Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 20(8): 14117–14135.
DOI: 10.3934/mbe.2023631
Received: 23 April 2023
Revised: 28 May 2023
Accepted: 16 June 2023
Published: 25 June 2023

Research article

PSO-MCKD-MFFResnet based fault diagnosis algorithm for hydropower units

Xu Li*, Zhuofei Xu and Yimin Wang

State Key Laboratory of Eco-hydraulics in Northwest Arid Region of China, Xi'an University of Technology, Xi'an 710048, Shaanxi, China

* **Correspondence:** Email: 43512116@qq.com.

Abstract: Due to the coupling effect of external environmental noise and vibration noise, the feature rate of the original hydroelectric unit fault signal is not prominent, which will affect the performance of fault diagnosis algorithms. To solve the above problems, this paper proposes a PSO-MCKD-MFFResnet algorithm for fault diagnosis of hydropower units (Particle swarm optimization, PSO; Maximum correlation kurtosis deconvolution, MCKD; Multi-scale feature fusion residual network, MFFResnet). In practical applications, the selection of key parameters in the traditional MCKD method is heavily dependent on prior knowledge. First, this paper proposes a PSO-MCKD enhancement algorithm for fault features, which uses the PSO algorithm to search for the influencing parameters of MCKD to enhance the features from the original fault signal. Second, a fault feature diagnosis algorithm based on MFFResnet is proposed to improve the utilization of local features. The multi-scale residual module is used to extract features at different scales and then put the enhanced signal into MFFResnet for training and classification. The experimental results show that our approach can accurately and effectively classify the fault types of hydropower units, with an accuracy rate of 98.85%. It is superior to other representative algorithms in different indicators and has a good stability.

Keywords: hydropower unit; fault diagnosis; maximum correlation kurtosis deconvolution; multi-scale feature fusion residual network

1. Introduction

Hydropower units are the core key equipment of hydropower stations, and their health condition directly affects the safety and stability of the entire hydropower stations. In recent years, as the size and capacity of hydropower units have increased, their structures have become more and more complex. The resulting problems in terms of safety and stability have become more and more prominent [1]. Vibration is the most common and important factor to measure the stable operation of a unit, and most

failures are caused by unit vibration [2]. Research on the vibration features of the unit, accurately extracting the features of the vibration signal, effectively evaluating the unit status and providing early warning of faults, and then achieving fault diagnosis and timely troubleshooting and maintenance, can extend the unit life and improve maintenance efficiency.

However, the troubleshooting of hydropower units faces many difficult problems. On the one hand, the multi-source noise interference from the external environment and equipment makes the signal signal-to-noise ratio low and the fault features weak. On the other hand, when local damage or defects occur in rotating machinery, a high-frequency impulse response is caused during rotation, and the inherent frequency of each component is easily excited, which leads to vibration noise coupling of components. Therefore, how to effectively enhance the weak fault feature of rotating machinery has received more and more attention [3, 4].

There are many traditional signal processing methods and their improvements, including Fourier transform [5], wavelet transform (WT) [6, 7] and empirical modal decomposition (EMD) [8, 9]. As a pre-processing for fault diagnosis, it highlights the periodic pulse of the fault signal, which is more conducive to the extraction of subsequent feature signals. However, when the vibration signal is affected by complex background noise and multiple faults in practical applications, the effectiveness of the above methods will be affected. The maximum correlation kurtosis deconvolution (MCKD) algorithm [10] takes the maximum correlation kurtosis value of the signal as the objective function. It can realize the deconvolution operation in an iterative manner, and highlight the periodic impact components in the fault signal. Li et al. [11] proposed a fault feature extraction method based on MCKD and improved empirical wavelet transform (EWT), which has been validated in fault feature extraction and enhancement. Hua et al. [12] combined maximum correlation kurtosis deconvolution and envelope spectrum to extract fault feature information of a gearbox and successfully realized fault pattern recognition of the gearbox in a strong noise environment. Wang et al. [13] used maximum correlation kurtosis deconvolution to denoise fault signals, and further empirical mode decomposition was performed on the obtained signals to extract feature vectors. Finally, combined with support vector machine (SVM), they successfully realized the recognition and classification of composite faults. However, there are still some problems with the above method. First of all, the input parameters of MCKD need to be set manually. Only by properly selecting the filter length parameter L and the deconvolution period parameter T can one guarantee the effectiveness of the algorithm. Secondly, the process of extracting fault features by traditional algorithms is complicated and requires certain prior knowledge. Moreover, the pattern recognition methods such as SVM have limited ability to express complex functions. It is difficult to fully mine the inherent information in fault signals, and the generalization ability of the model is weak.

Deep learning-based models as an efficient pattern recognition network have brought broad prospects for the development of intelligent fault diagnosis methods [14–16]. This type of method is data-driven and combines feature learning with intelligent recognition. Compared with traditional methods, it eliminates the dependence on signal preprocessing and expert knowledge, and it has significant advantages when analyzing massive monitoring data [17]. Deep belief network (DBN) [18], autoencoder (AE) [19] and convolutional neural network (CNN) [20], as representative deep learning methods, have shown their advantages in fault diagnosis of rotating machinery. Sun et al. proposed a feature extraction method with sparse stacked denoising autoencoders for fault diagnosis [21]. However, the model has many parameters, which makes the training process difficult and time-consuming.

Convolutional neural networks make the model complexity and the risk of overfitting greatly reduced due to the local perceptual field and weight-sharing strategies. Hoang et al. [22] established a two-dimensional CNN model for rolling bearing fault diagnosis based on preprocessing of one-dimensional vibration signals. Liao et al. [23] proposed a fault diagnosis system for hydropower units based on one-dimensional CNN and gated cyclic units based on sequential data structures. Wang et al. [24] used one-dimensional CNN to process industrial signal features and acoustic signal features. The model is used for fault diagnosis under different environmental noise conditions. Song et al. [25] used a strategy of expanding the convolution kernel to obtain a larger perceptual field in order to improve the recognition rate. This method adopts the wide kernels of the first two convolutional layers to quickly extract features to improve efficiency. The smaller convolution kernels are used for multi-layer nonlinear mapping to deepen the network and improve detection accuracy. Jia et al. [26] proposed a new end-to-end CNN model, called Gramian Time Frequency Enhancement Network (GTFE-Net), to address the lack of denoising structure in traditional CNNs. This method incorporates a Gramian-based noise reduction strategy, called Gramian noise reduction (GNR), for bearing fault diagnosis.

