display the operation status of each component of the oil pump and the corresponding position of the sensor parameters in real time, and can prompt whether the system has an alarm. The general refresh time of the screen is 5S. The function of the dynamic curve display module will display each sensor parameter in the form of a curve for each oil well. The dynamic curve has historical data curve display and a short-term operation trend graph in the near future. The built-in fault diagnosis algorithm of the system can predict whether the future will there will be a possibility of failure, which allows users to prepare in advance and prevent problems before they happen [35]. The sensor node parameter module can set a series of parameters such as the type of the system oil pump sensor, the working mode of the sensor, and the communication mode of the sensor network. The aggregation node parameter module can set a series of parameters such as the working mode of the aggregation nodes of each oil station in the system, the backup logic rule base, and the node network communication mode.

V. Design of Machine Learning Fault Diagnosis for Oil Well Pump

Fault Diagnosis System (FDS) [36-37] refers to a system that collects data, analyzes data, judges the cause of the fault, and predicts the development trend of the fault when the machine fails, so as to take corresponding measures to avoid or reduce the damage caused by the fault to the machine.

The working principle of the fault diagnosis system mainly includes the following steps [38]:

1) *Data collection:* The fault diagnosis system first needs to collect various data generated during the operation of the machine, such as temperature, pressure, flow, vibration and other data. This data can be collected by sensors, meters or other devices.

2) *Data analysis:* The fault diagnosis system analyzes the collected data to find out abnormal data. Anomaly data is an early warning sign that a machine is malfunctioning.

3) *Fault judgment:* The fault diagnosis system judges the fault type of the machine according to the abnormal data.

4) *Fault prediction:* The fault diagnosis system predicts the development trend of faults according to the types of faults.

5) *Troubleshooting:* The fault diagnosis system takes corresponding measures according to the fault prediction to avoid or reduce the damage caused by the fault to the machine.

6) *The fault diagnosis system can help enterprises improve the operating efficiency of machines, reduce maintenance costs and prolong the service life of machines.*

Following are some advantages of fault diagnosis system [39]:

It can improve the operating efficiency of the machine: the fault diagnosis system can prevent the occurrence or expansion of faults through early warning of faults, thereby improving the operating efficiency of the machine.

Can reduce the maintenance cost: the fault diagnosis system can accurately determine the cause of the fault and avoid false maintenance, thereby reducing the maintenance cost.

Can prolong the service life of the machine: the fault diagnosis system can prolong the service life of the machine by finding and handling faults in time.

A. Macro Design of Fault Diagnosis Expert System

Combining the operation characteristics of the oil well oil pump process and the user's fault diagnosis discovery and processing requirements, the fault diagnosis expert system structure shown in Figure 6 is designed.

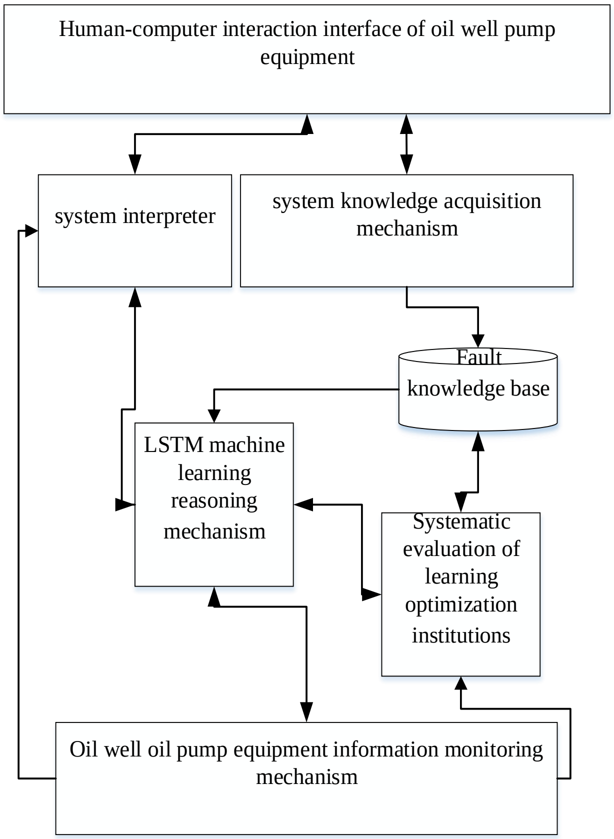

Figure 6. The structure of the expert system for oil well pump LSTM machine learning fault diagnosis

