Journal of Physics: Conference Series

Refined time-shift multiscale diversity entropy: a novel feature extraction algorithm for fault diagnosis of planetary gearbox

To cite this article: Shun Wang and Yongbo Li 2022 *J. Phys.: Conf. Ser.* **2184** 012010

Refined time-shift multiscale diversity entropy: a novel feature extraction algorithm for fault diagnosis of planetary gearbox

Shun Wang, Yongbo Li*
Northwestern Polytechnical University, Xian, 710072, China

Abstract. Planetary gearboxes play a critical role in aerospace and heavy industry fields, such as wind turbines, heavy vehicles and construction machines. Intelligent fault diagnosis is significant for safe operation and fault prevention of planetary gearboxes. Recently, multiscale diversity entropy and related entropy methods are proposed to extract features of time series and applied for the fault diagnosis. However, there are still some limitations in fault feature representation and stability for multiscale diversity entropy. To solve the problem, in this paper, a novel planetary gearboxes fault diagnosis method via refined time-shift multiscale diversity entropy (RTSMDE) is proposed. First, a novel entropy algorithm called RTSMDE is proposed to measure the complexity of time series and extract fault features of the vibration signals, which is robust and efficient in performance. Then, the obtained features are utilized to fulfil automatically the fault pattern identifications using support vector machine. To confirm the superiority of the RTSMDE-based fault diagnosis method, simulated signals and experimental studies are constructed and three used widely methods are employed to present a comprehensive comparison. The results indicate that RTSMDE performs best and obtains the highest accuracy.

Keywords: planetary gearbox, feature extraction, diversity entropy, and fault diagnosis.

1. Introduction

Planetary gearboxes play a critical role in aerospace, automotive and heavy industry fields, such as heavy trucks, aerospace, wind turbines, and complex mechanical transmission systems because of the large transmission capacity and strong load-bearing capacity. However, considering the influence of variable working conditions, the key parts of planetary gearboxes system such as the ring gear and planet gear are prone to faults including spalling, wear and pitting, which will lead to unstable operation or even unexpected accidents. Therefore, it is of significance to timely recognize the faults for prevention of catastrophic failure and ensure reliable operations of machinery and equipment [1].

Generally, when local faults occurs during the operation, the vibration signals of planetary gearbox show obvious non-linearity and non-stationarity, making linear analysis approaches useless in analysing these nonlinear and nonstationary signals. Meanwhile, vibration signals will be disturbed by environmental noise, which makes it difficult to extract fault features. Hence, feature extraction algorithms and intelligent fault pattern recognition of planetary gearbox become the current research hotspots [2], [3].

The widely used vibration-based signal processing methods mainly include frequency analysis, time analysis, time-frequency analysis and so on. Moreover, wavelets transform, empirical mode decomposition (EMD), full Fourier transform, and hidden Markov models are also seen as the signal processing tool for fault diagnosis. However, a common drawback of above algorithms is that they rely heavily on prior knowledge or require extensive expertise. In order to overcome this defect, the entropy-based approaches are introduced by conducting a quantitative complexity analysis for time series and diagnosing faults of the system. As a powerful nonlinear signal analysis technique, the entropy-based methods have been verified to be an effective method in fault diagnosis[4], which has its unique merits in real applications.

The most widely used entropy-based methods mainly include approximate entropy (AE), sample entropy (SE), fuzzy entropy (FE), and permutation entropy (PE), dispersion entropy (DisEn)[5]. AE was proposed by Pincus, which measures the complexity by computing the proportion of new states emerging. However, approximate entropy is a biased statistic, which lacks relative consistency in some cases. Therefore, SE was proposed, which relieved the bias caused by self-matching. Unfortunately, for SE, the Heaviside step function will cause the sudden changes of the similarity measurement of the template so that it leads to unreliable values for short signals and it is invalid in analysing short time series. Thus, Chen et al. introduced the fuzzy set theory to count the states of the orbits and proposed FE method, which can be seen as the improvement of SE. Unlike SE and FE method, PE was proposed by computing the state probability of the permutation, which has high calculation efficiency. However, it considers only the order of the amplitude values so that some information with regard to the amplitudes may be discarded. Thus, to overcome the disadvantages of SE and PE, dispersion entropy was proposed [6], [7] as a new indicator for complexity or orderliness measurement.

