Article citation info:
Ma D. Predictive maintenance technology for industrial production equipment using cloud platform. Diagnostyka. 2025;26(2):2025211. https://doi.org/10.29354/diag/205035.

DIAGNOSTYKA, 2025, Vol. 26, No. 2

e-ISSN 2449-5220
DOI: 10.29354/diag/205035

PREDICTIVE MAINTENANCE TECHNOLOGY FOR INDUSTRIAL PRODUCTION EQUIPMENT USING CLOUD PLATFORM

Dongbao MA *
College of Mechanical and Electrical Engineering, Beijing Polytechnic, Beijing 100176, China
* Corresponding author, e-mail: bpimadongbao@126.com

Abstract

To enhance the effectiveness of predictive maintenance for industrial production equipment, this work explores a cloud platform-based predictive maintenance system. Moreover, it designs an equipment fault diagnosis model using the One Dimensional Deep Residual Shrinkage Network (1DDRSN). The performance of the 1DDRSN-based equipment fault diagnosis model and the proposed predictive maintenance system is validated through bearing fault detection experiments. The results demonstrate that the 1DDRSN model significantly outperforms other models in equipment fault diagnosis, achieving an accuracy, precision, recall, and F1 score of 0.9886, 0.9796, 0.9684, and 0.974, respectively. Compared to other models, these metrics represent improvements of at least 0.66%, 0.56%, 0.76%, and 0.62%, respectively. This indicates that the 1DDRSN model offers higher robustness and better predictive performance for complex industrial equipment fault diagnosis tasks. Additionally, performance testing of the cloud platform-based predictive maintenance system demonstrates superior response time, system throughput, and data processing efficiency compared to traditional systems. This suggests the proposed system's ability to better support real-time maintenance needs in complex industrial environments. The findings of this work provide technical support for intelligent maintenance in industrial production and lay the foundation for future developments in the field of smart manufacturing.

Keywords: predictive maintenance; equipment fault diagnosis; deep learning; cloud platform; deep residual shrinkage network

1. INTRODUCTION

As industrialization continues to advance, the role of equipment in production and manufacturing becomes increasingly important. Equipment failures not only cause production stoppages and increase maintenance costs but can also lead to more serious safety accidents. Therefore, effectively predicting equipment failures and ensuring timely maintenance has become a critical issue in industrial production. Traditional maintenance models often rely on experience-based repairs or regular inspections. This passive maintenance approach not only fails to detect potential faults in a timely manner but also tends to increase maintenance costs and production downtime [1, 2]. As a result, predictive maintenance has emerged as a solution. Predictive maintenance leverages technologies such as sensor data, machine learning, and deep learning to predict equipment failures and take preventive actions in advance. It has become a key technology for improving production efficiency and equipment reliability in modern industry [3].

With the development of cloud computing, cloud platforms now offer powerful computational capabilities and flexible resource scheduling, making equipment fault diagnosis and prediction both possible and more efficient [4]. How to combine the cloud platform's strong computational power with equipment fault diagnosis models to optimize performance, improve prediction accuracy, and enhance system efficiency remains a hot research topic and challenge. Meanwhile, various deep learning-based models have been widely applied in fault detection and diagnosis tasks. These models can automatically extract features from vast amounts of sensor data, significantly improving the ability to identify and predict equipment failures. However, existing models still face challenges related to training efficiency, robustness, and adaptability in complex industrial environments [5, 6].

Based on this background, this work aims to study predictive maintenance technology for industrial production equipment supported by cloud platforms, focusing on the effectiveness of using deep learning models for equipment fault prediction. Specifically, this work constructs a cloud platform-based predictive maintenance system and designs and implements an equipment fault diagnosis model

Received 2025-01-13; Accepted 2025-05-12; Available online 2025-05-15

on the basis of a Deep Residual Shrinkage Network. Finally, the model's performance is validated through experiments. This work provides technical support for intelligent maintenance in industrial production, significantly reduces maintenance costs, improves production efficiency, and lays the foundation for the widespread utilization of intelligent manufacturing in the future. The main innovations of this work are as follows:

(1) A cloud-based predictive maintenance system for equipment is developed. This system fully leverages the efficient data processing capabilities of cloud computing to achieve real-time monitoring and prediction of equipment operating conditions, enhancing system reliability and scalability.

(2) A one-dimensional deep residual shrinkage network (1DDRSN) model for equipment fault diagnosis is proposed. This model integrates residual learning and a soft-thresholding shrinkage mechanism. It enhances feature extraction for complex industrial fault signals through adaptive denoising, and improves fault detection accuracy and robustness.

(3) Experimental and comparative studies validate the effectiveness of the proposed approach. The results indicate that the 1DDRSN model outperforms existing methods in fault diagnosis, while the cloud-based predictive maintenance system surpasses traditional systems across multiple performance metrics, providing support for practical industrial applications.

This work presents a novel technological solution for intelligent maintenance in industrial production. It contributes to reduced maintenance costs, improved equipment utilization, and laying the foundation for the deep integration of smart manufacturing and the Industrial Internet of Things (IoT).

2. RELATED WORK

2.1. Research status of predictive maintenance and fault diagnosis for equipment

With the advancement of the Industrial IoT, predictive maintenance systems that integrate cloud computing, big data analytics, and artificial intelligence (AI) have become a key direction for enhancing the intelligence of equipment management. The combination of the powerful computational capabilities of cloud platforms with the real-time data acquisition capabilities of IoT sensors offers new solutions for equipment fault diagnosis and maintenance. Maurya et al. (2024) explored the importance of fault diagnosis in rotating machinery, and emphasized how the integration of IoT, cloud computing, and AI technologies has driven the rapid development of intelligent fault diagnosis and condition monitoring systems. These technologies effectively enable predictive maintenance for mechanical equipment [7]. Similarly, Kannammal et al. (2023) proposed a predictive maintenance system for pumpjacks by integrating IoT and deep learning. By monitoring telemetry data from pumpjacks, the system detects anomalies in real time, prevents faults, and reduces downtime [8]. Sharma and Gurung (2024) provided practical guidance on collecting and integrating industrial asset data, applying machine learning algorithms, and deploying predictive maintenance systems. They highlighted that real-time sensor monitoring combined with deep learning technologies optimized maintenance processes, reduced unplanned downtime and costs, and enhanced overall equipment efficiency [9].

