



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642

<https://www.inderscience.com/ijict>

Fault diagnosis and remaining life prediction of key industrial equipment based on machine learning

Dan Li, Mengqin Yang, Fei Xie

DOI: [10.1504/IJICT.2025.10070798](https://doi.org/10.1504/IJICT.2025.10070798)

Article History:

Received:	07 February 2025
Last revised:	07 March 2025
Accepted:	07 March 2025
Published online:	08 May 2025

Fault diagnosis and remaining life prediction of key industrial equipment based on machine learning

Dan Li, Mengqin Yang* and Fei Xie

Hunan Railway Professional Technology College,
ZhuZhou, 412001, China

Email: li_danhntd@163.com

Email: yang_mengqing@163.com

Email: 18773155082@163.com

*Corresponding author

Abstract: Fault diagnosis (FD) has certain practical significance for enterprise operations and social and economic development. This paper focuses on the deep learning-based rotating machinery FD and remaining useful life (RUL) prediction methods to carry out research. The feature extraction module contained in the framework can fully capture the time dependence of the data in the time domain. The lightweight temporal convolutional network-broad learning system-fault diagnosis (LTCN-BLS-FD) framework has achieved the best diagnostic results, and the average values of the four indicators accuracy (Acc), mean precision(MP), mean recall (MR)and F1-score (MF)are all above 0.9773. The temporal convolutional autoencode temporal convolutional network (TCAETCN-Res Net) framework has achieved the best prediction results in the ablation experiment and the comparative experiment, and its root mean square error (RMSE) for RUL prediction is 0.1658, mean absolute error (MAE) is 0.1259, and R2 is 0.6567. The above verifies the effectiveness of TCAETCN-Res Net in RUL prediction. The wide learning classifier with strong nonlinear mapping ability and low complexity accurately identifies various types of faults through the feature fusion vector.

Keywords: machine learning; industrial equipment; FD; lifetime prediction.

Reference to this paper should be made as follows: Li, D., Yang, M. and Xie, F. (2025) 'Fault diagnosis and remaining life prediction of key industrial equipment based on machine learning', *Int. J. Information and Communication Technology*, Vol. 26, No. 13, pp.1–21.

Biographical notes: Dan Li received her BSc degree in Automation from the Hunan Institute of Technology, China. She obtained her MSc degree in Control Engineering from Guangxi University, China. Currently, she is an Associate Professor at Hunan Railway Professional Technical College. Her research interests are industrial automation and electronic engineering.

Mengqin Yang received her BSc degree in Electrical Engineering and Automation from Changsha University of Science and Technology, China. She obtained her MSc degree in Control Science and Control Engineering from Changsha University of Science and Technology, China. Currently, she is an Associate Professor at Hunan Railway Professional Technical College. Her research interests are electrical automation and electronic engineering.

Fei Xie received her BSc degree in Automotive Service Engineering from Central South University of Forestry and Technology China. She obtained her MSc degree in Transportation Engineering from Central South University. Currently, she is a Lecturer at Hunan Railway Professional Technical College. Her research interest is the aerodynamics of high-speed trains.

1 Introduction

The full name of FD is fault detection and diagnosis. It is a technology that detects and identifies equipment status based on sensor monitoring data. This technology integrates many fields such as control theory, artificial intelligence and mathematical statistics, and is one of the important guarantees for the safe and stable operation of industrial equipment. When a fault occurs, the FD technology can timely and accurately find the abnormal situation of the equipment and the location of the faulty components, which is conducive to targeted maintenance of the equipment. Relevant studies have shown that most industrial equipment failures do not appear suddenly, but go through a degradation process from health to failure (Manikandan and Duraivelu, 2021). In addition, after a component fails, the risk of failure of its surrounding related components will increase. Therefore, in addition to exploring how to effectively diagnose the faults that have occurred. Industrial equipment remaining useful life (RUL) prediction can accurately give the time experienced by the equipment from the current moment to the moment when the fault occurs (Tao et al., 2022), thus providing important equipment maintenance reference for operation and maintenance personnel and effectively reducing the incidence of accidents. To sum up, effective FD and RUL prediction methods are important methods to ensure the safe operation of industrial equipment.

Data-driven industrial equipment FD and RUL prediction is a general term for a class of methods that can relate historical monitoring data to equipment health status. This kind of method can accurately diagnose faults and predict RUL without fully understanding the complex mechanism of the target equipment structure. A large number of advanced sensors are used in various industrial equipment, which makes it possible to record and store all kinds of data during equipment operation. In addition, with the continuous development of hardware such as semiconductors, processors and embedded systems in the direction of low energy consumption and high computing power, the stored data can be calculated quickly (Wang et al., 2024).

From the above analysis that the existing fault diagnosis (FD) and remaining service life prediction models still have the following problems: first, some models rely too much on prior knowledge or assume that the data conform to the linear relationship, which is difficult to effectively deal with complex nonlinear data and multi parameter coupling effects under dynamic operating conditions. Secondly, most models do not fully consider the impact of long-term storage state on the life of equipment, resulting in significant deviation in the prediction of military equipment and other scenarios. In addition, although the data-driven deep learning method can extract features automatically, its performance is highly dependent on high-quality annotation data, and it often faces data noise, missing and multi-modal interference in practical engineering. Finally, the existing

model has complex structure and insufficient generalisation ability, which makes it difficult to balance the calculation efficiency and prediction accuracy, especially under time-varying conditions, it is difficult to achieve stable prediction across devices or scenarios.

This paper focuses on the deep learning-based rotating machinery FD and RUL prediction methods to carry out research, and establishes deep learning models for different problems and difficulties, so as to improve the FD and remaining life prediction effects of key industrial equipment.

2 Related works

2.1 FD based on traditional data-driven

Data-driven FD is a method that can accurately describe the health status of equipment by monitoring data without establishing complex mathematical models of equipment (Ji et al., 2021). Before the arrival of the era of big data, traditional data-driven FD methods are mainly summarised into FD based on signal analysis method and FD based on pattern recognition method.