Although CNN has achieved great success in the field of fault diagnosis, it is difficult to effectively train deep models because the gradients of CNN may disappear or explode during the process of model optimization. In addition, CNN cannot accurately identify complex nonlinear signals when training deep neural networks. The emergence of Resnet [27] avoids the shortcomings of deep networks in training. Resnet has now been widely used in fault diagnosis. Zhang et al. [28] proposed a Resnet-based rolling bearing fault diagnosis model, which improved the training process of deep neural networks. This model provides a new idea for the intelligent diagnosis of rotating machinery faults. Wang et al. [29] created 1D original signals as 2D images by continuous wavelet transform (CWT). The ResNet composed of 2D convolution was combined with CWT for diagnosis experiments. Jin et al. [30] proposed a decoupled attention Resnet using attention blocks to selectively enhance important feature maps, improving diagnosis accuracy. However, the residual structure of the original Resnet network using the above method only contains a single-size convolution, which cannot fully extract local features of the signal. To solve the above problems, we first use an improved MCKD method to enhance the original signal's features, and we then input the enhanced signal into a multi-scale feature fusion residual network for fault feature identification. In this paper, we abandon the design idea of a fixed size of convolutional kernels in traditional residual blocks and use different scale convolutional kernels to enhance the accuracy and robustness of the algorithm. In addition, Chen et al. [31] aimed at the problem of fault diagnosis under low signal-to-noise ratio, and a new fault diagnosis framework based on a Multi-scale Split Dual Calibration Network with Periodic Information (PI-MSDCN) was proposed. In the fault diagnosis framework, a periodic block is constructed to automatically learn the periodic information of vibration signals through the neural network. Chen et al. [32] proposed a dual-path mixed-domain residual threshold network (DP-MRTN), which aims to improve the accuracy of the rolling bearing fault diagnosis in a noisy environment. This method combines attentional mechanisms as well as residual structures, and the dilated convolution is introduced to establish a dual-path neural network. Subsequently, to enable deep learning frameworks to perform more flexible nonlinear transformations for different input data, Chen et al. [33] developed a new activation function, the parameter-free adaptively Swish (PASwish). PASwish formulates different activation schemes for different input data so that vibration data under different operating conditions can carry out adaptive nonlinear transformation, and the generalization ability of the whole network is improved.

The contributions of this paper are as follows:

- For the original signal affected by factors such as noise, there will be problems with inconspicuous fault features, and this paper proposes a PSO-MCKD fault feature enhancement algorithm. In order to improve the effectiveness and efficiency of MCKD algorithm, the PSO algorithm is used to globally optimize parameters to enhance the original signal. Then, the enhanced signal is input into the fault diagnosis algorithm to improve the effectiveness in features extraction.
- To further improve the accuracy of the fault diagnosis algorithm, a fault feature diagnosis algorithm is proposed based on a multi-scale feature fusion residual network (MFFResnet). This method is an end-to-end model based on deep learning. MFFResNet incorporates convolution at different scales on top of the residual structure. A jump connection is used between each layer to improve the local feature extraction ability of the single size residual structure, so as to improve the accuracy and stability of fault diagnosis.

2. Proposed methods

2.1. PSO-MCKD fault feature enhancement algorithm

2.1.1. MCKD algorithm

In the MCKD algorithm, it is assumed that the fault vibration signal of the hydropower unit is collected by sensors as in Eq (2.1).

$$y(n) = h(n) * x(n) + e(n) \tag{2.1}$$

where $y(n)$ is the collected original vibration signal; $h(n)$ is the response of system transmission; $x(n)$ is the vibration fault signal part; $e(n)$ is the system noise component.

The MCKD algorithm uses the maximum correlation kurtosis as an index to optimize and find an optimal filter $y(n)$ to achieve better recovery of the vibration fault signal $x(n)$. The filter expression is Eq (2.2).

$$y(n) = \sum_{k=1}^{L} f_k x_{n-k+1} \tag{2.2}$$

where $f = [f1, f2, f3, ..., fL]$; f is the filter coefficient, and L is the length of the filter.

The larger the correlation kurtosis, the more sensitive it is to the periodic characteristics of the signal and the more obvious the recovery of the shock characteristics of the signal. The correlation kurtosis formula is shown in Eq (2.3).

$$CK_M(T) = \frac{\sum_{n=1}^{N}(\prod_{m=0}^{M} y_{n-mT})^2}{(\sum_{n=1}^{N} y_n^2)^{M+1}} \tag{2.3}$$

where M is the number of shifts; T is the period of the impact signal; N is the number of samples of the input signal. When the relative kurtosis index is maximized through iteration, the obtained filter is optimal, and the recovery of the shock signal is better, as shown in Eq (2.4).

$$\max_f CK_M(T) = \max_f \frac{\sum_{n=1}^{N}(\prod_{m=0}^{M} y_{n-mT})^2}{(\sum_{n=1}^{N} y_n^2)^{M+1}} \tag{2.4}$$

According to Eq (2.3), perform a derivative operation on f_k so that the derivative value is 0, and get Eq (2.5):

$$\frac{d}{df_k}CK_M(T) = 0, (k = 1, 2, 3, ..., L) \tag{2.5}$$

The matrix form of the result obtained by continuously iteratively solving Eq (2.5) is Eq (2.6), which is the order of the optimal filter.

$$f = \frac{||y||^2}{2||\beta||^2}(x_0 x_0^T)^{-1} \sum_{m=0}^{M} X_{mT}\alpha_m \tag{2.6}$$

where each of these parameters is shown below.

$$\alpha_m = \begin{pmatrix} y_{1-mT}^{-1}(y_1^2 y_{1-T}^2 \cdots y_{1-MT}^2) \\ y_{2-mT}^{-1}(y_2^2 y_{2-T}^2 \cdots y_{2-MT}^2) \\ \vdots \\ y_{N-mT}^{-1}(y_2^2 y_{N-T}^2 \cdots y_{N-MT}^2) \end{pmatrix}, \beta = \begin{pmatrix} y_1 y_{1-T} \cdots y_{1-MT} \\ y_2 y_{2-T} \cdots y_{2-MT} \\ \vdots \\ y_N y_{N-T} \cdots y_{N-MT} \end{pmatrix}, X_r = \begin{pmatrix} X_{1-r} & X_{2-r} & \cdots & X_{N-r} \\ 0 & X_{1-r} & \cdots & X_{N-1-r} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_{N-L-r+1} \end{pmatrix} \tag{2.7}$$

Before the data is processed by the MCKD algorithm, the parameter T can be calculated from the signal sampling frequency and the fault feature frequency, and the formula is shown in Eq (2.8).

$$T = \frac{f_1}{f_2} \tag{2.8}$$

where f_1 is the sampling frequency of the signal, and f_2 is the fault feature frequency of the bearing.

2.1.2. PSO-MCKD algorithm

The PSO algorithm [34] is an evolutionary computing technology. The basic idea of the PSO algorithm is to find the optimal solution through cooperation and information sharing among individuals in the group. Particles only have two attributes: velocity v and position x. Each particle searches for the optimal solution separately in the search space. It judges whether the particle is good or bad according to the set fitness function, and records it as the current individual extremum (pbest). The model uses the crest factor of envelope spectrum (E_c), as the fitness function. Assuming that the sequence of envelope spectral amplitudes is $X(z)$, $(z = 1, 2, ..., Z)$, E_c can be expressed as:

$$E_c = \frac{max(X(z))}{\sqrt{\sum_z X(z)^2/Z}} \tag{2.9}$$

where $X(z)$ is the amplitude of the envelope spectrum in the whole frequency range.