Complementing research on intelligent maintenance, significant progress has also been made in developing equipment fault diagnosis models in recent years. For instance, Raparthy et al. (2023) combined time-series analysis and deep learning to propose a predictive maintenance framework for fault prediction for IoT devices. By leveraging sensor data, fault logs, and maintenance records, the framework employed recurrent neural network (RNN) and long short-term memory (LSTM) networks for predictions. Experimental results demonstrated its superior accuracy and efficiency [10]. Qureshi et al. (2024) reviewed the application of machine learning in solar power plants, particularly its role in improving equipment reliability, reducing maintenance costs, and maximizing energy production. They summarized the application of machine learning algorithms such as logistic regression, decision trees, and support vector machines (SVM), in predictive maintenance, and emphasized the importance of model interpretability and scalability [11]. Ghasemkhani et al. (2023) introduced a novel interpretable machine learning method, "Balanced K-Star," to address data imbalance issues in manufacturing equipment predictive maintenance. The experimental results showed that this method outperformed standard approaches in classification accuracy, achieving 98.75%, significantly higher than the current state-of-the-art method, which had an accuracy of 91.74% [12].

Despite these notable advancements in predictive maintenance and fault diagnosis, certain challenges remain. First, most existing methods rely on traditional machine learning models and lack deep modeling of complex equipment fault patterns. Second, while the application of cloud platforms has gained attention, leveraging the full potential of cloud computing resources to enhance real-time responsiveness in large-scale equipment environments remains an area requiring further research. To address these challenges, this work proposes a cloud-based predictive maintenance system that integrates the 1DDRSN for fault diagnosis. The system aims to improve maintenance efficiency and diagnostic accuracy, addressing current limitations in the prediction of faults in large-scale and complex equipment.

2.2. Classification and comparative analysis of predictive maintenance technologies

In recent years, predictive maintenance technologies have been extensively studied and applied in industrial production. Based on different technical approaches, existing predictive maintenance technologies can be broadly categorized into the following types: rule-based maintenance, statistical model-driven maintenance, machine learning-based maintenance, and deep learning-based maintenance. Each approach has its applicable scenarios and limitations. Fig. 1 illustrates the specific classification structure.

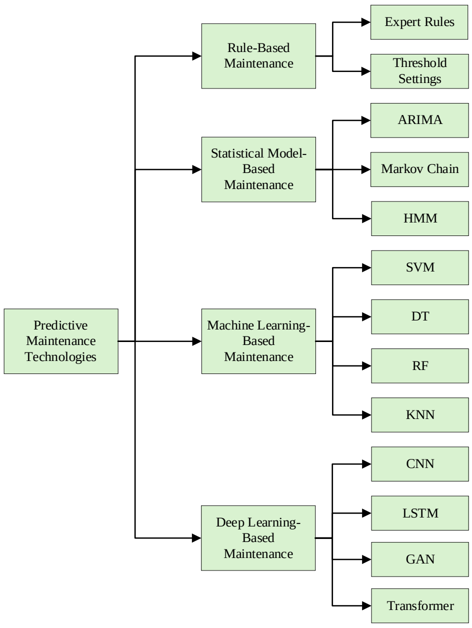

Fig. 1. Classification of predictive maintenance technologies

(1) Rule-based approaches primarily rely on expert knowledge, heuristic rules, or simple threshold settings to determine maintenance needs by analyzing equipment operating data such as temperature, vibration, and pressure. While easy to implement, these methods are prone to misjudgments or omissions when dealing with complex and dynamically changing equipment failures.

(2) Statistical models establish relationships between equipment operating conditions and failures using probabilistic statistics, time series analysis, and regression models. Examples include the AutoRegressive Integrated Moving Average (ARIMA) model, Markov chains, and the Hidden Markov Model (HMM). These methods are suitable for equipment data with stable patterns but may have limited predictive capability when dealing with complex conditions or external environmental influences.

(3) Machine learning approaches leverage data-driven learning capabilities to achieve more precise predictive maintenance through feature extraction and model training. Techniques such as SVM, Random Forest (RF), K-Nearest Neighbors (KNN), and Decision Trees (DT) have shown promising results in fault diagnosis and prediction. However, their performance depends on the quality of feature engineering, and computational overhead can be high when handling large-scale data.

(4) In recent years, deep learning has made significant advancements in predictive maintenance. Models such as the Convolutional Neural Network (CNN), LSTM, Generative Adversarial Network (GAN), and Transformer architectures can automatically learn data features, making them particularly suitable for multimodal data analysis of complex equipment. However, conventional deep learning models still face challenges such as high noise levels and data imbalance. To address these limitations, this work proposes a fault diagnosis model based on the 1DDRSN. This model incorporates residual learning, soft-thresholding shrinkage, and attention mechanisms to enhance robustness and accuracy in fault detection.

(5) Table 1 presents a comparative analysis of different predictive maintenance methods.

3. DESIGN OF AN EQUIPMENT FAULT DIAGNOSIS AND PREDICTIVE MAINTE-NANCE SYSTEM BASED ON A CLOUD PLATFORM AND DEEP LEARNING

3.1. Cloud platform-based predictive maintenance system for equipment

In the context of Industry 4.0, cloud platform-based predictive maintenance systems provide the manufacturing industry with a new approach to maintenance strategies. This system is primarily composed of five core modules: terminal devices, cloud computing services, equipment fault prediction, equipment remaining life prediction, and intelligent application services. These modules work in synergy to enable real-time monitoring and intelligent maintenance of industrial equipment, as illustrated in Fig. 2.

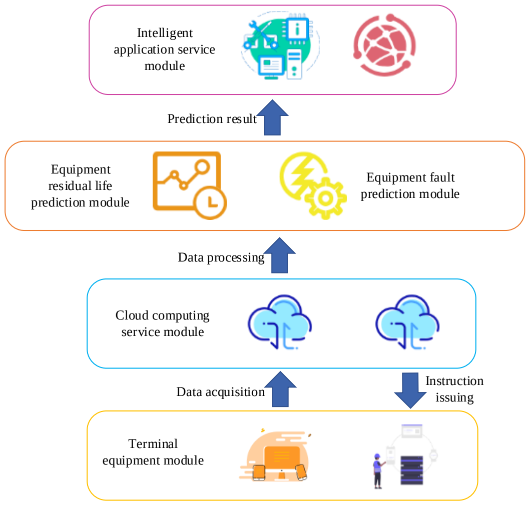

Fig. 2. The structure of a cloud platform-based predictive maintenance system for industrial equipment

Table 1. Comparison of different predictive maintenance methods

Method Category	Representative Methods	Applicable Scenarios	Advantages	Limitations
Rule-Based Maintenance	Heuristic rules, threshold setting	Simple equipment, traditional industrial systems	Easy to implement, low computational cost	Relies on human experience, poor generalization ability
Statistical Model-Driven Maintenance	ARIMA, HMM, Markov Chains	Linear systems, time series forecasting	Strong mathematical interpretability, can model certain fault patterns	Struggles with nonlinear, multivariate problems
Machine Learning	SVM, RF, KNN, DT	Complex equipment data analysis	Can automatically learn data features, strong predictive capability	Dependent on manual feature engineering, high computational cost
Deep Learning	CNN, LSTM, GAN, Transformer	Multimodal complex fault data	Suitable for large-scale data, can automatically extract features	High computational resource demand, some methods difficult to interpret
Proposed Method	1DDRSN	High noise, complex fault patterns	High robustness, strong feature extraction ability, automatic denoising	Needs optimization for computational efficiency and generalization capability

In the system depicted in Fig. 2, terminal devices form the foundational data acquisition layer, responsible for real-time monitoring of industrial equipment's operational status. Critical operational parameters, such as vibration, temperature, current, and rotational speed of devices are collected using sensors. To ensure data accuracy and completeness, terminal devices typically include high-precision sensors, data collectors, and embedded systems. These devices are capable of performing preliminary data processing and filtering locally, and they can also transmit the data to the cloud in real-time through IoT technology.

