Compound mechanical fault diagnosis based on CMDE

Guangwei Yu¹, Xianzhi Wang²,³ and Chunlin Da²

Abstract
The fault diagnosis technique is of important for the safety operation of the rotating machinery. In the fault diagnosis framework, the entropy-based method is a promising tool for the feature extraction and signal processing. Among the entropy-based methods, the diversity entropy has arisen increasing attention due to its merits of high consistency, strong robustness, and high calculation efficiency. However, it suffers the defect that the multiscale procedure leads to unstable complexity estimation at higher scales. This induces a poor cluster performance in analyzing the compound mechanical fault signals. To address this issue, this paper presents a novel feature extraction method called composite multiscale diversity entropy (CMDE). The proposed CMDE utilizes the mean complexity value of multiple sliding windows for each scale to enhance the stability, which enables the diversity entropy could dig richer fault information from deeper scales for the compound fault diagnosis of rotating machinery. Then, the stability of CMDE has been evaluated using synthetic gear signals. At last, the proposed CMDE has been applied in the compound mechanical fault diagnosis. The experimental results show that the CMDE achieves the highest diagnosis accuracy compared to the existing entropy-based feature extraction methods.

Keywords
Composite multiscale diversity entropy, fault diagnosis, compound mechanical fault, rotating machinery, feature extraction

Introduction
Fault diagnosis technique aims to provide reliable ensurence for the safety operation of the rotating machinery. Since the intelligent fault diagnosis technique has merits of fast response, easy to installation, and low cost, it has been widely used in the modern industry.¹,² In general, the intelligent fault diagnosis technique can be divided into three steps: data acquisition, feature extraction, and fault classification. Among these three steps, feature extraction is the most crucial step.³ To effectively extract features from the measured signal, the entropy-based method has become a hot topic due to the advantages of independence with prior knowledge, unnecessary of preprocessing, and easy to perform.⁴,⁵

¹School of Mechatronic Engineering, Xi’an Technological University, Xi’an, China
²School of Automation, Xi’an University of Posts and Telecommunications, Xi’an, China
³Xi’an Key Laboratory of Advanced Control and Intelligent Process, Xi’an University of Posts and Telecommunications, Xi’an, China

Corresponding author:
Xianzhi Wang, School of Automation, Xi’an University of Posts and Telecommunications, 618 West Chang’an Street, Chang’an District, Xi’an 710121, China.
Email: cagallilan@xupt.edu.cn
unidimensional feature provides insufficient information for the complex signal analysis. To solve this problem, Costa et al.\(^9\) developed multiscale sample entropy (MSE), which provided a measure called multiscale procedure to extend the unidimensional entropy method into a multidimensional feature extraction tool. In the multiscale procedure, the raw signal is decomposed into multiple scale time series by the Haar wavelet. The scale factor in the multiscale procedure is designed to control the decomposition layer of the Haar wavelet. Then, the MSE value can be obtained by calculating SE value under each scale factor. Inspired by MSE, permutation entropy and fuzzy entropy were also extended into multiscale entropy (MSE)\(^10\),\(^11\),\(^12\) and multiscale fuzzy entropy (MFE).\(^13\),\(^14\)

Recently, Wang et al.\(^15\) proposed a new entropy metric called multiscale diversity entropy (MDE). Different from the existing methods, the MDE utilize the diversity of the orbits in the phase space to quantify the dynamical complexity of time series. Compare to the MFE, MSE, and MPE, the MDE has merits as follows:

1. **High consistency.** The diversity entropy can estimate the dynamical complexity with high consistency, while PE, SE, and FE may yield inconsistent results in estimating some periodic systems.
2. **High calculation efficiency.** The diversity entropy has the lowest calculation complexity, which saves at least half of the computing time compared to PE, SE, and FE.
3. **Strong robustness.** Compared to PE, SE, and FE, the diversity entropy has the strongest robustness when the periodic signals are interfered by Gaussian white noise.

Although the MDE has above-mentioned advantages, it encounters challenge of extracting features from the mechanical compound fault. The compound fault means the machine occurs multiple fault synchronously, that is, the bearing and gear failed at the same time. In this circumstance, the measured vibrational signal consists of multiple features from different components, which requires the multiscale procedure to explore richer features from deeper scales. Actually, the multiscale procedure used in MDE is a Haar wavelet. As the scale increase, the fault information hidden in the different spectral resolution can be dig out. The richer fault information hidden in the deeper scale can provide sufficient fault features for the classifier to accurately recognize the compound fault (the detailed discussion can be seen in Section 4”Results and analysis”). However, the original multiscale procedure used in MDE has a certain defect that the multiscale time series will be greatly shortened at high scale. This makes the diversity entropy has insufficient information to estimate the dynamical complexity, which presents unstable behavior at high scale factor. This phenomenon restricts the application of diversity entropy in the compound fault diagnosis of rotating machinery.