The individual extremum is shared with other particles in the entire particle swarm, and find the optimal individual extremum is found as the current globally optimal particle of the whole particle population (gbest). The updating formulas of particle velocity v and position x are Eqs (2.10) and (2.11).

$$v_i^{k+1} = wv_i^k + c_1\eta_1(pbest_i^k - x_i^k) + c_2\eta_2(gbest_i^k - x_i^k) \tag{2.10}$$

$$x_i^{k+1} = x_i^k + v_i^k \tag{2.11}$$

where w is the inertia weight, and c_1, c_2 are the learning factors. η_1, η_2 are random numbers in the range of [0, 1].

When the inertia weight is greater than 1, the global search ability of the PSO algorithm is stronger, and when it is less than 1, the local search ability of the algorithm is stronger. The learning factor also has an impact on the searchability of the algorithm. In the early stage of the search, the global search ability of the particles is required to be strong, and the learning factor needs to be set larger. In the later stage of the search, it is necessary to set smaller learning factors to obtain stronger local search ability. Therefore, the inertia weight w and the learning factors c_1, c_2 need to be adjusted dynamically. The inertia weight w changes according to the change of the particle fitness value. The learning factors c_1 and c_2 are adjusted according to the number of iterations. The dynamic adjustment formulas are as Eqs (2.12) and (2.13).

$$w = \begin{cases} w_{min} - (w_{max} - w_{min})\dfrac{f - f_{min}}{f_{avg} - f_{min}}, & f \le f_{avg} \\ w_{max} \end{cases} \tag{2.12}$$

$$\begin{cases} c_1 = c_{1max} - c_{1min}\dfrac{t}{T} \\ c_2 = c_{2max} - c_{2min}\dfrac{t}{T} \end{cases} \tag{2.13}$$

where w_{min}, w_{max} are the maximum and minimum weights. f is the current fitness value, f_{avg} is the average fitness value, and f_{min} is the minimum fitness value. c_1 and c_2 are the range of the learning factor. T is the maximum number of iterations, and t is the current number of iterations.

Since the parameter deconvolution period T can be obtained from Eq (2.8), the filter length L and the shift number M are used as the parameters for particle optimization. When a fault occurs in a hydropower unit, a periodic shock signal is generated at a certain site, and the corresponding fault feature frequency and its multiplier frequency are seen in the frequency domain. The ratio of the energy of the fault feature frequency domain and its multiples to the total energy is used as the fitness function, and a larger ratio indicates that more fault features are extracted.

The flow chart of the PSO-MCKD algorithm is shown in Figure 1. First, initialize the parameters of the PSO algorithm. Set the inertia weight w, and the range of the learning factor c. Initialize the particle population, set the initial position and velocity of the particles, and use the two-dimensional array of parameters L and M as the properties of the particles. Then, calculate the fitness function of each particle, find out the locally optimal particle and the globally optimal particle, continuously update the weight w and learning factors c_1 and c_2 for iteration, and update the particle position and speed. If the iteration termination condition is met, the last globally optimal particle is the optimal parameter $[L, M]$.

2.2. Fault feature diagnosis algorithm based on multi-scale feature fusion residual network

In this paper, the Resnet network is used as the backbone network for fault feature diagnosis of hydropower units. Resnet designs a residual structure based on the convolutional neural network, which effectively solves the gradient disappearance problem of the deep convolutional neural network. The feature-enhanced vibration signal is input into the Resnet network to detect fault features, and to further improve the accuracy and stability of fault diagnosis.

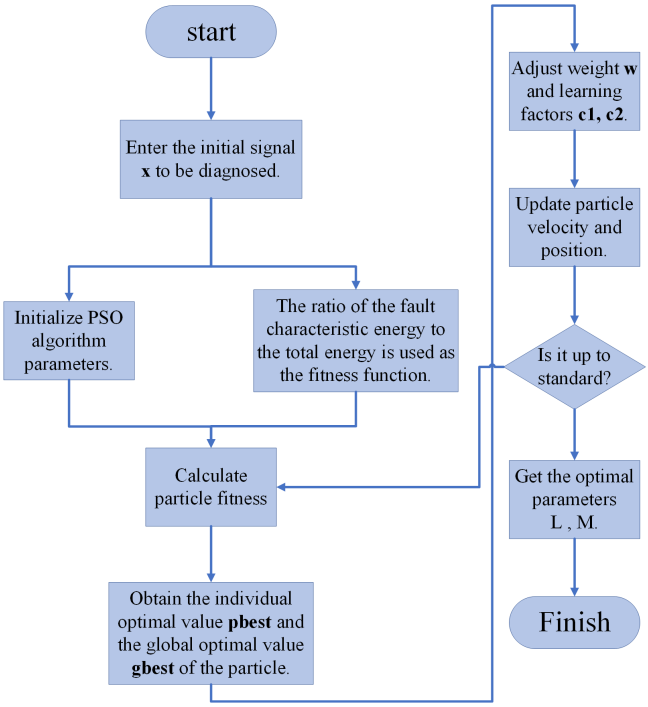

Figure 1. Flow chart of PSO-MCKD algorithm.

2.2.1. Residual network

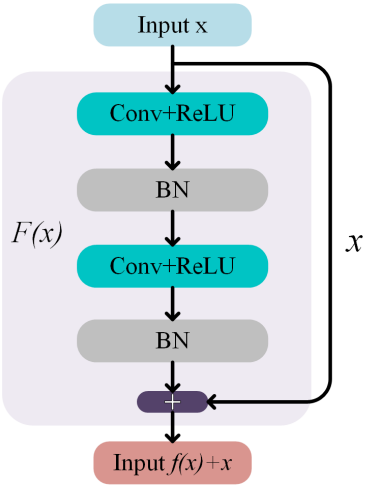

Figure 2. Structure diagram of the residual module.

The residual neural network introduces the concept of residual learning, which learns residual features by multiple residual blocks connected at the beginning and the end, and the residual block structure is shown in Figure 2, where x is the input of the residual network, $H(x)$ is the output, and $F(x)$ is the residual mapping function. The residual learning module is mainly composed of convolutional layer, BN layer and rectified linear unit (ReLU) activation function. The residual neural network adds

cross-layer fitting by learning the residual between the output and the input, and fitting them together, on top of the regular deep convolutional network. The fitted residual mapping function is

$$F(x) = H(x) - x. \tag{2.14}$$

Equation (2.14) is much easier than fitting a constant mapping function $H(x) = x$, and the underlying error can be passed to the upper layer through a shortcut connection during training. In addition to using the gradient of the objective function, the residual gradient is also added during the training process, so the residual neural network has a stronger feature learning ability while having a deeper number of layers.

2.2.2. Multi-scale feature fusion residual blocks

Fault signals of hydropower units are typically nonlinear and nonsmooth signals, and the fault information exists in different sensitive frequency bands of the signal. Among traditional feature extraction methods, multi-scale fault features can better characterize mechanical fault information, e.g., multi-scale fuzzy entropy [35], multi-scale morphological filtering [36], multi-scale wavelet analysis [37], etc. However, in the existing Resnet, the size of the convolution kernel in the residual block is fixed, so only the fault feature information at a specific scale can be extracted. The different scales of convolution operations correspond to the frequency resolutions of the corresponding results also being different. Therefore, by learning the convolutional layers at different scale sizes, more comprehensive feature information can be extracted, which in turn can characterize the fault features more accurately. To address the above drawbacks, we adopt a multi-scale residual structure, which is different from the multi-scale cascaded convolutional neural networks (MC-CNNs) proposed by Huang et al. [38]. Instead of placing the multi-scale convolutional layers on top of the network alone, the method introduces them into the residual module, replacing the original fixed-scale convolutional layers. At the same time, jump connections are used between features at different scales, enabling feature information to be shared and reused, taking into account the advantages of cross-layer constant mapping and multi-scale feature extraction.