The cloud computing service acts as the core processing layer of the entire system, handling the reception and storage of massive amounts of data from terminal devices. Using the strong computational capabilities of cloud computing, this service can quickly process and analyze large volumes of equipment data to recognize potential fault patterns and trends [13]. This module consists of sub-modules for data storage, data cleaning, feature extraction, and data analysis. By employing distributed computing technology, the cloud computing service maintains efficiency during large-scale data processing and uses big data analysis techniques to uncover hidden patterns in the equipment's operational data. This provides the necessary support for subsequent fault prediction and remaining useful life estimation. Additionally, the elastic scalability of the cloud allows the system to dynamically adjust resource allocation in response to changes in data volume and computational demand, ensuring efficient operation.

The equipment fault prediction module utilizes the processed data from the cloud computing service to predict the operational status of equipment using advanced deep learning models. This module integrates deep learning algorithms capable of capturing complex temporal and spatial features in the equipment data. Through model training, it can anticipate potential equipment failures and provide corresponding fault categories and estimated time of occurrence [14]. This module not only delivers highly accurate predictions but also dynamically updates the prediction models to adapt to changes in the operating environment and conditions of the equipment.

The equipment's remaining life prediction module aims to estimate the remaining useful life of equipment under current operating conditions. Relying on data from the cloud computing service, this module employs deep learning models, combining historical operational data with real-time monitoring data, to predict trends in the equipment's health status [15]. The equipment remaining life prediction module assists enterprises in planning maintenance and replacement schedules more effectively, maximizing equipment utilization, and extending its service life [16, 17].

Finally, the intelligent application services module functions as the decision support layer of the system. By integrating the results of the fault prediction and remaining life prediction, it provides enterprises with intelligent maintenance recommendations and decision support. This module can generate detailed equipment health reports, highlight potential risks, and suggest optimal maintenance strategies. Additionally, it includes sub-modules for alarm management, maintenance record management, and decision support, offering comprehensive support for equipment management. Seamlessly integrated with the enterprise's existing production management systems, this module ensures effective coordination between equipment maintenance and production planning, thereby enhancing overall productivity.

Through the collaborative operation of these modules, the cloud platform-based predictive maintenance system enables comprehensive lifecycle management of industrial equipment, offering enterprises a reliable and efficient maintenance solution. This system not only significantly reduces equipment failure rates and maintenance costs but also boosts production efficiency, helping enterprises maintain a competitive edge.

3.2. Deep learning-based equipment fault diagnosis model

In the cloud-based predictive maintenance system, fault diagnosis plays a crucial role in enabling predictive maintenance. With the purpose of enhancing the accuracy and robustness of fault diagnosis, this work proposes a model that combines residual networks, soft thresholding, and attention mechanisms-namely, the One Dimensional Deep Residual Shrinkage Network (1DDRSN) model [18, 19]. This model is designed to effectively extract and compress key features from industrial equipment operational data, thereby improving the ability to identify equipment faults.

Noise interference is a common issue in equipment fault diagnosis, so the model needs to minimize the impact of noise while maximizing the use of valid information. To address this challenge, an attention mechanism is incorporated into the deep residual shrinkage network. The attention mechanism, through a small subnetwork, automatically learns and assigns a set of weights to the feature maps, weighting each channel. This process enhances the useful feature channels and suppresses redundant ones [20]. By doing so, the model focuses on critical information, improving the detection capability of equipment faults.

The deep residual shrinkage network is an improved model based on the Squeeze-and-Excitation Network (SENet). It assigns appropriate weights to each feature channel through a weighting mechanism, where these weights are derived using soft thresholding. During the training process for each sample, the model learns adaptive thresholds, thereby enhancing training efficiency. The deep residual network introduces residual modules, effectively mitigating the gradient vanishing problem in deep learning, which improves the stability and performance of the model [21, 22]. In deep learning models like the CNN, the backpropagation algorithm often encounters gradient vanishing or explosion issues, making training difficult. The residual network addresses this issue by introducing skip connections, significantly improving model training. The process within the deep residual shrinkage network includes convolution operations, compression operations, and excitation operations, specifically:

Convolution Operation F_{con} : $U = [u_1, u_2, \cdots, u_c]$ represents the set of learned filter kernels, where u_c denotes the parameters of the c-th filter. The output feature map after the convolution operation is $Z = [z_1, z_2, \cdots, z_c]$.

$$A \to Z, A \in \mathbb{R}^{W' * H' * C'}, z \in \mathbb{R}^{W*H*C} \quad (1)$$

$$z_c = u_c \circ A = \sum_{s=1}^{c'} u_c^s \circ a^s \quad (2)$$

A represents the input feature map, $\circ$ denotes the convolution operation, and u_c^s is a convolution kernel with s channels. W and H represent the width and height of the input feature map, respectively, while C represents the number of channels in the input feature map. W' and H' represent the width and height of the output feature map after the convolution operation, which is typically calculated based on the kernel size, stride, and padding. C' represents the number of channels in the output feature map after convolution, which is usually equal to the number of filters.

Compression operation F_{te}: By applying global average pooling, the c channels are ultimately reduced to a real number sequence of size 1*1*c. The c-th channel of the output feature map q after the compression operation is calculated as follows:

$$q_c = F_{te}(z_c) = \frac{1}{H*W} \sum_{i=1}^{H} \sum_{j=1}^{W} z_c(i,j), q \in \mathbb{R}^C \quad (3)$$

H and W refer to the height and width of the input feature map, respectively.

Excitation operation F_{en} : By using gating to reduce the number of channels, the computational load is decreased, allowing the model to learn the nonlinear relationships between different channels.

$$s = F_{en}(q, w) = \sigma\big(g(q, w)\big) = \sigma(w_2(w_1 q)) \quad (4)$$

w is the weight matrix, while w_1 and w_2 are the weight matrices used for the excitation operation. σ refers to the sigmoid activation function, and $g(q, w)$ is the gating function used to calculate the weights.