Chu et al. (2022) proposed an improved double-tree complex wavelet transform for diagnosing gearbox faults, which can adaptively determine the signal decomposition level and extract the sub-band with the greatest signal correlation. this method has better fault feature extraction ability than empirical mode decomposition methods. Aiming at the bearing FD, Guo et al. (2023) proposed a FD method, this method can retain more useful fault information while reducing noise, and has good diagnostic performance for compound faults of bearings. Gao et al. (2022) proposed a bearing fault detection method which overcomes the shortcoming of strong sensitivity of SK to non-Gaussian noise. It is difficult to diagnose gear teeth and bearings when there are faults at the same time, Dong et al. (2023) proposed a compound FD method combining meshing resonance and SK. The method can identify bearing and gear faults at the same time. Because the interference factors such as noise contained in vibration signals reduce the performance of fault feature extraction methods such as SK to some extent, noise reduction of signals is also an important method to improve the accuracy of FD. Shi et al. (2021) proposed a false mode removal method based on wavelet threshold denoising for the problem of false modes in time domain signals caused by noise. The experimental results show that the false modes of signals are obviously reduced after denoising, thus reducing the difficulty of FD and extraction. Chen et al. (2022) proposed an adaptive wavelet denoising method based on whale optimisation algorithm. The wavelet denoising performance after parameter optimisation is obviously improved. Li et al. (2022) proposed an improved Hilbert-Huang transform (HHT) method, which effectively solves the problem of modal aliasing and false modes in intrinsic modal function components, and reduces the influence of noise on fault feature extraction.

Pattern recognition has been widely used in the field of FD to improve the level of intelligence. Similar to FD based on signal analysis, FD based on pattern recognition usually requires feature extraction of data, so as to obtain feature expressions that can reflect the differences between different health states of equipment (Kao and Chien, 2023). After obtaining effective fault features, the fault categories are obtained by classification through pattern recognition algorithms (Ge et al., 2022).

2.2 *FD based on deep learning*

Deep learning FD based on ‘end-to-end’ means that the data collected by sensors is not processed by domain transformation and other methods, but is directly input into the deep learning model to diagnose faults after simple data slicing, further reducing the dependence on expert experience. Common deep learning models in existing research include convolutional neural network (CNN) (Wang et al., 2021b) and long short-term memory network (LSTM). Zhou et al. (2023) proposed a deep learning model that can directly diagnose bearing faults through sensor signals and discussed in detail the influence of different sequence segmentation time lengths on FD results. The experimental results show that the increase of sequence segmentation time is beneficial to the improvement of FD accuracy.

Wang et al. (2021a) proposed a bearing FD model based on FFT and 1-DCNN, which considers both amplitude and phase in frequency domain data. Compared with only considering amplitude information, fault feature extraction of the model is more comprehensive, and the FD results show that the model still achieves an accuracy of more than 97% in the presence of noise. Hu and Deng (2022) proposed an industrial equipment FD method.

Because FD based on deep learning effectively reduces the negative impact caused by over-reliance on professional prior knowledge in diagnosis performance, this kind of method has become one of the mainstream research directions in the field of FD (Odeyar et al., 2022). At present, there are still many problems to be solved in this research direction, such as improving the comprehensiveness of data feature extraction, improving the feature mining ability and robustness, and realising model lightweight, which are still the research hotspots of FD based on deep learning.

3 **FD model**

The original sensor data usually contains a lot of redundant information, noise interference and nonlinear signals to simplify understanding, the complex process can be disassembled into the following key stages:

1 Data preprocessing (denoising and standardisation)

Noise filtering: eliminate high-frequency noise or abnormal values through digital filtering (such as low-pass filtering and Kalman filtering). Signal enhancement: amplify or normalise weak signals, such as baseline calibration of accelerometer data. Data segmentation: the continuous data stream is segmented according to the time window or event trigger to facilitate subsequent analysis (such as the segmentation of action segments in gesture recognition).

2 Feature extraction (mining key information from data)

Time domain features: extract statistics such as signal amplitude, mean value and variance (such as the motion intensity of the accelerometer). Frequency domain features: extract frequency components (such as rotation frequency of gyroscope signal) through Fourier transform. Combined features: integrate multi-dimensional data (such as acceleration+angular velocity joint analysis of motion mode).

3 Feature selection and optimisation

Correlation analysis: through principal component analysis (PCA) or mutual information method to screen the characteristics that are strongly related to the target. Dimensionality reduction: map high-dimensional features to low-dimensional space (such as t-sne visualisation) to reduce computational complexity.

4 Optimisation strategy in practical application

Multisensor fusion: cross validation of feature reliability combined with camera, radar and other data.

3.1 Proposal of LTCN-BLS FD framework

Set the number of data points to N_S , determine the value of N_S (Cen et al., 2022):

$$N_S = \frac{F_s}{S_r} \tag{1}$$

Among them, F_s and S_r represent the sampling frequency and the rotation speed of the device respectively.

$$\tilde{x} = 255 \times \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{2}$$

Among them, x is the input data, and x_{\max} and x_{\min} represent the maximum and minimum values of x , respectively.

Morlet wavelet function is used as the parent wavelet of CWT. The 2-DLDC module is shown in Figure 1. The detailed structures of the 2-D LTCN module 1 and module 2 are shown in Figure 2. The two modules consist of 3 2-D GC layers, 2 cropping layers, 2 BN layers, 2 ReLU layers, 2 CS layers, and 2 AMP layers. The clipping layer is only activated in the time axis direction in the two-dimensional feature map during the convolution process to ensure that the corner frequency information is not lost. The AMP layer in the 2-D LDCC module, the AMP layer 1 in the 2-DLTCN module 1 and the AMP layer 1 in the module 2 guarantee the feasibility of residual connection in the module by adjusting the size of the feature map output by the cropping layer. The AMP layer in the 2-D LDCC module, the 2-DLTCN module 1, and the AMP layer 2 in the module 2 reduce the dimension by downsampling the output of the respective modules, thereby preventing the occurrence of overfitting. The output size of the first two modules among these three modules is half of its own input data size, and the output size of the latter module is set to 1×1 .