Recently, diversity entropy (DE) was proposed by Wang et al. to measure the complexity and uncertainty of time series[8]. DE has the merits of high consistency, high calculation efficiency, and strong denoising ability. Meanwhile, DE is expanded to the multiscale framework, called multiscale diversity entropy (MDE). However, the coarse-graining multiscale analysis used in original MDE generally leads to large fluctuation in different scale factors. Moreover, with the increase of scale factors, the coarse grained time series becomes shorter and shorter.

In order to overcome the disadvantages of MDE, time-shift multiscale technique is introduced and refined diversity entropy (RTSMDE) is proposed in this paper. In the proposed RTSMDE method, the coarse-graining multiscale analysis is replaced by the new time-shift multiscale technique. Note that the time-shift multiscale analysis originates from the calculation process of Higuchi's fractal dimension, which can effectively preserve the important constructing information of the original data[9], [10]. On this basis, the proposed RTSMDE technology is applied jointly with support vector machine (SVM) and a new fault diagnosis flowchart is proposed. One case study of planetary gearbox is conducted by comparing it with MFE, MPE and MDE-based methods. The experimental results show that the proposed method obtains a higher accuracy in than the state-of-the-art entropy technologies.

The rest of this article is organized as follows. The original DE is reviewed and the proposed RTSMDE approach is introduced in Section 2. Meanwhile, simulated signals are given to verify the advantages of RTSMDE in Section 2. In Section 3, the RTSMDE and SVM-based fault diagnosis approach for planetary gearbox is proposed. Section 4 provides experimental validations using one case study of planetary gearbox. Finally, a conclusion is provided in the Section 5.

2. Method

2.1. Diversity entropy

For an arbitrary time series X={ $x_1,\ldots,x_i,\ldots,x_N$ } with data length N, the steps of DE can be described as follows[8]:

Step 1. According to Taken's phase space theory [11], phase space reconstruction is carried out and a series of vectors can be obtained denoted as $y_i(m)=\{x_i, x_{i+1},\cdots,x_{i+m}\}$ using the embedding dimension m as Eq.(1).

$$
\begin{aligned}
Y(m) &= \{y_1(m), y_2(m),\cdots, y_{N-m+1}(m)\} \\
&= \begin{Bmatrix} x_1 & x_2 & \cdots & x_m \\ x_2 & x_3 & \cdots & x_{1+m} \\ \vdots & \vdots & & \vdots \\ x_{N-m+1} & x_{N-m} & \cdots & x_N \end{Bmatrix}
\end{aligned} \tag{1}
$$

Step 2. Compute the cosine similarity between the adjacent vectors to obtain a series of cosine similarities $(d_1,\cdots,d_{N-m})$. The cosine similarity d between two orbits is defined as follows:

$$\begin{aligned} D(m) &= \{d_1,\cdots,d_{N-m}\} \\ &= \{d(y_1(m),y_2(m)),d(y_2(m),y_3(m)), \\ &\quad \cdots,d(y_{N-m-1}(m),y_{N-m}(m))\} \end{aligned} \tag{2}$$

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

Noted that the range of the cosine similarity d is [-1, 1].

Step 3. Partition the scope [-1, 1] into ε intervals denoted, which is expressed by $(I_1,I_2,\ldots,I_\varepsilon)$. Next, calculate the probability $(P_1,\ldots,P_\varepsilon)$, which represents the probability that cosine similarity falls on each interval.

Step 4. The *DE* of X={ $x_1,\ldots,x_i,\ldots,x_N$ } can be computed by Eq.(4).