Weighted features F_s : The sigmoid activation function is adopted to process the input features Z to obtain activation values that range from 0 to 1. Subsequently, the activation value for each channel is multiplied by the original features Z, resulting in the weighted features for each channel.

$$\tilde{a}_C = F_s(z_c, s_c) = s_c \cdot z_c \quad (5)$$

$\tilde{a}_C$ represents the weighted features, s_c is the excitation value for the c-th channel, and z_c is the feature map for the c-th channel.

Additionally, in the residual network, the threshold is obtained through the embedded small network. The nonlinear processing of signals is also crucial for improving model performance. To further denoise the input bearing vibration signals, soft thresholding is introduced and embedded into the improved residual block [23]. Currently, soft thresholding serves as a fundamental step in numerous denoising algorithms, as it can eliminate features with absolute values below the threshold and shrink features with absolute values above the threshold down to zero. Soft thresholding is a commonly used nonlinear processing method aimed at reducing noise interference by appropriately compressing and thresholding the signal, thereby enhancing signal quality. Specifically, the soft thresholding function is defined as:

$$y = \begin{cases} x - \varphi, x > \varphi \\ 0, -\varphi \le x \le \varphi \\ x + \varphi, x < -\varphi \end{cases} \quad (6)$$

x and y represent the input and output, respectively, while φ is the threshold. The threshold setting must satisfy two conditions: first, the threshold must be positive, and second, the threshold cannot exceed the maximum value of the input. Additionally, it is preferable to determine corresponding independent thresholds based on the input noise. The derivative of the soft thresholding function can be expressed as follows:

$$\frac{\partial y}{\partial x} = \begin{cases} 1, x > \varphi \\ 0, -\varphi \le x \le \varphi \\ 1, x < -\varphi \end{cases} \quad (7)$$

Equation (7) suggests that the function can only take values of 1 or 0, exhibiting the same properties as ReLU. Thus, soft thresholding not only mitigates noise interference but also aids in preventing the vanishing gradient problem within the model.

In summary, Fig. 3 reveals the network structure of the 1DDRSN.

By integrating the 1DDRSN into the cloud computing service, the cloud-based predictive maintenance system for equipment can analyze vast amounts of data in real time. It enables accurate predictions of equipment failures and provides reliable data support and a decision-making basis for the intelligent application service module.

3.3. Experimental design for performance validation of the equipment fault diagnosis model

This work focuses on bearing failures in industrial manufacturing equipment to validate the performance of the equipment fault diagnosis model based on the 1DDRSN. The Case Western Reserve University (CWRU) bearing dataset is utilized as the raw dataset for this experiment. The CWRU bearing dataset contains vibration data for bearings under various operating conditions, including normal states and various kinds of faults, like inner race faults, outer race faults, and rolling element faults [24]. Data collection is conducted using different load and speed settings to guarantee the data diversity and representativeness. Specifically, the CWRU dataset provides standardized vibration signal data, divided into training and testing sets to facilitate model training and validation. Specifically, the CWRU dataset includes multiple types of fault modes, with multiple sets of vibration signal data for each fault mode. For this experiment, the training set consists of vibration data from 10 different bearings, totaling approximately 6,000 sample data points. The test set includes data from 5 other bearings, with about 2,000 sample data points. All sample data are labeled as either normal or different fault types (such as inner race fault, outer race fault, and rolling element fault). These fault modes are combined with different operating conditions to ensure the diversity of the dataset. Specifically, the training set contains 4,000 normal samples and 2,000 fault samples, while the test set includes 1,500 normal samples and 500 fault samples. This data partitioning ensures that the model can fully learn the differences between fault patterns and normal states during training and effectively evaluate the model's diagnostic and generalization capabilities in real-world conditions during testing.

In the experiment utilizing the 1DDRSN for bearing fault diagnosis, the model training process includes the following steps: First, parameter settings are established for each layer; next, threshold determination is performed by calculating the product of the average absolute value of each feature channel and 0.01 to determine the threshold. Values exceeding the threshold undergo a threshold subtraction process, while values below the threshold are set to zero; then, classifier training is conducted, setting the number of classifier neurons according to the fault categories and fine-tuning the softmax classifier parameters using the gradient descent algorithm; finally, model parameter training is completed within the maximum training iterations. The model validation phase evaluates the performance of the trained model using the testing set to assess its applicability in real-world scenarios. In this experiment, Fig. 4 shows the pseudocode for the specific process of the 1DDRSN algorithm.

To validate the effectiveness of the equipment fault diagnosis model based on the 1DDRSN, comparative experiments are designed. The effectiveness of the 1DDRSN-based equipment fault diagnosis model is compared with that of some other models on the same CWRU bearing dataset to evaluate their performance. These models include SVM, CNN, Multiple Empirical Wavelet Transform