It consists of 7 parts, including: oil well pump man-machine interface, system interpretation mechanism, system knowledge acquisition mechanism, fault tree structure knowledge base, LSTM machine learning reasoning mechanism, system evaluation learning optimization mechanism and oil pump skid status information monitoring mechanism. The description of each part is as follows:

1) Man-machine interface of oil well oil pump: The terminal for displaying various information for the system operation and the operation interface for the user to interact with the system. On the one hand, the user inputs the original device fault diagnosis rules, logic rules and other knowledge into the system through this interface; on the other hand, the system outputs the diagnosis results through this interface and asks or answers reasoning questions to the user. In addition, the operator can input the manual intervention parameters of the diagnosis system through this interface at any time according to experience and knowledge, so as to achieve real-time manual correction of the diagnosis system, so that it can quickly enter the correct diagnosis track.

2) System explanation mechanism: Explain the cause of the fault and the reasoning process, and provide reasonable disposal suggestions to the user. It can answer various questions raised by users, and it can give necessary and clear explanations of the reasoning route and conclusion of the fault tree, so that users can understand the reasoning process. It is the main module to realize the transparency of reasoning. The system interpretation mechanism converts the information input by the user into the internal representation of the system, and then hands it over to the corresponding interpretation module for processing. The internal information of the system reasoning is also converted by the interpretation mechanism into an external form in the format understood by the user and displayed on the man-machine interface.

3) System knowledge acquisition organization: Domain expert knowledge engineers (mainly oilfield equipment experts) input all fault diagnosis experience knowledge and logical knowledge through the front man-machine interface, which will be processed by this organization and stored in the knowledge in a standardized form library to build fault trees.

In the fault diagnosis system, there are the original input rule knowledge, the rule knowledge generated by the system machine learning, and the intermediate rule knowledge generated during the reasoning process. Because there is an inevitable connection between this knowledge through the same attribute, it is necessary to eliminate their redundancy. That is to say, Jane. The basic idea of knowledge reduction is to use some measure to determine the importance of different attributes and construct the smallest subset of knowledge attributes [40]. The method of knowledge reduction based on rough set theory has good versatility, robustness and global search. There are two steps: one is to delete some columns from the decision table; the other is to delete redundant rows [41].

4) Fault knowledge base: Using production hierarchy rules to express the fault diagnosis reasoning logic function formulas of the phenomena and results of each component of the oil well oil pump. The knowledge in this knowledge base must be expressed in the normalized form of the fault tree hierarchy, which is the core of this expert diagnosis system and the key to whether the quality of the expert system is superior. The quality and quantity of knowledge in the knowledge base determine the quality of the expert system [42]. quality level. There are two main types of knowledge stored in this knowledge base: one is the open knowledge of oil well pumps, including definitions, facts and theories in the field; the other is the knowledge obtained by oil field experts in long-term business practice Practical experiential knowledge.

5) Machine learning reasoning mechanism: The machine learning reasoning mechanism refers to the reasoning process of the machine learning model, including forward propagation and back propagation. Forward propagation refers to the process of calculating the input data through the neural network model to obtain the output result; back propagation refers to the process of calculating the gradient according to the loss function, and then updating the model parameters through optimization algorithms such as gradient descent. In the expert system, the advantages of reasoning based on machine learning include: automatic feature extraction, adaptive adjustment of model parameters, and ability to deal with complex nonlinear relationships. Compared with traditional expert systems, machine learning-based expert systems can handle complex problems better because it can automatically extract features from data and adaptively adjust model parameters to suit different data distributions and tasks Require. In addition, expert systems based on machine learning can also deal with complex nonlinear relationships [43], which cannot be handled by traditional expert systems. The machine learning reasoning algorithm of LSTM is adopted in this system.

6) System evaluation learning optimization mechanism: During the operation of the system, the system periodically evaluates the effect of using the knowledge source in the fault tree knowledge base in the process of fault diagnosis and reasoning to solve problems. Status information, suspend the knowledge with poor diagnosis and processing effect, and give warnings and prompts to users for reasoning with insufficient diagnosis knowledge; in addition, according to the summary of system evaluation and optimization and the quality of actions after diagnosis and control instructions, it can monitor the system itself Constantly improve and innovate dynamically, and enrich the knowledge in the knowledge base in real time, and self-correct or delete the knowledge with poor performance and relatively large control deviation[44].