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

2.2. Refined time-shift multiscale diversity entropy

With the coarse-graining technique, DE is expanded to the multiscale framework, namely MDE. However, the multiscale analysis used in original MDE generally leads to large fluctuation in different scale factors. Meanwhile, with the increase of scale factors, the obtained time series becomes shorter and shorter by coarse-graining technique. Thus, in order to alleviate the impact, inspired by time-shift multiscale analysis, RTSMDE is proposed by replacing the original multiscale analysis. The specific computation steps of RTSMDE can be described as follows:

Step 1. For time series X={ $x_1,\ldots,x_i,\ldots,x_N$ }, let β and α be positive integers, where $\beta = 1, 2, \ldots, \alpha$, then α new time series can be constructed by:

$$u_\alpha^\beta = \left(x_\beta, x_{\beta+\alpha}, x_{\beta+2\alpha}, \ldots, x_{\beta+\left[\frac{N-\beta}{\alpha}\right]\alpha} \right) \tag{5}$$

where β and α represent the initial point and interval, respectively, $\left[\frac{N-\beta}{\alpha}\right]$ represents the nearest integer that does not exceed $\frac{N-\beta}{\alpha}$. For example, let $\alpha = 3$, α new sequences can be obtained after time-shift process, which can be displayed in Fig.1.

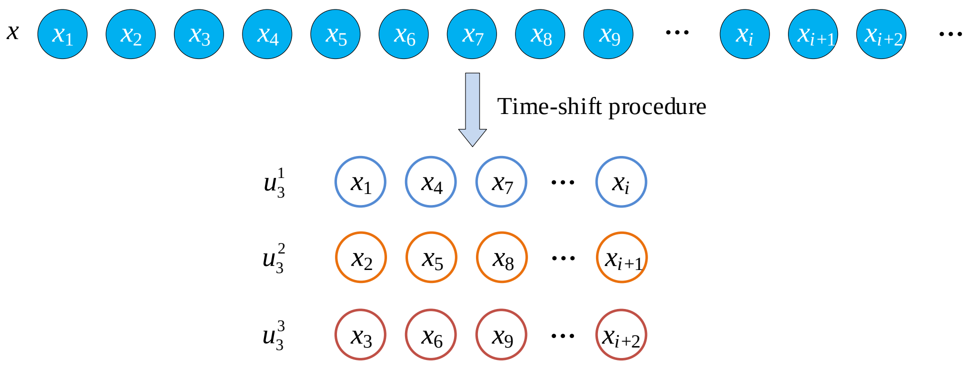

Fig.1. The illustration of time-shift multiscale analysis with $\alpha = 3$.

Step 2. For α new time series, the P_k^β of each u_α^β can be obtained according to step(1)-(3) of DE's computation. Then, we define $\bar{P}_k = \sum_{\beta=1}^{\alpha} P_k^\beta$.

Step 3. RTSMDE of time series X={ $x_1,\ldots,x_i,\ldots,x_N$ }can be defined as

$$\mathrm{RTSMDE}(X,m,\varepsilon,\tau)_\alpha = -\frac{1}{\ln\varepsilon}\sum_{k=1}^{\varepsilon}\bar{P}_k \ln \bar{P}_k \tag{6}$$

where τ is scale factor, $1 \le \alpha \le \tau$. For RTSMDE, time-shift multiscale analysis can avoid the neutralisation phenomenon caused by traditional coarse-graining analysis. At the same time, the refining technology is utilized to average the probability of each time-shift multiscale time series, which can increase the stability of time-shift multiscale analysis on a certain scale. There are three parameters needed to be set before calculation of RTSMDE method: embedding dimension *m*, number of intervals ε , and scale factor τ . According to Ref[8], *m* is fixed to 4 and ε is fixed to 30 in this study.

In summary, the calculation process of RTSMDE algorithm is shown in Fig. 1, and the pseudo code of RTSMDE is given in Algorithm 1.

