

# Cross Diversity Entropy-Based Feature Extraction for Fault Diagnosis of Rotor System

Yongbo Li , Member, IEEE, Zehang Jiao , Shun Wang , Ke Feng , Member, IEEE, and Zheng Liu , Senior Member, IEEE

**Abstract**—The rotor system in the fault state generally shows apparent nonlinear behavior and complex dynamic characteristics. In general, the data collected from multiple sensors (namely, multichannel signals) is required to achieve accurate condition monitoring of the rotor system. Even though the traditional multivariate entropy method can extract fault characteristics of the vibration signals from multiple sensors in the rotor system, some critical fault information might be lost when it is applied to process the multivariate information. In this article, a novel approach called cross diversity entropy (Cross-DE) is proposed to address the issue discussed above. More specifically, when processing the multichannel signals, the developed Cross-DE uses the diversity of orbits in different phase spaces to represent the complexity of the whole system. Moreover, in the process of calculating the diversity of orbits, Cross-DE calculates the cosine similarity between the orbits in the same phase space and different phase spaces, resulting in information interaction between different channels. Furthermore, the developed Cross-DE fuses the information from different channels to solve the problem of fault information loss during feature extraction. Therefore, the developed Cross-DE has the capability of accurately extracting complete fault features from multichannel, which benefits the health management of rotor system condition monitoring significantly. The 2-D displacement signal of the rotor is applied in this article to demonstrate the performance of the Cross-DE. Also, the effectiveness of Cross-DE is verified using both simulated signal and experimental data.

**Index Terms**—Cross diversity entropy (Cross-DE), fault diagnosis, feature extraction, rotor system.

Manuscript received 15 November 2022; revised 15 January 2023, 2 March 2023, 2 June 2023, and 21 August 2023; accepted 11 September 2023. Recommended by Technical Editor A. Alanis and Senior Editor R. Gao. (Corresponding author: Ke Feng.)

Yongbo Li, Zehang Jiao, and Shun Wang are with the School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: yongbo@nwpu.edu.cn; jzh1138119507@163.com; wangshun@mail.nwpu.edu.cn).

Ke Feng is with the Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore 117576 (e-mail: ke.feng@outlook.com.au).

Zheng Liu is with the School of Engineering, University of British Columbia, Kelowna, BC V1V 1V7, Canada (e-mail: zheng.liu@uottawa.ca).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TMECH.2023.3318633>.

Digital Object Identifier 10.1109/TMECH.2023.3318633

## I. INTRODUCTION

ROTATING machinery is widely used in a range of modern industries, such as transportation, power equipment, aerospace, vehicles, etc [1], [2], [3]. Due to the harsh working environments, the rotor system of rotating machinery is prone to failure, which will lead to high maintenance costs and even serious accidents. Therefore, it is of great significance for condition monitoring and fault diagnosis of rotating machinery, especially for the rotor system [4], [5], [6].

In recent decades, the feature extraction and fault recognition of rotor systems have been a research hotspot [7], [8], [9]. For instance, the authors in [10] proposed a modified stacked autoencoder (MSAE), which uses an adaptive Morlet wavelet to diagnose faults of rotating machinery automatically. With the development of machine learning, Chen et al. [11] developed a CNN-based duplet classifier for rotor and bearing fault diagnosis, which can accurately identify mixed faults. Xia [12] developed a fault diagnosis method for rotating machinery based on CNN, which considers multiple sensors information fusion to achieve higher diagnostic accuracy. Li et al. [13] proposed a novel online domain adaptation learning method based on deep reinforcement learning, which has excellent diagnostic performance and adaptability. Besides, some other scholars use variational mode decomposition and ensemble empirical mode decomposition for rotor fault diagnosis [14], [15].

In general, the nonlinear behavior in the rotor system will become stronger when faults occur, and the corresponding measured vibration signal contains rich fault information, which could significantly benefit the health management of the rotor system [16], [17]. Therefore, extracting useful features from vibration signals has become a crucial procedure for fault diagnosis [18], [19]. As for the operating rotor system, the motion of the rotor shaft is usually tracked using displacement signals in the horizontal and vertical directions. The signals in the horizontal and vertical directions are obtained by a pair of displacement sensors at 90° to each other. These are typical 2-D signals [20], [21]. Deep learning approaches, such as hybrid attention mechanism can extract the features of 2-D vibration signals and achieve good fault diagnosis results [22], [23]. However, it is important to consider the limitations of hybrid attention mechanisms and deep learning. These include challenges related to interpretability, data requirements, and computational resources.

In processing such vibration signals, the entropy-based fault characterization method has been a vital tool for feature extraction in recent years [24]. The entropy method is widely used because it does not require prior knowledge, signal preprocessing, and only needs to set fewer parameters when characterizing vibration signals [25]. Moreover, the entropy method provides advantages, such as simplicity, interpretability, low computational cost, and high data efficiency in handling complex data. The entropy-based method can extract features from vibration signals, so as to accurately identify the health state of the system. The principle of the entropy-based method is that when there is a higher complexity in the vibration signal, the corresponding entropy will be greater [26], [27]. The rotor vibration signal characteristics can be extracted by entropy algorithm to realize rotor condition monitoring and fault diagnosis [28].

Entropy theory originates from information entropy, which is a physical quantity used to characterize the complexity of information [29]. Approximate entropy (AE) uses entropy theory to measure the complexity of any time series [30]. Richman et al. [31] proposed a sample entropy (SE), which uses the correlation dimension to measure complexity. The fuzzy entropy (FE), which is an improvement of SE, was introduced in [32]. Bandt et al. [33] proposed a permutation entropy (PE), which uses the state probability in information entropy to measure complexity. Recently, a new entropy method, namely, diversity entropy (DE), is proposed, which introduces cosine similarity to measure the divergence of trajectory in phase space [34]. Compared with SE, FE, and PE, DE has three advantages: 1) High consistency, 2) robust-to-noise influence, and 3) calculation efficiency.

It should note that the single-scale entropy algorithm usually ignores the correlation of time series at different scales, and the amount of information is limited. To address this issue, Costa et al. [35] extended the SE to the multiscale space and proposed a multiscale sample entropy (MSE). Similarly, FE and PE are also extended to be a multiscale fuzzy entropy (MFE) [36], [37] and multiscale permutation entropy (MPE) [38], [39]. DE is also extended from single scale to multiscale space in [34], namely, the multiscale diversity entropy (MDE), and the MDE has a strong capability of extracting abundant fault information. However, it is difficult to accurately identify the failure mode of rotating machinery with a single-channel signal. Zografos et al. [40] extended information entropy to multivariate entropy for feature extraction of multisource information fusion, and proposed multivariate embedding theory. Combining multivariate embedding theory with MSE, MFE, and MPE, multivariate multiscale sample entropy (MVMSE), multivariate multiscale fuzzy entropy (MVMFE), and multivariate multiscale permutation entropy (MVMPE) were proposed in [41], [42], and [43]. In essence, these multivariate strategies construct the same phase space for each channel to realize the feature extraction of multichannel signals.

Through the combination of MDE and multivariate strategy, multivariate multiscale diversity entropy (MVMDE) and multichannel multiscale diversity entropy (MCMDE) were proposed in [44]. The calculation process of MVMDE is similar to MVMSE and MVMFE, and the process of MCMDE is similar

to MVMPE. However, the original multivariate strategy has difficulties in distinguishing single faults in different channels. In order to solve this problem, variational embedding multiscale diversity entropy (VEMDE) was proposed [44]. The variational embedding strategy constructs phase spaces with different embedding dimensions for different channels, improving fault characteristics' separability.

The main challenges and problems are as follows: For the condition monitoring and fault diagnosis of rotor systems, it is necessary to extract the features of 2-D displacement signals. However, take the MDE-based multivariate strategy as an example, the current multivariate strategies are essentially an independent calculation of multivariate phase space. Then accumulates the cosine similarity in the two channels, which can only reflect the information in the channels, and the correlation and coupling between the horizontal and vertical displacements are not reflected. These multivariate strategies may lead to the lack of information when processing signals and impair the performance of feature extraction and fault diagnosis accuracy.

This article proposes a novel multichannel strategy cross diversity entropy (Cross-DE) for 2-D signal processing to solve this problem. The main improvement of Cross-DE compared with other methods is to analyze the correlation between different channels. Specifically, after constructing the phase space for two channels, it calculates the cosine similarity of a single channel and the cosine similarity between two channels, which supplements the information on correlation and coupling between signals. In this way, the coupling information between channels is captured, and the relationship between signals is considered to avoid the loss of information in the multivariate fusion process. This method can extract more comprehensive information, significantly improving feature extraction ability and providing more accurate fault diagnosis results. It is worth noting that Cross-DE is different from cross-entropy methods [45], which are used to quantify the synchronism or similarity between two time series.

To conclude, the main contributions of this article can be summarized as follows.

- 1) A reliable multichannel strategy called Cross-DE is proposed to extract the 2-D signal. This method considers the information fusion between different channel signals, and effectively solves the problem of fault information loss in the process of feature extraction of multichannel signals.
- 2) Cross-DE has better feature extraction ability and higher classification accuracy than VEMDE, MVMDE, and MCMDE, and this has been proved by simulation and experimental studies. This method is an effective tool for typical fault diagnosis of the rotor systems.

The rest of this article is organized as follows. Section II describes the methodology, more specifically, the proposed Cross-DE method. Section III uses simulation and experiment to evaluate the performance of Cross-DE, and its effectiveness for fault identification is validated. Finally, the effectiveness of the proposed method and its possible applications are highlighted in Section IV. Finally, Section V concludes this article.