In addition, the end 1×1 convolutional layer can be used as a bottleneck layer to facilitate feature fusion, reduce computational complexity and ease the difficulty of training. As shown in Figure 3, our multi-scale residual structure consists of two parts: multi-scale feature fusion and local residual learning. The detailed parameters are shown in Table 1.

Multi-scale feature fusion: We construct a network of bilateral paths, with different convolutional kernels for different paths. In this way, the information between these paths can be shared to be able to detect image features at different scales. The structure can be expressed as follows:

$$\begin{aligned} X_1 &= \varepsilon(c^1_{3\times3} * f_{n-1} + b^1), \\ Y_1 &= \varepsilon(c^1_{5\times5} * f_{n-1} + b^1), \\ X_2 &= \varepsilon(c^1_{3\times3} * [X_1, Y_1] + b^2), \\ Y_2 &= \varepsilon(c^1_{5\times5} * [X_1, Y_1] + b^2), \\ X^{'} &= c^1_{1\times1} * [X_2, Y_2] + b^3 \end{aligned} \tag{2.15}$$

where c and b represent the weight and bias respectively, the superscript represents the number of layers they are in, and the subscript represents the size of the convolution kernel used for that layer. $\varepsilon(x)$ represents the ReLU activation function, and $[X, Y]$ represents the skip connection operation.

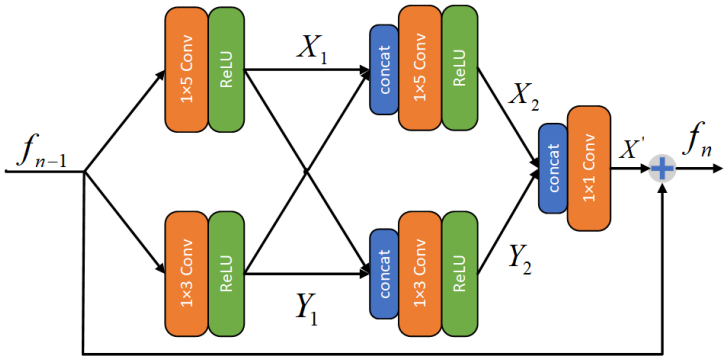

Figure 3. Structure diagram of multi-scale feature fusion residual module.

Local residual learning: To make the network more efficient, we used residual learning for each multi-scale module. We can formulate the multi-scale residual block (MSRB) as follows:

$$f_n = X^{'} + f_{n-1} \tag{2.16}$$

where f_n and f_{n-1} represent the input and output of the multi-scale residual module, respectively. It is worth mentioning that local residual learning is used, resulting in a significant reduction in computational complexity. At the same time, the performance of the network is also improved.

Table 1. Parameters of multi-scale feature fusion residual module.

Layer type	filters	size	padding
Conv1D(1-3)	32	3	✓
Conv1D(1-5)	32	5	✓
Conv1D(2-3)	64	3	✓
Conv1D(2-5)	64	5	✓
Conv1D(1-3)	64	1	✓

The proposed multi-scale feature fusion residual network structure in this paper is shown in Figure 4, consisting of a large-scale convolutional layer; AveragePooling layer; Convolution layer with ReLU activation function; MSFF module, a GAP (global average pooling) layer; and a Dense (dense classification) layer. The resonant waveform and shock period are important factors to characterize the failure state of the component, and a large-scale convolutional layer is used in the first layer of the network to extract the longer period fault shock features. The resonant waveform and shock period are important factors to characterize the failure state of the component, and a large-scale convolutional layer is used in the first layer of the network to extract the longer period fault shock features. This article uses 32 convolution kernels with a stride size of 1 and a size of 1×100 to perform convolution operations on the input data. For the impact of convolution kernels of different sizes, we will do further comparative analysis in the later part of the experiment. The detailed parameters of the network are shown in Table 2.

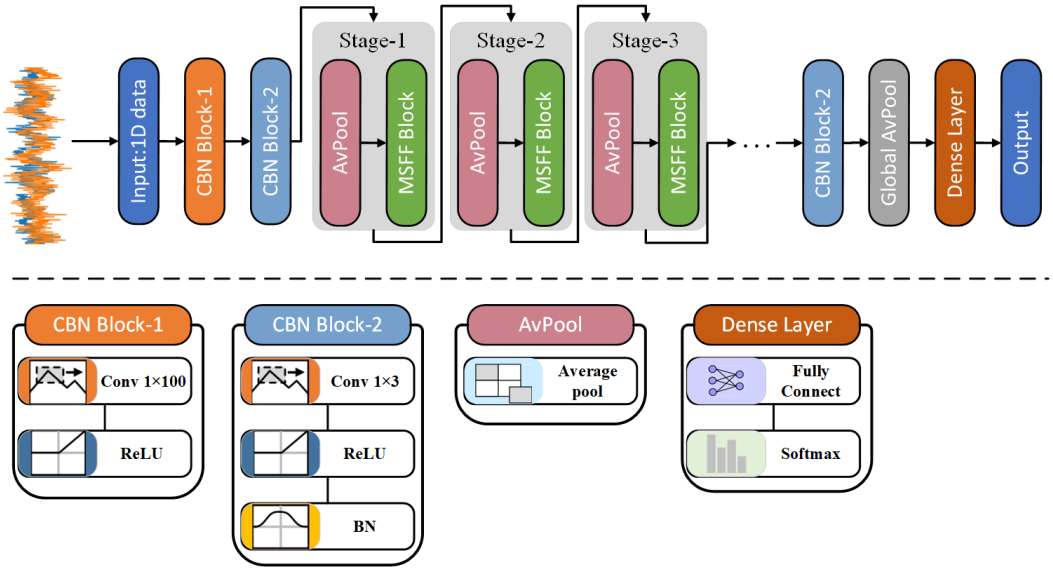

Figure 4. Structure diagram of multi-scale feature fusion residual network.

Table 2. Parameters of multi-scale feature fusion residual network.

Layer type	filters	size
Conv1D	32	100
Conv1D	64	3
Avpooling	-	2
MSFF Block	64	5
Conv1D	64	3
GAvpooling	-	-
Dense	-	256
Dropout	-	0.5

Subsequently, the network model adopts 5 MSFF modules. If the number of residual blocks is too small, the feature information of the data cannot be fully extracted. If the number is too large, the network structure will be complicated, and the operating efficiency will decrease. The residual network model proposed in this paper has strong flexibility in structure, and an appropriate number of residual modules can be selected to build the network model according to the different analyzed data.

3. Experiment

3.1. Datasets

In this paper, the publicly available experimental dataset of hydropower units from Wuhan University is used for analysis and discussion. This dataset uses the GTS3-TG series simulator to simulate the vibration signals of hydropower units. The vibration signals of the four rotor states were

set in the experiment as shown in Figure 5: (a) is normal, (b) is contact-rubbing, (c) is unbalanced, and (d) is misalignment.

