Modified Hierarchical Multiscale Dispersion Entropy and Its Application to Fault Identification of Rotating Machinery

FUMING ZHOU, JINXING SHEN, XIAOQIANG YANG, XIAOLIN LIU, AND WUQIANG LIU

College of Field Engineering, Army Engineering University, Nanjing 210007, China
Corresponding author: Xiaoqiang Yang (3113593699@qq.com)

ABSTRACT The rotating machinery possesses complicated structures and various fault types, whose health state monitoring is essential for the normal production and operation of the equipment. To distinguish different working states of rotating machinery efficiently and accurately, this article presents a novel approach for extracting fault features of vibration signals called modified hierarchical multiscale dispersion entropy (MHMDE). And on this basis, an innovative approach for fault diagnosis of rotating machinery based on MHMDE, multi-cluster feature selection (MCFS) and particle swarm optimization kernel extreme learning machine (PSO-KELM) is developed. Firstly, MHMDE is employed to extract the high-dimensional fault features of rotating machinery. This approach can effectively overcome the shortcomings that multi-scale entropy only focuses on the information in the low-frequency components but discards the high-frequency information, as well as the significant dropping of efficiency if the number of hierarchical layers of hierarchical entropy is large. Then MCFS is employed to screen the sensitive features from the high-dimensional fault features. Finally, the sensitive feature vectors are input into the PSO-KELM-based fault classifier to complete the rotating machinery fault diagnosis. It is proved that the presented approach can effectively identify different fault states of rotating machinery through three typical examples. Meanwhile, the presented approach is compared with multi-scale dispersion entropy (MDE) and hierarchical dispersion entropy (HDE), etc. The results show that the presented approach possesses more superior performance.

INDEX TERMS Rotating machinery, fault diagnosis, modified hierarchical multiscale dispersion entropy, multi-cluster feature selection, PSO-KELM.

I. INTRODUCTION

Rotating machinery is one of the most diffusely applied mechanical devices in the industry. It plays a significant role in lots of mechanical systems, so their working states are critical to the normal operation of the whole working system. However, rotating machinery usually possesses complicated structures; meanwhile, it works under heavy load and poor working conditions, so the probability of faults is also higher than other mechanical equipment [1], [2]. Once the rotating machinery fails, it will cause the equipment to stop production and bring huge economic losses or even casualties. In consequence, it is of great practical significance to monitor and diagnose the working states of the rotating machinery in real-time to ensure the normal and smooth operation of the equipment [3].

Vibration signals of mechanical equipment often contain information that is closely related to the working states of the equipment, so identifying different working states of rotating machinery through vibration analysis has become a widely applied approach [4], [5]. Owing to the impact of some nonlinear factors, such as background noise, friction, etc., the collected vibration signals of rotating machinery usually possess unstable and nonlinear characteristics [6]. The commonly applied signal time-frequency analysis methods such as empirical mode decomposition (EMD) [7], local mean decomposition (LMD) [8] and wavelet transform (WT) [9] exist some shortcomings of mode mixing, low calculation efficiency and poor adaptation, respectively. At the same time, the signal time-frequency analysis methods have certain
requirements on the experience and knowledge of equipment operators, and it is difficult to directly determine the work states of equipment for ordinary users who are without certain knowledge of time-domain plot and spectrum. Therefore, the above methods still have many limitations in practical application.

Entropy can effectively measure the dynamic features of complex time series, so it has been diffusely applied to identify different fault states of rotating machinery in recent years [10], [11]. Commonly applied entropy approaches include sample entropy (SE) [12], fuzzy entropy (FE) [13], and permutation entropy (PE) [14]. Nevertheless, these approaches all have certain drawbacks. SE is unstable when dealing with short time series; meanwhile, the calculation efficiency of it is low. FE is more stable than SE, but its calculation efficiency is still low. PE is computationally fast, however, it only considers the order of amplitude of each element in time series while ignores the difference between different amplitudes; meanwhile, its anti-noise performance is weak [15]. In order to solve the above defects, Rostaghi and Azami [16] presented a novel index to detect the dynamic features of complex series called dispersion entropy (DE) recently. This method divides the elements of the original time series into different classes, thereby can effectively improve the anti-noise performance; meanwhile, compared with PE, DE does not need a sorting process, so it enjoys higher calculation efficiency. Literature [17] verified the superiority of DE over SE, FE and PE in terms of calculation efficiency and stability through simulated signal and biological signal experiments.

Unfortunately, the aforementioned approaches all analyze the time series on a single scale, which discards the feature information that may exist on other scales. For instance, according to the definition of the regularity measurement of entropy, the larger the entropy value, the higher the randomness and complexity of the time series. When adopting the above approaches to analyze white noise and 1/f noise [18], the entropy value of white noise is significantly greater than 1/f noise. But in fact, 1/f noise is more complex than white noise because of the long-range correlation properties [19], indicating that the single-scale analysis still exists defects when detecting the dynamic features. To address this question, Costa et al. [20] presented the multi-scale entropy (MSE) approach, which comprehensively considers the information of time series on multiple scales, thereby can accurately reflect the complexity of time series. Based on MSE approach, multi-scale dispersion entropy (MDE) [21] and its modified forms, such as composite multi-scale dispersion entropy (CMDE) [22] and refined composite multi-scale dispersion entropy (RCMDE) [23], have been successively presented. The purposes of adopting composite form and refined composite form are to solve the entropy instability problem existing in the traditional multi-scale expansion. Although these two modified approaches improve the stability of multi-scale entropy by averaging entropy values or averaging probability values, there are still some outstanding defects. One is that their calculation efficiency is low, and the other is that the coarse-grained time series will become shorter with the increase of scale factor, resulting in the risk of unstable calculation results still exists [19]. Wu et al. [24] presented a modified multi-scale entropy (MMSE) approach, in which the origin time series is processed by the moving-average process instead of the traditional coarse-graining process. This approach solves the problem of short coarse-grained time series and possesses higher calculation efficiency. Inspired by this, this article presents modified multi-scale dispersion entropy (MMDE) which combines MMSE with DE.