Algorithm 1 Refined time-shift multiscale diversity entropy

Input:

1) X : the time series
2) *m*: the embedding dimension
3) ε : the number of intervals
4) τ : the scale factor

Procedure Begin:

```
for α =1,2,…,τ do
        Construct time shift analysis and obtain new time series u_α^β based on Eq.(5).
        Compute the P_k^β valve of each u_α^β, and obtain the P̄_k
        Compute RTSMDE_α based on Eq.(6).
end for
```

End procedure

Output: RTSMDE value

2.3. Simulated noisy signals

In the subsection, the simulated signals are utilized to verify the merits of RTSMDE in stability. For comparison purpose, MDE is also utilized to process the simulated signals. Without loss of generality, one hundred groups of white Gaussian noises (WGN) and 1/f noises as examples are to verify the stability of RTSMDE method. Here, each sample signal has the length of 5000 and the scale factor τ is set to 20 for RTSMDE and MDE methods. The obtained results are illustrated in Fig.2. Two conclusions can be drawn from Fig.2. First, the proposed RTSMDE curve is smoother and more stable than original MDE method. Second, RTSMDE method obtains smaller error bars, as illustrated in Fig.2, especially at larger scales. In conclusion, the numerical results have validated the proposed RTSMDE has the advantage of stability.

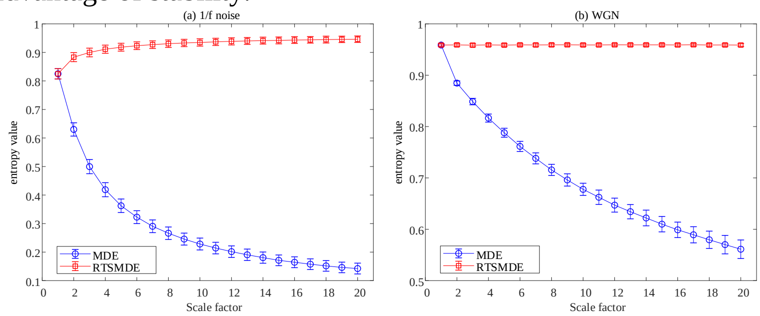

Fig.2. The comparisons of MDE and RTSMDE for WGN and 1/f noise: (a) analysis results of 1/f noise, (b) analysis results of WGN.

3. The Proposed RTSMDE-based fault diagnosis approach

The proposed RTSMDE technology is applied jointly with SVM and a new fault diagnosis flowchart is proposed. The proposed approach mainly includes the parts of data acquisition, RTSMDE-based feature extraction and SVM-based fault pattern recognition, which proceeds according to the following steps.

(1) The data acquisition under different health conditions is conducted.
(2) Segment data into 100 samples for each health condition and divide the obtained samples into the training set and testing set.
(3) RTSMDE is employed to quantify nonlinearity and fault information from the vibration signals.
(4) The obtained feature sets of the training data are fed into SVM to train a classifier.
(5) Test the trained classifier using the features of testing set and identify the fault types of planetary gearboxes automatically.

4. Engineering experiment

In this section, experiments are conducted on a planetary gearbox system to validate the effectiveness of proposed fault diagnosis approach, as shown Fig.3. The test rig is mainly composed of the driving motor, tachometer, planetary gearbox and magnetic damping. Here, an accelerometer was mounted on the top of the planetary gearbox casing for acquisition of vibration signals. Note that the sampling frequency is set to be 10240 Hz with rotation speed of 2400PRM. Meanwhile, the 5 Nm load is designed to simulate the real application scenario.

