

A New Bearing Fault Diagnosis Method based on Fine-to-Coarse Multiscale Permutation Entropy, Laplacian Score and SVM

Zhiqiang Huo^{1,3}, Yu Zhang¹, Lei Shu^{2,1*}, Michael Gallimore¹

¹School of Engineering, University of Lincoln, Lincoln, UK

²NAU-Lincoln Joint Research Center of Intelligent Engineering, Nanjing Agricultural University, China

³Guangdong Provincial Key Laboratory on Petrochemical Equipment Fault Diagnosis, Guangdong University of Petrochemical Technology, Maoming, China

Email: {zhuo, yzhang, lshu, mgallimore}@lincoln.ac.uk

Abstract—Fault diagnosis of rotating machinery is vital to identify incipient failures and avoid unexpected downtime in industrial systems. This paper proposes a new rolling bearing fault diagnosis method by integrating the Fine-to-Coarse Multiscale Permutation Entropy (F2CMPE), Laplacian Score (LS) and Support Vector Machine (SVM). A novel entropy measure, named F2CMPE, was proposed by calculating permutation entropy via multiple-scale fine-grained and coarse-grained signals based on wavelet packet decomposition. The entropy measure estimates the dynamic changes of time series from both low- and high-frequency components. Moreover, the F2CMPE mitigates the drawback of producing time series with sharply reduced data length via the coarse-grained procedure in the conventional Composite Multiscale Permutation Entropy (CMPE). The comparative performance of the F2CMPE and CMPE is investigated by analyzing synthetic and experimental signals for entropy-based feature extraction. In the proposed bearing fault diagnosis method, the F2CMPE is first used to extract entropy-based features from bearing vibration signals. Then, LS and SVM are used for selection of features and fault classification respectively. Finally, the effectiveness of the proposed method is verified for rolling bearing fault diagnosis using experimental vibration data sets, and the results have demonstrated the capability of the proposed method to recognize and identify bearing fault patterns under different fault states and severity levels.

Index Terms—Fault Detection and Diagnosis, Fine-to-Coarse Multiscale Permutation Entropy, Laplacian Score, Support Vector Machine.

I. INTRODUCTION

In industrial manufacturing plants, rolling bearings are usually operated under harsh and complicated working environment for pursuing higher profits and are inevitably subjected to incipient defects, which can potentially lead to energy waste and performance degradation of the whole industrial system [1]–[3]. Hence, in recent decades, fault detection and diagnosis of rolling bearing have attracted tremendous attention for accurately identifying and isolating faults and performing appropriate maintenance planning to avoid financial losses caused by unexpected downtime linked to bearing failures. For this purpose, lots of effort has been put into the

field of health monitoring of industrial systems by applying statistical measurements to characterize fault symptoms hidden in complex signals as early as possible [4], such as the use of time-domain and frequency-domain features [5]; nevertheless, both of them are most suitable for linear and stationary signals. As a result, traditional linear methods may not efficiently detect the dynamic change of complex and non-linear vibration signals [6].

Over the past decades, enormous entropy algorithms have been applied successfully in fault diagnosis applications of rotating machinery, providing useful statistical indicators as to measures of uncertainty and disorder of time series acquired from a physical system [7]–[10]. The commonly used entropy measures include single-scale entropy [11], such as Approximate Entropy (ApEn), Sample Entropy (SampEn), Fuzzy Entropy (FuzzyEn), and Permutation Entropy (PE). For instance, in [12], the ApEn was used as a non-linear feature indicator for discriminating different bearing conditions. In [13], a study was carried out to compare the performance of ApEn, SampEn, and FuzzyEn in analyzing simulation and experimental electromyographic signals respectively. Recently, PE was applied for extracting fault features from bearing vibration signals, and the results demonstrated that PE enjoys the advantages of simplicity, fast calculation and invariance for non-linear monotonous transformations [6], [14], [15].

Nonetheless, the performance of PE might be limited by analyzing time series with one single scale, which may neglect potential useful information associated with primary symptoms hidden in multiple scales. For overcoming this shortcoming, Multiscale Permutation Entropy (MPE) was first proposed by Aziz et al. [16], inspired by the concept of the coarse-grained procedure proposed in Multiscale Entropy (MSE) [17]. The main ideas behind MSE and MPE are producing coarse-grained time series with different scales and then applying single-scale entropy approaches (e.g., SampEn and PE) for the calculation of the complexity of coarse-grained time series with different scales. However, it was reported that MPE still encounters some drawbacks [7], [18]. First, the coarse-

grained procedure greatly reduces the data length of the time series with an increasing scale. Furthermore, the original time series is equally divided into non-overlapping fragments by the coarse-grained procedure, the results of which may yield inappropriate PE measure in the MPE. Concerning this limitation, Composite Multiscale Permutation Entropy (CMPE) was later proposed in [19], [20] by integrating information of multiple coarse-grained time series in one same scale. Comparatively, the CMPE could provide higher entropy reliability than the MPE while the use of high scale factors. Though the CMPE has improved from the MPE; the fundamental theory behinds the coarse-grained procedure is essentially a linear smoothing of the original time series. As a result, coarse-grained time series with increasing scale factors inevitably have sharply decreased data length which possibly yields inconsistent PE values. Moreover, only low-frequency components are maintained in the coarse-grained time series because of the coarse-grained procedure is similar to the sub-sampling operation. This may not adequately detect and identify incipient bearing degradation, especially when coherent fault symptoms are hidden in both low and high frequency components caused by the emergence of early failures in rolling bearings.