7) Oil well pump status information monitoring mechanism: This mechanism is mainly used to monitor the operating parameters of each equipment during the operation of the device through interface modules (such as I/O, A/D, D/A, and TCP/IP, etc., such as pressure, flow rate, etc.) , valve opening, temperature, etc., as well as the parameters of the components themselves, such as the frequency converter itself, such as current and voltage. The information is processed by the system data information rules and handed over to various institutions for use.

B. Principle of BP Neural Network in Fault Diagnosis

This system plans to use the LSTM algorithm, and its core is the BP neural network, so it is necessary to introduce the principle of the neural network in advance [45].

The BP neural network module used in the bottom layer of machine learning in this oil well oil pump fault diagnosis system is divided into four layers: input layer, fuzzy layer, fuzzy reasoning layer and output layer. The BP neural network input layer nodes correspond to the parameters that need to be controlled by the oil pump system, and the BP neural network output layer nodes correspond to the future measurement data predicted by the sensor. The function of the BP neural network module is to train and learn a large number of historical data of oil well pump sensors. The BP neural network controller can continuously optimize the real-time parameters of the machine learning fault diagnosis of the oil well oil pump by adjusting the weight coefficients of each layer, so as to realize the self-adaptive adjustment of fault diagnosis, prediction and alarm. The structure diagram of BP neural network is shown in Figure 7.

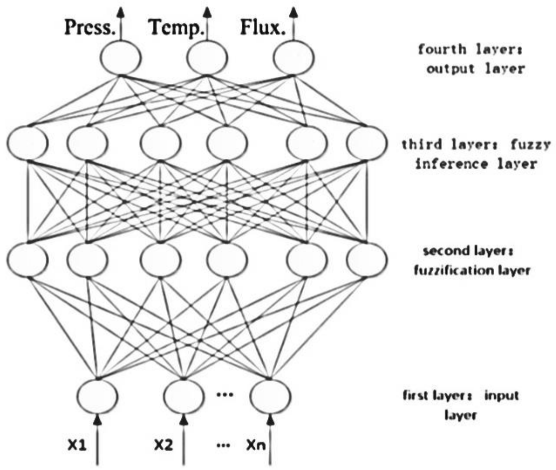

Figure 7. Neural network module structure diagram

In this network diagram, the first layer is the input layer of the BP neural network module. Each neuron node in this layer represents an input variable of the system such as oil well pressure, flow rate, temperature and other information. Since the system needs to control the oil well pump Inlet oil pressure, outlet oil pressure and output oil flow, so the neural network module of the input layer fuzzy neural network LSTM machine learning control is composed of three neurons, that is, three nodes, and the main function of this layer is that the neurons will input the variable The value is passed to the neurons in the fault diagnosis and prediction layer through the action function, and the input and output in this layer are as follows:

$$I_i^{(1)} = x_i \tag{1}$$

$$O_{ij}^{(2)} = x_i \tag{2}$$

The second layer in the BP neural network is the fuzzy layer of fault diagnosis prediction. Each neuron in this layer is used to simulate the membership function of the input variables of fault diagnosis prediction. The input and output of fault diagnosis prediction in this layer are expressed as:

$$I_{ij}^{(2)} = -\frac{(x_i - c_{ij})^2}{b_{ij}^{\ 2}} \tag{3}$$

$$O_{ij}^{(2)} = \exp(I_{ij}^{(2)}) \tag{4}$$

C. LSTM Machine Learning Model Design in Fault Diagnosis

LSTM is short for Long Short-Term Memory Neural Network. It is a variant of Recurrent Neural Networks (RNN) that excels at processing sequential data. LSTM was proposed by Hochreiter and Schmidhuber in 1997. LSTMs work by using gates to control the flow of information in the network. A gate is a structure in a neural network that controls the flow of information. LSTM has three gates: input gate, forget gate and output gate. Input gates control the entry of new information into the network. Forget gates control the deletion of old information from the network. The output gate controls the information output by the network. The structural principle of the LSTM machine learning model is shown in Figure 8.