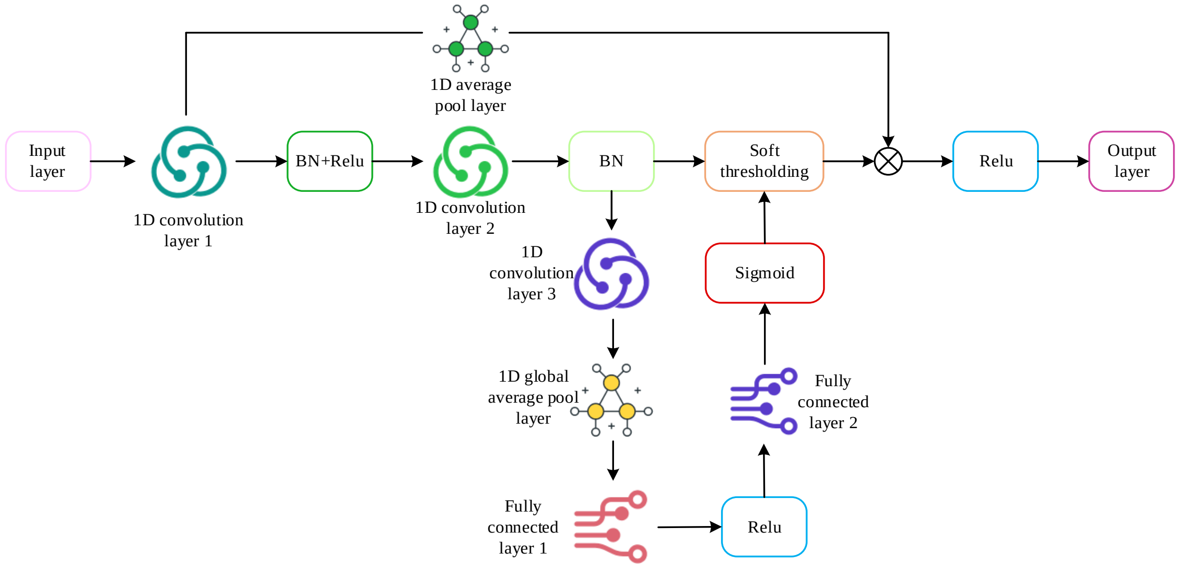

Fig. 3. Network structure of 1DDRSN

-CNN (MEWT-CNN), RNN, and One Dimensional CNN-LSTM (1DCNN-LSTM). Additionally, the performance of the cloud-based predictive maintenance system is assessed and compared with traditional systems, focusing on response time, system throughput, and data processing efficiency as the main evaluation metrics.

```
Input: Low-resolution signal (input_signal)
    1. Input Processing:
        processed_signal = Preprocess(input_signal)
    2. Feature Extraction:
        feature_map = Convolution_Layers(processed_signal)
        feature_map = Activation_Functions(feature_map)
    3. Feature Shrinkage:
        shrunk_feature_map = Shrinkage_Layers(feature_map)
    4. Feature Expansion:
        expanded_feature_map =Expansion_Layers(shrunk_feature_map)
    5. Residual Connection:
        high_resolution_signal = Residual_Connection(processed_signal, expanded_feature_map)
    6. Activation Function:
        activated_signal = Activation_Function(high_resolution_signal)
    7. Output Processing:
        final_output = Postprocess(activated_signal)
Return: final_output
```

Fig. 4. Pseudocode for the specific process of the 1DDRSN algorithm

To ensure the effectiveness of model training and the reliability of optimization results, hyperparameter optimization is performed. Initially, a grid search and cross-validation tests are conducted to select the best hyperparameter combination. To avoid model overfitting and improve generalization, cross-validation is applied on the training set using K-fold cross-validation (K=5) to select the optimal hyperparameters for each model. Through grid search, an appropriate learning rate is chosen, ranging from 0.001 to 0.1, with 0.01 selected as the optimal learning rate. Different batch sizes are tested, namely 32, 64, and 128, with 128 chosen as the optimal batch size. The regularization coefficient is adjusted within the range of 0.0001 to 0.01, with 0.001 selected as the optimal value. Experiments are also conducted to adjust the convolution kernel width to 64 and the stride to 1, as these two parameters significantly influence the performance of the 1DDRSN model. Table 2 shows the experimental environment and parameter settings.

Table 2. Experimental environment and parameter settings

Hardware/Parameter Name	Parameter/Value
Operating System	Windows10
Central Processing Unit (CPU)	AMD R7-5800H
Clock Speed	3.2GHz
Graphics Processing Unit (GPU)	RTX3060
Memory	16GB
Storage	512G SSD
Learning Rate	0.01
Number of Epochs	100
Batch Size	128
Regularization Coefficient	0.001
Convolution Kernel Width	64
Convolution Kernel Stride	1

Furthermore, accuracy, precision, recall, as well as F1-score are used as evaluation metrics for the equipment fault diagnosis model. Accuracy is defined as the proportion of correctly predicted samples to the total number of samples:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

TP represents the number of true positives. TN indicates the number of true negatives. FP refers to the number of false positives. FN denotes the number of false negatives. Precision refers to the proportion of actual positives among all samples predicted as positives by the model.

$$Pre = \frac{TP}{TP+FP} \tag{9}$$

Recall suggests the ratio of actual positives that are correctly predicted as positives by the model among all samples that are actually positive.

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

The F1 score is the harmonic mean of precision and recall, offering a single metric that takes into account both precision and recall. The equation for calculating the F1 score is as follows:

$$F1 = 2 * \frac{Pre*Rec}{Pre+Rec} \tag{11}$$

During the experiment, several practical issues are encountered regarding the application of the 1DDRSN-based equipment fault diagnosis model. One major challenge is the data imbalance problem. In the CWRU bearing dataset, there is an imbalance in the number of samples across different fault categories, which leads to poor recognition of minority classes during training. To address this issue, over-sampling and under-sampling techniques are employed to either increase the number of minority class samples or reduce the number of majority class samples, thus balancing the dataset. Additionally, a weighted loss function is applied, assigning different weights to different classes, further enhancing the model's sensitivity to the minority classes. Another challenge is the computational resource issue during model training. Due to the complex network structure and the computational demands of deep learning, the 1DDRSN model consumes significant hardware resources, particularly during multiple iterations, leading to longer training time. To overcome this, GPU acceleration is used to speed up the training process via parallel computing. Moreover, model pruning is implemented to remove redundant parameters in the network, reduce the computational load and maintain model performance. The data noise issue also impacts the model’s diagnostic accuracy. In real-world equipment operations, varying degrees of vibration interference lead to noisy data, which affects the model's training effectiveness. To tackle this, data preprocessing methods such as denoising filters, standardization, and normalization are applied to ensure data quality and the stability of the model. Through these measures, the issues encountered during the experiment are effectively addressed, improving the accuracy, efficiency, and applicability of the 1DDRSN-based equipment fault diagnosis model.

Furthermore, to ensure the reliability of the experimental results and the model's stability, the impact of initialization weights on training results is considered during the training of the 1DDRSN model. Since the training results of neural networks are often influenced by initial weights, multiple experiments are conducted to verify the stability and reproducibility of the model's results. In the comparative experiments, five independent trials are performed, with random initialization of weights in each experiment. The training and testing sets remain the same across all trials, and the same parameter settings are used for the training process. By comparing the results of different experiments, the stability of the model performance and the influence of initial weights on the results are assessed. In all five experiments, the variations in the model's accuracy, precision, recall, and F1-score are minimal, ranging from 0.9850 to 0.9886, with a change of less than 0.5%. This indicates that while the randomness of the initial weights may have a slight impact on the training results, the model's performance is highly stable, and the differences in results across multiple experiments are small, further validating the reliability of the model.

4. EXPERIMENTAL RESULTS OF THE EQUIPMENT FAULT DIAGNOSIS AND PREDICTIVE MAINTENANCE SYSTEM BASED ON CLOUD PLATFORM AND DEEP LEARNING

4.1. Performance evaluation of the equipment fault diagnosis model

The 1DDRSN, 1DCNN, and ResNet networks are each incorporated into the equipment fault diagnosis model to compare the number of epochs required for each model to achieve an accuracy of 90%. Fig. 5 shows the results.

Fig. 5 shows that in the equipment fault diagnosis model, the 1DDRSN network demonstrates a significant advantage over both 1DCNN and ResNet in terms of the number of training epochs required to reach 90% accuracy. Specifically, the 1DDRSN network achieves 90% accuracy in just 17 epochs, which is notably fewer than the 59 epochs required by 1DCNN and the 38 epochs required by ResNet. This indicates that the 1DDRSN network excels in training efficiency, reaching high accuracy more quickly and reducing both training time and computational resource demands. Furthermore, the highest accuracy achieved by the 1DDRSN network after reaching 90% accuracy is 98.86%, which is higher than 1DCNN's 96.24% and ResNet's 95.68%. This result further validates the superior performance of the 1DDRSN network in equipment fault diagnosis, as it not only reaches high accuracy with fewer training epochs but also achieves a final accuracy significantly better than the other two networks.