Figure 1 Detailed structure of 2-D DLDC module

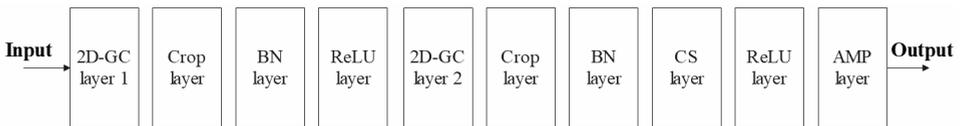
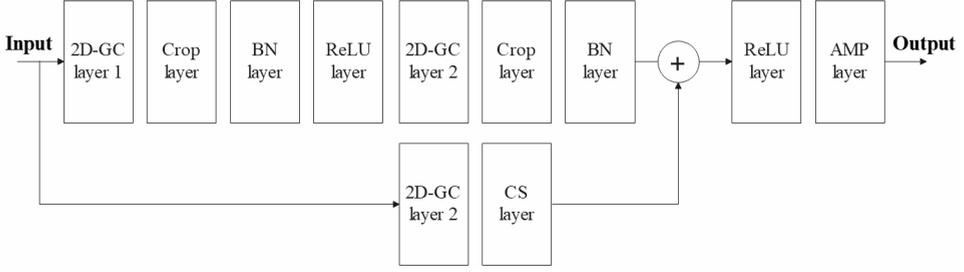
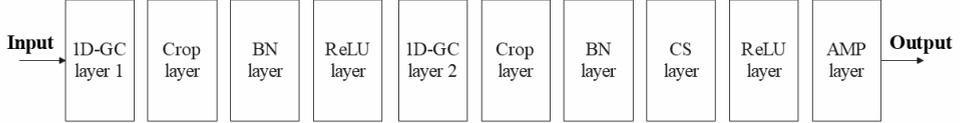
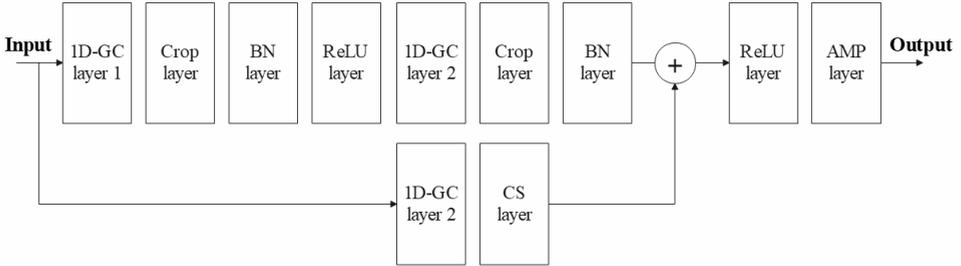


Figure 2 Detailed structure of 2-D LTCN module 1 and module 2**Figure 3** Detailed structure of 1-D LDCC module**Figure 4** Detailed structure of 1-D LTCN module 1 and module 2

The 1-D LTCN can extract the time-dependent features of different categories of faults from the original vibration data normalised to $[0, 1]$ range, as shown in Figure 3. As shown in Figure 4, the clipping layer removes the information from the future time in the process of convolution of one-dimensional time series features, ensuring that the output of the 1-DLTCN branch is not affected by the input from the future time. The AMP layer in the 1-DLDC module and 1-DLTCN module 1 downsamples the output, effectively reducing the risk of overfitting.

3.2 Proposal of TCAETCN-ResNet RUL prediction framework

A deep learning framework named TCAETCN-ResNet-Life prediction (TCAETCN-ResNet-LP) is proposed for industrial equipment RUL prediction. The framework consists of two main parts

- 1 TCAE for constructing health degradation indicators
- 2 RUL prediction model based on multi-angle time series feature extraction and fusion.

Among them, TCAE includes an encoder and a decoder based on inflated causal convolution. The overall structure of the TCAETCN-ResNet framework is shown in Figure 5.

Figure 5 Overall structure of 1TCAETCN-ResNet framework

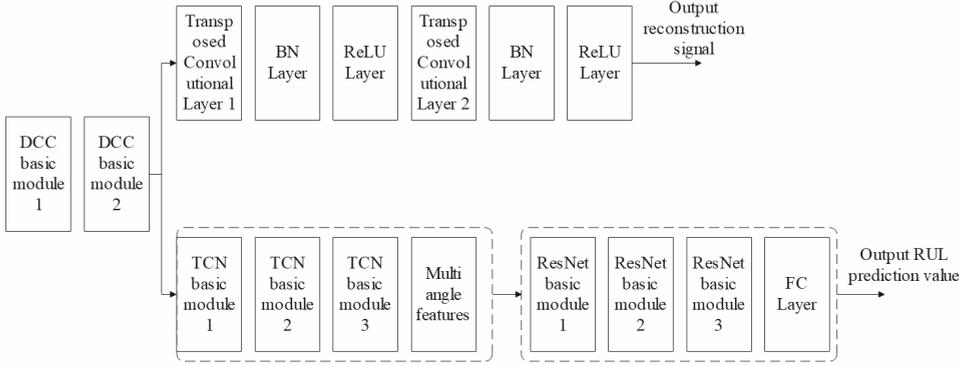
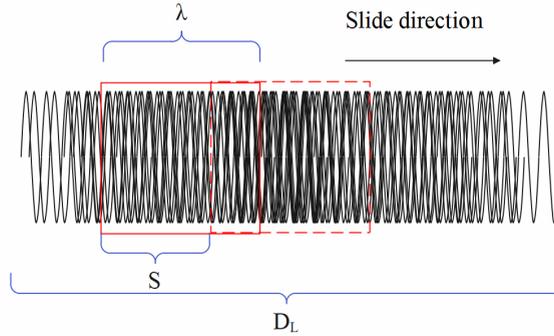


Figure 6 Schematic diagram of sliding window method (see online version for colours)



Before the feature extraction of the vibration signal, the sliding window method as shown in Figure 6 is used to segment the original data with overlap. In the figure, λ , s and D_L represent the window width, sliding step length and data length respectively. Then, the number of samples n can be expressed as formula (3). In addition, to eliminate the influence of the difference of vibration amplitude order in the segmented data segments on RUL prediction results, the data segments are normalised to the range of $[0, 1]$ and then the feature extraction module is input as samples (Wang et al., 2021c).

$$mum = \left(\frac{D_L - \lambda}{s} \right) + 1 \quad (3)$$

ML models learn from historical data, so they may not detect new or rare failure modes. Therefore, this paper sets that through incremental learning, new failure data is regularly added to the training set, and model parameters are updated through continuous learning technology, so as to avoid catastrophic forgetting of old knowledge, and update new failure information in time to improve the training and prediction ability of the model.