To make up this defect, composite multiscale diversity entropy (CMDE) is proposed in this paper. In the proposed CMDE method, a moving averaging window is used to generate several composite multiple scale time series under each scale factor. Then, the mean complexity value of composite multiple scale time series is calculated as the CMDE value. With the composite multiscale procedure, the stability of diversity entropy at high scale has been greatly enhanced. This enables the diversity entropy could dig richer fault information from deeper scales for the compound fault diagnosis of rotating machinery.

Based on CMDE and k-nearest neighbor (kNN) classifier, an intelligent fault diagnosis frame for the mechanical compound-fault diagnosis has been proposed. The effectiveness of the proposed method is evaluated through simulated and experimental signals. The results show that the proposed CMDE has the best feature extraction ability compared to MSE, MPE, MFE, MDE, CMSE, CMPE, and CMFE. Moreover, the results also show that the composite multiscale procedure can effectively improve the stability of the diversity entropy at high scales, which provides more reliable dynamical complexity estimation for feature extraction.

The rest of this paper is organized as follows. Section “Methodology” introduces the basic definition of the proposed method. Section “Simulation evaluation” briefly presents the simulated gear signals and evaluates the stability of the proposed CMDE. Section “Experiment evaluation” evaluates the performance of the proposed fault diagnosis frame using the experimental signals. Last, Section “Conclusions” summarizes the conclusions.

**Methodology**

**Diversity entropy**

For an arbitrary time series \(X = \{x_1, x_2, \ldots, x_i, \ldots, x_N\}, i \in [1, N]\), \(N\) indicates the data length. The diversity entropy (DE) can be computed by four steps.

**Step 1.** Reconstructed the time series \(X\) into a series of orbits with embedding dimension \(m\) as expressed in equation (1).

\[
Y(m) = \left\{y_{1}(m), y_{2}(m), \ldots, y_{N-m+1}(m)\right\}
\]

\[
= \left\{\begin{array}{c}
x_1 \ x_2 \ \cdots \ x_m \\
x_2 \ x_3 \ \cdots \ x_{1+m} \\
\vdots \ \vdots \ \ \vdots \\
x_{N-m} + 1 \ x_{N-m} \ \cdots \ x_N
\end{array} \right\} \quad (1)
\]
Step 2. Construct the similarity set \( D(m) \) as equation (2). The \( D(m) \) can be computed by calculating the cosine similarity between each adjacent orbits as equation (3).

\[
D(m) = \{d_1, \ldots, d_{N-m}\} = \{d(y_1(m), y_2(m)), d(y_2(m), y_3(m)), \ldots, d(y_{N-m-1}(m), y_N(m))\} \tag{2}
\]

\[
d(y_j(m), y_j(m)) = \frac{\sum_{k=1}^{m} y_j(k) \times y_j(k)}{\sqrt{\sum_{k=1}^{m} y_j(k)^2} \times \sqrt{\sum_{k=1}^{m} y_j(k)^2}} \tag{3}
\]

Step 3. Divide the range \([-1, 1]\) into \( \varepsilon \) intervals. The \( \varepsilon \) is called the number of symbols in diversity entropy. Then count the cumulative frequency of \( D(m) \) falling into each interval to obtain the state probability \( (P_1, \ldots, P_\varepsilon) \). Obviously, \( \sum_{k=1}^{\varepsilon} P_k = 1 \). The physical meaning of this step is that symbolize the similarity set \( D(m) \) with \( \varepsilon \) symbols, then the infinite state of the orbits can be numerical.

Step 4. The diversity entropy value can be calculated as equation (4).

\[
DE(m, \varepsilon) = -\frac{1}{\ln \varepsilon} \sum_{k=1}^{\varepsilon} P_k \ln P_k \tag{4}
\]

As the diversity entropy defined the last step as the Shannon entropy, it presents the same mathematical properties: when DE value approaches 1, the dynamical complexity of the time series is high; when DE value approaches 0, the dynamical complexity of the time series is low. During the range of \([0, 1]\), the diversity entropy value monotone increases with the dynamical complexity of the time series.