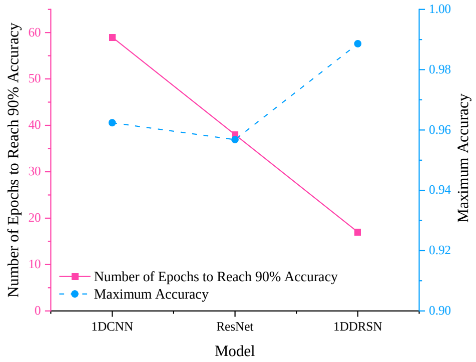

Fig. 5. Comparison of fault diagnosis results of various models

4.2. Comparison of Performance with Other Models

The effectiveness of the 1DDRSN model is evaluated against five other models: SVM, CNN, MEWT-CNN, RNN, and 1DCNN-LSTM. Fig. 6 displays the results.

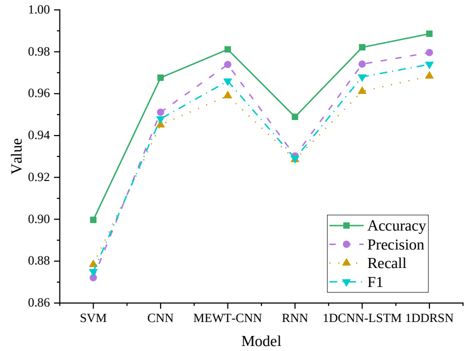

Fig. 6. Comparison of fault diagnosis results of various models

The experimental results in Fig. 6 indicate that the 1DDRSN model significantly outperforms the other five models in terms of fault diagnosis performance, including SVM, CNN, MEWT-CNN, RNN, and 1DCNN-LSTM. Specifically, the 1DDRSN model achieves an accuracy of 0.9886, precision of 0.9796, recall of 0.9684, and an F1 score of 0.974, all of which are the best among the models. In comparison, the second-ranked 1DCNN-LSTM model exhibits metrics that are lower by 0.66%, 0.56%, 0.76%, and 0.62%, respectively. These results suggest that the 1DDRSN model demonstrates greater robustness and superior predictive performance when handling complex industrial equipment fault diagnosis tasks.

To further evaluate the performance of the 1DDRSN model in equipment fault diagnosis, its training time and resource consumption are examined, as shown in Fig. 7.

Fig. 7 reveals that the 1DDRSN model outperforms others in terms of training efficiency and resource consumption. The 1DDRSN requires only 17 training epochs, with a training time of 45 minutes, making it the fastest among all models. Specifically, compared to the 1DCNN-LSTM model (150 minutes), 1DDRSN saves 105 minutes of training time. Additionally, the CPU usage of 1DDRSN is 39%, GPU usage is 78%, memory consumption is 4GB, and GPU memory usage is 5.6GB. These are significantly lower than other more complex models, such as 1DCNN-LSTM (GPU usage of 86%, memory of 8 GB, and GPU memory usage of 6.5 GB). Therefore, while achieving high accuracy, 1DDRSN not only reduces training time but also saves computational resources, demonstrating higher training efficiency and resource utilization.

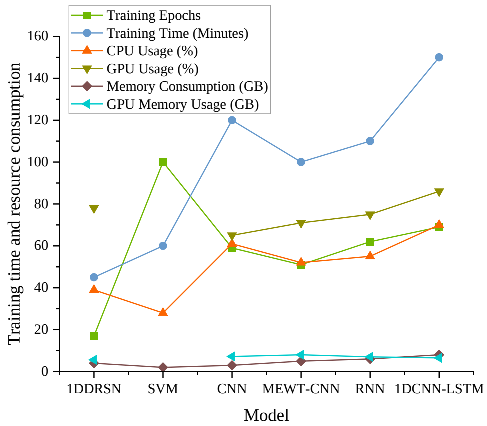

Fig. 7. Comparison of training time and resource consumption across models

4.3. Performance evaluation of the predictive maintenance system for equipment

The performance of the cloud-based predictive maintenance system for equipment is evaluated under three different loads. Moreover, its response time, system throughput, and data processing efficiency are compared with those of traditional systems. Fig. 8 displays the results.

The data in Fig. 8 demonstrate that the cloud-based predictive maintenance system for equipment outperforms traditional systems in terms of response time, system throughput, and data processing efficiency. Under light load conditions, the response time of the cloud-based system is 120 milliseconds, significantly lower than the 300 milliseconds of the traditional system. The system throughput is 1000 requests per hour, surpassing the 500 requests per hour of the traditional system. The system's data processing efficiency is 30 milliseconds per packet, which is more efficient than the traditional system's 50 milliseconds per packet. In heavy load conditions, the response time, throughput, and data processing efficiency of the cloud-based system are 400 milliseconds, 400 requests per hour, and 100 milliseconds per packet, respectively. They showcase a clear advantage compared to the traditional system's 1000 milliseconds, 150 requests per hour, and 200 milliseconds per packet. These results indicate that the cloud-based system is superior to the traditional system in performance and efficiency, better supporting the real-time maintenance needs in complex industrial environments.

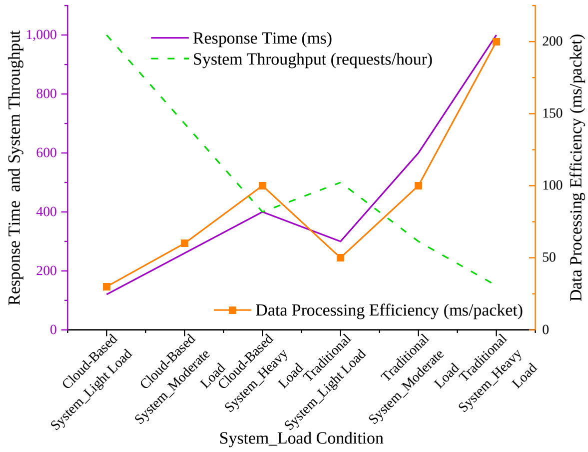

Fig. 8. Performance comparison of the cloud-based predictive maintenance system and traditional systems