Figure 7 Detailed structure of DCC base block (see online version for colours)

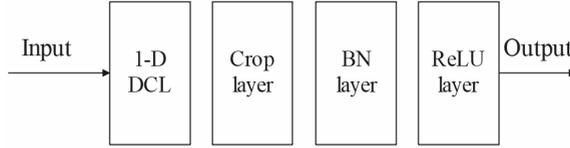
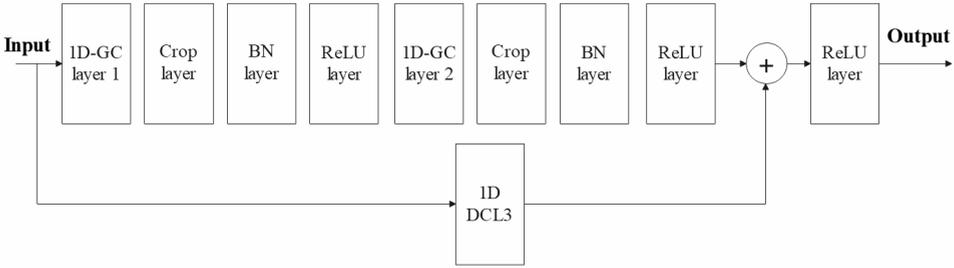


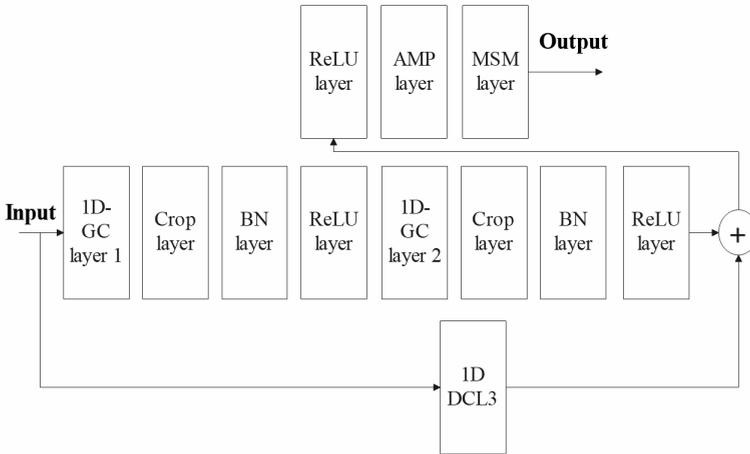
Table 1 Specific parameters of TCAE

<i>Name</i>	<i>Input channel</i>	<i>Output channel</i>	<i>Convolution kernel</i>	<i>Step size</i>	<i>Filling</i>	<i>Coefficient of expansion</i>	
DCC base block 1	1-D DCL	1	8	19	1	18	1
DCC base block 2	1-D DCL	8	16	3	1	4	2
Transposed convolutional layer 1	16	8	3	1	1	/	/
Transposed convolutional layer 2	8	1	3	1	1	/	/

Figure 8 Detailed structure of TCN base block in feature extraction module, (a) shows the detailed structure of TCN base blocks 1 and 2 (b) shows the detailed structure in the TCN base block 3



(a)



(b)

In order to strengthen the time-series feature characterisation of vibration signals to improve the prediction accuracy of RUL, an adaptive extraction method of health degradation indicators called TCAE is proposed based on one-dimensional inflated causal convolution. Among them, the encoder consists of two DCC base blocks, and each DCC base block includes a 1-D DCL, 1 cropping layer, 1 BN layer and 1 ReLU layer. The detailed structure of the DCC base block is shown in Figure 7. The decoder consists of 2 transposed convolutional layers, 2 BN layers, and 2 ReLU layers. The inflated causal convolution structure in the encoder ensures that it has strict time constraints and sufficient receptive fields in the process of obtaining vibration signal feature representations, thereby strengthening the time series feature representations in health degradation indicators and helping the feature extraction module to better capture the time dependence that can describe the component HDP. In addition, the step size of DCL in the encoder is set to 1, that is, the health degradation index is not dimensionally reduced compared with the input sample, thus avoiding information loss as much as possible. Detailed parameters of TCAE are shown in Table 1.

Based on the health degradation index established by TCAE, a RUL prediction model based on multi-angle time series feature extraction. As shown in Figure 8, the feature extraction module contains three TCN base blocks. Among them, TCN base block 1 and TCN base block 2 are composed of 3 1-DDCLs, 2 clipping layers, 2 BN layers, and 3 ReLU activation function layers. The TCN base block 3 includes 1 AMP layer and 1 MSM layer in addition to the components in the TCN base block 1 and the TCN base block 2. 1-DDCL3 in the base block is directly replaced by the input of the base block. The structure of three TCN basic blocks in series can adaptively extract time-dependent information from health degradation indicators. In addition, the inflated convolution in these basic blocks makes up for the shortcomings of the insufficient receptive field of the feature map caused by the convolution layer step size of 1, and ensures that the feature extraction module has sufficient capabilities to mine global features and high-level semantic information in the sample. The convolution kernel of 1-D DCL layer 1 and 1-D DCL layer 2 in TCN base block 1 is 19, and the larger convolution kernel is beneficial to capture the long-term dependency in the data, thus enriching the characterisation information of the component degradation process. The convolution kernel of 1-D DCL layer 1 and 1-DDCL layer 2 in TCN base block 2 and base block 3 is 3, and the smaller convolution kernel is beneficial to extract detailed information in features. The AMP layer plays the role of reducing the feature dimension, and its output dimension is set to the maximum integer that satisfies the feature dimension input to the MSM layer and can be divided by the number of heads. The role of the MSM layer is to perform multi-feature angle weight allocation on the time-dependent information output by the TCN basic block 3 to improve the diversity and comprehensiveness of the feature representation of the component degradation process. That is, the feature is self-attention calculated in N_h (the number of heads of MSM) subspaces, and then the multi-angle feature (F_{MSM}) shown in Figure 5 is output, where F_{MSM} can be expressed as formula (4) (Jiang et al., 2021).

$$F_{MSM} = Concat(F_1, F_2, \dots, F_i), i = 1, 2, \dots, N_h \quad (4)$$

The feature fusion module includes 3 ResNet basic blocks and 1 FC layer. Its function is to further fuse the multi-angle time series features output by the feature extraction module and conduct high-level semantic mining and output RUL prediction values. These 3 ResNet base blocks are composed of three one-dimensional convolution layers

(1-DCL), 2 BN layers, and 3 ReLU layers, and their structures are shown in Figure 9. Different from the TCN base block in the feature extraction module, the step size of 1-D CL1 and 1-D CL2 in these 3 ResNet base blocks is 2, thus forming a continuous downsampling structure to increase the receptive field of the output feature map of the feature fusion module, which is beneficial to the global feature mining of FMsm. The BN layer in the module plays the role of accelerating the convergence process and optimising the model training. The characterisation ability of time series features to the degradation process is further improved. Subsequently, the output of the ResNet base block is flattened and input to the FC layer, and a nonlinear relationship between the feature and the RUL is established to realise the prediction of the RUL. The specific parameters are shown in Table 2.