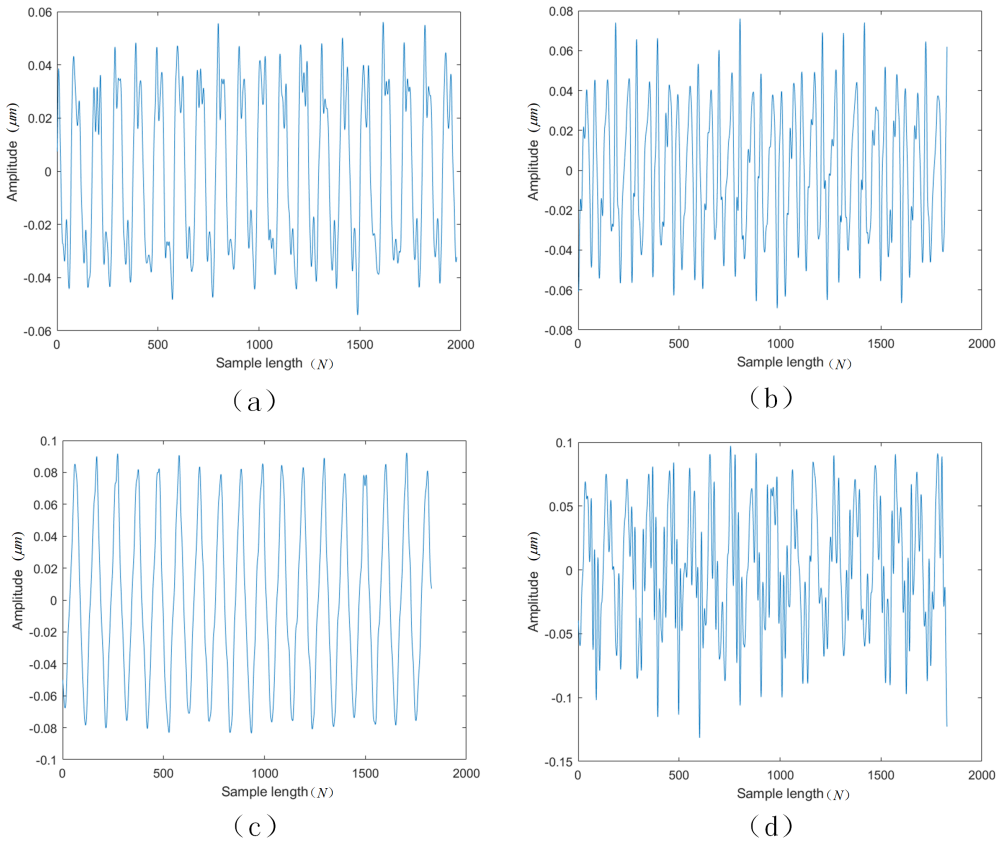

Figure 5. Example diagram of the vibration signal of the hydropower unit.

For each state, we performed 45 repeated experiments, and the stationary vibration signals collected for 1 s were recorded in each experiment to form a data set for analysis.

Table 3. Data set of the four hydropower unit signals.

State	Training set	Test set	Category label
Normal	2048	400	1
Contact-rubbing	2048	400	2
Unbalanced	2048	400	3
Misalignment	2048	400	4

The training set and test set of the four hydropower unit signals are shown in Table 3. We used 2048 sampling points as training data, and 400 sampling points were reserved for testing data to verify the accuracy of the classification. The CPU used in this experiment was an Intel(R) Core(TM) i9-9820XCPU@3.30 GHz, and the GPU was an Nvidia GeForce RTX 2080Ti.

3.2. Fault feature enhancement algorithm experiment

Initialize the PSO algorithm parameters, the population size $N = 30$; the weight range $[w_{min}, w_{max}] = [0.4, 0.9]$; $[c_{min}, c_{max}] = [0.5, 2]$; the initial value $c_1 = c_2 = 2$; the number of iterations $K = 20$; the search space of parameters L and T to be optimized is [100, 1000], [85, 142], and the displacement number is set to $M = 3$. The optimal parameters of the four signals obtained by PSO-MCKD algorithm are shown in Table 4.

Table 4. Optimization results of PSO-MCKD algorithm.

State	L	T	M	K
Normal	366	87	3	20
Contact-rubbing	165	99	3	20
Unbalanced	296	102	3	20
Misalignment	253	102	3	20

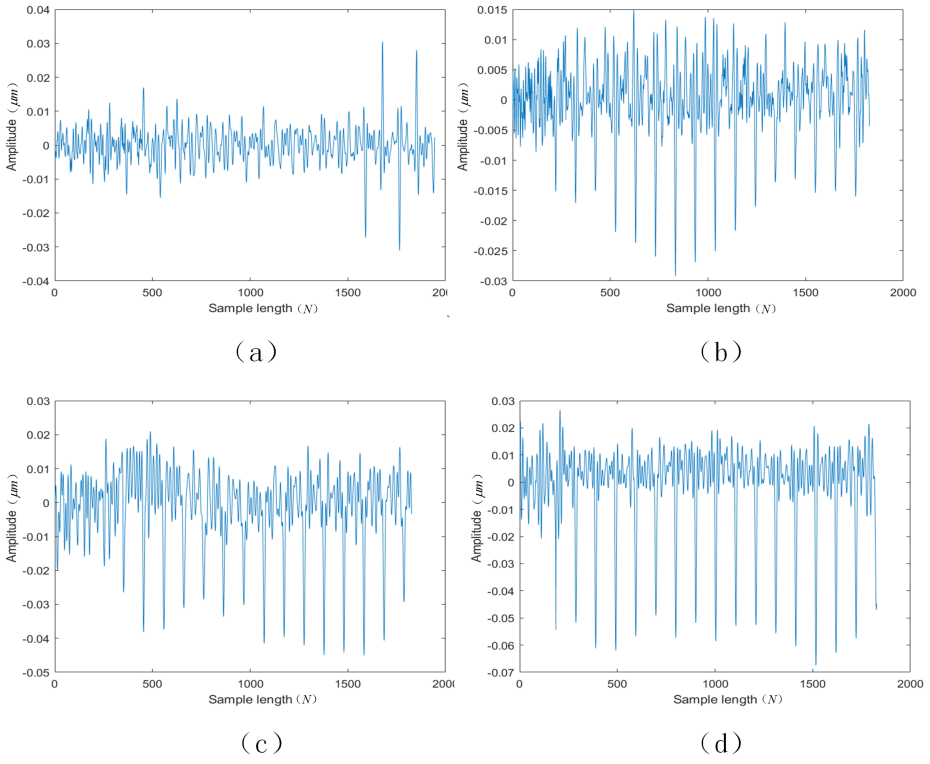

Figure 6. PSO-MCKD algorithm effect example diagram.

The optimized MCKD algorithm was used to process the vibration data of the hydropower unit, and the processed vibration signals are shown in Figure 6, where the ones in (a)–(d) are the four types of enhanced vibration signals. Compared with Figure 5, it can be seen that the vibration shock signal is more obvious, and the algorithm effectively solves the shortcomings of the original signal features not being obvious. The enhanced signal is more conducive to the subsequent feature extraction work.