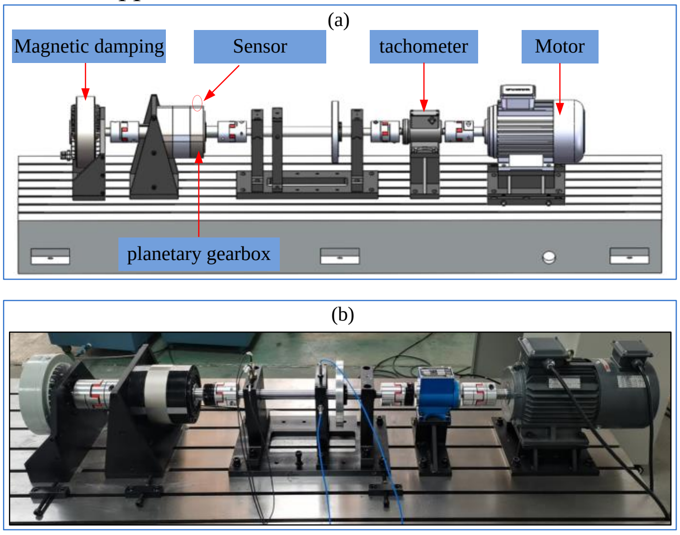

Fig.3. The experimental planetary gearbox system: (a) the layout of the test rig, (b) the test rig.

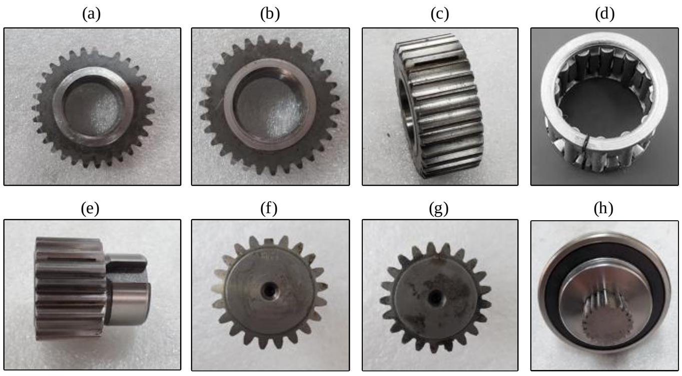

Fig.4. (a) planet gear with a missing tooth (PGMT), (b) planet gear with a broken tooth (PGBT), (c) planet gear with a spalling tooth (PGST), (d) fracture of bearing cage (FBC), (e) sun gear with a spalling tooth (SGST), (f) sun gear with a broken tooth (SGBT), (g) sun gear with a cracked tooth (SGCT), (h) sun gear with a missing tooth (PGMT).

In the experiment, nine working conditions of planetary gear system are considered, including healthy gear, planet gear with a missing tooth (PGMT), planet gear with a broken tooth (PGBT), planet gear with a spalling tooth (PGST), fracture of bearing cage (FBC), sun gear with a spalling tooth (SGST), sun gear with a broken tooth (SGBT), sun gear with a cracked tooth (SGCT), sun gear with a missing tooth (PGMT), as illustrated in Fig.4. It is noticed that there are 100 samples for each health condition and this case study contains total 900 samples. The detailed description of experimental data is given in Table 1. Here, the time-domain waveforms of vibration signals under different working conditions are illustrated in Fig.5.