Ying et al. [25] believed that MSE only considered the low-frequency information in the measured series, while the information hidden in the high-frequency components might be discarded. Therefore, they presented the hierarchical entropy (HE) approach and verified its effectiveness through experiments. Nevertheless, HE cannot make full use of the information of the time series. When the number of hierarchical layers is too large, the entropy values will be unstable. For this reason, on the basis of HE, Li et al. [26] presented a modified hierarchical entropy (MHE) approach, where the original hierarchical decomposition process is replaced by moving-averaging and moving-difference processes, thereby effectively improves the information utilization of the time series. Based on MHE, this article presents the modified hierarchical multi-scale dispersion entropy (MHMDE). Unfortunately, both multiscale analysis [21]–[24] and hierarchical analysis [25], [26] have certain inherent defects. Specifically, for hierarchical analysis, with the increase of hierarchical layers, the calculation efficiency will be significantly reduced, which limits the ability of HE to extract fault features. If the number of layers is small, the number of fault features is small, which may lead to the loss of fault information. In contrast, if the number of layers is large, the calculation efficiency must be sacrificed, and the practicability will be reduced. For multiscale analysis, it only considers the low-frequency information of the time series but discards the high-frequency information. To solve the aforementioned inherent defects, this article presents a new time series processing approach called the modified hierarchical multi-scale approach. And the concept of DE is introduced to present the modified hierarchical multi-scale dispersion entropy (MHMDE). This proposed approach integrates the advantages of MHE with MMDE. It can not only extract fault information in both high-frequency and low-frequency components but also surmount the weakness of limited features number in hierarchical analysis. The most important point is that this approach can extract enough fault features in a very short time, which overcomes the loss of fault information effectively and is more practical.

The fault features extracted using MHMDE are high-dimensional, which not only contain rich fault information but also inevitably contain some redundant information. It is not recommended to process the extracted high-dimensional fault features directly. Multi-cluster feature
selection (MCFS) [27] is a novel feature selection approach recently presented by Cai et al., which possesses good clustering and classification performance. After MHMDE is adopted to extract fault features of rotating machinery, MCFS is applied to select sensitive features from obtained high-dimensional fault features to construct sensitive feature vectors. Then the sensitive feature vectors are input into the fault classifier constructed by particle swarm optimization kernel extreme learning machine (PSO-KELM) to complete fault identification of rotating machinery. Kernel extreme learning machine (KELM) [28] was presented by combining kernel functions with extreme learning machine (ELM). While maintaining the high efficiency of ELM, KELM has better generalization performance when dealing with linear inseparability problems by introducing kernel functions. Moreover, there is no need to set the number of hidden layer nodes and the overall performance of KELM is better consequently. Inspired by the above, this article presents a new approach for fault diagnosis of rotating machinery based on MHMDE, MCFS and PSO-KELM. In the following, we compare this approach with MDE, MMDE, CMDE, RCMDE, hierarchical dispersion entropy (HDE) and MHDE, proving that the presented approach possesses better performance.

The main contributions of this article can be summarized as follows:

1. MMDE and MHDE are proposed by integrating MMSE and MHE with the recently presented DE with excellent performance, respectively. So the hierarchical entropy and multi-scale entropy approaches are improved to some extent.

2. To address the inherent defects of hierarchical analysis and multi-scale analysis, MHMDE is proposed by integrating MMDE with MHDE, which possesses the merits of both approaches.

3. MHMDE is introduced into the field of rotating machinery fault diagnosis to extract fault features of vibration signals, and the concept of average Euclidean distance is introduced to determine the best parameters of MHMDE. Based on this, combining MHMDE, MCFS and PSO-KELM, a new fault diagnosis approach for rotating machinery is presented, whose effectiveness is verified through three examples, and the superiority is verified through some comparison experiments.

The structure of this article is as follows: Section II mainly explains the principle and parameter selection of the MHMDE algorithm. Section III mainly explains the presented approach in this article in detail. Then, the superiority of this approach is reflected through three typical examples and comparative experiments in Section IV. Conclusions are summarized in Section V.

II. MODIFIED HIERARCHICAL MULTISCALE DISPERSION ENTROPY
A. DISPERSION ENTROPY
For random series $X = \{x_1, x_2, \ldots, x_N\}$, the principle of DE [16] algorithm is summarized as follows:

1. All elements in $X$ are assigned to $c$ classes from 1 to $c$ by Equation (1) and Equation (2), and the time series $Z = \{z_1, z_2, \ldots, z_N\}$ is obtained.

$$y_i = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{x_i} e^{-\frac{(x-y)^2}{2\sigma^2}} \, dt \quad (1)$$

$$z_i = R(c \cdot y_i + 0.5) \quad (2)$$

where $\mu$ is expectation, $\sigma^2$ is variance, $c$ is an integer and $R$ is the rounding function.

2. Based on embedding dimension $m$ and time delay $\lambda$, $Z$ is reconstructed by Equation (3).

$$Z_{m,\lambda,k}^{m,\lambda,c} = \{z_{k}, z_{k+\lambda}, \ldots, z_{k+(m-1)\lambda}\}, \quad k = 1, 2 \ldots N - (m - 1)\lambda \quad (3)$$

Each $Z_{m,\lambda,k}^{m,\lambda,c}$ obtained by reconstruction has its corresponding dispersion pattern $\pi_{v0^0v1\ldots v_{m-1}}(1 \leq v \leq c)$, and $Z_{k,1}^{m,\lambda,c} = v_0, Z_{k,2}^{m,\lambda,c} = v_1, \ldots, Z_{k,m}^{m,\lambda,c} = v_{m-1}$. The length of $Z_{k}^{m,\lambda,c}$ is $m$ and each element of which may be an integer from 1 to $c$, so there are $c^m$ possible dispersion patterns, and the probability of each dispersion pattern can be calculated by:

$$p(\pi_{v0^0v1\ldots v_{m-1}}) = \frac{\text{Number(\pi_{v0^0v1\ldots v_{m-1}})}}{N - (m - 1)\lambda} \quad (4)$$

3. DE can be obtained by Equation (5).

$$DE(X, m, c, \lambda) = - \sum_{\pi=1}^{c^m} p(\pi_{v0^0v1\ldots v_{m-1}}) \ln p(\pi_{v0^0v1\ldots v_{m-1}}) \quad (5)$$