### Multiscale diversity entropy

To extend the unidimensional entropy method into a multidimensional feature extraction tool, the diversity is combined with the multiscale procedure as shown in Figure 1. The multiscale diversity entropy can be computed by two steps:

**Step 1.** For a given time series \( X = \{x_1, x_2, \ldots, x_t, \ldots, x_N\} \), segment the \( X \) into multiscale time series as equation (5):

\[
y_j = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{\tau} x_i \tag{5}
\]

**Step 2.** Calculate the entropy value of each multiscale time series as equation (6).

\[
MDE(X, m, \varepsilon) = DE(y_j^\tau, m, \varepsilon) \tag{6}
\]

where the \( \tau \) represents the scale factor in the multiscale analysis, which should be a positive integer. For \( \tau = 1 \), the time series \( \{y_1^{(1)}\} \) is the original time series.

### Composite multiscale diversity entropy

The original multiscale procedure used in MDE has a certain defect that the multiscale time series will be greatly shortened at high scale. This makes the diversity entropy has insufficient information to estimate the dynamical complexity, which presents unstable behavior at high scale. This phenomenon restricts the application of diversity entropy in the mechanical compound-fault diagnosis. To make up this defect, composite multiscale diversity entropy (CMDE) is proposed in this paper. The stability of these two procedures will be detailed explored and discussed in Section “Simulation evaluation.” The composite multiscale diversity entropy can be calculated as following steps:

**Step 1.** For a given time series \( X = \{x_1, x_2, \ldots, x_t, \ldots, x_N\} \), segment it into composite multiple scale time series as equation (7):

![Figure 1. Multiscale procedure.](image-url)
Figure 2. Composite multiscale procedure.

\[
y_{k,j} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau+k-1} x_i, \quad 1 \leq j \leq N/\tau, \quad 1 \leq k \leq \tau
\]  

where the \( \tau \) represents the scale factor which has the same meaning as multiscale procedure in Section “Multiscale diversity entropy.” For \( \tau = 1 \), the time series \( \{y^{(1)}\} \) is the original time series.

**Step 2.** Compute the diversity entropy value of each composite multiple scale time series.

\[
CMDE(X, m, \varepsilon) = \frac{1}{\tau} \sum_{k=1}^{\tau} DE(y_{k}^{*}, m, \varepsilon)
\]

where the \( m \) and \( \varepsilon \) indicate the embedding dimension and number of symbols, which has the same meaning as in Section “Diversity entropy.”

The composite multiscale procedure can be seen in Figure 2. Seen that the CMDE uses repeated estimation of DE from composite multiple time series at high scale, which can avoid the unreliable estimation of shorten time series used in MDE. This can make up the shortage of original multiscale procedure used in MDE. Noted that the scale factor \( \tau \) has the same meaning as MDE, this indicates each scale factor \( \tau \) corresponding to an individual DE value. The composite multiscale procedure only enhances the stability of the diversity entropy, while the physical meaning is unchanged.

**KNN classifier**

The kNN is a commonly-used supervised learning model which has merits of less parameters, easy to perform, and high calculation efficient. In this paper, the kNN is used as a classifier to accomplish the fault diagnosis frame, and the detailed discussion of kNN is not involved. The principle of kNN is shown in Figure 3. The main calculation steps of kNN classifier are briefly reviewed as follows:

**Step 1.** Calculate the distance between the test sample and each training sample.

**Step 2.** Rank distances in ascending order.

**Step 3.** Select \( k \) samples with the smallest distance.

**Step 4.** Count the occurrence frequency of the label to the first \( k \) samples.

**Step 5.** The label with the highest frequency in the first \( k \) samples is returned as the prediction classification of the test sample.

**Compound fault diagnosis frame based on CMDE and kNN**

In the practical engineering, the multiple components of rotating machinery may fail at the same time. In this circumstance, the measured vibrational signal consists of multiple features from different components, which requires the multiscale procedure to explore richer features from deeper scales. However, the original MDE has poor stability at high scales. Thus, this paper proposed a novel compound fault diagnosis frame based on CMDE and kNN. The proposed fault diagnosis frame can be seen in Figure 4. It can be implemented by three stages as follows.

**Stage 1.** Data acquisition: use the accelerometers mounted on the machine to capture the vibrational signals under different health states.