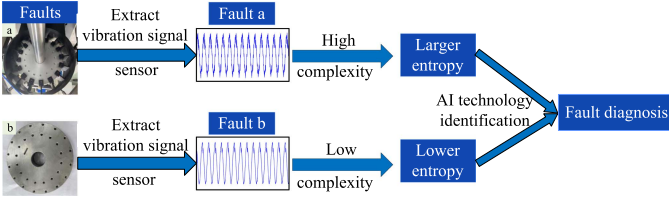


Fig. 1. Brief illustration of fault diagnosis using entropy-based methods.

## II. METHODOLOGY

In this section, first, the principle of fault diagnosis method based on entropy is briefly explained. Then, the five steps of calculating Cross-DE will be introduced in detail. The disadvantages of other multivariate methods and the improvements made by Cross-DE are also presented. After that, the phase space embedding dimension is discussed, and the developed fault diagnosis framework based on Cross-DE and extreme learning machine (ELM) is introduced.

The principle of fault diagnosis of rotor system based on entropy is shown in Fig. 1. First, the vibration signals of different faults should be extracted from the sensors. Then, these signals will be used for feature extraction by the entropy algorithm. Finally, artificial intelligence technology is used for feature recognition and fault diagnosis. Feature extraction is a vital step in the process of fault diagnosis. The effect of fault features extracted from vibration signals directly affects the accuracy of diagnosis results. Entropy method is an excellent feature extraction method. Entropy can characterize the complexity of the signals. The larger the complexity of the vibration signal, the larger the feature value extracted by Cross-DE. Therefore, entropy methods can effectively distinguish different health conditions and be used for fault diagnosis.

### A. Proposed Cross Diversity Entropy

1) *Theory of Cross Diversity Entropy*: In this section, a novel method, called Cross-DE, is proposed and introduced. It is used to describe the complexity of 2-D time series. For a 2-D time series  $X$ , it can be written as

$$X = \begin{bmatrix} \mathbf{X}^1 \\ \mathbf{X}^2 \end{bmatrix} = \begin{bmatrix} x_1^1 x_2^1 \cdots x_N^1 \\ x_1^2 x_2^2 \cdots x_N^2 \end{bmatrix}. \quad (1)$$

$\mathbf{X}^c (c = 1, 2)$  represents two signal channels, where  $N$  denotes the data length of time series in each channel.

The proposed Cross-DE calculation method is introduced as follows.

*Step 1*: Divide the signals of the two channels into multiscale time series in the same way. Segment the channel  $\mathbf{X}^c (c = 1, 2)$  into multiscale time series  $\mathbf{Y}^c = \{\mathbf{Y}_1^c, \mathbf{Y}_2^c, \dots, \mathbf{Y}_\tau^c\}$  which is a positive integer, representing the scale factor in the multiscale analysis. For different scale factors  $\tau$ , the time series after segmentation is  $\mathbf{Y}_\tau^c = [y_{1,\tau}^c, y_{2,\tau}^c, \dots, y_{k,\tau}^c, \dots, y_{j,\tau}^c]$ ,  $j = N/\tau$ ,

$k = 1, 2, \dots, j$ ,  $y_{k,\tau}^c$  can be calculated as follows:

$$y_{k,\tau}^c = \frac{1}{\tau} \sum_{i=(k-1)\tau+1}^{\tau k} x_i^c. \quad (2)$$

*Step 2*: According to the phase space embedding theory [46], [47], the time series at each scale factor  $\mathbf{Y}_\tau^c$  can be reconstructed into a series of orbits, the embedding dimension is  $m$ , and the structure of the phase space is demonstrated as follows:

$$\mathbf{Y}_\tau^c(m) = [\mathbf{Y}_{1,\tau}^c, \mathbf{Y}_{2,\tau}^c, \dots, \mathbf{Y}_{j-m+1,\tau}^c] \\ = \begin{bmatrix} y_{1,\tau}^c & y_{2,\tau}^c & \cdots & y_{j-m+1,\tau}^c \\ y_{2,\tau}^c & y_{3,\tau}^c & \cdots & y_{j-m+2,\tau}^c \\ \vdots & \vdots & \ddots & \vdots \\ y_{m,\tau}^c & y_{m+1,\tau}^c & \cdots & y_{j,\tau}^c \end{bmatrix}. \quad (3)$$

$\mathbf{Y}_\tau^c(m)$  is the phase space reconstructed by the time series in channel  $c$  under the scale factor  $\tau$  after embedding in dimension  $m$ .  $\mathbf{Y}_\tau^c(m)$  is a  $(m) \times (j - m + 1)$  matrix.

*Step 3*: First, calculate the cosine similarity between adjacent orbits in two channels  $\mathbf{D}_a^c(m) = \{d_1^c, d_2^c, \dots, d_{j-m}^c\}$ ,  $c = 1, 2$ . After that, by considering the correlation between the 2-D time series, Cross-DE will calculate the cosine similarity between the corresponding orbits of the two channels  $\mathbf{D}_b(m) = \{d_1, d_2, \dots, d_{j-m+1}\}$ .

The cosine similarity is calculated as follows:

$$\mathbf{D}_a^c(m) = \{d_1^c, d_2^c, \dots, d_{j-m}^c\} \\ = \{d^c(\mathbf{Y}_1^c, \mathbf{Y}_2^c), d^c(\mathbf{Y}_2^c, \mathbf{Y}_3^c) \\ \cdots, d^c(\mathbf{Y}_{j-m}^c, \mathbf{Y}_{j-m+1}^c)\} \quad (4)$$

$$d^c(\mathbf{Y}_i^c, \mathbf{Y}_{i+1}^c) = \frac{\sum_{k=1}^m y_{i+k-1}^c \times y_{i+k}^c}{\sqrt{\sum_{k=1}^m (y_{i+k-1}^c)^2} \times \sqrt{\sum_{k=1}^m (y_{i+k}^c)^2}} \quad (5)$$

$$\mathbf{D}_b(m) = \{d_1, d_2, \dots, d_{j-m+1}\} \\ = \{d_1(\mathbf{Y}_1^1, \mathbf{Y}_1^2), d_2(\mathbf{Y}_2^1, \mathbf{Y}_2^2), \\ \cdots, d_{j-m+1}(\mathbf{Y}_{j-m+1}^1, \mathbf{Y}_{j-m+1}^2)\} \quad (6)$$

$$d(\mathbf{Y}_s^1, \mathbf{Y}_s^2) = \frac{\sum_{k=1}^m y_{s+k-1}^1 \times y_{s+k-1}^2}{\sqrt{\sum_{k=1}^m (y_{s+k-1}^1)^2} \times \sqrt{\sum_{k=1}^m (y_{s+k-1}^2)^2}} \quad (7)$$

$$\mathbf{D} = \mathbf{D}_a^1(m) + \mathbf{D}_a^2(m) + \mathbf{D}_b(m) \quad (8)$$

where  $c = 1, 2$ ,  $i = 1, 2, \dots, j - m$ ,  $s = 1, 2, \dots, j - m + 1$ . Under the scale factor  $\tau$ ,  $\mathbf{D}_a^c(m)$  represents the cosine similarity calculated by the  $c$  channel, and  $\mathbf{D}_b(m)$  represents the cosine similarity calculated between the two channels. Therefore, a total of  $2(j - m) + (j - m + 1)$  cosine similarities will be generated at each scale. The cosine similarity of two orbits in phase space represents the similarity between them, and the range of cosine similarity is  $[-1, 1]$ . When the calculated cosine similarity tends to 1, it means that the two orbits tend to be periodic and deterministic dynamical behavior. On the contrary,

**Algorithm 1:** Cross Diversity Entropy.

**Input:**  $x$ : the time series,  
 $m$ : the embedding dimension,  
 $t$ : the time delay value (it is usually equal to 1),  
 $\tau$ : the scale factor,  
 $\varepsilon$ : the number of symbols

**Output:**  $Cross - DE$

- 1: **for**  $s = 1$  to  $\tau$  **do**
- 2: obtain the time series  $Y_s$  at scale  $s$ ;
- 3: **for**  $c = 1$  to 2 **do**
- 4: reconstruct the phase space  $Y_s^c(m)$ ;
- 5: calculate cosine similarity of the same channel  $D_a^c(m)$ ;
- 6: **end for**
- 7: calculate the cosine similarity between the two channels  $D_b(m)$ ;
- 8: obtain all cosine similarity  $D = [D_a^1(m), D_a^2(m), D_b(m)]$ ;
- 9: calculate the number that  $D$  falling into each subinterval;
- 10: calculate the state probability  $P_k = (P_1, P_2, \dots, P_\varepsilon)$ ;
- 11: calculate  $Cross - DE_s = -\frac{1}{\ln \varepsilon} \sum_{k=1}^{\varepsilon} P_k \ln P_k$ ;
- 12: **end for**
- 13:  $Cross - DE = \{Cross - DE_1, \dots, Cross - DE_\tau\}$
- 14: **return**  $Cross - DE$

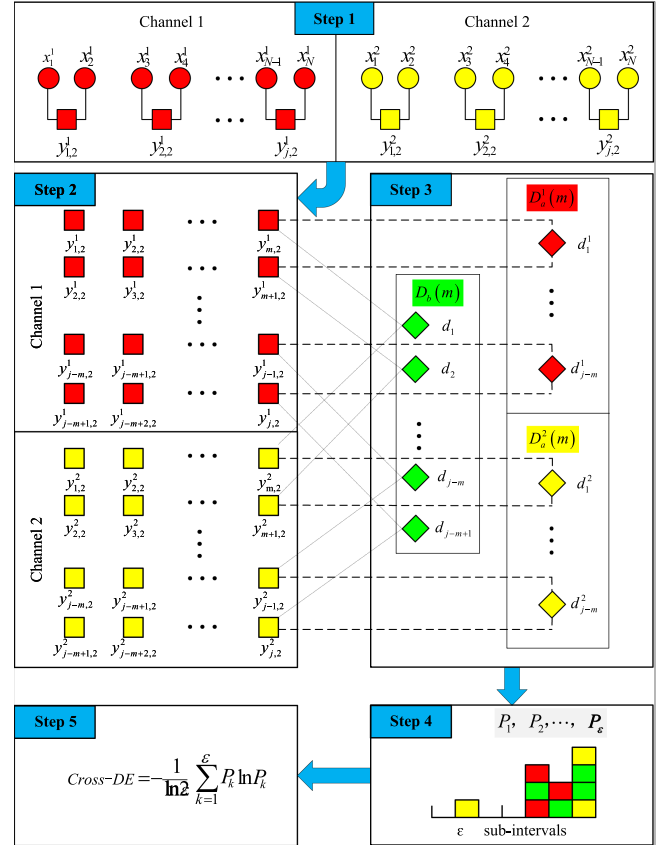


Fig. 2. Schematic diagram of the calculation steps for the Cross-DE.

when the calculated cosine similarity tends to -1, it means that the two orbits tend to have chaotic and stochastic dynamical behavior.