B. MODIFIED HIERARCHICAL MULTISCALE DISPERSION ENTROPY
MHDE and MMSE are improved on the basis of HE and MSE, respectively. They all improve the information utilization of the initial series and improve the stability of HE and MSE. This article integrates the two approaches and presents a modified hierarchical multiscale signal processing approach, which associates the merits of MHE and MMSE while compensating for their shortcomings. Combining modified hierarchical multiscale approach with DE, a new dynamic features index called modified hierarchical multiscale dispersion entropy is presented. Its principle can be divided into two parts, namely modified hierarchical decomposition process and modified multiscale process, which are summarized as follows:

1. For random series $X = \{x_1, x_2, \ldots, x_N\}$, we define the operators $Q_0(x)$ and $Q_1(x)$ as follows [25]:

$$Q_0(x) = \frac{x_i + x_{i+1}}{2}, \quad Q_1(x) = \frac{x_i - x_{i+1}}{2}, \quad i \in [1 \ldots N-1] \quad (6)$$

$Q_0(x)$ and $Q_1(x)$ contain the low-frequency information and high-frequency information of $X$, respectively.

2. The matrix form of the operator $Q_t^p(t = 0 \text{ or } 1$) corresponding to the $t$-th layer can be expressed as follows:
The hierarchical decomposition when $k = 3$.

The equation can be derived, as shown at the bottom of the page:

(3) Construct the vector $[v_1, v_2, \ldots, v_k]$ of length $k$, where $v_k = 0$ or 1, and the integer $e$ can be calculated as follows:

$$ e = \sum_{m=1}^{k} 2^{k-m} v_m \quad (7) $$

$v_m (m = 1, 2, \ldots, k)$ represents the average or difference operators of the $m$-th layer. According to Equation (7), for a given non-negative integer $e$, there is a unique vector corresponding to it.

(4) Based on the vector $[v_1, v_2, \ldots, v_k]$, the hierarchical components of series $X$ can be obtained using Equation (8).

$$ X_{k,e} = \mathcal{Q}_{v_k} \cdot \mathcal{Q}_{v_{k-1}}^{1} \cdot \ldots \cdot \mathcal{Q}_{v_1}^{1} \cdot X \quad (8) $$

$X_{k,e}$ represents the hierarchical component at the node $e$ of the $k$-th layer of series $X$. When the number of hierarchical layers is 3, the modified hierarchical decomposition process of $X$ is depicted in Fig. 1, where $X_{1,0}$ is the low-frequency component of $X$ when the number of hierarchical layers is 1, while $X_{1,1}$ is the high-frequency component. $X_{3,2}$ represents the hierarchical component at the node 2 of the 3-rd layer, whose corresponding unique vector is $[0, 1, 0]$. Then according to Equation (8), $X_{3,2} = \mathcal{Q}_2 \cdot \mathcal{Q}_1^{1} \cdot \mathcal{Q}_1^{1} \cdot X$.

(5) After the hierarchical component $X_{k,e}$ is obtained, it is processed by the modified coarse-graining approach, and the coarse-grained time series under various scale factors are obtained, as shown in Equation (9)

$$ X_{k,e,t}^{i} = \frac{1}{\tau} \sum_{j=i}^{j+\tau-1} X_{k,e}^{i} \quad 1 \leq j \leq N - 2^{k-\tau} + 2 \quad (9) $$

where $\tau$ represents the scale factor.

(6) According to Equations (1) - (5), the MHE values corresponding to the hierarchical component $X_{k,e}$ of the original random series $X$ can be calculated by Equation (10).

$$ \text{MHE}(X, k, e, m, c, \lambda, \tau) = \text{DE}(X_{k,e,t}, m, c, \lambda) \quad (10) $$

It can be seen from the principle of MHEMDE that this approach combines MHE and MMSE. Through the multi-scale expansion of each component series obtained by hierarchical decomposition, as many features as possible can be obtained with less time consumption. The method effectively overcomes the defect that the calculation efficiency of hierarchical entropy decreases significantly when the number of layers is too large, which leads to a limited number of features. Fig. 2 shows the flowchart of MHEMDE.

C. PARAMETERS SELECTION OF MHEMDE

The main parameters of MHEMDE algorithm include embedding dimension $m$, class $c$, time delay $\lambda$, hierarchical layer

$$ Q^k_l = \begin{bmatrix}
\frac{1}{2} & 0 & \ldots & 0 \\
0 & \frac{1}{2} & \ldots & 0 \\
\vdots & \ddots & \ldots & \vdots \\
0 & 0 & \ldots & 0
\end{bmatrix} \frac{(-1)^{l'}}{2^{k-l'-1}} \begin{bmatrix}
0 & \ldots & 0 \\
0 & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & 0
\end{bmatrix} \frac{(-1)^{l'}}{2^{k-l'-1}} \begin{bmatrix}
0 & \ldots & 0 \\
0 & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & 0
\end{bmatrix} \frac{(-1)^{l'}}{2^{k-l'-1}} \begin{bmatrix}
0 & \ldots & 0 \\
0 & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & 0
\end{bmatrix} \frac{(-1)^{l'}}{2
\end{bmatrix} (N-2^{k+l'})(N-2^{k+l'+1})
k and scale factor \( \tau \). Time delay \( \lambda \) has little impact on the calculation results, which is usually 1 [26, 29]. If the number of layers \( k \) is too small, the frequency band allocation of the original series is not enough detailed to obtain sufficient high-frequency and low-frequency information [30]. If \( k \) is too large, the calculation cost will be too high and the practicability will be reduced. After careful consideration, this article takes \( k = 3 \) [26, 30]. If the scale factor \( \tau \) is too small, the feature information of time series cannot be fully extracted, while if \( \tau \) is too large, the calculation efficiency will decrease and redundant information will be generated. In this article, we set \( \tau_{\text{max}} = 8 \). Consequently, after the hierarchical decomposition of the raw time series, 8 hierarchical series can be obtained, and after the multiscale expansion of each hierarchical series, 8 coarse-grained series can be obtained. The DE value of each coarse-grained series can be calculated to obtain a total of \( 8 \times 8 = 64 \) features.

In order to select the best embedding dimension \( m \) and class \( c \), we introduce the concept of average Euclidean distance (AED) to describe the distance between the feature vectors of different states. The greater the distance, the better the separability of different states. In other words, MHMDE is more capable of distinguishing the fault features of different states, and the corresponding parameters are the best. The specific parameter selection process can be summarized as follows:

Suppose there are \( N \) different working states of rotating machinery, while each state possesses \( M \) samples and the sample length is \( L \).