Figure 9 Detailed structure of ResNet base block in feature fusion module

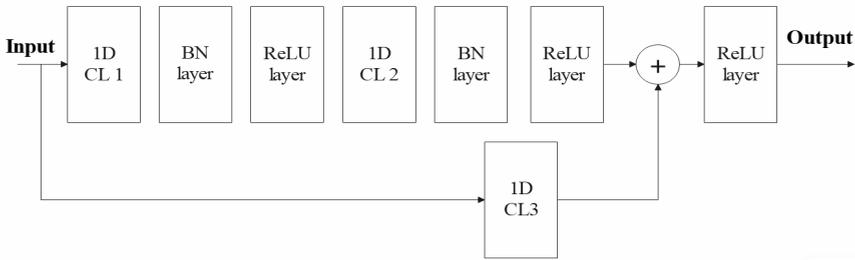
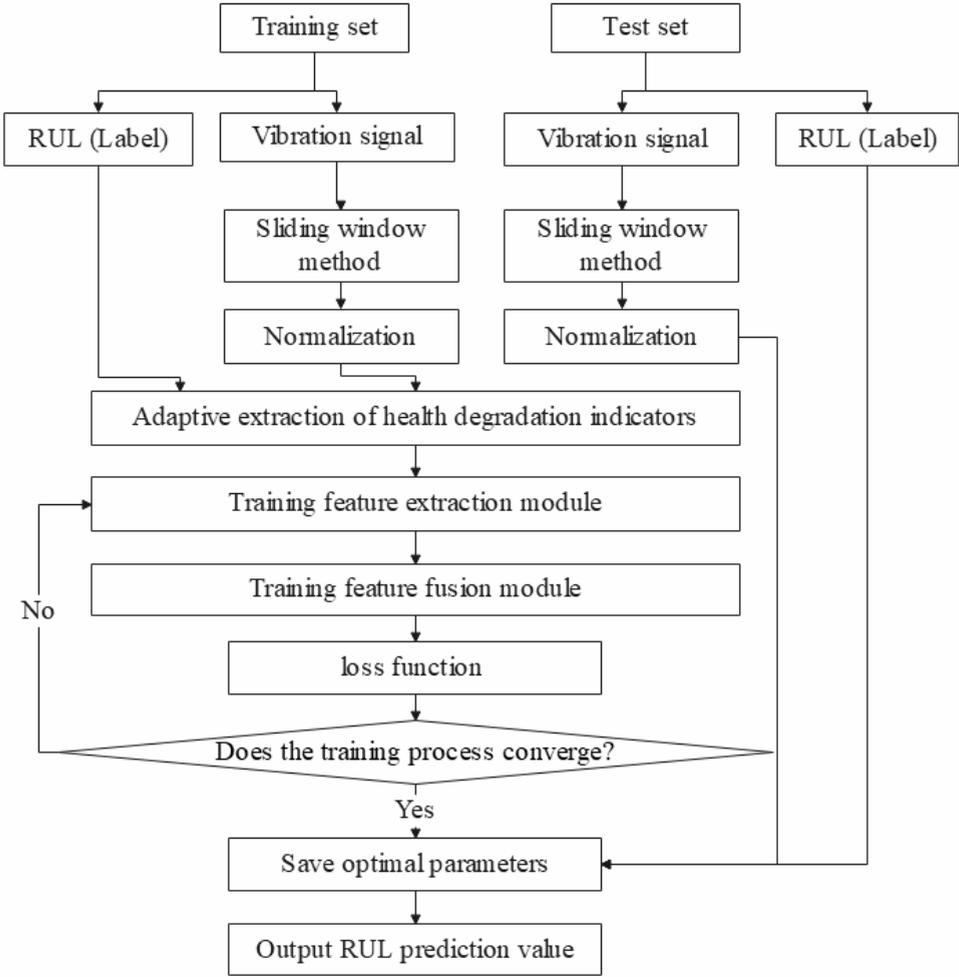


Table 2 Specific parameters

Name		Input channel	Output channel	Convolution kernel	Step length	Filling	Coefficient of expansion
TCN base block 1	1-D DCL1	16	16	19	1	18	1
	1-D DCL2	16	16	19	1	18	1
	1-D DCL3	16	16	1	1	0	1
TCN base block 2	1-D DCL1	16	32	3	1	4	2
	1-D DCL2	32	32	3	1	4	2
	1-D DCL3	16	32	1	1	0	1
TCN base block 3	1-D DCL1	32	32	3	1	8	4
	1-D DCL2	32	32	3	1	8	4
	1-D DCL3	32	32	1	1	0	1
ResNet base block 1	1-D CL1	32	64	3	2	1	/
	1-D CL2	64	64	3	2	1	/
	1-D CL3	32	64	1	4	0	/
ResNet base block 2	1-D CL1	64	128	3	2	1	/
	1-D CL2	128	128	3	2	1	/
	1-D CL3	64	128	1	4	0	/
ResNet base block 3	1-D CL1	128	256	3	2	1	/
	1-D CL2	256	256	3	2	1	/
	1-D CL3	128	256	1	4	0	/
FC layer		8,192	1	/	/	/	/

Figure 10 RUL prediction flow chart of TCAETCN-ResNet framework



The RUL prediction process of the TCAETCN-ResNet framework is divided into data preprocessing stage (stage 1), adaptive extraction stage of health degradation indicators (stage 2), and RUL prediction stage (stage 3). In stage 1, the sliding window method as shown in Figure 6 and formula (3) is used to segment the original signal with overlap, and each data segment is normalised. In addition, the label corresponding to each sample is defined as the percentage of RUL in the time elapsed by the degradation process of component health state, and its mathematical expression is shown in formula (5) (Ferreira and Gonçalves, 2022).

$$L_{RUL} = \frac{T_{HDP} - T_C}{T_{HDP}} \tag{5}$$

Among them, T_{HDP} and T_C are the time experienced by the component HDP and the current time predicted by the RUL, respectively. In stage 2, the TCAE adaptively extracts the health deterioration index from the data processed in stage 1. In stage 3, the feature fusion module in the TCAETCN-ResNet framework performs multi-angle feature extraction on the health degradation index and inputs it to the feature fusion module to obtain the RUL prediction value.

In the training process of TCAETCN-ResNet framework, a loss function as shown in formula (6) is proposed. Reduce the sensitivity of the training process of the model to outliers (Cao et al., 2021).

$$Loss = \omega_1 Loss_1 + \omega_2 Loss_2 \quad (6)$$

$$Loss_1 = \sum_{x \in X} Q(x) \log \frac{Q(x)}{E(x)} \quad (7)$$

$$Loss_2 = \frac{1}{n} \sum_{i=1}^n \begin{cases} \frac{(l_i - y_i)^2}{2}, & \text{if } |l_i - y_i| < 1 \\ |l_i - y_i| - \frac{1}{2}, & \text{other} \end{cases} \quad (8)$$

ω_1 and ω_2 are the weight coefficients of $Loss_1$ and $Loss_2$ respectively, $Q(x)$ and $E(x)$ are the two probability distributions of the random variable X , l_i and y_i are the i^{th} data label. The Adam optimiser minimises the loss function, and saves the framework parameters when the loss function reaches the minimum value, thus completing the training of TCAETCN-ResNet framework. Finally, the RUL prediction performance of the trained framework is verified by the test set data processed in stage 1. The RUL prediction flowchart of the TCAETCN-ResNet framework is shown in Figure 10.