*Step 4:* Divide the interval  $[-1, 1]$  into  $\varepsilon$  subintervals as  $(I_1, I_2, \dots, I_\varepsilon)$ . Then, all the calculated cosine similarities are classified into these subintervals, and the state probability  $(P_1, P_2, \dots, P_\varepsilon)$  that the cosine similarities fall into each subinterval is obtained.

*Step 5:* Substitute the state probability into the following equation, the Cross-DE of the 2-D time series when the scale factor is  $\tau$  can be calculated

$$Cross - DE = -\frac{1}{\ln \varepsilon} \sum_{k=1}^{\varepsilon} P_k \ln P_k. \quad (9)$$

Like other multivariate MDE methods, the calculated range of Cross-DE is  $[0, 1]$ . Cross-DE will increase monotonically with the increase of dynamical complexity. When the Cross-DE approaches 1, the more complicated the time series is, which represents a more chaotic and irregular phenomenon. Conversely, when the Cross-DE approaches 0, the lower the complexity of the time series is, which represents a more periodic and deterministic phenomenon. The pseudo-code of Cross-DE is given in Algorithm 1.

In order to more clearly explain the calculation steps of Cross-DE, the algorithm diagram is executed, by taking scale  $\tau=2$  as an example (as shown in Fig. 2). Step 1 is to segment the channel  $X^c$  ( $c = 1, 2$ ) into time series  $Y_2^c = [y_{1,2}^c, y_{2,2}^c, \dots, y_{j-m+1,2}^c]$  with scale  $\tau=2$ . Step 2 is to reconstruct the time series into

a series of orbits  $Y_2^c(m) = [Y_{1,2}^c, Y_{2,2}^c, \dots, Y_{j-m+1,2}^c]$ , and the embedding dimension is  $m$ . Step 3 is to calculate the cosine similarity  $D_a^c(m)$  of the orbits in the same channel and the cosine similarity  $D_b(m)$  of the orbits between the two channels. Step 4 is to calculate the probability  $(P_1, P_2, \dots, P_\varepsilon)$  of cosine similarity falling into  $\varepsilon$  subinterval. Step 5 is to substitute the obtained probability into (9) and calculate the Cross-DE.

Entropy is a physical quantity that can quantify the dynamic complexity of 2-D time series. The multivariate entropy methods based on DE use the cosine similarity between orbits to calculate the state probability, and the entropy obtained can be regarded as the expectation of the diversity between orbits in the whole system. The original multivariate entropy methods only calculate the cosine similarity between adjacent orbits in two-phase spaces, which is  $D_a^c(m)$ , and the number is  $2(j - m)$ . These methods cannot reflect the correlation and coupling between 2-D signals, and do not correlate and synthesize the information between the two channels.

However, in order to quantify the complexity of the system, the diversity between orbits needs to be considered more comprehensively. The diversity of orbits in different phase spaces also represents the complexity of the whole system. Therefore, Cross-DE increases the calculation of cosine similarity between orbits in different phase spaces, which is  $D_b(m)$ , and the number of cosine similarities increases by  $j - m + 1$ . In



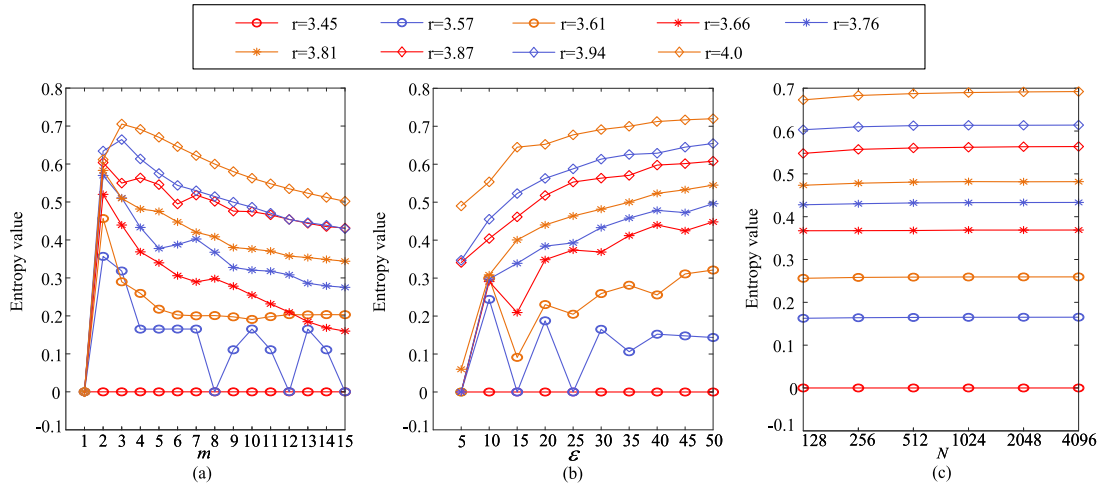


Fig. 3. For systems of different complexity, the entropy value calculated by DE with different parameters choices. (a) Fixed  $\varepsilon = 30$ ,  $N = 2048$ , and varied  $m$  (1 to 15). (b) Fixed  $m = 4$ ,  $N = 2048$ , and varied  $\varepsilon$  (5 to 50). (c) Fixed  $m = 4$ ,  $\varepsilon = 30$ , and varied  $N$  (128 to 4096).

this way, the correlation and coupling between 2-D signals are considered. Cross-DE considers the information fusion between two channels, which can extract the features of 2-D signals more accurately.

**2) Parameter Analysis of Cross Diversity Entropy:** The proposed Cross-DE is a multivariate multiscale method based on DE. In order to guarantee the optimal feature extraction effect, the most crucial three parameters for DE feature extraction should be considered first: the embedding dimension  $m$ , the number of symbols  $\varepsilon$ , and the signal length  $N$ .

The parameter analysis of DE will use logistic map [48], [49]. The basic iterative equation of the logistics map is as follows:

$$x_{n+1} = rx_n(1 - x_n) \quad (10)$$

where  $x_0$  is any number within  $[0, 1]$ , and  $r$  indicates a variable parameter. When  $r \in (1, 3)$ , after multiple iterations, a stable value of  $x$  can always be obtained; when  $r \in (3, 3.448)$ , two different solutions will be obtained; when  $r > 3.5699$ , the system enters a chaotic state; when  $r$  continues to increase to  $1 + \sqrt{8}$ , the system returns stability and presents three solutions; when  $r$  continues to increase, the system returns to chaos. Changing different parameters  $r$  will change the complexity of the system. The steps to discuss the embedding dimension  $m$  using logistic map are presented as follows.

**Step 1:** Choose a set of parameters  $r$  that can make the system chaos be different.  $x_0 = 0.1$  and the time series is 410 000 points generated by 410 000 iterations in this article. A total of nine  $r$  values are taken, and the complexity of the system increases along the increase of  $r$ , where  $r \in \{3.45, 3.57, 3.61, 3.66, 3.76, 3.81, 3.87, 3.94, 4.0\}$ .

**Step 2:** Change the values of the three parameters, respectively, and use DE to extract features of the time series in step 1, and distinguish systems of different complexity. The time-series input to DE is divided into 100 samples. The change value of signal length is set to  $N \in [128, 4096]$ , the embedding dimension is set to  $m \in [1, 15]$ , and the number of symbols is set to  $\varepsilon \in [5, 50]$ .

**Step 3:** Change the value of a parameter to compare the effect of DE in distinguishing different complexity systems. The parameter value with the best effect is most suitable for DE feature extraction.

The calculation result is shown in Fig. 3. Different polylines in the figure represent different complexity. Any point represents the entropy value calculated by DE for the current complexity system when the parameter takes the corresponding abscissa value. Moreover, the complexity increases with the increase of  $r$ , so the entropy value should also increase accordingly. Under the same parameter, the more significant the difference in entropy value corresponding to different  $r$ , the better the DE can distinguish systems of different complexity.

Fig. 3(a) shows the selection of  $m$ , the other two parameters are fixed as  $\varepsilon = 30$  and  $N = 2048$ . It can be found that when  $m = [4, 6]$ , DE has the best effect in distinguishing systems with different complexity. When  $m$  takes other values, the difference in entropy values corresponding to systems of different complexity is too small, and even violates the law that the entropy value increases with the increase of  $r$ . Fig. 3(b) is the selection of  $\varepsilon$ , the other two parameters are fixed as  $m = 4$  and  $N = 2048$ . Based on the same criteria, it can be seen that the optimal value of  $\varepsilon$  is  $[30, 50]$ . At this time, the differentiation of different complexity is most obvious. Fig. 3(c) shows the selection of  $N$ , the other two parameters are fixed as  $m = 4$  and  $\varepsilon = 30$ . It shows that the change in signal length  $N$  cannot have a significant impact on DE performance. Based on the above discussion, set the signal length as  $N = 2048$ . The number of symbols is set to  $\varepsilon = 30$ . The embedding dimension is set to  $m = 4$ .