1. According to the recommendation presented in [17], we set \( c \in [4, 8] \), and \( m \) is determined according to the criterion \( c^m < L \). MHMDE of all samples is calculated with the initial parameter \((c, m)\) to obtain the corresponding feature vectors \([\text{MHMDE}(j), 1 < j < k \cdot \tau_{\text{max}}]\).
2. The AED value of the \( m \)-th and the \( n \)-th states is calculated by Equation (11).

\[
AED(m, n) = \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1}^{k \cdot \tau_{\text{max}}} (\text{MHMDE}_{m,j}(i) - \text{MHMDE}_{n,i}(j))^2 \quad (11)
\]

3. Calculate the AED value corresponding to the parameter \((c, m)\).

\[
AED = \sum_{m=1}^{N} \sum_{n=1, n \neq m}^{N} AED(m, n) \quad (12)
\]

4. Update the parameters and repeat steps (1)—(3) to calculate the AED values under different \((c, m)\). The \((c, m)\) corresponding to the highest AED value is the best \((c, m)\).

III. MHMDE BASED FAULT DIAGNOSIS FOR ROTATING MACHINERY

A. MULTI-CLUSTER FEATURE SELECTION

It can be seen from the above that after using MHMDE to extract the fault features of rotating machinery, the high-dimensional feature vectors with a length of 64 are obtained, which contain rich fault information and inevitably exist some redundant information. If all the features are input into the classifier, it will not only increase the training time, reduce efficiency, but also affect the accuracy of fault recognition. In consequence, it is necessary to reduce the dimension of the high-dimensional fault features extracted by MHMDE. MCFS [27] is an effective feature selection approach presented recently, which measures the correlation between different features using spectral analysis. It can implement supervised or unsupervised dimensionality reduction of high-dimensional features. Cai et al. verified the superior performance of MCFS in clustering and classification compared with Laplacian score and Q-\( \alpha \) approaches. The specific principle of MCFS can be referred to [27], and this article does not describe it in detail.

B. PSO-KELM

The classifier should possess the merits of high operation efficiency and strong generalization ability. ELM is a machine learning approach based on single-hidden layer feedforward neural network (SLFN). Compared with support vector machine (SVM), single layer perceptron (SLP) and other approaches, ELM has advantages in computing efficiency and generalization ability [31]. KELM is a new algorithm presented by introducing kernel functions on the basis of ELM. While retaining the advantages of ELM, KELM does not need to set the number of hidden layer nodes, which avoids the overfitting problems caused by too many hidden layer nodes when dealing with some classification or fitting problems; meanwhile, KELM has better performance when dealing with linear inseparability problems. Owing to the existence of kernel functions, KELM is more sensitive to parameter settings. Particle swarm optimization (PSO) [32] enjoys good global search capabilities, so this article adopts PSO to optimize KELM to construct the fault classifier. The principles of KELM and the steps of PSO-KELM can be summarized as follows:

The SLFN model is:

\[
f(x) = h(x)\beta = H\beta \quad (13)
\]

where \( H \) and \( h(x) \) are the output matrix of the hidden layer, \( f(x) \) is the output vector, which represents label vector in classification problems, and \( \beta \) is the connection weight matrix between the hidden layer and the output layer, which is defined as follows:

\[
\beta = H^T (HH^T + \lambda I)^{-1} T
\]

where \( \lambda \) is the penalty factor.

When \( h(x) \) is unknown, define the kernel matrix \( \Omega = HH^T, \Omega(i,j) = h(x_i) \cdot h(x_j) = K(x_i, x_j) \). Then,

\[
f(x) = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix} (\Omega + \lambda I)^{-1} T \quad (15)
\]
The RBF kernel function is usually adopted as the kernel function $K(x_i, x_j)$ as follows:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2g^2}\right) \quad (16)$$

Consequently, the main parameters of KELM are penalty factor $\lambda$ and kernel function parameter $g$. Choosing appropriate parameters has an important impact on the performance of KELM. PSO is applied to iteratively optimize to determine the best combination of $\lambda$ and $g$, and the fault classification accuracy obtained in each run is the fitness value. The specific steps of PSO-KELM are as follows:

1. Set the initial parameters of PSO algorithm, and take $[\lambda, g]$ as the particle position to generate the initial population.
2. Calculate the classification accuracy corresponding to all particles in the initial population, with the highest accuracy as the fitness value.
3. Update the position and velocity of particles, calculate the corresponding classification accuracy, and update the fitness value.
4. Iterate until the maximum number of loops is met, and output the best parameters and the corresponding accuracy.

C. THE PRESENTED FAULT DIAGNOSIS APPROACH

To distinguish different fault states of rotating machinery accurately and efficiently, a new fault diagnosis approach based on MHMDE, MCFS and PSO-KELM is presented. The specific steps of this approach can be summarized as follows, and Fig. 3 shows the flowchart of the presented approach.

1. $N$ samples are taken from the collected vibration signal of each working state, and the length of each sample is $L$.
2. Randomly select $M$ samples as training samples and the remaining $N - M$ samples as testing samples. MHMDE is employed to extract fault features of all samples to obtain the corresponding high-dimensional fault feature vectors. We set the parameters of MHMDE to $k = 3$, $\tau = 8$, $\lambda = 1$, while $m$ and $c$ need to be determined according to the AED corresponding to different rotating machinery.
3. After using MHMDE to extract high-dimensional fault features, MCFS is adopted to sort the fault features of training samples according to the importance. In this article, the first 16 elements of the training sample fault features which are processed by MCFS are used to indicate the sensitive fault features of the testing samples. That is, the elements in the sensitive feature vectors of the training samples and the testing samples have the same position in the original high-dimensional feature vectors.
4. Input the sensitive feature vectors of the training samples into the fault classifier constructed based on PSO-KELM for training, and then input the sensitive feature vectors of the testing samples for recognition to complete fault diagnosis of rotating machinery.