4 Test

4.1 Test methods

The datasets in this paper are mainly CWRU datasets. Ablation experiments are designed in all FD cases. In terms of software and hardware, the codes of all models in the experiments involved are written using PyTorch 1.5. 1 and run on four RTX2080Ti graphics cards.

According to formula (1), each sample in this case contains 425 data points. On the premise of not disrupting and reusing the data.

Four evaluation indexes, Acc, MP, MR and MF, are used to comprehensively evaluate the classification performance of the model.

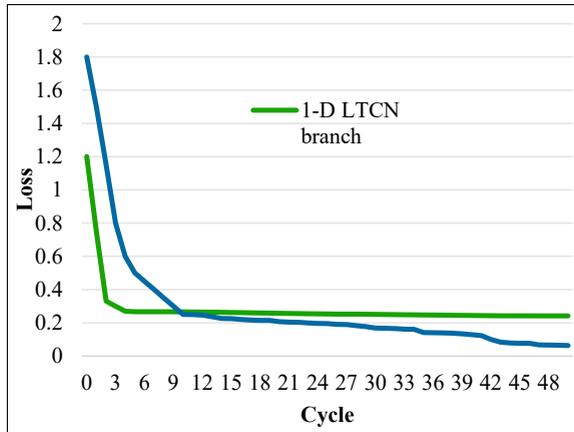
4.2 Results

Figure 11 shows the loss curves of the training process of the 1-DLTCN branch and the 2-DLTCN branch on the CWRU dataset. Table 3 summarises the mean and standard deviation (SD) of the ablation experimental results of different models on the CWRU dataset, and demonstrates the contribution of key structures in the lightweight temporal convolutional network-broad learning system-fault diagnosis (LTCN-BLS-FD) framework. Table 4 shows the FD results and the number of trainable parameters of different models on the CWRU dataset.

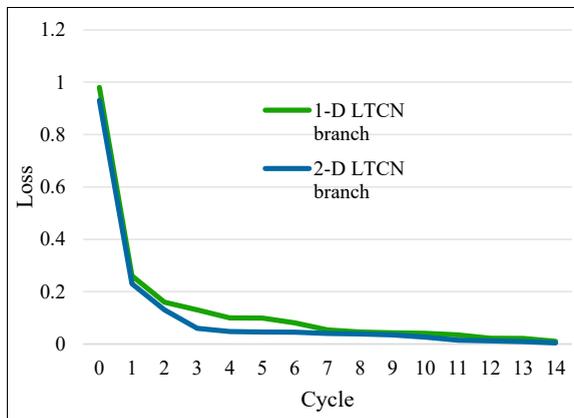
Table 5 shows the ablation experimental results of different models on the CWRU dataset with introduced noise, and Table 6 shows the FD results of different models on the CWRU dataset with introduced noise.

Table 7 shows the root mean square error (RMSE), mean absolute error (MAE), R2 of the reminding useful life (RUL) prediction results of different models.

Figure 11 Loss curves of training process of 1-DLTCN branch and 2-DLTCN branch on CWRU dataset, (a) loss curve before introducing noise (b) loss curve after introducing noise (see online version for colours)



(a)



(b)

Table 3 Ablation experimental results of different models on CWRU dataset

<i>Models</i>	<i>ACC</i>		<i>MP</i>		<i>MR</i>		<i>MF</i>	
	<i>Mean value</i>	<i>SD</i>						
OLTCN-BLS	0.9388	0.0052	0.9400	0.0054	OLTCN-BLS	0.9388	0.0052	0.9400
TLTCN-BLS	0.9487	0.0036	0.9510	0.0031	TLTCN-BLS	0.9487	0.0036	0.9510
OLTCN-TLTCN	0.9682	0.0042	0.9690	0.0039	OLTCN-TLTCN	0.9682	0.0042	0.9690
LTCN-BLS-FD	0.9776	0.0034	0.9783	0.0030	LTCN-BLS-FD	0.9776	0.0034	0.9783

Table 4 FD results of different models on CWRU dataset

Models	ACC		MP		MR		MF		Parameter quantity (M) mean value
	Mean value	SD							
R-O-BLS	0.9019	0.0020	0.9040	0.0030	0.9022	0.0040	0.9019	0.0020	0.9040
WDCNN	0.8824	0.0040	0.8889	0.0026	0.8819	0.0026	0.8824	0.0040	0.8889
ODCNN	0.8480	0.0035	0.8597	0.0045	0.8475	0.0045	0.8480	0.0035	0.8597
1-D DRN	0.8906	0.0036	0.8959	0.0023	0.8920	0.0023	0.8906	0.0036	0.8959
CWT-CNN	0.7883	0.0047	0.8012	0.0103	0.7878	0.0103	0.7883	0.0047	0.8012
Alex Net	0.8483	0.0045	0.8596	0.0031	0.8475	0.0031	0.8483	0.0045	0.8596
LTCN-BLS-FD	0.9315	0.0017	0.9363	0.0021	0.9308	0.0021	0.9315	0.0017	0.9363

Table 5 The ablation experimental results of different models on the CWRU dataset with introduced noise

<i>Models</i>	<i>ACC</i>		<i>MP</i>		<i>MR</i>		<i>MF</i>	
	<i>Mean value</i>	<i>SD</i>						
OLTCN-BLS	0.8592	0.0034	0.8615	0.0029	0.8592	0.0034	0.8615	0.0034
TLTCN-BLS	0.8430	0.0040	0.8474	0.0041	0.8430	0.0040	0.8474	0.0040
OLTCN-TLTCN	0.8913	0.0044	0.9128	0.0042	0.8913	0.0044	0.9128	0.0044
LTCN-BLS-FD	0.9315	0.0017	0.9363	0.0021	0.9315	0.0017	0.9363	0.0017

Table 6 The FD results of different models on the CWRU dataset with introduced noise