## B. Cross Diversity Entropy Based Fault Diagnosis

First, the fault features are extracted from the 2-D signal describing the trajectory of the rotor shaft. Then, the ELM classifier is used to identify different types of faults. Fig. 4 is a schematic diagram of the proposed fault diagnosis method. Specific steps are introduced as follows.

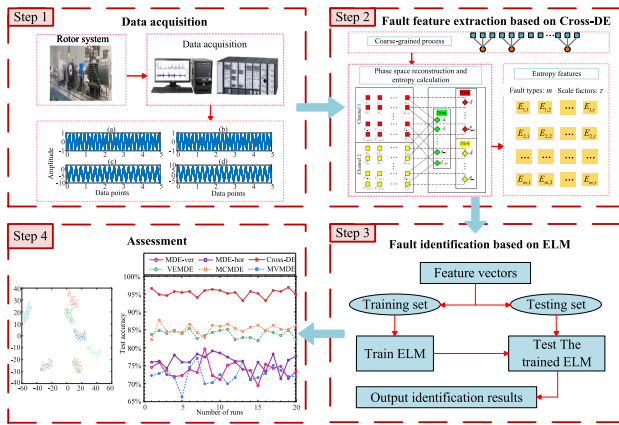


Fig. 4. Schematic diagram of the proposed fault diagnosis method.

*Step 1:* Measure the vibration signal of the rotor axis orbit under different health conditions (horizontal and vertical directions).

*Step 2:* Cross-DE is used for feature extraction, and the measured signal is quantified as an entropy value, representing the complexity of the signal in different health conditions.

*Step 3:* Use half of the fault features as a training sample set to train ELM. Then use the remaining fault features as a test set to test the trained ELM to identify different health conditions.

*Step 4:* Step 3 will be repeated 20 times, and the mean testing accuracy is taken as the fault diagnosis results.

### III. RESULTS AND VALIDATION

In this section, simulation and experiment are used to validate the effectiveness and superiority of the proposed method. Simulation can help evaluate the fault conditions that are difficult to realize in practice; also, it can simply simulate different degrees of faults. The simulation will not cause false verification caused by the actual assembly and other environmental noise. It is an effective research method with generality. The simulation simulates the fault signals in the ideal environment, but there is a lot of interference in the practical working environment, so the effectiveness of the method needs to be further verified in the practical environment, which can be achieved in the experimental study. Therefore, simulation and experiment are complementary and indispensable, and both of them are applied to validate the effectiveness of the developed method for fault diagnosis.

#### A. Simulation Evaluation

1) *Rotor Simulation System:* The rotor model used in this simulation is a single rotor model with two supporting bearings, which are 6205 deep groove ball bearings. Fig. 5 is a 3-D model of the rotor test bench based on SolidWorks. Fig. 6 shows the simplified rotor model. The system is divided into eight units and six nodes. The rotor shaft is simplified as an Euler–Bernoulli beam [50], corresponding to ①, ②, ③, ⑤, and ⑥ units; the blade disk rotor is represented by the lumped-mass method, corresponding to unit ④; the support bearing is simplified as the

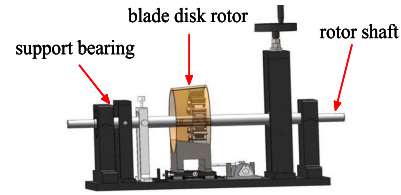


Fig. 5. 3-D model of the rotor test bench.

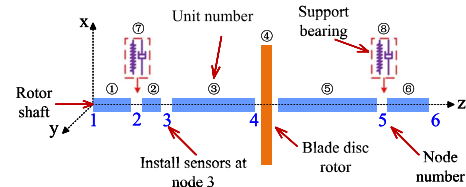


Fig. 6. Simplified rotor model: rotor shaft (corresponding to ①, ②, ③, ⑤, and ⑥ units), blade disc rotor (corresponding to ④), support bearing (corresponding to ⑦ and ⑧ units).

TABLE I  
MAIN PARAMETERS OF THE ROTOR TEST BENCH

Parameters of rotor system	Value	Parameters of rotor system	Value
Shaft length(m)	0.65	Disc weight(kg)	1.202
Shaft weight(kg)	3.183	Disc thickness(m)	0.032
Shaft radius1(m)	0.022	Disc outer radius(m)	0.179
Shaft radius2(m)	0.03	Disc inner radius(m)	0.03
Shaft density(kg/m <sup>3</sup> )	78500	Length of unit ① (m)	0.0825
Elastic modulus of shaft (GPa)	205	Length of unit ② (m)	0.0475
Shaft Poisson's ratio	0.29	Length of unit ③ (m)	0.165
Bearing support stiffness(N/m)	5.00E+07	Length of unit ④ (m)	0.2725
Bearing support damp(N*s/m)	0	Length of unit ⑥ (m)	0.0825

parallel connection of the damper and the spring, corresponding to ⑦ and ⑧ units. Among them, node 3 is the signal point for measuring the vibration displacement.

The main parameters of the rotor test bench are shown in Table I. The finite element method is applied to model the simplified rotor shaft, blade disk rotor, and rotor system, and finally obtain the general dynamic equation [51], [52]. According to the first-order differential equation finally established in the Appendix of this article, the response analysis of the rotor can be solved. The simulation in this section introduces different generalized force vectors to simulate the response analysis of rotor imbalance and friction.

2) *Fault Diagnosis Results of the Simulated Rotor System:* This simulation introduces different generalized force vectors to the dynamic equations of the rotor system. The three health conditions of rotor imbalance (RI), shaft friction (SF), and full-periphery friction (FPF) are simulated. White Gaussian noise is added to the obtained simulation signal, and the signal-noise ratio of the noise is 20 db. The simulated vertical displacement signals (VDS) and horizontal displacement signals (HDS) of the three health conditions are shown in Fig. 7.

The Cross-DE, VEMDE, MCMDE, and MVMDE methods are used for comparison. These methods are used to extract features at 20 scales from the three types of simulated signals, and take 50 samples for each health condition. The main parameters

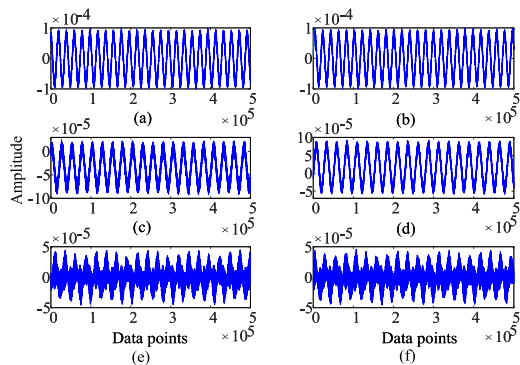


Fig. 7. Simulated displacement signals of the three health conditions. (a) VDSRI. (b) HDSRI. (c) VDSSF. (d) HDSSF. (e) VDSFPF. (f) HDSFPF.

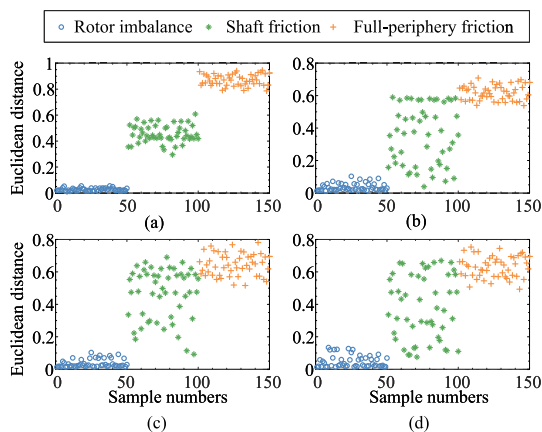


Fig. 8. Features extracted from simulated signal: four entropy methods are used: (a) Cross-DE. (b) VEMDE. (c) MCMDE. (d) MVMDE.

of all methods are the same: The embedding dimension  $m = 4$ , the scale factor  $\tau = 20$ , and the number of symbols  $\varepsilon = 30$  [44].

The feature extraction results of the simulation signal are shown in Fig. 8. The horizontal axis of Fig. 8 represents the sample numbers. The horizontal axis is divided into three parts: The first fifty samples are from the condition of rotor imbalance; the middle fifty samples are from the shaft friction; the last fifty samples are from the full-periphery friction. The mean value of the features of the first ten samples is taken as the reference vector, and the vertical axis represents the Euclidean distance between the samples and the reference vector. The evaluation criteria are as follows: The closer the Euclidean distance between the same fault sample and the reference vector, the better the stability of the entropy method; the more obvious the difference of Euclidean distance between different fault samples, the better the performance of feature extraction and fault classification of the entropy method.

From Fig. 8, it can be found that the features extracted by VEMDE, MCMDE, and MVMDE from the shaft friction samples have many overlaps with the features of other conditions. There is a massive gap between the features of the shaft friction samples. From Fig. 8(a), the features extracted by Cross-DE can clearly distinguish these three faults. And the features of the samples for each type of fault are close.

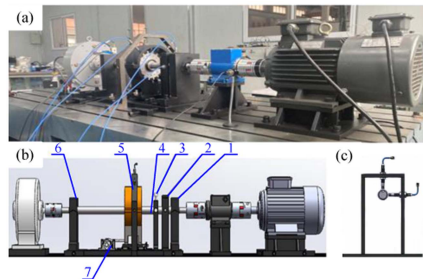


Fig. 9. (a) Physical diagram of experimental system. (b) Schematic diagram of the main components. (c) Horizontal and vertical displacement sensors installed radially on the shaft.

This phenomenon can be attributed to the fact that both VEMDE, MCMDE, and MVMDE only independently calculate the cosine similarity of the horizontal and vertical displacement signals. These three entropy methods do not consider the connection between the vertical and horizontal displacement signals. However, Cross-DE increases the cosine similarity between the signals of the two channels, and then considers the correlation between the 2-D signals. Therefore, Cross-DE has better stability and feature extraction effect.