IV. EXPERIMENTAL VALIDATION

To verify the practicability of the presented approach, we take three typical vibration signals such as rolling bearing, hydraulic pump and gearbox, as examples for experimental analysis. Among them, the purposes of experiment 1 and experiment 2 are to check the performance of the presented approach when distinguishing the faults of a single component (rolling bearings and sliding shoes of the hydraulic pump), and the purpose of experiment 3 is to check the performance of the presented approach when distinguishing the compound faults. The presented approach is compared with existing approaches such as MDE, HDE, MMDE and MHDE to verify the superiority of it. We adopt the fault recognition accuracy and calculation efficiency as indicators to judge the performance of different diagnosis approaches.

A. EXPERIMENT 1: ROLLING BEARING FAULT DIAGNOSIS

In experiment 1, the bearing fault data of CWRU bearing data center is adopted to reflect the performance of the presented approach [33]. Fig. 4 shows the experimental device, which mainly consists of an induction motor, a torque transducer and a dynamometer. In the experiment, electro-discharge...
machine (EDM) technology was applied to process normal bearings to simulate bearing faults, and accelerometers installed at the fan end and the drive end were applied to collect vibration data corresponding to various bearing fault types under different loads and speeds. In this experiment, the vibration data collected by the accelerometer at the drive end is analyzed when the load was 0 and the motor speed was 1797r/min, while the sampling frequency was 12kHz. There are three types of bearing faults (inner race fault, outer race fault and ball fault). Each fault type takes three fault diameters of 0.1778, 0.3556 and 0.5334mm, so a total of 10 kinds of bearing vibration signals are obtained. The waveforms are shown in Fig. 5. For vibration data of each state, 50 samples with a length of 2048 are taken, of which 30 samples are randomly selected as the training samples, and the remaining 20 samples are the testing samples. A total of 300 training samples and 200 testing samples are obtained. For simplicity, the inner race faults of bearings with 0.1778, 0.3556 and 0.5334mm are simplified as “IR1”, “IR2” and “IR3”, respectively, and other fault types is simplified by the same form, while normal state is simplified as “Nor”. The detailed information of ten bearing states is depicted in Table 1.

According to Fig. 3, in order to identify different fault states of rolling bearings, the best parameters of MHMDE should be determined firstly. This article takes AED as the index to select the best \([c, m]\) of MHMDE. Take 20 samples for each working state of the rolling bearing, and calculate the AED of all samples when taking different \([c, m]\), where \(4 < c < 8\), and \(m\) is determined according to the criterion \(c^m < L\). \(L = 2048\) in this article. The results are shown in Fig. 6. As we can see, when \(c = 4, m = 5\), the AED achieves the maximum value, so we set the parameters of MHMDE to \(k = 3, \tau = 8, \lambda = 1, m = 5, c = 4\) when
extracting the fault features of rolling bearings. The average high-dimensional fault features of training samples corresponding to the ten states extracted by MHMDE are shown in Fig. 7. MCFS is applied to reduce the high-dimensional fault features of the training samples. The average feature value curves after dimensionality reduction are depicted in Fig. 8, where the abscissa represents the order of sensitive features in initial high-dimensional features. The sensitive feature vectors of 300 training samples are input into the fault classifier based on PSO-KELM for training, and then the sensitive feature vectors of 200 testing samples are input for identification. Parameter settings of PSO-KELM are shown in Table 2, where $G$ is the maximum number of iterations and $N$ is the number of particles. Fig. 9 shows the identification results. Obviously, the presented approach can effectively identify different fault types and fault degrees of rolling bearings, while the recognition accuracy is 100%.

In order to compare the presented approach with other approaches to reflect the superiority of the presented approach, MDE, MMDE, CMDE, RCMDE, HDE and MHDE are employed to extract the rolling bearing fault features, respectively. In order to test the efficiency of different approaches for extracting fault features, the time consumed for extracting the same number of features by different approaches is calculated, respectively. The results are depicted in Table 3, where “CPU time” refers to the time taken to extract all sample fault features. It is observed that when the number of fault features is 64, the calculation efficiency of the presented approach is significantly higher. To compare the recognition accuracy of several approaches; meanwhile, without loss of generality, five different fault diagnosis approaches all run 50 times. To improve the experimental efficiency, the number of fault features extracted by the four comparison approaches is 16, which is the same as the number of sensitive fault features obtained by the
presented approach, so dimension reduction is not required. The experimental results are depicted in Table 4 and Fig. 10. As we can see, the fault diagnosis approach based on MDE and HDE takes a shorter time, but the corresponding accuracy is significantly lower. The accuracy of MMDE, CMDE and RCMDE approaches is similar, but MMDE is more efficient in calculating, therefore, more practical. Meanwhile, the accuracy of the presented approach for 50 run times is all 100%, while the time consumed is also shorter than MMDE, CMDE, RCMDE and MHDE approaches. Considering the two indexes of calculation efficiency and fault identification accuracy, the presented approach significantly has better performance.

To study the influence of the size of training/testing samples on the performance of different methods, the size of training/testing samples is divided into four cases of 10/40, 20/30, 30/20, and 40/10. The average recognition accuracy after 50 runs is shown in Fig. 11. As we can see, the presented approach is less affected by the size of training/testing samples, while the average accuracy is 100% in each case. The classification accuracy of other approaches increases with the increase of training samples, however, too many training samples will lead to a decrease in training efficiency. Considering

![FIGURE 10. The accuracy of different approaches.](image1)

![FIGURE 11. The accuracy of various training/testing samples.](image2)
comprehensively, the presented approach possesses distinct performance advantages.

B. EXPERIMENT 2: HYDRAULIC PUMP FAULT DIAGNOSIS

Hydraulic pump is a typical high-speed rotating machinery that is prone to faults. To verify the universal applicability of the presented approach, experiment 2 adopts the presented approach to distinguish different fault states of hydraulic pump.

Vibration data of the hydraulic pump were collected from the hydraulic pump experiment platform, as depicted in Fig. 12. In the experiment, the model type of the hydraulic pump was SY-10MCY14-1EL, of which the number of plungers was 7, and the operating frequency was 25Hz. In the experiment, loose slipper plunger and slipper wear plunger were used to replace the normal plunger to simulate hydraulic pump faults, as shown in Fig. 13(b), and a total of four working states of the hydraulic pump are obtained: (a) normal state; (b) single plunger loose slipper; (c) double plungers loose slipper; (d) piston shoes wear. During the experiment, the pressure of the main overflow valve was 10Mpa. Vibration data were obtained by the 603C01 piezoelectricity acceleration sensor installed on the end cover of the hydraulic pump with a sampling frequency of 20kHz. The waveforms of the four working states of the hydraulic pump are shown in Fig. 14. 50 samples of vibration data of each state are taken, where 30 samples are randomly selected as training samples, and the remaining 20 samples are testing samples. The length of each sample is 2048. The details of the four states are depicted in Table 5.