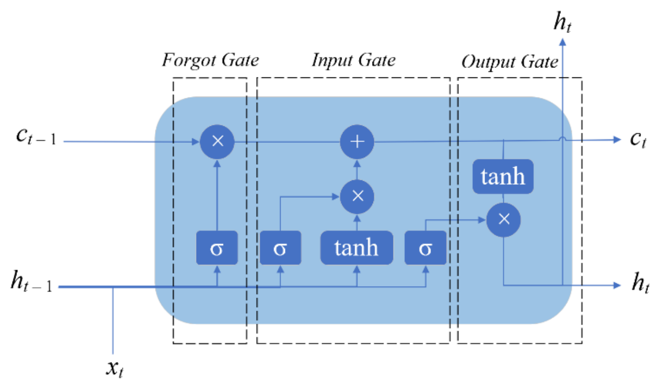

Figure 8. LSTM machine learning model structure principle

LSTMs process sequence data through the following steps:

1) The input information is fed into the LSTM network.

2) Input gates control the entry of new information into the network.

3) Forget gates control the deletion of old information from the network.

4) The output gate controls the information output by the network.

5) *LSTM can learn parameters through the backpropagation algorithm.* The backpropagation

algorithm updates the parameters of the network by calculating the error of the network.

LSTM has been widely used in natural language processing, machine translation, speech recognition, time series prediction and other fields. LSTM is a very powerful neural network model that has achieved excellent performance in many tasks.

Here are some advantages of LSTMs:

- Can handle long sequence data
- Can learn long-term dependencies
- Achieved excellent performance in many tasks
- Here are some disadvantages of LSTMs:

Many parameters require a large amount of data for training, easy to overfit, Overall, LSTM is a very powerful neural network model that achieves excellent performance in many tasks.

The data basis formula adopted by the LSTM algorithm is as follows:

$$f_t = \sigma\left(U_f h_{t-1} + W_f x_t + b_f\right) \tag{5}$$

$$i_t = \sigma\left(U_i h_{t-1} + W_i x_t + b_i\right) \tag{6}$$

$$\acute{c}_t = tanh\ \left(U_c h_{t-1} + W_c x_t + b_c\right) \tag{7}$$

$$c_t = f_t c_{t-1} + i_t \acute{c}_t \tag{8}$$

$$o_t = \sigma\left(U_o h_{t-1} + W_o x_t + b_o\right) \tag{9}$$

$$h_t = o_t tanh\left(c_t\right) \tag{10}$$

In Figure 8, the cell state ct – 1 and hidden state ht – 1 at the previous time step is tuned by the forget gate switch to generate the corresponding states ct and ht at the current time step, where the forget gate vector ft is expressed in Formula 5.

In Formula 5, σ denotes the sigmoid function layer which determines the parameters to be updated, U and W denote the weight matrices to be applied to the hidden state vector ht – 1 at the previous time step and input vector xt at the current time step, respectively, and b denotes the bias vector to reduce inaccuracy during training.

In order to update the cell state, the input gate utilizes the sigmoid function σ and tanh function to process the cell state vector ct – 1 at the previous time step to generate the cell state vector ct at the current time step. The working expression of the sigmoid function σ and tanh function are illustrated in Formula 6 and Formula 7, respectively.

Then, the cell state ct – 1 at the previous time step is processed by forget gate, and the product of tanh layer output and sigmoid layer output are added together to update to the cell state ct at current time step, as illustrated in Formula 8.

Finally, after being processed by the same sigmoid function and tanh function in the output gate to determine the output parameters and the weighted value in Formula 9, the hidden state ht at the current time step is output by an LSTM cell to transfer to the next one or use as the final output in Formula 10.

D. Using LSTM to predict oil well pump pressure experiment in fault diagnosis

This system builds a data set based on a large amount of historical data such as temperature, pressure, and valve opening monitored online by all sensors of the oil well oil pump. Through the LSTM machine learning neural network prediction model and effective training, the future of the oil well oil pump can be obtained more accurately. The occurrence temperature, pressure, valve opening trend curve and the probability of failure data, once the abnormality is predicted by the fault diagnosis machine learning reasoning mechanism, an alarm will be issued immediately.

In the process of system experiment, the construction of back-end server is a crucial step.