## B. Experiment Evaluation

The physical diagram of the experimental rotor system is shown in Fig. 9(a). The schematic diagram of the main components and their names are shown in Fig. 9(b), where the names of the components from 1 to 7 are: Support bearing pedestal, displacement sensor bracket, friction assembly and bracket, shaft, casing friction support and blade disc, test bearing pedestal, and worm gear and worm.

The 2-D time series processed in this experiment are the signals of the horizontal displacement sensor and the vertical displacement sensor. The two sensors are installed radially on the rotor shaft, position 2 in Fig. 9(b). And the specific location of the displacement sensor is shown in Fig. 9(c).

In order to evaluate the recognition ability of Cross-DE to a single fault and compound fault, six single faults, six composite faults, and normal conditions are designed in this experiment. The designed six single faults include full-periphery friction (FPF), blade crack 1 (BC1), blade crack 2 (BC2), leaf disc crack (LDC), coupling failure (CF), and shaft crack (SC). The designed six compound faults include leaf disc cracks and full-periphery friction (LDC-FPF), leaf disc cracks and bearing failures (LDC-BF), shaft friction and coupling failure (SF-CF), shaft cracks and full-periphery friction (SC-FPF), shaft cracks and leaves disc crack (SC-LDC), full-periphery friction and coupling failures and bearing failures (FPF-CF-BF). The sampling frequency was 10240 Hz. The rotating speed was 1000 r/min. These single faults are shown in Fig. 10.

This experiment is divided into two parts to evaluate the recognition ability of Cross-DE for single fault and composite fault, respectively. First, 2-D displacement signals with different health conditions are extracted. The vertical displacement signals (VDS) and horizontal displacement signals (HDS) of

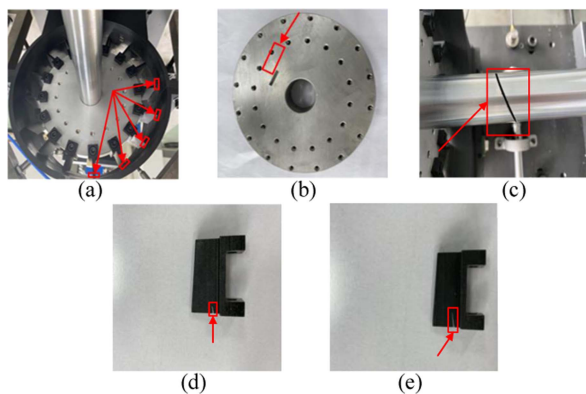


Fig. 10. Fault states of the experimental rotor system. (a) Full-periphery friction. (b) Leaf disc crack. (c) Shaft crack. (d) Blade crack 1. (e) Blade crack 2.

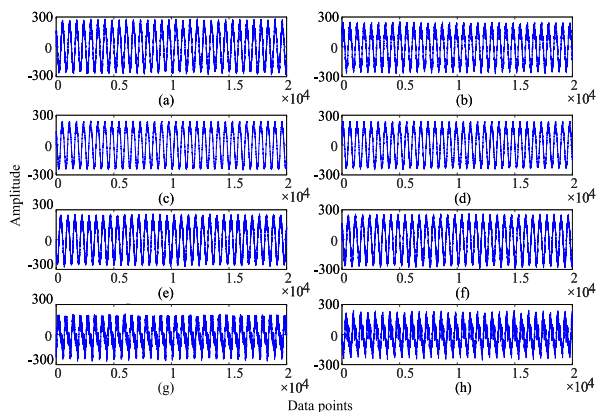


Fig. 11. Displacement signals of four example faults.

the four example faults are shown in the Fig. 11. Then the 2-D signals of each health condition are divided into 100 samples for feature extraction. The obtained features are then used as input to the ELM classifier. In addition, 50 samples of each health condition were randomly selected to construct a training dataset, and the remaining 50 samples were used to construct a test dataset.

In order to demonstrate the superiority of Cross-DE, the existing multivariate entropy methods are used for comparison, including VEMDE, MCMDE, and MVMDE to extract fault features. In addition, to demonstrate the superiority of multichannel analysis, we also compare Cross-DE with MDE single-channel signal. The single-channel signal is to perform feature extraction on the vertical or horizontal displacement signal separately. The main parameters of all methods are the same: the embedding dimension  $m = 4$ , the scale factor  $\tau = 20$ , and the number of symbols  $\varepsilon = 30$ .

The above entropy methods are used to extract the features of designed single and compound faults. The obtained features are then used as input to the ELM classifier. Each method is run 20 times and the average test accuracy is taken as the result to reduce randomness. The line chart of classification results is shown in Fig. 12. Tables II and III, respectively, show the average

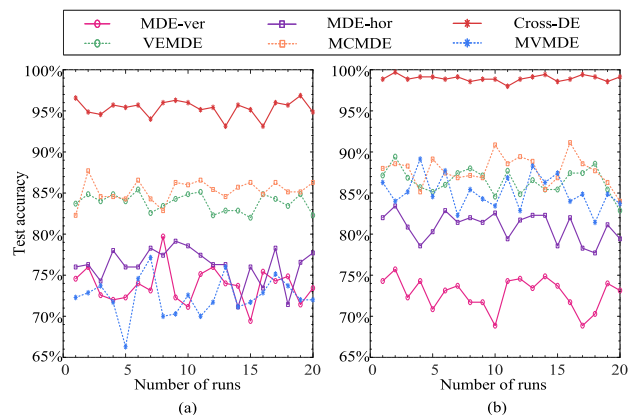


Fig. 12. (a) Results of single fault experiment. (b) Results of compound fault experiment.

TABLE II  
CLASSIFICATION RESULTS OF THE SINGLE FAULT EXPERIMENT

Method	MDE-ver	MDE-hor	Cross-DE	VEMDE	MCMDE	MVMDE
Average test accuracy(%)	73.77	76.23	<b>95.31</b>	83.84	85.29	72.39
Variance(%)	2.15	2.14	<b>0.98</b>	1.05	1.28	2.33

The bold values represent best method and the optimal parameters.

TABLE III  
CLASSIFICATION RESULTS OF THE COMPOUND FAULT EXPERIMENT

Method	MDE-ver	MDE-hor	Cross-DE	VEMDE	MCMDE	MVMDE
Average test accuracy(%)	72.77	80.91	<b>98.96</b>	86.46	87.76	85.14
Variance(%)	1.89	1.65	<b>0.36</b>	1.52	1.74	1.99

The bold values represent best method and the optimal parameters.

test accuracy and variance of various methods for single fault and compound fault identification.

From Fig. 12, it can be found that multivariate methods are better than single-channel-based methods. This is because the information contained in single channel signal is limited. The multivariate method comprehensively considers the information of the two channels, so the effect of extracting features is better. It can be found from Tables II and III that the average test accuracy of Cross-DE for single fault and compound fault identification is the highest, 95.31% and 98.96%, respectively. In addition, Cross-DE has the lowest variance, which is 0.98% and 0.36%, respectively. This shows that Cross-DE has the best feature extraction ability and stability in recognizing single faults or compound faults. This is because VEMDE, MCMDE, and MVMDE do not correlate and synthesize the information between the two channels, resulting in the loss of information. Cross-DE considers the information fusion between two channels, which can extract complete fault information from 2-D signals. So Cross-DE has better feature extraction ability.

In order to compare the effect of feature extraction more clearly, we visualized the features extracted by different methods. The feature dimension is reduced to two dimensions by t-SNE visualization method [53]. The visual results of single fault feature extraction by various entropy methods are shown in Fig. 13. The smaller the intracluster spacing of sample points



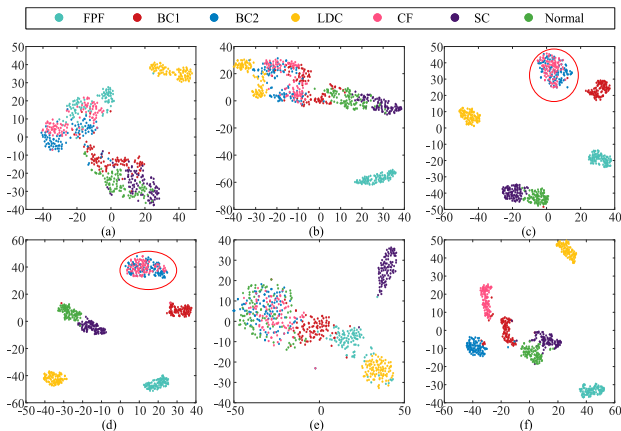


Fig. 13. Visualization of features extracted by six entropy methods. (a) MDE-ver. (b) MDE-hor. (c) VEMDE. (d) MCMDE. (e) MVMDE. (f) Cross-DE.

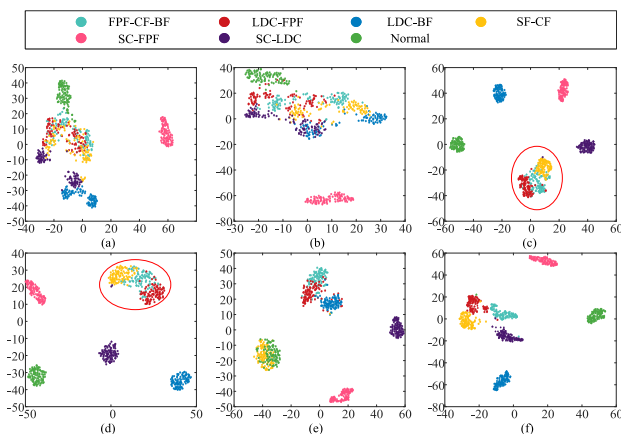


Fig. 14. Visualization of features extracted by six entropy methods. (a) MDE-ver. (b) MDE-hor. (c) VEMDE. (d) MCMDE. (e) MVMDE. (f) Cross-DE.

of each type of fault, the larger the intercluster spacing, which indicates that the better the classification effect and the higher the feature extraction ability.