**Stage 2.** Feature extraction: utilize the proposed CMDE method to extract the fault features from the measured vibrational signals in Stage 1.

**Stage 3.** Fault identification: use the extract features from Stage 2 to train the kNN classifier. Finally, the trained classifier can be used to diagnosis the compound fault in the rotating machinery.
Simulation evaluation

Simulation setting

In this section, the synthetic gear signals are used to evaluate the stability of the proposed CMDE method. The synthetic gear signals simulate the normal meshing and pitting faults of two spur gears as shown in Figure 5(a) and (b). The synthetic signals are generated as Liang et al.\textsuperscript{16} In this paper, we briefly review the basic assumption and the dynamic model of the synthetic gear signals. The basic assumptions of the simulation are given as follows.

1. The two gears are standard spur gears, and the meshing between gears is simplified as spring-damping structure.
2. The friction is ignored.
3. The support of shaft and bearing is simplified as spring-damper structure.
4. The centralized parameter method is adopted, which regards each component of the system as a mass point, and the connection between each component is a spring-damper structure.
5. The vibration in the $x$ direction is free response and will disappear due to inherent damping. In this paper, we focus only on the motion in the $y$ direction.

The dynamic model is given as Figure 6. The vertical motion in the $y$ direction of the input/output gear are expressed as equations (9)–(10).

$$m_1\ddot{y}_1 = -F_k - F_c - F_{k1} - F_{c1}$$
$$m_2\ddot{y}_2 = F_k + F_c - F_{k2} - F_{c2}$$

where the $m_1/m_2$ indicates the mass of input/output gear. $y_1/y_2$ indicates the linear displacement of input/output gear in $y$ direction. $F_k/F_c$ indicates the stiffness/damping between gears. $F_{k1}/F_{c1}$ indicates the stiffness/damping of input bearing. $F_{k2}/F_{c2}$ indicates the stiffness/damping of output bearing.

The rotary motion of the input/output gear are expressed as equations (11) and (12)
\[ I_1\dot{\theta}_1 = M_{pk} + M_{pc} - R_{b1}(F_k + F_c) \]  
\[ I_2\dot{\theta}_2 = R_{b2}(F_k + F_c) - M_{gk} + M_{pc} \]

where \( I_1/I_2 \) indicates the rotational inertia of input/output gear. \( \theta_1/\theta_2 \) indicates the angular displacement of input/output gear. \( M_{pk}/M_{pc} \) indicates the stiffness/damping moment of input couplings. \( R_{b1}/R_{b2} \) indicates the base circle radius of input/output gear. \( M_{gk}/M_{pc} \) indicates the stiffness/damping moment of output couplings.

The rotary motion of the motor/load are expressed as equations (13) and (14).

\[ I_m\ddot{\theta}_m = M_1 - M_{gk} - M_{pc} \]  
\[ I_b\ddot{\theta}_b = M_{gk} + M_{gc} - M_2 \]

where \( I_m/I_b \) indicates the rotational inertia of motor/load. \( \theta_m/\theta_b \) indicates the angular displacement of motor/load. \( M_1/M_2 \) indicates the torque of the input/output.

The values of forces and moments are expressed as equations (15)–(24).

\[ F_k = k_1 y_1 \]  
\[ F_c = c_1 \dot{y}_1 \]  
\[ F_k = k_2 y_2 \]  
\[ F_c = c_2 \dot{y}_2 \]  
\[ M_{pk} = k_p(\theta_m - \theta_1) \]  
\[ M_{pc} = c_p(\dot{\theta}_m - \dot{\theta}_1) \]  
\[ M_{gk} = c_g(\dot{\theta}_2 - \dot{\theta}_b) \]  
\[ M_{gc} = c_g(\theta_2 - \theta_b) \]  
\[ F_k = k_t(R_{b1}\dot{\theta}_1 - R_{b2}\dot{\theta}_2 + y_1 - y_2) \]  
\[ F_c = c_t(R_{b1}\ddot{\theta}_1 - R_{b2}\ddot{\theta}_2 + \dot{y}_1 - \dot{y}_2) \]

where \( k_1/k_2 \) indicates the vertical radial stiffness of input/output bearings. \( c_1/c_2 \) indicates the vertical radial damping of input/output bearings. \( k_p/k_g \) indicates the torsional stiffness of input/output flexible coupling. \( c_p/c_g \) indicates the damping coefficient of input/output flexible coupling. \( k_t \) indicates the total mesh stiffness. \( c_t \) indicates the mesh damping coefficient.