The back-end server is responsible for receiving and processing front-end requests, performing data processing, implementing business logic, and returning the results to the front-end. In this design, we need to build a reliable, efficient, and secure back-end server to support various functions and requirements of the system. We chose Python as the back-end development language, combined with the development of the Flask framework. Python is a concise and powerful programming language with rich library and framework support, suitable for rapid development and deployment. Flask is a lightweight web framework with flexible routing management and template rendering functions, easy to learn and use.

LSTM normalization refers to normalizing the input and output in the LSTM network to improve the training speed and performance of the network. Commonly used normalization methods include, but are not limited to: Batch Normalization, Layer Normalization, and Instance Normalization, etc.

Among them, Batch Normalization is a commonly used normalization method, which can normalize the input on each mini-batch to reduce the internal covariate displacement, thereby improving the training speed and performance of the network. Layer Normalization is another commonly used normalization method, which can normalize the input on each layer to reduce the internal covariate displacement, thereby improving the training speed and performance of the network. Instance Normalization is a normalization method for image processing tasks, which can normalize the input on each channel to reduce the internal covariate displacement, thereby improving the training speed and performance of the network.

The following takes oil well oil pump pressure as an example to carry out fault diagnosis and prediction. The algorithm steps are as follows:

Step 1: Obtain the historical data of oil well pump pressure to form a training data set;

Step 2: Normalize the data in the dataset;

Step 3: Divide the training set into the training set for machine learning LSTM model training to obtain model parameters;

Step 4: The verification set divided into the training set is verified by the machine learning LSTM model to obtain error data;

Step 5: The machine learning LSTM model is used to predict the pressure of the oil pump for a period of time in the future. If it exceeds the range, it will warn that there will be a failure in the future;

The algorithm formula used for normalization in this system is as follows:

$$\mathrm{x}_{\mathrm{standard}} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{11}$$

$$x_{scaled} = x_{standard} * \left(max ? min\right) + min \tag{12}$$

In Formula 11 and Formula 12, max represents the maximum value, and min represents the minimum value.

In the LSTM prediction experiment of the oil well pump pressure in this system, the system uses the normalized root mean square error (n-RMSE) to characterize the pressure prediction performance of the oil well pump for different models given different lengths of input windows and prediction windows. In order to obtain the indicator data n-RMSE in Python, it is necessary to sum the deviation squares between the predicted value of the oil well pump pressure and the actual value of the oil well pump pressure, and then divide the sum by the size of the oil well pump pressure test sample to calculate the MSE. Then, the indicator RMSE of oil well delivery pump pressure machine learning can be calculated by taking the root of MSE. Finally, the metric nRMSE for machine learning of oil well pump pressure is calculated by dividing the RMSE by the mean value of the real dataset. Usually, a lower index nRMSE of oil well oil pump pressure machine learning will indicate that the model prediction performance is better, and at this time it can be used for real-time diagnosis of oil well oil pump pressure faults.

The main technical index parameters of the LSTM machine learning algorithm for oil well pump pressure failure are shown in formula 13, formula 14 and formula 15:

MSE (Mean Squared Error):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2 \qquad (13)$$

RMSE (Root Mean Squared Error):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2} \qquad (14)$$

n-RMSE (Normalized Root Mean Squared Error):

$$nRMSE = \frac{RMSE}{\bar{y}} \qquad (15)$$

This experiment is based on the Python development environment. The oil well pump pressure sensor recorded in the project is used as the data set, and the data is cleaned, and then normalized. There are more than 3,000 pieces of data in this dataset, and the first 20 pieces of data are shown in Table 1.

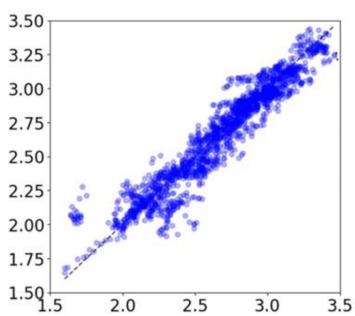

Figure 9. Fault diagnosis and prediction scatter diagram of oil well pump oil delivery pressure training set

Then, the LSTM machine learning model is used for training to obtain the effective parameters of the model, and finally the prediction is made. The error of the training collection and distribution point is shown in Figure 9, the error of the testing collection and distribution point is shown in Figure 10, and the actual curve and predicted curve of the oil well pump are shown in Figure 11.