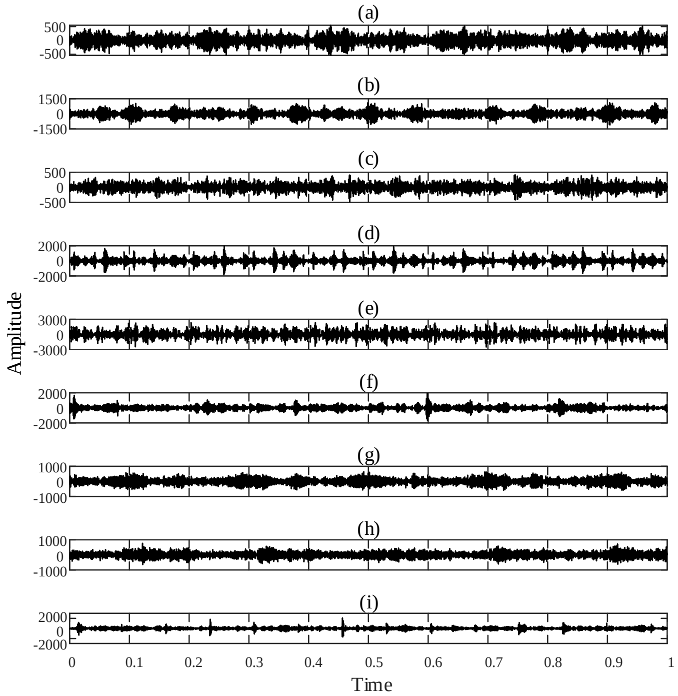

Fig.5. The time-domain waveforms of vibration signals: (a) healthy gear, (b) sun gear with a spalling tooth (SGST), (c) sun gear with a cracked tooth (SGCT), (d) sun gear with a broken tooth (SGBT), (e) sun gear with a missing tooth (PGMT), (f) planet gear with a broken tooth (PGBT), (g) fracture of bearing cage (FBC), (h) planet gear with a spalling tooth (PGST), (i) planet gear with a missing tooth (PGMT).

Table 1. The detailed description of numbers of experimental planetary gearbox data sets.

Fault class	Class label	Number of training data	Number of testing data
Healthy gear	1	50	50
planet gear with a missing tooth (PGMT)	2	50	50
planet gear with a broken tooth (PGBT)	3	50	50
planet gear with a spalling tooth (PGST)	4	50	50
fracture of bearing cage (FBC)	5	50	50
sun gear with a spalling tooth (SGST)	6	50	50
sun gear with a broken tooth (SGBT)	7	50	50
sun gear with a cracked tooth (SGCT)	8	50	50
sun gear with a missing tooth (PGMT)	9	50	50

In the proposed fault diagnosis methodology, at first, RTSMDE is utilized to extract fault features with scale factor τ = 20. Hence, twenty features can be obtained. Secondly, the classifier SVM is employed to accomplish pattern identification. Moreover, MDE, MFE and MPE are also utilized to accomplish fault diagnosis for comparison. In order to reduce the impact of randomness on the results, twenty trials are conducted in the study. The obtained diagnosis results are shown in Table 2.

Table 2. The diagnosis results of experimental data sets.

Method	Mean accuracy	Standard deviation	Average time(s)
RTSMDE	**94.21%**	**0.82%**	**33.29**
MDE	91.9%	1.13%	70.28
MFE	91.66%	1.4%	338.52
MPE	90.38%	1.32%	474.12

Seen from Table 2, it can be observed that RTSMDE obtains the highest classification accuracy and the smallest standard deviation among all the methods, which confirms the advantages of time shift analysis in stability and feature extraction. Meanwhile, it can be found that among the time consumption comparison of four methods RTSMDE has the highest calculation efficiency. Moreover, confusion matrix of the RTSMDE-based fault diagnosis approach under nine working conditions is illustrated in Fig.6. The results show that the recognition accuracy of each fault state is above 85%, and three fault types can be identified completely, which further validates the feature extraction ability of RTSMDE.

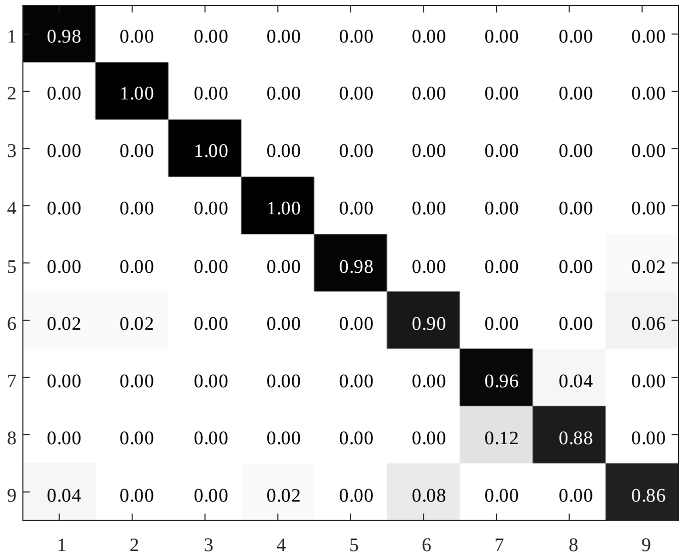

Fig.6. Confusion matrices of the proposed method under nine working conditions.