Models	ACC		MP		MR		MF		Parameter quantity (M)	
	Mean value	SD	mean value							
R-O-BLS	0.9019	0.0020	0.9040	0.0030	0.9022	0.0023	0.9009	0.0018	0.0346	
WDCNN	0.8824	0.0040	0.8889	0.0026	0.8819	0.0047	0.8790	0.0040	0.0153	
ODCNN	0.8480	0.0035	0.8597	0.0045	0.8475	0.0034	0.8444	0.0031	0.0429	
1-D DRN	0.8906	0.0036	0.8959	0.0023	0.8920	0.0040	0.8887	0.0044	0.0206	
CWT-CNN	0.7883	0.0047	0.8012	0.0103	0.7878	0.0048	0.7806	0.0064	5.3532	
Alex Net	0.8483	0.0045	0.8596	0.0031	0.8475	0.0046	0.8420	0.0047	56.4744	
LTCN-BLS-FD	0.9315	0.0017	0.9363	0.0021	0.9308	0.0016	0.9283	0.0022	0.0148	

Table 7 RMSE values of RUL predicted by different models

<i>Models</i>	<i>RMSE</i>	<i>MAE</i>	<i>R2</i>
TCAE-Res Net	0.2745	0.2238	0.0766
TCAETCN	0.1968	0.1405	0.5207
TCN-Res Net	0.1698	0.1391	0.6406
DCNN	0.1765	0.1505	0.6123
CNN-LSTM	0.1672	0.1359	0.6509
TLSTM-A	0.1681	0.1352	0.6474
TCAETCN-Res Net	0.1658	0.1259	0.6567

4.3 Analysis and discussion

In Figure 11(a), the 1-DLTCN branch and the 2-DLTCN branch converge quickly on the CWRU dataset before noise is introduced, and in Figure 11(b), the convergence speed of the 1-DLTCN branch is faster than that of the 2-DLTCN branch after noise is added.

In Table 3, the LTCN-BLS-FD framework achieves the best diagnostic results, and the average values of the four indicators all reaches above 0.9773. Compared with OLT-CN-TLTCN, which achieves the second-best result, the values of Acc, MP, MR, and MF are higher: 0.0094, 0.0093, 0.0091, and 0.0094. The diagnostic results of OLT-CN-BLS are the worst, and it only achieves an average of 0.9388 Acc, 0.94 MP, 0.9393 MR, and 0.9387 MF. Moreover, the SD values of each index are the highest compared with the other three models, all exceeding 0.0050.

The ITCN-BLS framework proposed in Table 4 achieves the best diagnostic results and has the lowest parameter quantity (0.0149 M). Compared with 1-DDRN, which achieves the worst diagnostic result, the four indicators of LTCN-BLS-FD are higher by 0.0409, 0.0405, 0.0388, and 0.0396, respectively.

In Table 5, only the mean Acc of LTCN-BLS-FD exceeds 0.9300 and has the lowest SD value. The diagnostic results of OLT-CN-TLTCN are second only to those of LTCN-BLS-FD, and it achieves an average of 0.8913 Acc, 0.9128 MP, 0.8915 MR, and 0.8794 MF, which are 0.0321, 0.0513, 0.0320, and 0.0225 higher than those of OLT-CN-BLS, respectively. In addition, the diagnostic results of TLTCN-BLS have the lowest indicators, and the means of the four indicators are 0.843, 0.8474, 0.8473 and 0.8368, respectively.

Compared with the results shown in Table 4 without introducing noise, the average indexes of the diagnosis results of each model in Table 6 have decreased, especially the average values of Acc, MR and MF of the CWT-CNN model have decreased below 0.8000, and the SD of its MP value is 0.0103, which has a high dispersion. In the presence of noise, the LTCN-BLS-FD framework still achieves the best FD results. Compared with the CWT-CNN model, the mean values of the four evaluation indexes are 0.1431, 0.1351, 0.1430 and 0.1477 higher, respectively.

The TCAETCN-Res Net framework in Table 6 achieves the best prediction results in ablation experiments and comparative experiments as a whole, and the RMSE predicted for RUL is 0.1658, MAE is 0.1259, and R2 is 0.6567. The above verifies the effectiveness of TCAETCN-Res Net in RUL prediction.

The FD performance of LTCN-BLS-FD is better than that of OLT-CN-BLS and TLTCN-BLS, which proves that the time series features and time-frequency features

extracted from 1-DLTCN branch and 2-DLTCN branch make essential contributions to the correct fault classification. In addition, the results of ablation experiments show that the mean values of evaluation indexes of diagnosis results of OLTCN-BLS and TLTCN-BLS are very similar, which can further conclude that the deep semantic information contained in time series features and time-frequency features is equally important for different types of fault descriptions. The results of ablation experiments show that the FD performance of LTCN-BLS-FD is better than that of LTCN-TLTCN, which proves that the ILAEN-based BLS classifier is more conducive to reducing the model complexity than the traditional FC classifier, and has better nonlinear mapping ability, which can more accurately establish the relationship between data features and labels. At the same time, by using noise-introduced data to conduct ablation experiments, the results show that LTCN-BLS-FD achieves the optimal FD structure. This demonstrates that the framework structure is effective in reducing the negative impact of external disturbances on model performance, thereby further supporting the above analysis.

It is difficult to run complex ml models on industrial IOT devices with limited processing capacity. Therefore, in the process of practical application, it can be modified by tinyml quantitative model (tensorflow Lite micro) + hardware acceleration (NPU) on the device side, combined with data caching and preprocessing through lightweight reasoning services at the edge, and parameter aggregation through large model training and federated learning in the cloud Through the above improvements, the operation effect of ML model on industrial IOT devices can be improved.

To adapt the machine learning fault prediction model across industries, the following strategies need to be adopted:

- 1 data standardisation and feature engineering, through unified data cleaning, normalisation processing, and extraction of industry common features and industry-specific features, to build a transferable data expression
- 2 transfer learning and modular design, based on the pre training general model, use fine-tuning technology to adapt to the target industry data, and disassemble the model into pluggable modules to reduce the reconstruction cost
- 3 dynamic parameter optimisation, combined with the difference of equipment failure mechanism in the industry, automatically adjusts the super parameters of the model through reinforcement learning
- 4 a cross industry validation framework was established to establish a test set containing multi domain fault scenarios.

The generalisation ability of the model was improved through confrontation training, and the industry prediction bias was corrected iteratively using boosting technology.

5 Conclusions

This paper conducts an in-depth exploration of the problems that need to be solved in the existing research on FD and life prediction of industrial equipment. From the perspective of engineering practice, this paper proposes a series of methods based on deep learning theory to solve the difficulties in sample imbalance between different fault categories,

one-sidedness in fault feature extraction, and optimisation of deep learning models. The wide learning classifier with strong nonlinear mapping ability and low complexity accurately identifies various types of faults through the feature fusion vector. In addition, this paper proposes the 1-D LTCN and 2-D LTCN model structures for the first time, which effectively reduces the complexity of the framework. The robustness of the framework is verified by experiments conducted by introducing additional noise into the data.