From Fig. 13(a) and (b), it can be found that the features extracted using the displacement signal of a single channel are ineffective. Similarly, it can be seen in Fig. 13(e) that the feature extraction effect of MVMDE is also poor. From Fig. 13(c) and (d), among the features extracted with VEMDE and MCMDE, the features of BC2 and CF are obviously mixed. So VEMDE and MCMDE are not able to distinguish all health conditions. In contrast, it can be seen from Fig. 13(f) that Cross-DE has the best feature extraction and can distinguish different single health conditions well. Consistent with the previous discussion results, Cross-DE has better feature extraction ability than other multivariate methods.

Similarly, the dimension of compound fault features extracted by various entropy methods is reduced to two dimensions through t-SNE, and the visualization results are shown in Fig. 14.

From Fig. 14(a) and (b), it can be found that the features extracted by analyzing the single-channel signal are inferior

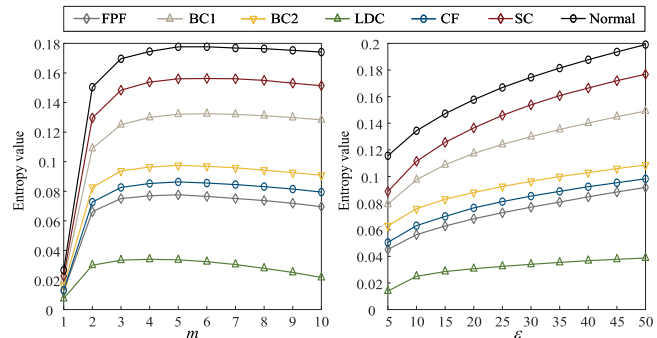


Fig. 15. Entropy value when selecting different parameters.

and cannot distinguish most of the compound faults. As shown in Fig. 14(e), MVMDE cannot distinguish between SF-CF and normal conditions. Moreover, the features of LDC-FPF, FPF-CF-BF, and LDC-BF are also mixed with many points, and the distinction between different clusters is not clear. From Fig. 14(c) and (d), it can be found that although the feature extraction effect of VEMDE and MCMDE is better than that of MVMDE, the clusters of SF-CF, FPF-CF-BF, and LDC-FPF are obviously mixed. The classification effect is still not satisfactory. In contrast, the features extracted with Cross-DE in Fig. 14(f) perform best. Clusters of different compound health conditions can be more clearly distinguished. This is because Cross-DE extracts the information in the 2-D signal more thoroughly, and it is easier to distinguish compound faults.

In addition, the single fault signals are employed to further verify that the parameters set are optimal. Fig. 15 shows the entropy curves with the change of parameters embedding dimension  $m$  and the number of symbols  $\varepsilon$ . The embedding dimension is set to  $m \in [1, 10]$ , and the number of symbols is set to  $\varepsilon \in [5, 50]$ . It can be found from Fig. 15 that the entropy curves tend to be stable when  $m > 2$ , and does not affect the separability of the entropy values of different faults with the increase of  $m$ . The value of  $\varepsilon$  also has no noticeable impact on the fault classification performance of DE. Therefore, the recommended parameter values  $m = 4$  and  $\varepsilon = 30$  can achieve the desired effect when processing experimental data, which are still the optimal values. Moreover, the entropy features of different parameter selection are input into ELM. The dimensionality of the input features of ELM is fixed to 20, and the result is shown in Fig. 16. Fig. 16 shows that when the embedding dimension  $m > 4$  and the number of symbols  $\varepsilon > 25$ , the classification accuracy exceeds 90%. Furthermore, when the parameters take the preset values  $m = 4$  and  $\varepsilon = 30$ , the classification accuracy exceeds 95%, which again verifies that the preset parameter is the optimal value.

Then, Table IV shows the standard deviation (std) of entropy values of fault samples under different scales. It can be seen from Table IV that the std of entropy value diminishes with the scale increase. When the scale is around 20, the std reaches the minimum value. When the scale continues to rise, the std shows an upward trend. This indicates that when the scale factor is around 20, the stability of Cross-DE is better and the

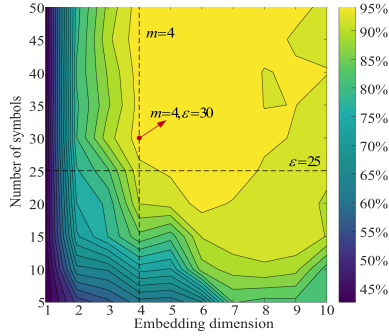


Fig. 16. Classification accuracy when selecting different parameters.

TABLE IV  
STD OF CROSS-DE UNDER PARTIAL SCALES ( $\times 10^{-3}$ )

scale	1	5	10	15	20	25	30
FPF	9.233	8.49	8.03	8.55	<b>6.752</b>	7.171	8.074
BC1	8.728	6.729	7256	6.155	<b>5.927</b>	6.28	8.933
BC2	7.488	5.954	4.485	4.666	<b>3.669</b>	6.522	6.533
LDC	5.989	4.255	4.676	4.449	<b>4.048</b>	5.402	7.967
CF	6.079	5.137	4.988	6.968	<b>4.344</b>	5.201	7.083
SC	13.391	13.043	10.165	9.37	<b>6.604</b>	8.539	8.051
Normal	8.09	8.013	6.27	7.181	<b>5.979</b>	6.896	10.392

The bold values represent best method and the optimal parameters.

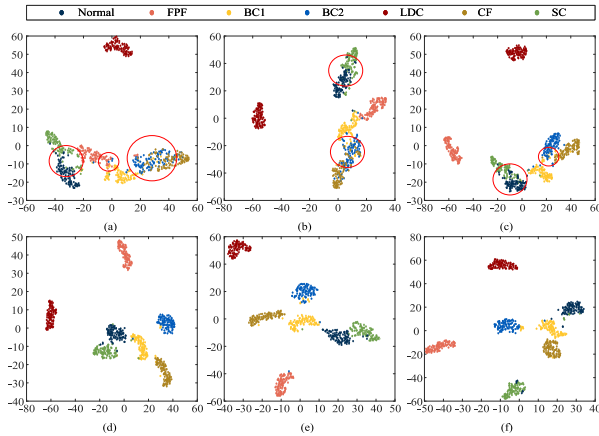


Fig. 17. Visualization results under different scales.

entropy estimation is more accurate. This can effectively prevent misclassification of faults caused by large entropy fluctuation. Therefore, the recommended scale factor is  $\tau = 20$ . In addition, Fig. 17 shows the visualization results after the dimensionality reduction of entropy features under different scales. It can be seen from Fig. 17 that when the scale factor  $\tau = 5$ , the fault classification effect is poor, and other fault samples except LDC have evident mixing. With the increase of scale, the impact of fault classification becomes better. When  $\tau = 20$ , the best classification effect can be achieved and most fault types can be distinguished. After that, the increasing scale will not change the classification effect significantly. The recommended scale factor is set to  $\tau = [20, 30]$ . However, large scale will cause expensive calculation costs. Therefore, the optimal value of scale factor is  $\tau = 20$ .

## IV. CONCLUSION

In this article, Cross-DE is proposed to solve the problem of fault information loss caused by the traditional multivariate entropy method when dealing with multichannel information. When processing multichannel signals, Cross-DE calculates the cosine similarity between orbits in the same and different phase spaces and considers the information fusion between different channels. Therefore, Cross-DE can extract fault information from multichannel signals comprehensively, which can help achieve more accurate fault diagnosis and benefit the health management of the rotor system. The effectiveness of Cross-DE is verified by simulation signals and experimental data. Thus, the developed Cross-DE is of high practical value in fault diagnostics to various industrial practices.

In this preliminary study, the proposed Cross-DE can be effectively applied to the feature extraction of rotor displacement vibration signals. In future work, we will use Cross-DE for fault diagnosis of common components, such as bearings, which are very common in mechatronic equipment. Moreover, in practice, the rotor system often requires more sensors for condition monitoring. Therefore, in future work, the effectiveness of Cross-DE in processing higher dimensional signals will be investigated.

## APPENDIX

The first model of the Euler–Bernoulli beam element. The left and right nodes of each shaft unit each have two generalized displacement vectors of vertical and horizontal planes, which are represented as  $u_{xs}$  and  $u_{ys}$ . For the elastic axis section of the Euler–Bernoulli beam, the translational inertia matrix of the shaft  $M_M^Z$ , the moment of inertia matrix of the shaft (including the element diameter moment of inertia matrix  $M_T^Z$ , the element pole moment of inertia matrix  $J_S^Z$ ), and the stiffness matrix  $K_S^Z$  of the shaft element. The dynamic equation of the elastic shaft section of the Euler–Bernoulli beam can be represented as follows:

$$\begin{aligned} (M_M^Z + M_T^Z) \ddot{u}_{xs} + \Omega J_S^Z \dot{u}_{xs} + K_S^Z u_{xs} &= Q_{xs} \\ (M_M^Z + M_T^Z) \ddot{u}_{ys} - \Omega J_S^Z \dot{u}_{ys} + K_S^Z u_{ys} &= Q_{ys} \end{aligned} \quad (11)$$

where  $\Omega$  is the rotational angular velocity of the shaft, and  $Q_{xs}$  and  $Q_{ys}$  are the generalized force vectors.