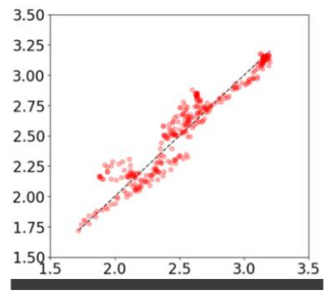

Figure 10. Oil well pump oil delivery pressure test set fault diagnosis and prediction scatter diagram

TABLE I. TOP 20 RECORDS IN THE DATA SET USED BY OIL WELL PUMP MACHINE LEARNING

Code	Date	Time	Oil Pressure (Mpa)
1	6/1	0:00	1.69
2	6/1	1:00	1.81
3	6/1	2.00	2
4	6/1	300	1.97
5	6/1	4:00	1.84
6	6/1	5:00	2.01
7	6/1	6:00	1.62
8	6/1	7:00	1.89
9	6/1	8:00	2.14
10	6/1	9.00	2
11	6/1	10:00	2.06
12	6/1	1100	1.8
13	6/1	12:00	1.65
14	6/1	13:00	1.67
15	6/1	14:00	2.11
16	6/1	15.00	1.88
17	6/1	16:00	2.12
18	6/1	17:00	1.54
19	6/1	18:00	2.08
20	6/1	19:00	2.02

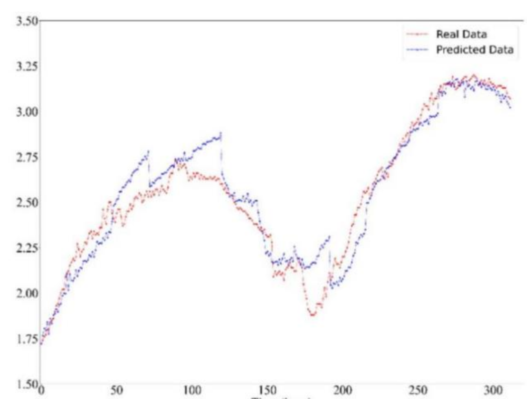

Figure 11. Comparison chart of machine learning actual value and predicted value of oil delivery pressure of oil well delivery pump

It can be seen from Figure 11 that the prediction accuracy of the LSMT machine learning algorithm is very high, which fully meets the needs of the oil well oil delivery pump for late pressure over-limit fault diagnosis and alarm.

According to the aforementioned evaluation indicators, the data obtained by machine learning operations are shown in Table 2.

TABLE II. EVALUATION INDEX DATA OBTAINED BY MACHINE LEARNING OPERATIONS

Type	Train	Test
MAE	1.91	5.59
MSE	9.33	11.35
RMSE	5.42	7.41
nRMSE (%)	4.03	8.85

It can be seen from Table 2 that the machine learning model proposed in this paper has a very small error in the fault diagnosis and prediction of oil well pumps.

In the actual use of the system, when it is predicted that the pressure of the oil pump in the oil well will exceed the safety limit in the future, the oil pump will be diagnosed in advance and the engineer will be reminded to deal with it on site, so as to truly prevent problems before they happen.

VI. CONCLUSION

The author used the theoretical framework of the Internet of Things system proposed in this paper to successfully develop and experiment in the remote fault prediction and monitoring project of oil well digital oilfield of a subsidiary oilfield enterprise of PetroChina. The system can switch and control the output of four well groups distributed in different regions oil pump. The system's IoT central computer software is developed using the object-oriented integrated programming tool Embarcadero RAD Studio XE Community Edition, the database system uses SQL Server 2008, and the remote desktop and mobile terminal use Visual Studio 2020 to develop a B/S architecture Web system. Considering the safety of the special equipment controlled by the system, the mobile terminal can only browse the operating parameters and data of each oil pump, and cannot remotely control it. The system has been in trial operation for more than half a year. Compared with the early traditional manual inspection and management of oil pumps, the system is easy to operate, has a high degree of intelligence, and accurate and timely diagnosis and early warning of equipment operation faults. Strong guarantee. The experiment proves that the LSTM machine learning model has better prediction accuracy of oil well pressure failure.

The follow-up research recommendations of this project are as follows:

1) It is necessary to increase the complexity of the system, investigate the sensor data format and characteristics of most oil well pumps in China, connect the sensor data to the system in a standardized manner, expand and improve the monitoring software, in order to realize wide and comprehensive monitoring Comprehensive monitoring of the faults of oil well oil pumps, and the accumulation of data sets of many oil well oil pump sensors to provide a sufficient amount of data sets for machine learning.