Additionally, the synthetic signals are added with white Gausses noise, and the signal-noise-ration is 10 dB. The data length is set to be 1024 points.

**Results and analysis**

After acquiring the synthetic gear signals as described in Section “Simulation setting,” the MDE and CMDE are used to estimate the dynamical complexity of the synthetic gear signals. Each method has been run 20 times and the standard deviations of the diversity entropy under each scale are plotted in Figure 7. To be
The MDE and CMDE are set with the same parameters: embedding dimension $m = 4$, symbols $e = 30$, and scale factor $\tau = 30$.

In Figure 7(a) and (b), it can be seen that the standard deviation of MDE increases with the scale factor. This is coincident with the discussion in Section “Methodology”: since the averaging window increases with the scale factor, the multiscale time series will be greatly shortened. The short multiscale time series cannot provide sufficient information for DE to estimate the dynamical complexity. Thus, the diversity entropy value will fluctuate at high scale, which presents unstable behavior in Figure 7. Obviously, the entropy value with large standard deviation will make the extracted features unreliable at high scale, which restricts the wilder application of diversity entropy.

Conversely, the CMDE presents stable dynamical complexity at high scale. Can be seen from both Figure 7(a) and (b), the standard deviation of CMDE maintains no more than 2% even at scale 30. Benefiting from the composite multiscale procedure, the CMDE can repeatedly estimate the dynamical complexity to obtain a constant diversity entropy value. From classification perspective, a stable multiscale feature is more conducive to categorization.

Experiment evaluation

Experiment setting

To evaluate the performance of the proposed CMDE-kNN method, a compound mechanical fault experiment was conducted on the test rig called Machinery Fault Simulator produced by SpectraQuest. The test rig and its schematic diagram are shown in Figure 8(a) and (b) respectively. This experiment utilized an electric motor to drive a three-way gearbox. The rotating speed of motor was set as 3000 rpm. The three-way gearbox consists of one pair of straight cut bevel gears and the gear ratio is 18:27 (seen Figure 8(b)). The load was set as 5 in-lbs of torque. The accelerometer was mounted on the top of the gearbox to acquire the vibrational signal. The sampling frequency was 12,800 Hz.

Noted that since the rotating speed of motor was set as 3000 rpm, it takes 0.02 s to make a revolution. That means a sample with data length 1024 points (0.08 s) will contain four revolutions. Theoretically, for periodic signals, the data length should be greater than one period. That means a sample with data length 1024 points contains sufficient fault information for analyzing.

The different health conditions were realized by replacing the test bearing and test gear with artificial...
damage, as shown in Figure 9. The testing bearing type was 6202 deep groove ball bearing. Six types of faults were designed in this experiment: pitting in the driving gear (PIT, seen Figure 9(a)), broken tooth in the driving gear (BRO, seen Figure 9(b)), missing tooth in the driving gear (MIS, seen Figure 9(c)), ball fault (BF, seen Figure 9(d)), inner race fault (IRF, seen Figure 9(e)), and outer race fault (ORF, seen Figure 9(f)). Totally, there will be 16 health conditions as Table 1. There are has 100 samples in each health state and 1600 samples in total. In each test, we randomly choose 50% samples from each health conditions as training samples, and the residual samples are treated as training samples. Meanwhile, the data length is 1024 points. The waveforms of 16 health conditions are showed in Figure 10.

To compare, MPE, MFE, MSE, CMPE, CMFE, and CMSE were used to identify the health condition with the same procedure as CMDE. The parameters of each entropy based method were set as their best parameters as seen in Table 2. Noted to be fair, the scale factor of eight methods were the same, and the classifier of each method was kNN.

### Results and Analysis

The final classification results are shown in Figure 11. From Figure 11, three conclusions can be made as follows:

First, all the composite multiscale based entropy methods outperform the multiscale based entropy methods. For example, the diagnosis accuracy of CMPE (84.76%) is higher than MPE (52.96%). This can attribute to the stable entropy estimation generated by the composite multiscale procedure: the moving averaging procedure can repeatedly estimate the dynamical complexity to obtain a constant entropy value. Thus, the kNN classifier can distinguish the fault type much easier.