3.3. Fault feature diagnosis algorithm experiment

In this study, experiments were conducted using the same hydropower unit dataset. We used the four indicators Precision, Recall, F1score and Accuracy to evaluate the fault diagnosis ability of different methods [39]. The formulas are expressed as:

$$Precision = \frac{TP}{TP + FP} \tag{3.1}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.2}$$

$$F1score = \frac{2TP}{2TP + FP + FN} \tag{3.3}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.4}$$

where TP is the number of true positive samples, FP is the number of false positive samples; FN is the number of false negative samples, and TN is the number of true negative samples.

The first convolutional layer in CNN is equivalent to a set of bandpass filters. Convolution kernels of different sizes can extract feature information in different frequency bands in the signal. The larger the convolution kernel size is, the wider the frequency band that can be extracted. Therefore, this paper analyzes the influence of the size of the convolution kernel in the first layer convolution of the proposed network on the performance of the network. The first-layer convolution kernel size was set to 1×3, 1×5, 1×9, 1×25, 1×49, 1×100, 1×200, 1×300, and the test results are shown in Table 5.

Table 5. Experimental results of first-layer convolution sizes affecting diagnosis accuracy.

Conv Size	Precision	Recall	F1score	Accuracy
1×3	97.15	97.10	96.25	96.13
1×9	96.92	96.83	96.71	96.85
1×25	97.70	97.68	97.61	97.87
1×49	98.25	98.12	98.01	98.21
1×100	**98.89**	**98.86**	**98.78**	**98.85**
1×200	97.61	97.57	97.51	97.67
1×300	96.79	96.76	96.58	96.75

As shown in Figure 7, with the increase of the convolution size of the first layer, the diagnosis accuracy shows a trend of increasing slowly. When the size reaches 1×100, the various indicators of the network reach the peak. If we continue to increase the size of the convolution, the accuracy will not continue to increase, but there will be a downward trend.

To demonstrate the effectiveness of the PSO-MCKD fault feature enhancement method, the method was applied to CNN, Resnet, and our MFFResnet method for further validity analysis, and the results are shown in Table 6.

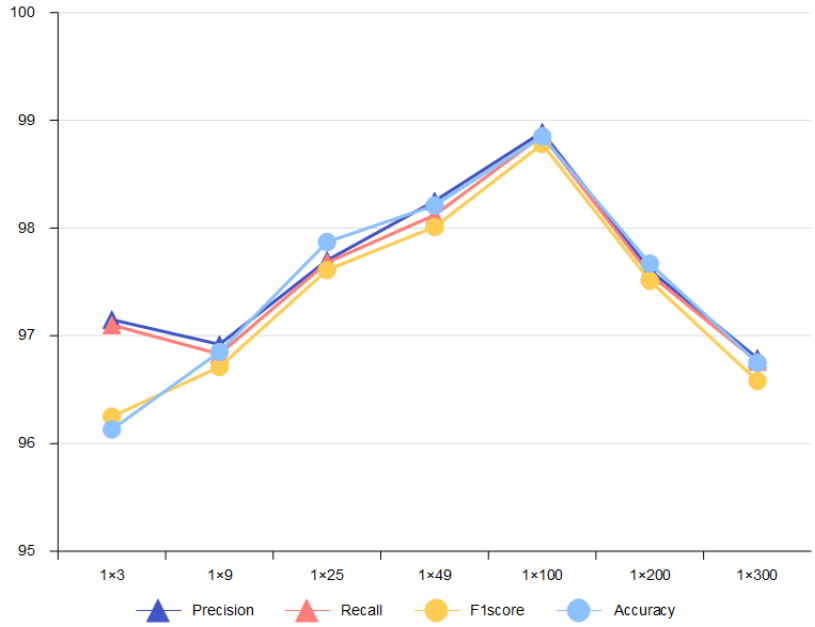

Figure 7. Diagnosis accuracy curves for different first-layer convolution sizes.

Table 6. Effectiveness of PSO-MCKD fault feature enhancement algorithm.

Method	Precision	Recall	F1score	Accuracy
CNN	93.70	93.28	92.11	92.87
PSO-MCKD-CNN	94.15	93.89	93.51	93.91
ResNet	95.82	95.61	95.18	96.78
PSO-MCKD-ResNet	97.29	97.26	97.15	97.28
MFFResNet	97.75	97.66	97.28	97.65
PSO-MCKD-MFFResNet	**98.89**	**98.86**	**98.78**	**98.85**

Table 7. Comparative experiments of diagnosis results on various models.

Method	Precision	Recall	F1score	Accuracy
RF [35]	80.13	83.78	80.01	80.11
SAE [19]	91.61	91.05	89.66	90.50
CNN [20]	93.70	93.28	92.11	92.87
WKCNN [25]	96.80	96.28	95.71	96.67
GTFE-Net [26]	98.21	97.82	97.71	98.10
ResNet [28]	95.82	95.61	95.18	96.78
CWT-Resnet [29]	98.01	97.81	97.63	97.97
A-ResNet [30]	97.89	97.76	97.75	97.85
PSO-MCKD-MFFResNet (ours)	**98.89**	**98.86**	**98.78**	**98.85**

It can be seen from the above table that different methods use the signal data processed by the PSO-MCKD feature enhancement method for training to effectively improve the accuracy of the final fault diagnosis. CNN increased by 1.04%, Resnet increased by 0.5%, and our approach improved by 1.2%.

This paper compares experimental results with eight other algorithms, including RF [40], SAE [19], CNN [20], WKCNN [25], GTFE-Net [26], ResNet [28], CWT-Resnet [29] and A-ResNet [30]. The comparative experiments of diagnosis results for various models are shown in Table 7. In this dataset, the diagnosis accuracy of RF, SAE, and CNN is poor. With the introduction of the attention mechanism, the diagnosis accuracy of A-Resnet is slightly higher than that of Resnet. Resnet with CWT preprocessed data improves overall diagnosis accuracy. Our approach takes the data through fault enhancement and then feeds it into a multiscale feature fusion residual network, which achieves the best diagnosis performance on the dataset.

Figure 8 shows the visualization figures of the four metrics for different methods, and it can be seen that our approach is higher than the other six algorithms for different metrics.