4.4. Discussion

In the field of fault prediction, numerous studies in recent years have proposed various machine learning and deep learning models to improve the accuracy and efficiency of equipment fault prediction. The 1DDRSN-based equipment fault diagnosis model proposed performs excellently across multiple performance metrics. Therefore, it is necessary to compare it with other related research to highlight its relative advantages. Zamzam et al. (2023) developed a machine learning model for predicting medical equipment failures, focusing on predicting first failure events, failure-to-repair ratios, and repair groups. The experiment showed that SVM was suitable for first-failure prediction, decision trees for failure-to-repair ratio prediction, and artificial neural networks for repair group prediction, with accuracies of 96.9%, 83.9%, and 76.7%, respectively [25]. Although SVM performs well in first-failure prediction, its accuracy is lower than the results obtained from the 1DDRSN model, indicating the advantages of deep learning models in complex fault diagnosis tasks. Specifically, the high accuracy and robustness of the 1DDRSN model provide a significant advantage when dealing with multiple fault types. Xu et al. (2023) proposed a multi-stage fault prediction model for short-term and long-term fault prediction of continuous casting rollers. The results showed that the model significantly improved short-term fault prediction accuracy and effectively predicted future long-term fault trends [26]. Although their model achieved excellent results in short-term fault prediction, its performance under complex and diverse fault types was not as universally effective as the 1DDRSN model proposed. The 1DDRSN model, by combining one-dimensional deep residual networks and shrinkage mechanisms, demonstrates higher accuracy and lower computational overhead in diagnosing multiple fault types, making it more suitable for diverse fault diagnosis in industrial equipment. Shaheen et al. (2023) proposed a data-driven fault prediction method for mechanical components' fault prediction and remaining useful life estimation by combining artificial neural network architecture and an improved training algorithm. The results indicated that this method had high prediction accuracy and success rates in complex manufacturing systems [27]. However, although the method performed well in certain specific applications, the 1DDRSN model presented enhanced fault diagnosis capability in various fault scenarios by combining deep residual networks, soft-threshold processing, and attention mechanisms. It also demonstrated high-accuracy performance, especially in low-noise environments, making it more advantageous in the widespread application of industrial equipment.

Overall, the 1DDRSN-based equipment fault diagnosis model proposed has significant advantages over existing models. Whether in terms of equipment fault type diversity, diagnostic accuracy, or computational efficiency, the 1DDRSN model demonstrates strong competitiveness. Additionally, by integrating cloud platform technology, it realizes efficient fault prediction and maintenance, and further enhances real-time capabilities and scalability. This highlights the potential of the model for application in industrial production and smart manufacturing.

5. CONCLUSION

This work proposes an equipment fault diagnosis model based on 1DDRSN through the optimization of deep learning models. It explores a predictive maintenance system that integrates the cloud platform with the 1DDRSN model. Through experimental evaluations, the following conclusions are drawn: (1) In comparison with other models, the effectiveness of the 1DDRSN model in equipment fault diagnosis significantly surpasses that of five other models. It achieves accuracy, precision, recall, and F1 score of 0.9886, 0.9796, 0.9684, and 0.974, respectively. (2) Compared to the second-ranked 1DCNN-LSTM model, the 1DDRSN model achieves improvements of 0.66%, 0.56%, 0.76%, and 0.62% in accuracy, precision, recall, and F1 score, respectively. Furthermore, while achieving high accuracy, 1DDRSN is able to save computational resources in less time, offering higher training efficiency and resource utilization. (3) Under moderate load conditions, the cloud-based predictive maintenance system demonstrates significant enhancements in response time (260 milliseconds), system throughput (700 requests/hour), and data processing efficiency (60 milliseconds/packet) when compared to traditional systems. This suggests the superior effectiveness and efficiency of the cloud-based approach.

Furthermore, the proposed 1DDRSN model and cloud platform-based predictive maintenance system have broad application prospects in renewable energy stations and their energy storage systems. Renewable energy devices such as wind turbines and photovoltaic inverters often operate under complex and changing environmental conditions. Key components like bearings, gearboxes, and power electronic modules are prone to degradation and failure due to factors such as temperature, humidity, and mechanical stress. Traditional maintenance strategies typically rely on periodic inspections or simple threshold-based alarm mechanisms, which struggle to achieve precise predictions and efficient maintenance. By integrating the 1DDRSN model, the accuracy of fault diagnosis can be significantly improved, especially in handling complex nonlinear signals, where its feature extraction capabilities are particularly strong. Meanwhile, the cloud platform-based predictive maintenance system can collect and analyze operational data in real-time from distributed energy devices, enabling remote fault diagnosis and health status assessment through deep learning algorithms. This enhances the operational reliability and maintenance efficiency of the equipment. Additionally, in energy storage systems such as lithium-ion battery packs and supercapacitors, predicting the health status and remaining useful life of batteries is crucial for ensuring the stable operation of energy management systems. The 1DDRSN model can more accurately identify degradation patterns in energy storage units. Combined with the real-time computational power of the cloud platform, it can provide optimized maintenance strategies and scheduling plans for smart grids. Thus, the proposed method is not only applicable to predictive maintenance for traditional industrial equipment but also provides theoretical support and technical insights for the intelligent operation and maintenance of renewable energy stations and energy storage systems. This further promotes the integration of smart manufacturing and intelligent energy management.

Despite significant progress in the field of equipment fault diagnosis and predictive maintenance, some limitations still exist. First, the experimental dataset only used the CWRU-bearing data, lacking fault data from other types of industrial equipment. This could potentially affect the model's generalization ability. To enhance the model's applicability, future research will expand the dataset to include more data from different types of equipment, particularly those covering complex operating conditions and various fault types, thereby further improving the model's generalization and robustness. Next, while the 1DDRSN model has shown excellent performance in diagnostic accuracy and training efficiency, its performance under more complex fault types and extreme operating conditions still needs further validation. Future studies will test the model's adaptability by introducing more diverse fault types and simulating extreme working environments. Additionally, as industrial equipment continues to evolve, fault modes may become more complex. Therefore, further optimization of the 1DDRSN model to address unknown fault modes will be an important direction for future work. Moreover, the cloud platform-based predictive maintenance system has demonstrated superior performance. However, in order to handle more complex industrial applications, future work will focus on integrating more edge computing and distributed technologies to enhance the system's real-time capability and stability. Particularly in large-scale industrial IoT environments, where data volumes are massive and real-time requirements are high, developing distributed architectures with stronger real-time processing capabilities and lower latency will be key to improving system performance. Finally, with the rapid development of AI technologies, future research could explore the application of advanced techniques like deep reinforcement learning and GAN in equipment fault diagnosis to further improve fault prediction accuracy and the level of intelligent diagnosis. Furthermore, integrating equipment fault diagnosis with equipment health management systems and promoting the comprehensive application of intelligent maintenance solutions will provide more comprehensive and intelligent support for industrial production.