Second, the standard deviation of composite multiscale based entropy methods is less than that of the multiscale based entropy methods. This can also attribute to the stability of the composite multiscale procedure: the small standard deviation at high scales can make cluster boundary much clearer, which makes the kNN classifier much easier to category the samples around the boundary. The detailed discussion of cluster boundaries will be made later.

Last, the CMDE achieves the highest accuracy among eight methods. Benefiting from the three advantages of diversity entropy, the MDE achieves the highest accuracy among the four multiscale based entropy methods. Besides, CMDE achieves the highest accuracy among the four the four composite multiscale based entropy methods. This suggests that the final classification accuracy of multiscale based method depends on the entropy value method itself: the stronger the feature extraction ability of the entropy method, the higher the clustering ability of the fault features.

Additionally, the relationship between scale and accuracy has been investigated and the result is shown on Figure 12. Figure 12 shows the test accuracy of MDE and CMDE with different number of features (scales). From Figure 12, for MDE, it can be seen that the accuracy increase as the scale increase when the scale is less than 16. This means the measured vibration signal from compound fault consists of multiple features from different components, which requires the multiscale procedure to explore richer features from deeper scales. Actually, the multiscale procedure used in MDE is a Haar wavelet. As the scale increase, the fault information hidden in the different spectral resolution can be dig out. However, the accuracy decreases when the scale exceeds 16. This is because the
multiscale time series will be greatly shortened at high scale. This makes the diversity entropy has insufficient information to estimate the dynamical complexity, which presents unstable behavior at high scale factor.

The proposed CMDE using mean complexity value of composite multiple scale time series to enhance the stability of diversity entropy at high scale, provides reliable fault information at high scales. Thus, the accuracy of CMDE keeps increasing when the scale is less than 25. When scale exceeds 25, accuracy of CMDE maintains around 90%. This means the fault features become redundant when scale exceeds 25.

To show the feature extraction ability of CMDE more intuitively, Fisher score (FS) and visualized

Figure 10. The waveforms of 16 health conditions: (a) BRO + BF, (b) BRO + IRF, (c) BRO + ORF, (d) MIS + BF, (e) MIS + IRF, (f) MIS + ORF, (g) BF, (h) IRF, (i) ORF, (j) PIT + BF, (k) PIT + IRF, (l) PIT + ORF, (m) BRO, (n) MIS, (o) normal, and (p) PIT.
feature are adopted in this section. Fisher score is a simple criterion to judge the separability of the extracted features. FS uses the ratio of inter-class distance to inner-class distance as equation (25).

\[ \text{FS} = \frac{S_B}{S_W} = \frac{\text{tr} \left( \frac{1}{N} \sum_{i=1}^{C} N_i (m_i - m_0)(m_i - m_0)^T \right)}{\text{tr} \left( \frac{1}{N} \sum_{i=1}^{C} \sum_{j=1}^{N_i} (x_{i,j} - m_i)(x_{i,j} - m_i)^T \right)} \]  

where \( S_B \) is the distance between different health conditions, and \( S_W \) is the distance within a single condition. When the ratio of the \( S_B \) and the \( S_W \) is maximized, the samples can be separated easily. \( N \) indicates the total number of the samples. \( C \) represents the total number of health conditions. \( N_i \) indicates the number of the samples in health condition \( i \), \( m_i \) is the mean vector of samples in condition \( i \). \( m_0 \) is the mean vector of all samples. \( x_{i,j} \) is the two-dimensional feature vector of \( j \)th sample in health condition \( i \).

The visualized features and FS of each methods are displayed in Figure 13. Four conclusions can be made as follows:

First, among the multiscale based methods (Figure 13(a) MDE, Figure 13(b) MPE, Figure 13(c) MFE, Figure 13(a) MSE), only MDE (FS = 3.72) has explicit class centers. Thus, the final diagnosis accuracy of MDE is higher than other multiscale based methods. From FS we can clearly see that MDE (3.72) > MPE (3.27) > MFE (0.77) > MSE (0.10). This indicates that MDE has the best feature extraction ability among the four multiscale based methods in this test.

Second, the class centers of the composite multiscale based methods are much clearer than the multiscale based methods. The most obvious example is MFE and CMFE: for MFE, we can hardly observe the class centers in Figure 13(c); but for MFE, each types of sample forms a clear cluster in Figure 13(g). This proves the effectiveness of the composite multiscale procedure in the feature extraction.