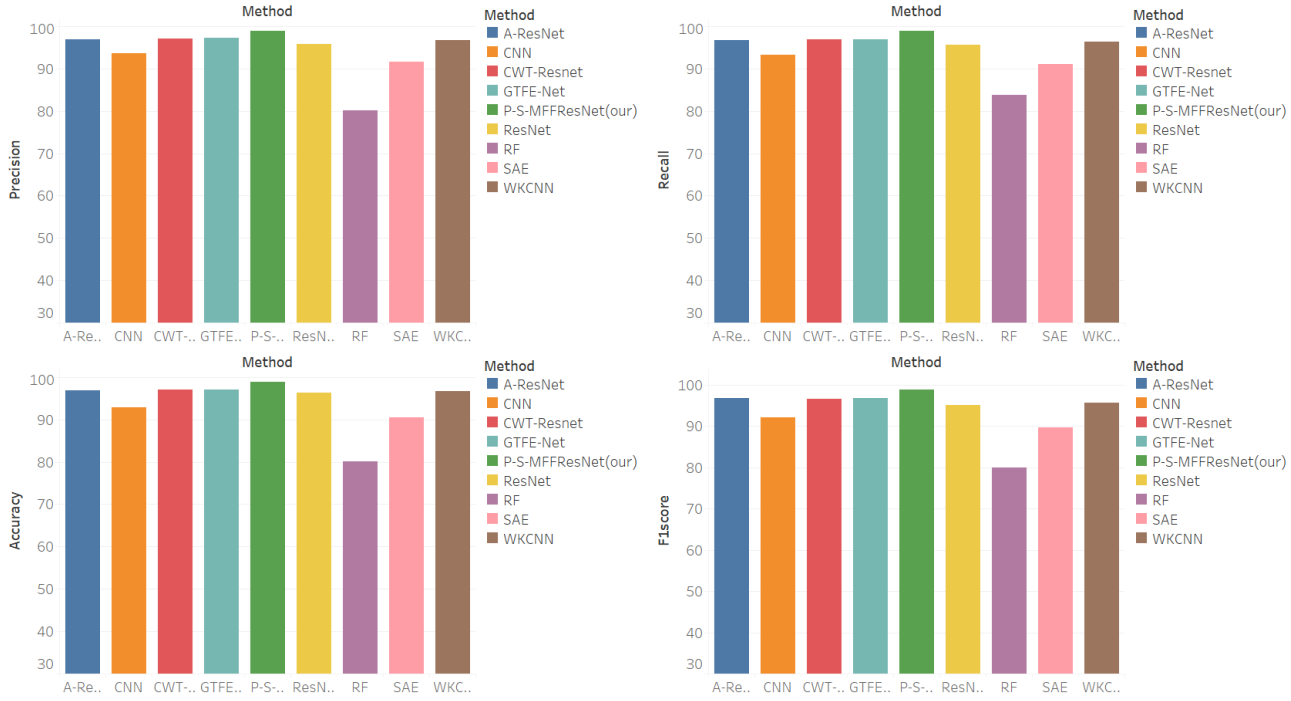

Figure 8. Comparison results on the Rain800 datasets.

4. Conclusions

The fault signal of hydropower units may have weak fault features due to the interference of various factors such as noise, which affects the accuracy and stability of the subsequent fault diagnosis algorithm. To solve this problem, this paper proposes a PSO-MCKD-MFFResnet algorithm for fault diagnosis of hydropower units. First, this paper proposes a PSO-MCKD fault feature enhancement algorithm, which uses the PSO algorithm to search for the influencing parameters of MCKD to enhance the features of the original fault signal. Second, a fault feature diagnosis algorithm based on MFFResnet is proposed to improve the utilization of local features. The multi-scale residual module is used to

extract features at different scales and then put the enhanced signal into MFFResnet for training and classification.

The method proposed in this paper applies to four common states of hydropower units. Since mechanical failures will have other rare and complex failure situations, accumulating different situations for multi-task learning to improve performance is also worth exploring in future research [41].

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by the Key Project of Shaanxi Provincial Department of Science and Technology's Joint Fund for Introducing Han to Wei Rivers (No. 2022JC-LHJJ-02). This work was also supported by the Project of the Dadu River Maintenance and Installation Company (No. JX-QT-2022-03).

References

1. R. Santis, M. A. Costa, Extended isolation forests for fault detection in small hydroelectric plants, *Sustainability*, **12** (2020), 6421. https://doi.org/10.3390/su12166421
2. W. Liu, Y. Zheng, Z. Ma, B. Tian, Q. Chen, An intelligent fault diagnosis scheme for hydropower units based on the pattern recognition of axis orbits, *Meas. Sci. Technol.*, **34** (2022). https://doi.org/10.1088/1361-6501/ac97ff
3. Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, A. K. Nandi, Applications of machine learning to machine fault diagnosis:a review and roadmap, *Mech. Syst. Signal Process.*, **138** (2022), 106587. https://doi.org/10.1016/j.ymssp.2019.106587
4. K. Xu, Y. Li, C. Liu, X. Liu, X. Hao, J. Gao, Advanced data collection and analysis in data-driven manufacturing process, *Chin. J. Mech. Eng.*, **33** (2020), 40–60. https://doi.org/10.1186/s10033-020-00459-x
5. A. R. Al-Obaidi, Detection of cavitation phenomenon within a centrifugal Pump based on vibration analysis technique in both time and frequency domains, *Exp. Tech.*, **44** (2020), 329–347. https://doi.org/10.1007/s40799-020-00362-z
6. J. Lin, C. Dou, Q. Wang, Comparisons of MFDFA, EMD and WT by neural network, mahalanobis distance and SVM in fault diagnosis of gearboxes, *Sound Vib.*, **52** (2018), 11–15.
7. L. Bai, W. Xi, Early fault diagnosis of rolling bearing based empirical wavelet transform and spectral kurtosis, in *2018 IEEE International Conference on Prognostics and Health Management (ICPHM)*, (2018), 1–6. https://doi.org/10.1109/ICPHM.2018.8448997
8. J. S. Cheng, D. Yu, J. Tang, Y. Yang, Application of SVM and SVD technique based on EMD to the fault diagnosis of the rotating machinery, *Shock Vib.*, **16** (2019), 89–98. https://doi.org/10.3233/SAV-2009-0457