Similar to experiment 1, we firstly need to determine the best \([c, m]\) of MHMDE for extracting fault features of the hydraulic pump. Take 20 samples of each working state of the hydraulic pump, and calculate the AED corresponding to different \([c, m]\), as shown in Fig. 15, where the selection criteria of \([c, m]\) is the same as experiment 1. It can be seen that AED reaches the maximum value when \(c = 5\) and \(m = 4\). Therefore, we set the parameters of MHMDE to \(k = 3\), \(\tau = 8\), \(\lambda = 1\), \(m = 4\), \(c = 5\). The average high-dimensional fault features of the training samples of the four hydraulic pump working states extracted by MHMDE are shown in Fig. 16. The sensitive feature curves obtained after MCFS processing are shown in Fig. 17. 120 training samples are input into the fault classifier for training, and then 80 testing samples are input for fault identification. Table 2 shows the parameter settings of PSO-KELM, and the results are
TABLE 5. Detailed description of different hydraulic pump fault types.

| Hydraulic pump states              | Abbreviation | Size of training samples | Size of testing samples | Labels |
|------------------------------------|--------------|--------------------------|-------------------------|--------|
| Normal state                       | N            | 30                       | 20                      | 1      |
| Single plunger loose slipper       | S            | 30                       | 20                      | 2      |
| Double plungers loose slipper      | D            | 30                       | 20                      | 3      |
| Piston shoes wear                  | P            | 30                       | 20                      | 4      |

TABLE 6. The parameter settings and accuracy of different approaches.

| Different approaches               | Parameter settings | Classification accuracy (%) | CPU time(s) |
|------------------------------------|--------------------|-----------------------------|-------------|
|                                    |                    | Max | Min | Mean |                  |
| MHMDE+MCFS+PSO-KELM                | $k = 3, \tau = 8, \lambda = 1, m = 4, c = 5$ | 100 | 100 | 100 | 128.64 |
| MDE+PSO-KELM                       | $\tau = 16, \lambda = 1, m = 4, c = 5$ | 96  | 89.5 | 94.78 | 61.20 |
| MMDE+PSO-KELM                      | $\tau = 16, \lambda = 1, m = 4, c = 5$ | 99.5 | 96  | 98.95 | 155.58 |
| CMDE+PSO-KELM                      | $\tau = 16, \lambda = 1, m = 4, c = 5$ | 100 | 96.25 | 98.55 | 272.81 |
| RCMDE+PSO-KELM                     | $\tau = 16, \lambda = 1, m = 4, c = 5$ | 100 | 96.25 | 98.98 | 277.15 |
| HDE+PSO-KELM                       | $k = 4, \lambda = 1, m = 4, c = 5$ | 97.5 | 90  | 93.23 | 48.63 |
| MHDE+PSO-KELM                      | $k = 4, \lambda = 1, m = 4, c = 5$ | 100 | 100 | 100 | 159.04 |

FIGURE 16. High-dimensional fault features of four states.

FIGURE 17. Sensitive fault features of four states.

depicted in Fig. 18. It is observed that the presented approach can identify different fault types of the hydraulic pump well, and the recognition accuracy is 100%.

FIGURE 18. Recognition results of the presented approach.

Similar to experiment 1, to compare the presented approach with other approaches and avoid the contingency of diagnosis results, different approaches are adopted to
distinguish the fault types of the hydraulic pump. Each approach runs 50 times. Parameter settings and final results of different approaches are shown in Table 6 and Fig. 19. It is observed that both the presented approach and MHDE approach possess the highest accuracy, but the calculation efficiency of the presented approach is significantly higher. Considering comprehensively, the presented approach possesses superior performance in hydraulic pump fault types identification.

The same as experiment 1, we divided the size of training/testing samples into 10/40, 20/30, 30/20, 40/10 to study the impact of the number of training and testing samples on the performance of different approaches. The average accuracy of 50 runs for each approach is shown in Fig. 20. It is observed that the presented approach possesses the highest accuracy in all four cases. In addition, with the increase of training samples, the accuracy of MMDE, CMDE, RCMDE
and MHDE approaches can also reach a higher level, but it is evident that the presented approach has a wider range of application, so the performance is more superior.

C. EXPERIMENT 3: GEARBOX COMPOUND FAULTS DIAGNOSIS

Experiment 1 and experiment 2 adopt the presented approach to distinguish different fault types of rotating machinery, respectively. The results are satisfactory, which prove the application potential of the presented approach in the field of rotating machinery fault diagnosis. However, experiment 1 and experiment 2 are all to diagnose the faults of a single component. Experiment 1 adopts the presented approach to diagnose different faults types of rolling bearings, while experiment 2 adopts the presented approach to analyze different slipper fault types of the hydraulic pump. Nevertheless, during the actual operation of rotating machinery, the fault types are diverse and may be compound faults of multiple component faults. To verify the performance of the presented approach when diagnosing compound faults of rotating machinery, experiment 3 is conducted.

Experiment 3 employs the compound fault data set of the 2009PHM Challenge gearbox [34]. Fig. 21 is the gearbox structure, which mainly consists of three shafts, four gears, and six bearings. The experiment tested two types of spur gears and helical gears. In this experiment, the test data set corresponding to the helical gears is adopted, in which the corresponding number of teeth of the helical gear 1, 2, 3, 4 is 16, 48, 24, 40, respectively. The speed of the input shaft was 30Hz, and the load was high. The vibration data was collected by two accelerometers installed on the input shaft side and the output shaft side, respectively, and set as accelerometer 1 and accelerometer 2, as shown in Fig. 22. This article uses the vibration data collected by accelerometer 2 for analysis, in which the sampling frequency was 66.7kHz, and the sampling time was 4s. There are six gearbox working states, and the detailed information of each state [35] and the corresponding waveforms are shown in Table 7 and Fig. 23. 50 samples of vibration data of each state are taken, where 30 samples are randomly selected as training samples, and the remaining 20 samples are testing samples. The length of each sample is 2048.