The hyperparameters of deep learning models, such as the number of neural network layers, are mostly determined based on empirical values and cannot be explained by theoretical methods. Therefore, exploring how to determine the hyperparameters of deep learning from a theoretical perspective is a major direction worthy of study.

Declarations

All authors declare that they have no conflicts of interest.

Acknowledgements

This research was supported by the Scientific Research and Innovation Team Construction Project of Hunan Railway Professional Technology College (No. KYTD202407) and the Scientific Research Project of the Hunan Provincial Department of Education (No. 24B1047).

References

- Cao, Y., Ding, Y., Jia, M. and Tian, R. (2021) 'A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings', *Reliability Engineering & System Safety*, Vol. 215, No. 1, pp.107813–107824.
- Cen, J., Yang, Z., Liu, X., Xiong, J. and Chen, H. (2022) 'A review of data-driven machinery fault diagnosis using machine learning algorithms', *Journal of Vibration Engineering & Technologies*, Vol. 10, No. 7, pp.2481–2507.
- Chen, M., Zhou, Z., Zhang, B., Hu, G. and Cao, Y. (2022) 'A novel combination belief rule base model for mechanical equipment fault diagnosis', *Chinese Journal of Aeronautics*, Vol. 35, No. 5, pp.158–178.
- Chu, J., Wang, Q., Liu, Y., Pan, J., Yuan, H., Yang, A. and Rong, M. (2022) 'Fault diagnosis of SF 6-insulated equipment by micro gas sensor array', *IEEE Transactions on Power Delivery*, Vol. 38, No. 1, pp.222–230.
- Dong, J., Su, D., Gao, Y., Wu, X., Jiang, H. and Chen, T. (2023) 'Fine-grained transfer learning based on deep feature decomposition for rotating equipment fault diagnosis', *Measurement Science and Technology*, Vol. 34, No. 6, pp.065902–065914.
- Ferreira, C. and Gonçalves, G. (2022) 'Remaining useful life prediction and challenges: a literature review on the use of machine learning methods', *Journal of Manufacturing Systems*, Vol. 63, No. 1, pp.550–562.
- Gao, Y., Gao, L., Li, X. and Cao, S. (2022) 'A hierarchical training-convolutional neural network for imbalanced fault diagnosis in complex equipment', *IEEE Transactions on Industrial Informatics*, Vol. 18, No. 11, pp.8138–8145.

- Ge, Y., Zhang, F. and Ren, Y. (2022) 'Adaptive fault diagnosis method for rotating machinery with unknown faults under multiple working conditions', *Journal of Manufacturing Systems*, Vol. 63, No. 2, pp.177–184.
- Guo, Z., Hao, Y., Shi, H., Wu, Z., Wu, Y. and Sun, X. (2023) 'A fault diagnosis algorithm for the dedicated equipment based on the CNN-LSTM mechanism', *Energies*, Vol. 16, No. 13, pp.5230–5243.
- Hu, J. and Deng, S. (2022) 'Rolling bearing fault diagnosis based on wireless sensor network data fusion', *Computer Communications*, Vol. 181, No. 1, pp.404–411.
- Ji, D., Wang, C., Li, J. and Dong, H. (2021) 'A review: data driven-based fault diagnosis and RUL prediction of petroleum machinery and equipment', *Systems Science & Control Engineering*, Vol. 9, No. 1, pp.724–747.
- Jiang, W., Wang, C., Zou, J. and Zhang, S. (2021) 'Application of deep learning in fault diagnosis of rotating machinery', *Processes*, Vol. 9, No. 6, pp.919–930.
- Kao, S.X. and Chien, C.F. (2023) 'Deep learning-based positioning error fault diagnosis of wire bonding equipment and an empirical study for IC packaging', *IEEE Transactions on Semiconductor Manufacturing*, Vol. 36, No. 4, pp.619–628.
- Li, D., Liu, S., Gao, F. and Sun, X. (2022) 'Continual learning classification method and its application to equipment fault diagnosis', *Applied Intelligence*, Vol. 52, No. 1, pp.858–874.
- Manikandan, S. and Duraivelu, K. (2021) 'Fault diagnosis of various rotating equipment using machine learning approaches – a review', *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, Vol. 235, No. 2, pp.629–642.
- Odeyar, P., Apel, D.B., Hall, R., Zon, B. and Skrzypkowski, K. (2022) 'A review of reliability and fault analysis methods for heavy equipment and their components used in mining', *Energies*, Vol. 15, No. 17, pp.6263–6275.
- Shi, L., Zhu, Y., Zhang, Y. and Su, Z. (2021) 'Fault diagnosis of signal equipment on the lanzhou-Xinjiang high-speed railway using machine learning for natural language processing', *Complexity*, Vol. 2021, No. 1, pp.9126745–9126756.
- Tao, H., Cheng, L., Qiu, J. and Stojanovic, V. (2022) 'Few shot cross equipment fault diagnosis method based on parameter optimization and feature metric', *Measurement Science and Technology*, Vol. 33, No. 11, pp.115005–115017.
- Wang, C., Sun, Y. and Wang, X. (2024) 'Image deep learning in fault diagnosis of mechanical equipment', *Journal of Intelligent Manufacturing*, Vol. 35, No. 6, pp.2475–2515.
- Wang, G., Liu, Y., Chen, X., Yan, Q., Sui, H., Ma, C. and Zhang, J. (2021a) 'Power transformer fault diagnosis system based on internet of things', *Eurasip Journal on Wireless Communications and Networking*, Vol. 2021, No. 3, pp.1–24.
- Wang, J., Wang, D., Wang, S., Li, W. and Song, K. (2021b) 'Fault diagnosis of bearings based on multi-sensor information fusion and 2D convolutional neural network', *IEEE Access*, Vol. 9, No. 3, pp.23717–23725.
- Wang, Y., Yan, J., Sun, Q., Zhao, Y. and Liu, T. (2021c) 'ShuffleNet-based comprehensive diagnosis for insulation and mechanical faults of power equipment', *High Voltage*, Vol. 6, No. 5, pp.861–872.
- Zhou, H., Chen, W., Shen, C., Cheng, L. and Xia, M. (2023) 'Intelligent machine fault diagnosis with effective denoising using EEMD-ICA-FuzzyEn and CNN', *International Journal of Production Research*, Vol. 61, No. 23, pp.8252–8264.