9. Y. Hu, Q. Li, An adjustable envelope based EMD method for rollingbearing fault diagnosis, in *IOP Conference Series: Materials Science and Engineering*, **1043** (2021), 032017. https://doi.org/10.1088/1757-899X/1043/3/032017
10. Y. He, H. Wang, H. Xue, T. Zhang, Research on unknown fault diagnosis of rolling bearings based on parameter-adaptive maximum correlation kurtosis deconvolution, *Rev. Sci. Instrum.*, **92** (2021), 055103. https://doi.org/10.1063/5.0046113
11. Z. Li, A. Ming, W. Zhang, T. Liu, F. Chu, Y. Li, Fault feature extraction and enhancement of rolling element bearings based on maximum correlated kurtosis deconvolution and improved empirical wavelet transform, *Appl. Sci.*, **9** (2019), 1876. https://doi.org/10.3390/app9091876
12. W. Hua, C. Luo, J. Leng, Z. Wang, Mine gearbox fault diagnosis based on multiwavelets and maximum correlated kurtosis deconvolution, *J. Vibroeng.*, **19** (2017), 4185–4197. https://doi.org/10.21595/jve.2017.17497
13. F. Wang, C. Liu, W. Su, Z. Xue, Q. Han, H. Li, Combined failure diagnosis of slewing bearings based on MCKD-CEEMD-ApEn, *Shock Vib.*, **2018** (2018), 1070–9622. https://doi.org/10.1155/2018/6321785
14. P. Wang, Y. Miao, Multi classification ERT flow pattern recognition method based on deep learning, *J. Phys. Conf. Ser.*, **2181** (2022). https://doi.org/10.1088/1742-6596/2181/1/012010
15. D. T. Hoang, H. J. Kang, A survey on deep learning based bearing fault diagnosis, *Neurocomputing*, **335** (2019), 327–335. https://doi.org/10.1016/j.neucom.2018.06.078
16. D. Yao, H. Liu, J. Yang, J. Zhang, Implementation of a novel algorithm of wheelset and axle box concurrent fault identification based on an efficient neural network with the attention mechanism, *J. Intell. Manuf.*, **32** (2021), 0956–5515. https://doi.org/10.1007/s10845-020-01701-y
17. B. Peng, Y. Bi, B. Xue, M. Zhang, S. Wan, A survey on fault diagnosis of rolling bearings, *Algorithms*, **15** (2022), 357. https://doi.org/10.3390/a15100347
18. H. Shao, H. Jiang, X. Li, T. Liang, Rolling bearing fault detection using continuous deep belief network with locally linear embedding, *Comput. Ind.*, **96** (2018), 27–39. https://doi.org/10.1016/j.compind.2018.01.005
19. B. Liu, C. Liu, Y. Zhou, D. Wang, Y. Dun, An unsupervised chatter detection method based on AE and merging GMM and K-mean, *Mech. Syst. Signal Process.*, **186** (2023), 109861. https://doi.org/10.1016/j.ymssp.2022.109861
20. B. Ma, W. Cai, Y. Han, G. Yu, A novel probability confidence CNN model and its application in mechanical fault diagnosis, *IEEE Trans. Instrum. Meas.*, **70** (2021), 1–11. https://doi.org/10.1109/TIM.2021.3077965
21. M. Sun, H. Wang, P. Liu, S. Huang, P. Fan, A sparse stacked denoising autoencoder with optimized transfer learning applied to the fault diagnosis of rolling bearings, *Measurement*, **146** (2019), 305–314. https://doi.org/10.1016/j.measurement.2019.06.029
22. D. Hoang, H. Kang, Rolling element bearing fault diagnosis using convolutional neural network and vibration image, *Cognit. Syst. Res.*, **53** (2019), 42–50. https://doi.org/10.1016/j.cogsys.2018.03.002

23. G. Liao, W. Gao, G. Yang, M. Guo, Hydroelectric generating unit fault diagnosis using 1-D convolutional neural network and gated recurrent unit in small hydro, *IEEE Sens. J.*, **19** (2019), 9352–9363. https://doi.org/10.1109/JSEN.2019.2926095

24. X. Wang, D. Mao, X. Li, Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network, *Measurement*, **173** (2021), 108518. https://doi.org/10.1016/j.measurement.2020.108518

25. X. Song, Y. Cong, Y Song, Y. Chen, P. Liang, A bearing fault diagnosis model based on CNN with wide convolution kernels, *J. Ambient Intell. Hum. Comput.*, **13** (2022), 4041–4056. https://doi.org/10.1007/s12652-021-03177-x

26. L. Jia, T. W. S. Chow, Y. Yuan, GTFE-Net: A Gramian time frequency enhancement CNN for bearing fault diagnosis, *Eng. Appl. Artif. Intell.*, **119** (2023), 105794. https://doi.org/10.1016/j.engappai.2022.105794

27. N. Sakli, H. Ghabri, B. O. Soufiene, F. Almalki, H. Sakli, O. Ali, et al., Resnet-50 for 12-Lead electrocardiogram automated diagnosis, *Comput. Intell. Neurosci.*, **2022** (2022), 1–16. https://doi.org/10.1155/2022/7617551

28. W. Zhang, X. Li, Q. Ding, Deep residual learning-based fault diagnosis method for rotating machinery, *ISA Trans.*, **95** (2019), 295–305. https://doi.org/10.1016/j.isatra.2018.12.025

29. Y. Wang, J. Liang, X. Gu, D. Ling, H. Yu, Multi-scale attention mechanism residual neural network for fault diagnosis of rolling bearings, *Proc. Inst. Mech. Eng., Part C: J. Mech. Eng. Sci.*, **236** (2022), 10615–10629. https://doi.org/10.1177/09544062221104598

30. Y. Jin, C. Qin, Y. Huang, C. Liu, Actual bearing compound fault diagnosis based on active learning and decoupling attentional residual network, *Measurement*, **173** (2021), 108500. https://doi.org/10.1016/j.measurement.2020.108500

31. Y. Chen, D. Zhang, H. Ni, J. Cheng, H. R. Karimi, Multi-scale split dual calibration network with periodic information for interpretable fault diagnosis of rotating machinery, *Eng. Appl. Artif. Intell.*, **123** (2023), 106181. https://doi.org/10.1016/j.engappai.2023.106181

32. Y. Chen, D. Zhang, H. Zhang, Q. Wang, Dual-path mixed-domain residual threshold networks for bearing fault diagnosis, *IEEE Trans. Ind. Electron.*, **69** (2022), 13462–13472. https://doi.org/10.1109/TIE.2022.3144572

33. Y. Chen, D. Zhang, K. Zhu, R. Yan, An adaptive activation transfer learning approach for fault diagnosis, *IEEE/ASME Trans. Mechatron.*, **2023** (2023), 1–12. https://doi.org/10.1109/TMECH.2023.3243533

34. N. H. Phong, A. Santos, B. Ribeiro, PSO-convolutional neural networks with heterogeneous learning rate, *IEEE Access*, **10** (2022), 89970–89988. https://doi.org/10.1016/10.1109/ACCESS.2022.3201142

35. Z. Jiang, J. Zheng, H. Pan, Sigmoid-based refined composite multiscale fuzzy entropy and t-SNE based fault diagnosis approach for rolling bearing, *Measurement*, **129** (2018), 332–342. https://doi.org/10.1016/j.measurement.2018.07.045

36. J. Yu, C. Xiao, T. Hu, Y. Gao, Selective weighted multi-scale morphological filter for fault feature extraction of rolling bearings, *ISA Trans.*, **132** (2022), 544–556. https://doi.org/10.1016/j.isatra.2022.06.003

37. X. Dong, G. Li, Y. Jia, K. Xu, Multiscale feature extraction from the perspective of graph for hob fault diagnosis using spectral graph wavelet transform combined with improved random forest, *Measurement*, **176** (2021), 109178. https://doi.org/10.1016/j.measurement.2021.109178

38. W. Huang, J. Chen, Y. Yang, G. Guo, An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis, *Neurocomputing*, **395** (2019), 77–92. https://doi.org/10.1016/j.neucom.2019.05.052

39. B. Zhao, X. Zhang, Z. Zhan, Q. Wu, Deep multi-scale adversarial network with attention:a novel domain adaptation method for intelligent fault diagnosis, *J. Manuf. Syst.*, **59** (2021), 565–576. https://doi.org/10.1016/j.jmsy.2021.03.024

40. R. Fezai, K. Dhibi, M. Mansouri, M. Trabelsi, M. Hajji, K. Bouzrara, et al., Effective random forest-based fault detection and diagnosis for wind energy conversion systems, *IEEE Sens. J.*, **21** (2021), 6914–6921. https://doi.org/10.1109/JSEN.2020.3037237

41. J. Xu, S. Liang, X. Ding, R. Yan, A zero-shot fault semantics learning model for compound fault diagnosis, *Expert Syst. Appl.*, **221** (2023), 119642. https://doi.org/10.1016/j.eswa.2023.119642

AIMS Press