Source of funding: *This research received no external funding.*

Author contributions: *The author confirms being the sole contributor of this work and has approved it for publication.*

Declaration of competing interest: *the author declares no conflict of interest.*

REFERENCES

1. Kou L, Li Y, Zhang F, et al. Review on monitoring, operation and maintenance of smart offshore wind farms. Sensors. 2022;22(8):2822. https://doi.org/10.3390/s22082822.
2. Rediske G, Michels L, Cezar Mairesse Siluk J, et al. Management of operation and maintenance practices in photovoltaic plants: Key performance indicators. International Journal of Energy Research. 2022;46(6): 7118-7136. https://doi.org/10.1002/er.7737.
3. Jambol DD, Sofoluwe OO, Ukato A, et al. Transforming equipment management in oil and gas with AI-Driven predictive maintenance. Computer Science & IT Research Journal. 2024;5(5):1090-1112. https://doi.org/10.51594/csitrj.v5i5.1117.
4. Teoh Y K, Gill S S, Parlikad A K. IoT and fog-computing-based predictive maintenance model for effective asset management in Industry 4.0 using machine learning. IEEE Internet of Things Journal. 2021;10(3):2087-2094. https://doi.org/10.1109/JIOT.2021.3050441.
5. Mnyakin M. Applications of ai, iot, and cloud computing in smart transportation: A review. Artificial Intelligence in Society. 2023;3(1):9-27. https://www.researchberg.com/index.php/ai/article/view/108
6. Kolasani S. Revolutionizing manufacturing, making it more efficient, flexible, and intelligent with Industry 4.0 innovations. International Journal of Sustainable Development Through AI. 2024;3(1):1-17. https://ijsdai.com/index.php/IJSDAI/article/view/46.
7. Maurya M, Panigrahi I, Dash D, et al. Intelligent fault diagnostic system for rotating machinery based on IoT with cloud computing and artificial intelligence techniques: a review. Soft Computing.2024; 28(1): 477- 494. https://doi.org/10.1007/s00500-023-08255-0.
8. Kannammal A, Guhanesvar M, Venketesz R R. Predictive maintenance for remote field IoT devices-A deep learning and cloud-based approach. Mobile Computing and Sustainable Informatics: Proceedings of ICMCSI 2023. Singapore: Springer Nature Singapore. 2023:567-585. https://doi.org/10.1007/978-981-99-0835-6_40.
9. Sharma R, Gurung S. Implementing big data analytics and machine learning for predictive maintenance in manufacturing facilities in South Korea. AI, IoT and the Fourth Industrial Revolution Review. 2024;14(2): 1-17. https://scicadence.com/index.php/AI-IoT-REVIEW/article/view/36.
10. Raparthy M, Dodda B, Maruthi S. Predictive maintenance in IoT devices using time series analysis and deep learning. Dandao Xuebao/Journal of Ballistics. 2023;35(3):01-10. https://doi.org/10.52783/dxjb.v35.113.
11. Qureshi M S, Umar S, Nawaz M U. Machine learning for predictive maintenance in solar farms. International Journal of Advanced Engineering Technologies and Innovations. 2024;1(3):27-49. https://ijaeti.com/index.php/Journal/article/view/228.
12. Ghasemkhani B, Aktas O, Birant D. Balanced k-star: An explainable machine learning method for internet-of-things-enabled predictive maintenance in manufacturing. Machines. 2023;11(3):322. https://doi.org/10.3390/machines11030322
13. Nunes P, Santos J, Rocha E. Challenges in predictive maintenance - A review. CIRP Journal of Manufacturing Science and Technology. 2023: 40: 53-67. https://doi.org/10.1016/j.cirpj.2022.11.004.
14. Achouch M, Dimitrova M, Ziane K, et al. On predictive maintenance in industry 4.0: Overview, models, and challenges. Applied Sciences. 2022; 12(16): 8081. https://doi.org/10.3390/app12168081.
15. Resende C, Folgado D, Oliveira J, et al. TIP4. 0: industrial internet of things platform for predictive maintenance. Sensors. 2021;21(14):4676. https://doi.org/10.3390/s21144676.
16. Calabrese F, Regattieri A, Bortolini M, et al. Predictive maintenance: A novel framework for a data-driven, semi-supervised, and partially online prognostic health management application in industries. Applied Sciences. 2021;11(8):3380. https://doi.org/10.3390/app11083380.
17. Mourtzis D, Angelopoulos J, Panopoulos N. Design and development of an IoT enabled platform for remote monitoring and predictive maintenance of industrial equipment. Procedia Manufacturing. 2021; 54:166-171. https://doi.org/10.1016/j.promfg.2021.07.025.
18. Yang P, Geng H, Wen C, et al. An intelligent quadrotor fault diagnosis method based on novel deep residual shrinkage network. Drones. 2021;5(4):133. https://doi.org/10.3390/drones5040133.

19. Tong J, Tang S, Wu Y, et al. A fault diagnosis method of rolling bearing based on improved deep residual shrinkage networks. Measurement; 2023;206:112282. https://doi.org/10.1016/j.measurement.2022.112282.
20. Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. Neurocomputing. 2021; 452:48-62. https://doi.org/10.1016/j.neucom.2021.03.091.
21. Gao Z, Tong L, Zhou J, et al. Stochastic depth residual network for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing. 2021;60:1-13. https://doi.org/10.1109/TGRS.2021.3090429.
22. Shafiq M, Gu Z. Deep residual learning for image recognition: A survey. Applied Sciences. 2022;12(18): 8972. https://doi.org/10.3390/app12188972.
23. Mahaur B, Mishra K K, Singh N. Improved Residual Network based on norm-preservation for visual recognition. Neural Networks. 2023;157:305-322. https://doi.org/10.1016/j.neunet.2022.10.023.
24. Hendriks J, Dumond P, Knox D A. Towards better benchmarking using the CWRU bearing fault dataset. Mechanical Systems and Signal Processing. 2022; 169:108732. https://doi.org/10.1016/j.ymssp.2021.108732.
25. Zamzam A H, Hasikin K, Wahab A K A. Integrated failure analysis using machine learning predictive system for smart management of medical equipment maintenance. Engineering Applications of Artificial Intelligence. 2023;125:106715. https://doi.org/10.1016/j.engappai.2023.106715.
26. Xu E, Zou F, Shan P. A multi-stage fault prediction method of continuous casting machine based on Weibull distribution and deep learning. Alexandria Engineering Journal. 2023;77:165-175. https://doi.org/10.1016/j.aej.2023.06.079.
27. Shaheen B, Kocsis Á, Németh I. Data-driven failure prediction and RUL estimation of mechanical components using accumulative artificial neural networks. Engineering Applications of Artificial Intelligence. 2023;119:105749. https://doi.org/10.1016/j.measurement.2022.112282

APPENDIX

Table. Experimental parameter settings

Parameter Name	Value
Learning Rate	0.01
Number of Epochs	100
Batch Size	128
Regularization Coefficient	0.001
Convolution Kernel Width	64
Convolution Kernel Stride	1

Dongbao MA, Associate Professor, graduated from Beijing Jiaotong University with a major in Mechanical Engineering. He currently serves as the Deputy Director of the Engineering Training Center at Beijing Polytechnic. His main research areas include Digital Twin Technology and Predictive Maintenance Technology.
e-mail: bpimadongbao@126.com