2) The background of the system should use a variety of machine learning models other than LSTM to train and predict the oil well pump data set, and obtain the accuracy values predicted by different machine learning models, compare these values, and select the appropriate machine the learning model is finally used in this system.

ACKNOWLEDGMENT

The Research is supported by the college student innovation training project. (Financing projects No: S202210702082).

REFERENCES

[1] Ma Hongjun. The research and application of key technologies and applications of offline monitoring and diagnosis of long-voltage pipeline oil pumps [J]. Water pump technology, 2023 (03): 29-35.

[2] Xu Jia, Hu Jiancun, Qin Ciwei, etc. Optimized high-pressure oil pump diagnosis based on parameter optimization VMD and distributed entropy [J]. Internal combustion engine Journal, 2023, 41 (02): 166-174.

[3]]Zhang Zongke, Dai Longxing. The application of the oil gas pump in the oil and gas collection system [J]. Comprehensive corrosion control, 2022, 36 (09): 34-35.

[4] Li Yaping, Li Sujie, Ma Bo, etc. A method of failure diagnosis and health assessment of the oil pump unit [J].

[5] Teacher Chaoguang. Based on the labView oil pump unit running noise noise, three -dimensional boundary sound field simulation research [D]. Northeast Petroleum University, 2022.

[6] Liu Zhaoting. Study on the failure diagnosis method of the oil pump unit based on running noise analysis [D]. Northeast Petroleum University, 2022.

[7] Zhang Xing, Zhang Yikun, Lin Mingchun, etc. The diagnosis and treatment of the resonance failure of the oil pump motor [J]. Chinese petroleum and chemical standards and quality, 2021, 41 (17): 47-49.

[8] Confucius. Oil pump unit fluid power noise noise acoustic characteristics and their detection methods Study [D]. Northeast Petroleum University, 2021.

[9] Fang Zhen. The design and realization of the chemical pump fault diagnosis system [d]. Anhui University, 2021.

[10] Wei Lixin, Chu Yongqiang, Liu Fang, etc. Oilfield station oil pumps fuzzy security evaluation [J]. Flow machinery, 2021, 49 (03): 68-73.

[11] [11] Xu Jia. Under the condition of complex working conditions, the diagnosis of fault diagnosis of high -pressure oil pumps for ships for ships under complex working conditions [D]. Nanjing University of Aeronautics and Astronautics, 2021.

[12] Xu Kai, Fushimi, Chen Shijun equivalent. Steaming generator status monitoring system based on LSTM neural network [J]. Application technology, 2020, 47 (06): 96-100.

[13] Wang Yizheng, Wang Yanwu, Gao Hongbin. Diagnosis of ship fuel pump-based fuel pump-based monitoring [J]. Mechanical and Electrical Engineering Technology, 2021, 50 (01): 163-165.

[14] Li Baisong, Su Jianfeng, Zhang Xing, etc. The diagnosis method of the oil pump inlet and export pipes [J]. Oil and gas storage and transportation, 2021, 40 (01): 21-25.

[15] Ning Shuangsheng. Pipeline pumps safety risk evaluation based on fuzzy FMECA [J]. Safety, 2020, 41 (08): 13-19.

[16] Xu Jing. The diagnosis of aircraft fuel pumps based on wavelets and migration learning [J]. Hydraulic and pneumatic, 2020 (06): 183-188.

[17] Guo Rui. ANSYS -based oil pump unit electromagnetic noise production mechanism and simulation research [D]. Northeast Petroleum University, 2020.

[18] Zhang Xinbo. Inquire about the vibration influencing factors and response strategies during the operation of the oil pump [J]. Chemical management, 2019 (33): 120-121.

[19] Zhang Dawei. The performance test exploration of the oil pump [J]. Technology Information, 2019, 17 (30): 70+72.

[20] Qin Jianan, Li Wenjin, Cheng Zhenxin, etc. The oil pump unit predicts the establishment and research of the maintenance decision model [J]. Equipment management and maintenance, 2019 (02): 182-184.

[21] Liu Zhaoquan. Oil pumps typical failure analysis and maintenance [J]. Chemical management, 2019 (03): 40-41.