Firstly, we need to select the best \([c, m]\) of MHMDE when extracting the gearbox fault features. The selection process is the same as experiment 1 and experiment 2. Fig. 24 shows the results. It is observed that when \(c = 6, m = 4\), the maximum AED value is obtained, so we set the parameters of MHMDE to \(k = 3, \tau = 8, \lambda = 1, m = 4, c = 6\) when extracting the fault features of the gearbox. Fig. 25 and Fig. 26 show the high-dimensional fault features of the training samples extracted by MHMDE and the sensitive features obtained by MCFS screening, respectively. Input 180 training samples to the classifier for training, and then input 120 testing samples. The parameter settings of PSO-KELM are the same as Table 2, and Fig. 27 shows the results. As we can see, three testing samples in state 1 are misrecognized, and the accuracy rate is 85%, while the recognition accuracy of testing samples in several other states reaches more than 90%. A total of 6 samples of 120 testing samples are misrecognized, with an overall recognition accuracy of 95%, proving that the presented approach can effectively identify different compound fault states of the gearbox.
To compare the performance of different approaches on identifying different compound fault types of gearbox, similar to experiment 1 and experiment 2, each approach runs 50 times to avoid contingency. Parameter settings and final results of different approaches are depicted in Table 8 and Fig. 28. Fig. 29 shows the average accuracy of each approach running 50 times with different training/test samples. Obviously, the identification accuracy of the presented approach is the highest under any cases, and the calculation efficiency is higher than MMDE and MHDE approaches. Combining the two indexes of calculation efficiency and identification accuracy, the presented approach enjoys a noticeable performance advantage in diagnosing the compound fault types of the gearbox.

In this section, we take three typical vibration signals such as rolling bearing, hydraulic pump and gearbox as examples to discuss the practicability of the presented approach in the field of rotating machinery fault diagnosis. Experiments 1 and 2 reflect the good performance of the presented approach in faults diagnosis of a single component, while experiment 3 proves that the presented approach is still very practical when diagnosing complex compound faults of rotating machinery. Analyzing the results of the three experiments, it is observed that the approach presented in this article is a
new approach for fault diagnosis of rotating machinery with a wide application range and excellent performance. And it can effectively identify single component faults or compound faults of different rotating machinery. Compared with other existing approaches, the presented approach possesses obvious advantages.

V. CONCLUSION AND PROSPECTS

This article presents a new approach for fault diagnosis of rotating machinery using MHMDE, MFCs and PSO-KELM. Firstly, in the feature extraction stage, integrating the merits of MHDE and MMDE, MHMDE is proposed and employed to extract the fault features of vibration signals. To exert the best performance of MHMDE, the concept of AED is introduced to determine the best parameters of MHMDE. Then MCFS is employed to screen sensitive features from the obtained high-dimensional features to form sensitive feature vectors. Finally, in the fault recognition stage, PSO is adopted to optimize KELM to build a classifier with excellent performance, and the sensitive feature vectors of training samples and testing samples are input into PSO-KELM-based fault classifier for training and recognition. Through the analysis of three examples, it is proved that this approach can effectively identify different fault types of rotating machinery, and possesses apparent advantages in comparison with other approaches.

Although the presented approach has been proved to be an excellent approach for fault diagnosis of rotating machinery, there are still some improvements that can be made to make the presented approach more practical. Our future research should include the following aspects:

1. The three examples cited in this article are all under constant speed and load. It is unknown whether the presented approach still possesses strong practicality when solving the problem of variable load and speed. Therefore, in the next step, we will verify the performance of the approach under the above conditions. If necessary, we will consider improving the presented approach to adapt to different situations.

2. MHMDE cannot extract fault features from multi-channel vibration data, which limits the performance of MHMDE to some extent. We will try to improve MHMDE into a multi-channel analysis approach to process vibration data from multiple directions of rotating machinery simultaneously.

REFERENCES

[1] R. Liu, B. Yang, E. Zio, and X. Chen, “Artificial intelligence for fault diagnosis of rotating machinery: A review,” Mech. Syst. Signal Process., vol. 108, pp. 33–47, Aug. 2018.
[2] Y. Lei, Z. He, and Y. Zi, “A new approach to intelligent fault diagnosis of rotating machinery,” Expert Syst. Appl., vol. 35, no. 4, pp. 1593–1600, Nov. 2008.
[3] D. Dou, J. Yang, J. Liu, and Y. Zhao, “A rule-based intelligent method for fault diagnosis of rotating machinery,” Knowl. Based Syst., vol. 36, pp. 1–8, Dec. 2012.
[4] J. Chen, Z. Li, J. Pan, G. Chen, Y. Zi, J. Yuan, B. Chen, and Z. He, “Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review,” Mech. Syst. Signal Process., vols. 70–71, pp. 1–35, Mar. 2016.
[5] J. Ben Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, “Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals,” Appl. Acoust., vol. 89, pp. 16–27, Mar. 2015.
[6] Z. Feng, X. Chen, and T. Wang, “Time-varying demodulation analysis for rolling bearing fault diagnosis under variable speed conditions,” J. Sound Vibrat., vol. 400, pp. 71–85, Jul. 2017.
[7] Y. Lei, J. Lin, Z. He, and M. J. Zuo, “A review on empirical mode decomposition in fault diagnosis of rotating machinery,” Mech. Syst. Signal Process., vol. 35, nos. 1–2, pp. 108–126, Feb. 2013.
[8] J. Cheng, Y. Yi, and Y. Yu, “A rotating machinery fault diagnosis approach based on local mean decomposition,” Digit. Signal Process., vol. 22, no. 2, pp. 356–366, 2012.
[9] R. Yan, R. X. Gao, and X. Chen, “Wavelets for fault diagnosis of rotary machines: A review with applications,” Signal Process., vol. 96, pp. 1–15, Mar. 2014.
[10] S. Luo, W. Yang, and Y. Luo, “Fault diagnosis of a rolling bearing based on adaptive sparsest narrow-band decomposition and refined coarse multiscale dispersion entropy,” Entropy, vol. 22, no. 4, p. 375, Mar. 2020.
[11] Y. Li, M. Xu, R. Wang, and W. Huang, “A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy,” J. Sound Vibrat., vol. 360, pp. 277–299, Jan. 2016.
[12] J. S. Richman and J. R. Moorman, “Physiological time-series analysis using approximate entropy and sample entropy,” Amer. J. Physiol.-Heart Circulatory Physiol., vol. 278, no. 6, pp. H2039–H2049, Jun. 2000.
[13] J. Zheng, J. Cheng, and Y. Yang, “A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy,” Mechanism Mach. Theory, vol. 70, pp. 441–453, Dec. 2013.