5. Conclusions

In this paper, a new entropy indicator, namely RTSMDE, is proposed and has been proven to be stability and highly efficient using multiple simulated signals. Moreover, based on RTSMDE algorithm and SVM classifier, a novel planetary gearboxes fault diagnosis approach is proposed, and one case study is conducted to demonstrate its advantage in classifying various faults of planetary gearboxes by comparing with MDE, MFE and MPE. From the experimental results, it can be observed that the proposed RTSMDE can effectively recognise planetary gearbox fault states, and achieves the highest recognition with 94%. Meanwhile, the proposed RTSMDE-based approach obtains the highest calculation efficiency. In conclusion, the proposed RTSMDE is an effective approach for complexity measure of time series and feature extraction for planetary gearbox. In the future work, the effectiveness of RTSMDE will be studied in a more realistic environment.

Acknowledgments

This research was funded by the National Natural Science Foundation of China under grant 51805434.

References

[1] H. Shao, J. Lin, L. Zhang, D. Galar, and U. Kumar, "A novel approach of multisensory fusion to collaborative fault diagnosis in maintenance," *Inf. Fusion*, vol. 74, pp. 65–76, Oct. 2021, doi: 10.1016/j.inffus.2021.03.008.

[2] Y. Lei, J. Lin, M. J. Zuo, and Z. He, "Condition monitoring and fault diagnosis of planetary gearboxes: A review," *Measurement*, vol. 48, pp. 292–305, Feb. 2014, doi: 10.1016/j.measurement.2013.11.012.

[3] T. Wang, Q. Han, F. Chu, and Z. Feng, "Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review," *Mech. Syst. Signal Process.*, vol. 126, pp. 662–685, Jul. 2019, doi: 10.1016/j.ymssp.2019.02.051.

[4] Z. Huo, M. Mart ńez-García, Y. Zhang, R. Yan, and L. Shu, "Entropy Measures in Machine Fault Diagnosis: Insights and Applications," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 2607–2620, Jun. 2020, doi: 10.1109/TIM.2020.2981220.

[5] Y. Li, X. Wang, S. Si, and S. Huang, "Entropy Based Fault Classification Using the Case Western Reserve University Data: A Benchmark Study," *IEEE Trans. Reliab.*, 2019, doi: 10.1109/TR.2019.2896240.

[6] M. Rostaghi and H. Azami, "Dispersion Entropy: A Measure for Time-Series Analysis," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 610–614, May 2016, doi: 10.1109/LSP.2016.2542881.

[7] M. Rostaghi, M. R. Ashory, and H. Azami, "Application of dispersion entropy to status characterization of rotary machines," *J. Sound Vib.*, vol. 438, pp. 291–308, Jan. 2019, doi: 10.1016/j.jsv.2018.08.025.

[8] X. Wang, S. Si, and Y. Li, "Multiscale diversity entropy: a novel dynamical measure for fault diagnosis of rotating machinery," *IEEE Trans. Ind. Inform.*, pp. 1–1, 2020, doi: 10.1109/TII.2020.3022369.

[9] T. D. Pham, "Time-Shift Multiscale Entropy Analysis of Physiological Signals," *Entropy*, vol. 19, no. 6, Art. no. 6, Jun. 2017, doi: 10.3390/e19060257.

[10] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Phys. Nonlinear Phenom.*, vol. 31, no. 2, pp. 277–283, Jun. 1988, doi: 10.1016/0167-2789(88)90081-4.

[11] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, vol. 898, Springer, pp. 366–381.