For the blade disk rotor unit, the displacement vectors of the vertical and horizontal planes are  $u_{xp}$  and  $u_{yp}$ . The quality matrix of the disc unit is  $M_p$ , the moment of inertia matrix of the disc is  $J_{pe}$ . The dynamic equation of the disc element can be expressed as follows:

$$\begin{aligned} M_p \ddot{u}_{xp} + \Omega J_{pe} \dot{u}_{yp} &= Q_{xp} \\ M_p \ddot{u}_{yp} - \Omega J_{pe} \dot{u}_{xp} &= Q_{yp} \end{aligned} \quad (12)$$

where  $\Omega$  is the rotational angular velocity of the shaft, and  $Q_{xp}$  and  $Q_{yp}$  are the generalized force vectors.

For any node  $i$  of the entire rotor system, the displacement vector in the vertical plane is  $u_{i1}$ , and the displacement vector in the horizontal plane is  $u_{i2}$ . The dynamic equation of the rotor

system is as follows:

$$\begin{aligned} M_i \ddot{u}_{i1} + \Omega J_i \dot{u}_{i2} + K_i u_{i1} &= Q_{i1} \\ M_i \ddot{u}_{i2} - \Omega J_i \dot{u}_{i1} + K_i u_{i2} &= Q_{i2} \end{aligned} \quad (13)$$

where  $\Omega$  is the rotational angular velocity of the shaft,  $M_i$  is the mass matrix,  $J_i$  is the moment of inertia matrix,  $K_i$  is the stiffness matrix, and  $Q_{i1}$  and  $Q_{i2}$  are the generalized force vectors.

Analysis of generalized force vector: the disc structure is complex, and the simplified disc model has an imbalance between its center of mass and the center of the circle. Therefore, add horizontal and vertical unbalanced forces at the leaf disc nodes as follows:

$$\begin{aligned} Q_{i1}(h) &= m_p e_p \omega^2 \sin(\Omega t + \beta_p) \\ Q_{i2}(h) &= m_p e_p \omega^2 \cos(\Omega t + \beta_p) \end{aligned} \quad (14)$$

where  $h$  is the node of the disc,  $e_p$  is the unbalance of the disc, and  $\beta_p$  is the initial phase of the disc.

According to the analysis of the friction, this phenomenon is mainly caused by the displacement of the rotor or the shaft vibration, which exceeds the friction gap, so the main consideration is the collision force and friction force

$$\begin{aligned} P_N &= -k_{ia}(s_{fa} - \sigma_a), s_{fa} > \sigma_a \\ P_T &= f_{ira} P_N \end{aligned} \quad (15)$$

where  $P_N$  is the impact force,  $P_T$  is the friction force,  $k_{ia}$  is the friction stiffness,  $\sigma_a$  is the friction gap,  $f_{ira}$  is the coefficient of friction,  $s_{fa}$  is the radial displacement of the disc rotor in a relatively static state.

In addition, the model simplified the bearing into a parallel connection of a damper and a spring. Therefore, the damping matrix  $C_i$  must be added to the equation, and the support stiffness of the bearing should be added to the stiffness matrix. So far, the rotor model is basically completed. The final kinetic equation is as follows:

$$M_z \ddot{u}_z + (\Omega J_z + C_z) \dot{u}_z + K_z u_z = Q_z. \quad (16)$$

Substituting each matrix to get the following equation:

$$\begin{aligned} &\left( \Omega \begin{bmatrix} 0 & J_i \\ -J_i & 0 \end{bmatrix} + \begin{bmatrix} C_i & 0 \\ 0 & C_i \end{bmatrix} \right) \begin{bmatrix} \ddot{u}_{i1} \\ \ddot{u}_{i2} \end{bmatrix} \\ &+ \begin{bmatrix} M_i & 0 \\ 0 & M_i \end{bmatrix} \begin{bmatrix} \dot{u}_{i1} \\ \dot{u}_{i2} \end{bmatrix} + \begin{bmatrix} K_i & 0 \\ 0 & K_i \end{bmatrix} \begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix} = \begin{bmatrix} Q_{i1} \\ Q_{i2} \end{bmatrix} \end{aligned} \quad (17)$$

where 0 is the homogeneous zero matrixes of the corresponding matrix. Convert the differential equation to the first-order form

$$\begin{bmatrix} M_z & 0 \\ 0 & K_z \end{bmatrix} \begin{bmatrix} \ddot{u}_z \\ \dot{u}_z \end{bmatrix} + \begin{bmatrix} \Omega J_z + C_z & K_z \\ -K_z & 0 \end{bmatrix} \begin{bmatrix} \dot{u}_z \\ u_z \end{bmatrix} = \begin{bmatrix} Q_z \\ 0 \end{bmatrix}. \quad (18)$$

## REFERENCES

- [1] Y. Zhang, Z. Ren, S. Zhou, K. Feng, K. Yu, and Z. Liu, "Supervised contrastive learning-based domain adaptation network for intelligent unsupervised fault diagnosis of rolling bearing," *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 6, pp. 5371–5380, Dec. 2022.
- [2] B. Hou, D. Wang, T. Yan, Y. Wang, Z. Peng, and K.L. Tsui, "Gini indices II and III: Two new sparsity measures and their applications to machine condition monitoring," *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 3, pp. 1211–1222, Jun. 2022.
- [3] J. Li, R. Huang, G. He, Y. Liao, Z. Wang, and W. Li, "A two-stage transfer adversarial network for intelligent fault diagnosis of rotating machinery with multiple new faults," *IEEE/ASME Trans. Mechatron.*, vol. 26, no. 3, pp. 1591–1601, Jun. 2021.
- [4] S. Lu, R. Yan, Y. Liu, and Q. Wang, "Tachless speed estimation in order tracking: A review with application to rotating machine fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 7, pp. 2315–2332, Jul. 2019.
- [5] L. Wang, S. Ma, and Q. Han, "Enhanced sparse low-rank representation via nonconvex regularization for rotating machinery early fault feature extraction," *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 5, pp. 3570–3578, Oct. 2022.
- [6] S. TayebiHaghighi and I. Koo, "Fault diagnosis of rotating machine using an indirect observer and machine learning," in *Proc. IEEE Int. Conf. Inf. Commun. Technol. Convergence*, 2020, pp. 277–282.
- [7] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, 2019.
- [8] H. Jin et al., "Small fault diagnosis with gap metric," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 53, no. 9, pp. 571–5728, Sep. 2023.
- [9] I. Aljarah, M. Mafarja, A. A. Heidari, H. Faris, Y. Zhang, and S. Mirjalili, "Asynchronous accelerating multi-leader salp chains for feature selection," *Appl. Soft Comput.*, vol. 71, pp. 964–979, 2018.
- [10] H. Shao, M. Xia, J. Wan, and C. W.de Silva, "Modified stacked autoencoder using adaptive Morlet wavelet for intelligent fault diagnosis of rotating machinery," *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 1, pp. 24–33, Feb. 2022.
- [11] S. Chen, Y. Meng, H. Tang, Y. Tian, N. He, and C. Shao, "Robust deep learning-based diagnosis of mixed faults in rotating machinery," *IEEE/ASME Trans. Mechatron.*, vol. 25, no. 5, pp. 2167–2176, Oct. 2020.
- [12] M. Xia, T. Li, L. Xu, L. Liu, and C. W.De Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE/ASME Trans. Mechatron.*, vol. 23, no. 1, pp. 101–110, Feb. 2018.
- [13] G. Li, J. Wu, C. Deng, X. Xu, and X. Shao, "Deep reinforcement learning-based online domain adaptation method for fault diagnosis of rotating machinery," *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 5, pp. 2796–2805, Oct. 2022.
- [14] R. Zhu and X. Xia, "Application of VMD in fault diagnosis for rotor-bearing system with rub-impact," in *Proc. IEEE 3rd Int. Conf. Inf. Commun. Signal Process.*, 2020, pp. 502–505.
- [15] Y. Yu and H. Lang, "Fault diagnosis of rotor rub based on ensemble EMD," in *Proc. IEEE 9th Int. Conf. Electron. Meas. Instruments*, 2009, pp. 2–144.
- [16] X.-B. Wang, Z.-X. Yang, and X.A. Yan, "Novel particle swarm optimization-based variational mode decomposition method for the fault diagnosis of complex rotating machinery," *IEEE/ASME Trans. Mechatron.*, vol. 23, no. 1, pp. 68–79, Feb. 2017.
- [17] R. Yan and R. X. Gao, "Complexity as a measure for machine health evaluation," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 4, pp. 1327–1334, Aug. 2004.
- [18] 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, 2014.
- [19] R. Yan and R. X. Gao, "Rotary machine health diagnosis based on empirical mode decomposition," *J. Vib. Acoust.*, vol. 130, no. 2, 2008.
- [20] W. Huachun, Z. Jian, X. Chunhu, Z. Jiyang, and H. Yiming, "Two-dimensional time series sample entropy algorithm: Applications to rotor axis orbit feature identification," *Mech. Syst. Signal Process.*, vol. 147, 2021, Art. no. 107123.
- [21] T. Liu, M. Lyu, Z. Wang, and S. Yan, "An identification method of orbit responses rooting in vibration analysis of rotor during touchdowns of active magnetic bearings," *J. Sound Vib.*, vol. 414, pp. 174–191, 2018.
- [22] X. Li, J. Wu, D. Chang, W. Huang, Z. Ma, and J. Cao, "Mixed attention mechanism for small-sample fine-grained image classification," in *Proc. IEEE Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2019, pp. 80–85.
- [23] H. Hu, Q. Li, Y. Zhao, and Y. Zhang, "Parallel deep learning algorithms with hybrid attention mechanism for image segmentation of lung tumors," *IEEE Trans. Ind. Inform.*, vol. 17, no. 4, pp. 2880–2889, Apr. 2021.
- [24] Y. Li, X. Wang, Z. Liu, X. Liang, and S. Si, "The entropy algorithm and its variants in the fault diagnosis of rotating machinery: A review," *IEEE Access*, vol. 6, pp. 66 723–66 741, 2018.