Third, although MDE (Figure 13(a)) can forms explicit class centers, but the inter-class distance is every close. In other words, each cluster is close. And we can see parts of samples are far from the class center, that is, condition 15, the pink samples in Figure 13(a). In Figure 13(e) CMDE, not only the inter-class distance is larger, but also the samples outer center are fewer.

Last, the CMDE achieves the highest FS (7.87), which indicates the CMDE owns the best feature extraction ability among the eight methods. This result coincides with the classification accuracy (Figure 11). This proves the proposed CMDE has the best feature extraction ability and the CMDE is independent with the classifier.

Conclusions

The conclusions of this paper are summarized as follows:

(1) A novel feature extraction method called composite multiscale diversity entropy (CMDE) has
been proposed. The CMDE utilizes the mean complexity value of composite multiple scale time series to enhance the stability of diversity entropy at high scale, which enables the diversity entropy could extract richer fault information from deeper scales for the compound fault diagnosis of rotating machinery.

(2) The simulation results prove that the stability of diversity entropy has been greatly improved, which can dig richer fault information from deeper scales.

(3) The experimental results show that the proposed CMDE has the best feature extraction ability in compound mechanical fault diagnosis compared with the existing entropy-based methods.

---

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

**ORCID iDs**

Guangwei Yu https://orcid.org/0000-0002-3441-2498
Xianzhi Wang https://orcid.org/0000-0002-7997-8348

**References**

1. Lei Y, Yang B, Jiang X, et al. Applications of machine learning to machine fault diagnosis: a review and road-map. *Mech Syst Sig Process* 2020; 138: 106587.
2. Liu H, Song W, Zhang Y, et al. Generalized Cauchy degradation model with long-range dependence and maximum Lyapunov exponent for remaining useful life. *IEEE Trans Instrum Meas* 2021; 70: 1–12.
3. Lei Y, Lin J, Zuo MJ, et al. Condition monitoring and fault diagnosis of planetary gearboxes: a review. *Measurement* 2014; 48: 292–305.
4. Li Y, Wang X, Liu Z, et al. The entropy algorithm and its variants in the fault diagnosis of rotating machinery: a review. *IEEE Access* 2018; 6: 66723–66741.
5. Huo Z, Martinez-Garcia M, Zhang Y, et al. Entropy measures in machine fault diagnosis: insights and applications. *IEEE Trans Instrum Meas* 2020; 69(6): 2607–2620.
6. Richman JS and Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol* 2000; 278(6): H2039–H2049.
7. Chen W, Wang Z, Xie H, et al. Characterization of surface EMG signal based on fuzzy entropy. *IEEE Trans Neural Syst Rehabil Eng* 2007; 15(2): 266–272.
8. Bandt C and Pompe B. Permutation entropy: a natural complexity measure for time series. *Phys Rev Lett* 2002; 88(17): 174102.
9. Costa M, Goldberger AL and Peng C-K. Multiscale entropy analysis of complex physiologic time series. *Phys Rev Lett* 2002; 89(6): 068102.
10. Wu S-D, Wu P-H, Wu C-W, et al. Bearing fault diagnosis based on multiscale permutation entropy and support vector machine. *Entropy* 2012; 14(8): 1343–1356.
11. Gao Y, Villecco F, Li M, et al. Multi-scale permutation entropy based on improved LMD and HMM for rolling bearing diagnosis. *Entropy* 2017; 19(4): 176.
12. Huang J, Wang X, Wang D, et al. Analysis of weak fault in hydraulic system based on multi-scale permutation entropy of fault-sensitive intrinsic mode function and deep belief network. *Entropy* 2019; 21(4): 425.
13. Zheng J, Cheng J, Yang Y, et al. A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and
variable predictive model-based class discrimination. Mech Mach Theory 2014; 78: 187–200.
14. Zhao H, Sun M, Deng W, et al. A new feature extraction method based on EEMD and multi-scale fuzzy entropy for motor bearing. Entropy 2017; 19(1): 14.
15. Wang X, Si S and Li Y. Multiscale diversity entropy: a novel dynamical measure for fault diagnosis of rotating machinery. IEEE Trans Ind Inf 2021; 17(8): 5419–5429.
16. Liang X-H, Liu Z-L, Pan J, et al. Spur gear tooth pitting propagation assessment using model-based analysis. Chin J Mech Eng 2017; 30(6): 1369–1382.
17. Li Y, Wang X, Si S, et al. Entropy based fault classification using the case western reserve university data: a benchmark study. IEEE Trans Reliab 2020; 69(2): 754–767.