[22] Xiao Bo ship. Explore the application of oil pump failure diagnosis technology in the refined oil pipeline system [J]. Contemporary chemical research, 2018 (10): 59-60.

[23] Zhao Jianghui. The application of status monitoring and fault diagnosis system in the oil pump unit [J]. Petrochemical technology, 2018,25 (08): 259.

[24] Baoxia L, Chenchen X, GANG L, et al. Application of Frequency Domain Analysis Method in Vibration Analysis and Fault Diagnosis of Oil Transfer Pump UNIT [J]. Journal of Physics: Conference Series, 2023, 2437 (1).

[25] liangliang d, qian x, yanjie j, et al. Review of reSearch on intelligent diagnosis of our transfer pump malfunction [J]. Petroleum, 2023,9 (2).

[26] Yao Sen, Dai Xingzheng, Cai Liang, etc. The research on the risk evaluation technology of the oil pump unit based on multi-level fuzzy analysis method [J]. Comprehensive corrosion control, 2021,35 (10): 53-56.

[27] He Min, Wang Xingru. Analysis and countermeasures for the causes of failure pump pump pump pumps in the long passing pipeline [J]. Equipment management and maintenance, 2021 (10): 55-57.

[28] Lin Yujun, Liu Yong, Zhao Shengsheng. Oil pump online monitoring mobile app design and development [J]. Science and education guidance (early journal), 2020 (28): 43-45.

[29] Ou Yonghong, Xu Lidong, Wang Lihong, etc.. Analysis and countermeasure study of the oil pipe oil pump failure pump [J]. Chemical management, 2020 (01): 181.

[30] Qiao Shijin, Qiao Linan. Analysis of the oil pump failure pump in the aircraft supply system [C] // Aviation Industry measurement and control technology development center, China Aviation Society testing technology branch, status monitoring special sensing technology Aviation Technology Key Laboratory. Tenth tenth tenth Sixth China Aviation Control and Control Technology Annual Conference Collection. [Unknown publishers], 2019: 455-456+461.

[31] Xu Guoping. Oil pipeline data collection and monitoring control (SCADA) system oil pump failure analysis of failure pumps [J]. Petroleum and gas stations, 2018,27 (05): 16-19+6.

[32] Analysis of Hou Nana. Oil pump failure diagnosis mode [J]. Chemical design communication, 2018,44 (04): 91+106

[33] Wu Peng. Analysis and countermeasures of the cause of the failure of the oil pipe oil pump pump [J]. Chemical management, 2018 (22): 128.

[34] liangliang d, qian x, yanjie j, et al. Review of reSearch on intelligent diagnosis of our transfer pump malfunction [j]. Petroleum, 2023,9 (2).

[35] Li Xiaoqing. Aquinus neural network and application in fault diagnosis [J]. Digital technology and application, 2022,40 (05): 150-152.

[36] Wang Lei. Diagnosis and health prediction of ore drainage pumps based on multi -source perception [D]. Anhui University of Science and Technology, 2022.

[37] Chen Yuhao. Study and application of typical industrial equipment health status prediction methods [D]. Chongqing University, 2022.

[38] Luo Shenghua. Research on the diagnosis and health state prediction method of hydraulic pump [D]. Jiangsu University of Science and Technology, 2022.

[39] Yin Jingtian. Filling in the diagnosis and health prediction of bearing fault diagnosis and health of the packaging system [D]. Jiangsu University, 2021.

[40] Shan Yahui. Research on the trend of vibration failure and health performance trends of hydropower units [D]. Huazhong University of Science and Technology, 2021.

[41] Ye Chunwei. Research on Early Equipment Diagnosis and Maintenance Strategy [D]. Nanjing University, 2021.

[42] Guli, Zou Yongsheng, Li Kaihong, etc. The ARMA model in the ARMA model in the research on the ectopic value trend of the oil pump vibration [J]. Fluid machinery, 2021, 49 (01): 22-28.

[43] Xu Kai. Research on the monitoring system based on the status characteristics [D]. Harbin Engineering University, 2020.

[44] Liu Jialong. Study on the diagnosis and health prediction of high -speed train steering racks based on data -driven high -speed trains [D]. Northeast University, 2020.

[45] Qiao Ningguo. High -speed train transmission systems based on multi -sensor data fusion prediction and health prediction [D]. Jilin University, 2019.