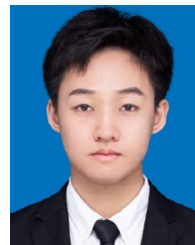


- [25] R. Yan and R. Gao, "Machine health diagnosis based on approximate entropy," in *Proc. IEEE 21st Instrum. Meas. Technol. Conf.*, 2004, pp. 2054–2059.
- [26] R. Yan and R. X. Gao, "Approximate entropy as a diagnostic tool for machine health monitoring," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 824–839, 2007.
- [27] R. Yan, Y. Liu, and R. X. Gao, "Permutation entropy: A nonlinear statistical measure for status characterization of rotary machines," *Mech. Syst. Signal Process.*, vol. 29, pp. 474–484, 2012.
- [28] M. Singh and A. G. Shaik, "Entropy-based broken rotor-bar fault detection and estimation of its severity in a three-phase induction motor," in *Proc. IEEE Int. Conf. Power Electron. Drives Energy Syst.*, 2020, pp. 1–6.
- [29] C. E. Shannon, "A mathematical theory of communication," *Bell System Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
- [30] H. Zhang and S.S. He, "Analysis and comparison of permutation entropy, approximate entropy and sample entropy," in *Proc. IEEE Int. Symp. Comput. Consum. Control*, 2018, pp. 209–212.
- [31] 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, pp. 2039–2049, 2000.
- [32] W. Chen, Z. Wang, H. Xie, and W. Yu, "Characterization of surface EMG signal based on fuzzy entropy," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 2, pp. 266–272, Jun. 2007.
- [33] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, no. 17, 2002, Art. no. 174102.
- [34] X. Wang, S. Si, and Y. Li, "Multiscale diversity entropy: A novel dynamical measure for fault diagnosis of rotating machinery," *IEEE Trans. Ind. Inform.*, vol. 17, no. 8, pp. 5419–5429, Aug. 2021.
- [35] M. Costa, A. L. Goldberger, and C.K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.*, vol. 89, no. 6, 2002, Art. no. 068102.
- [36] T. Han, C. C. Shi, Z. B. Wei, and T. R. Lin, "Analysis of complex time series using a modified multiscale fuzzy entropy algorithm," in *Proc. IEEE Int. Conf. Identification Inf. Knowl. Internet Things*, 2016, pp. 45–51.
- [37] R. Zhou, X. Wang, J. Wan, and N. Xiong, "EDM-fuzzy: An Euclidean distance based multiscale fuzzy entropy technology for diagnosing faults of industrial systems," *IEEE Trans. Ind. Inform.*, vol. 17, no. 6, pp. 4046–4054, Jun. 2021.
- [38] A. Humeau-Heurtier, C.W. Wu, and S.-D. Wu, "Refined composite multiscale permutation entropy to overcome multiscale permutation entropy length dependence," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2364–2367, Dec. 2015.
- [39] W. Chaobing, W. Rongzhen, Z. Long, C. Binghuan, Y. Lewei, and Y. Wenhao, "Extended intelligent recognition of rolling bearing early faults using multiscale permutation entropy," in *Proc. IEEE Prognostics System Health Manage. Conf.*, 2019, pp. 1–7.
- [40] K. Zografos and S. Nadarajah, "Expressions for Rényi and Shannon entropies for multivariate distributions," *Statist. Probability Lett.*, vol. 71, no. 1, pp. 71–84, 2005.
- [41] Z. Wu, W. Jin, and N. Qin, "Fault feature analysis of high-speed train suspension system based on multivariate multi-scale sample entropy," in *Proc. IEEE 35th Chin. Control Conf.*, 2016, pp. 3913–3918.
- [42] J. Zheng, D. Tu, H. Pan, X. Hu, T. Liu, and Q. Liu, "A refined composite multivariate multiscale fuzzy entropy and Laplacian score-based fault diagnosis method for rolling bearings," *Entropy*, vol. 19, no. 11, 2017, Art. no. 585.
- [43] W. Ying, J. Tong, Z. Dong, H. Pan, Q. Liu, and J. Zheng, "Composite multivariate multi-scale permutation entropy and Laplacian score based fault diagnosis of rolling bearing," *Entropy*, vol. 24, no. 2, 2022, Art. no. 160.
- [44] X. Wang, S. Si, and Y. Li, "Variational embedding multiscale diversity entropy for fault diagnosis of large-scale machinery," *IEEE Trans. Ind. Electron.*, vol. 69, no. 3, pp. 3109–3119, Mar. 2022.
- [45] A. Jamin and A. Humeau-Heurtier, "(multiscale) cross-entropy methods: A review," *Entropy*, vol. 22, no. 1, 2019, Art. no. 45.
- [46] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*, Warwick, RI, USA: Springer, 1981, pp. 366–381.
- [47] L. Noakes, "The Taken's embedding theorem," *Int. J. Bifurcation Chaos*, vol. 1, no. 04, pp. 867–872, 1991.
- [48] G.C. Wu and D. Baleanu, "Discrete fractional logistic map and its chaos," *Nonlinear Dyn.*, vol. 75, no. 1, pp. 283–287, 2014.
- [49] B. Yang and X. Liao, "Some properties of the logistic map over the finite field and its application," *Signal Process.*, vol. 153, pp. 231–242, 2018.
- [50] J.-M. Wang, G.-Q. Xu, and S.P. Yung, "Riesz basis property, exponential stability of variable coefficient Euler–Bernoulli beams with indefinite damping," *IMA J. Appl. Math.*, vol. 70, no. 3, pp. 459–477, 2005.
- [51] X. Ma, H. Ma, J. Zeng, and Y. Piao, "Rubbing-induced vibration response analysis of dual-rotor-casing system," *Trans. Nanjing Univ. Aeronaut. Astronaut.*, vol. 35, no. 1, pp. 101–108, 2018.
- [52] Y. Yang, D. Cao, T. Yu, D. Wang, and C. Li, "Prediction of dynamic characteristics of a dual-rotor system with fixed point rubbing—theoretical analysis and experimental study," *Int. J. Mech. Sci.*, vol. 115, pp. 253–261, 2016.
- [53] K. Y. Wong and F.-I. Chung, "Visualizing time series data with temporal matching based t-SNE," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2019, pp. 1–8.



**Yongbo Li** (Member, IEEE) received the master's degree in mechanical design and theory from Harbin Engineering University, Harbin, China, in 2012, and the Ph.D. degree in general mechanics from the Harbin Institute of Technology, Harbin, in 2017.

He was a Visiting Student in mechanical engineering with the University of Alberta, Edmonton, AB, Canada. In 2017, he joined Northwestern Polytechnical University, Xi'an, China, where he is currently an Associate Professor with the School of Aeronautics. He has authored or coauthored 54 papers. His research interests include signal processing, fault feature extraction, and fault pattern identification.



**Zehang Jiao** received the B.Eng. degree in flight vehicle propulsion engineering from Civil Aviation University of China, Tianjin, China, in 2021. He is currently working toward the master's degree in mechanical engineering from Northwestern Polytechnical University, Xi'an, China.

His research interests include entropy, fault diagnosis, condition monitoring, and pattern identification.



**Shun Wang** received the B.Eng. degree in aircraft control and information engineering, in 2020, from Northwestern Polytechnical University, Xi'an, China, where he is currently working toward the master's degree in aerospace science and technology.

His research interests include entropy, signal processing, fault diagnosis, condition monitoring, and pattern identification.





**Ke Feng** (Member, IEEE) received the Ph.D. degree in mechanical engineering from the University of New South Wales, Sydney, NSW, Australia, in 2021.

He was with the University of British Columbia, Vancouver, BC, Canada, in 2022, and the National University of Singapore, Singapore, in 2023. He is currently a Marie Curie Fellow with Imperial College London, London, U.K., and Brunel University, London, U.K. His research interests include digital twins, vibration

analysis, structural health monitoring, dynamics, tribology, signal processing, and machine learning.

Dr. Feng is a Fellow of Vebleo. He was the recipient of the Emerging Leader Award in 2023 by the Measurement Science and Technology journal. He is the Editor and Guest Editor for several journals, including *Mechanical Systems and Signal Processing*, *Engineering Applications of Artificial Intelligence*, IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, *Measurement*, IEEE SENSORS JOURNAL, MEASUREMENT SCIENCE AND TECHNOLOGY, etc.



**Zheng Liu** (Senior Member, IEEE) received the doctorate degree in engineering (earth resources) from Kyoto University, Kyoto, Japan, in 2000, and the Ph.D. degree in electrical engineering from the University of Ottawa, Ottawa, ON, Canada, in 2007.

From 2000 to 2001, he was a Research Fellow with the Nanyang Technological University, Singapore. He then joined the National Research Council of Canada (Ottawa, Ontario) as a Governmental Laboratory Visiting Fellow

nominated by NSERC, in 2001. In 2002, he became a Research Officer associated two research institutes of NRC [Aerospace (IAR) and Construction (IRC)]. From 2012 to 2015, he was a full Professor with Toyota Technological Institute, Nagoya, Japan. He is currently a Professor with the Faculty of Applied Science, University of British Columbia-Okanagan, Kelowna, BC, Canada. His research interests include predictive maintenance, data/information fusion, computer/machine vision, machine learning, smart sensor and industrial IoT, and non-destructive inspection and evaluation.

Dr. Liu is a Fellow of SPIE. He holds a Professional Engineer license in both British Columbia and Ontario. He serves on the editorial boards for journals including IEEE JOURNAL OF RADIO FREQUENCY IDENTIFICATION, *Information Fusion* (Elsevier), *Machine Vision and Applications* (Springer), *Canadian Journal of Electrical and Computer Engineering*, *Intelligent Industrial Systems* (Springer), and *IET/CAA Transactions on Intelligence Technology*. He also serves on the Standards & Interoperability Committee of Research Data Canada. He is co-chairing the IEEE Instrumentation and Measurement Society technical committee (TC-1).