[14] C. Bandt and B. Pompe, “Permutation entropy: A natural complexity measure for time series,” Phys. Rev. Lett., vol. 88, no. 17, Apr. 2002, Art. no. 174102.

[15] S.-D. Wu, C.-W. Wu, and A. Humeau-Heurtier, “Refined scale-dependent permutation entropy to analyze systems complexity,” Phys. A. Stat. Mech. Appl., vol. 450, pp. 454–461, May 2016.

[16] M. Rostaghi and H. Azami, “Dispersion entropy: A measure for time-series analysis,” IEEE Signal Process. Lett., vol. 23, no. 5, pp. 610–614, May 2016.

[17] H. Azami and J. Escudero, “Amplitude- and fluctuation-based dispersion entropy,” Entropy, vol. 20, no. 3, p. 210, Mar. 2018.

[18] H. Azami and J. Escudero, “Improved multiscale permutation entropy for biomedical signal analysis: Interpretation and application to electroencephalogram recordings,” Biomed. Signal Process. Control, vol. 23, pp. 28–41, Jan. 2016.

[19] H. Azami and J. Escudero, “Coarse-graining approaches in univariate multiscale sample and dispersion entropy,” Entropy, vol. 20, no. 2, p. 138, Feb. 2018.

[20] M. Costa, A. L. Goldberger, and C.-K. Peng, “Multiscale entropy analysis of complex physiologic time series,” Phys. Rev. Lett., vol. 89, no. 6, pp. 705–708, Jul. 2002.

[21] H. Azami, E. Kinney-lang, A. Ebied, A. Fernandez, and J. Escudero, “Multiscale dispersion entropy for the regional analysis of resting-state magnetoencephalogram complexity in Alzheimer’s disease,” in Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2017, pp. 3182–3185.

[22] X. Yan and M. Jia, “Intelligent fault diagnosis of rotating machinery using improved multiscale dispersion entropy and mRMR feature selection,” Knowl.-Based Syst., vol. 163, pp. 450–471, Jan. 2019.

[23] H. Azami, M. Rostaghi, D. Abásolo, and J. Escudero, “Refined composite multiscale dispersion entropy and its application to biomedical signals,” IEEE Trans. Biomed. Eng., vol. 64, no. 12, pp. 2872–2879, Dec. 2017.

[24] S.-D. Wu, C.-W. Wu, K.-Y. Lee, and S.-G. Lin, “Modified multiscale entropy for short-term time series analysis,” Phys. A. Stat. Mech. Appl., vol. 392, no. 23, pp. 5865–5873, Dec. 2013.

[25] Y. Jiang, C.-K. Peng, and Y. Xu, “Hierarchical entropy analysis for biological signals,” J. Comput. Appl. Math., vol. 236, no. 5, pp. 728–742, Oct. 2011.

[26] Y. Li, G. Li, Y. Yang, X. Liang, and M. Xu, “A fault diagnosis scheme for planetary gearboxes using adaptive multi-scale morphology filter and modified hierarchical permutation entropy,” Mech. Syst. Signal Process., vol. 105, pp. 319–337, May 2018.

[27] D. Cai, C. Zhang, and X. He, “Unsupervised feature selection for multi-cluster data,” in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), 2010, pp. 333–342.

[28] G.-B. Huang, “An insight into extreme learning machines: Random neurons, random features and kernels,” Cognit. Comput., vol. 6, no. 3, pp. 376–390, Sep. 2014.

[29] H. Azami, S. E. Arnold, S. Sanei, Z. Chang, G. Sapiro, J. Escudero, and A. S. Gupta, “Multiscale fluctuation-based dispersion entropy and its applications to neurological diseases,” IEEE Access, vol. 7, pp. 68718–68733, 2019.

[30] E. Song, Y. Ke, C. Yao, Q. Dong, and L. Yang, “Fault diagnosis method for high-pressure common rail injector based on LFOA-VMD and hierarchical dispersion entropy,” Entropy, vol. 21, no. 10, p. 923, Sep. 2019.

[31] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: Theory and applications,” Neurocomputing, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.

[32] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in Proc. IEEE ICNN, vol. 4. Nov./Dec. 1995, pp. 1942–1948.

[33] Case Western Reserve University. Bearing Data Center. Accessed: Oct. 24, 2019. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/pages/download-data-file

[34] Phm Data Challenge 2009. Accessed: Apr. 24, 2016. [Online]. Available: https://www.phmsociety.org/competition/PHM/09

[35] L. Jing, M. Zhao, P. Li, and X. Xu, “A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox,” Measurement, vol. 111, pp. 1–10, Dec. 2017.

**Fuming Zhou** was born in Zhengzhou, China. He is currently pursuing the master’s degree in mechanical engineering with the College of Field Engineering, Army Engineering University. His research interests include machine learning, signal processing, and fault diagnosis of rotating machinery.

**Jinxing Shen** was born in Xuzhou, China. He is currently a Lecturer with the College of Field Engineering, Army Engineering University. His research interests include fault diagnosis of mechanical equipment and machine learning.

**Xiaoqiang Yang** was born in Taiyuan, China. He received the B.S. and Ph.D. degrees in mechanical engineering from the PLA University of Science and Technology, Nanjing, China, in 1989 and 1994, respectively. He is currently a Professor with the College of Field Engineering, Army Engineering University. His research interests include fault diagnosis of mechanical equipment and machine learning.

**Xiaolin Liu** was born in Yulin, China. He is currently pursuing the master’s degree in mechanical engineering with the College of Field Engineering, Army Engineering University. His research interests include machine learning and fault diagnosis of rotating machinery.

**Wuqiang Liu** was born in Jingmen, China. He is currently pursuing the master’s degree in mechanical engineering with the College of Field Engineering, Army Engineering University. His research interests include machine learning and fault diagnosis of rotating machinery.