In this paper, a novel entropy measure, named Fine-to-Coarse Multiscale Permutation Entropy (F2CMPE), is put forward to overcome the shortcomings in the coarse-grained procedure mentioned above and generate reliable PE measure even with high scale factors. The F2CMPE method enjoys the advantages of extracting the dynamic change of time series from both low-frequency and high-frequency components. Besides, by using the Wavelet Packet Decomposition (WPD) analysis and F2C procedure, the F2C signals generated from reconstructed wavelet coefficients have the same data length of original time series, which improves the reliability of PE calculation, especially when high scale factors are used for feature extraction.

A new rolling bearing fault diagnosis method is proposed based on the F2CMPE, LS, and SVM, the flow chart of which is presented in Fig. 1. More specifically, the F2CMPE is first applied to characterize non-linear dynamic change associated with fault symptoms from bearing vibration signals. After having obtained fault feature vectors, we apply LS and SVM to select salient features and classify bearing patterns respectively. Finally, the applicability of the proposed F2CMPE method for analyzing synthetic signals with different noises was studied and compared with the CMPE in terms of different sources of noises (i.e., pink noise and Gaussian white noise) and Signal-to-Noise Ratios (SNRs). Then, the efficiency of the F2CMPE for bearing diagnosis was verified via real bearing experimental validation, and a comparative study between the proposed F2CMPE and the CMPE was also carried out for diagnosing bearing defects. The main contributions in this paper are concluded below:

- A novel non-linear complexity measure, named F2CMPE, is proposed to estimate the dynamic change of rolling bearing vibration signals. The F2CMPE not only takes low- and high-frequency components into account but

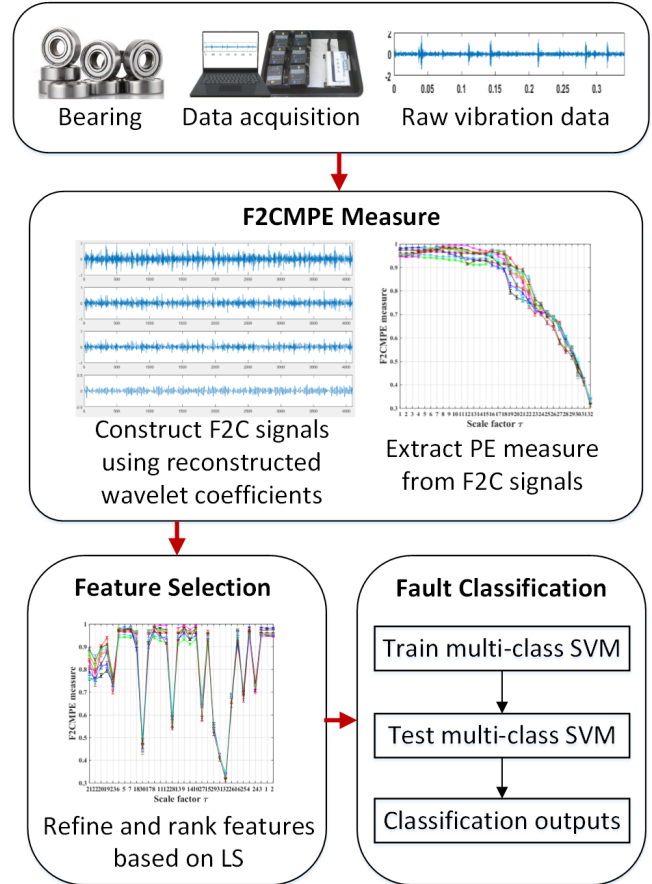


FIGURE 1. Flowchart of the proposed fault diagnosis method.

also mitigates the influence of decreased data length in signals with high multiple scale factors based on the WPD analysis and the F2C procedure.

- A new rolling bearing fault diagnosis method is proposed by integrating the F2CMPE, LS, and SVM methods for identifying and classifying incipient bearing failures.
- Both synthetic and experimental signals are analyzed to verify the effectiveness of the F2CMPE for feature extraction, and its performance is compared with that of the CMPE for characterizing the dynamic changes of time series. In the experimental validation, ten conditions of rolling bearing are applied in terms of different bearing components and fault severity levels. The analysis results have demonstrated that the proposed rolling bearing fault diagnosis method could guarantee high reliability and robustness in identifying bearing performance deterioration.

The rest of this paper is organized as follows: Section II reviews the principles of the PE, MPE and CMPE. Section III first reviews the WPD analysis and then presents the proposed F2CMPE method. The parameter selection in the calculation of the F2CMPE is also described and discussed. Section IV presents the proposed bearing fault diagnosis method based on the F2CMPE, LS and SVM. Section V shows the compar-

ative study between the F2CMPE and CMPE algorithms for analyzing synthetic signals with different noise levels. Section VI presents the experimental analysis to study and verify the effectiveness of the proposed method for rolling bearing fault diagnosis. Finally, conclusions are drawn in Section VII.

II. UNDERLYING PRINCIPLES OF PE, MPE AND CMPE

This section introduces the theoretical background of the PE, MPE, and CMPE algorithms.

A. Permutation Entropy (PE)

Bandit and Pompe proposed PE measure in 2002 for estimating the complexity of time series based on permutation patterns by comparing the neighboring values of the time series [21]. PE enjoys the advantages of simplicity, fast calculation and invariance concerning non-linear monotonous transformations [22]. Thus, PE is suitable for analyzing non-stationary time series measured from complex industrial system [6]. The principle of PE is briefly described below:

Given a time series $x(i)$ of length N , the time delay λ and the embedding dimension m , the phase space of a time series can be reconstructed as:

$$\mathbf{X}_i = \{x(i), x(i + \lambda), \dots, x(i + (m - 1)\lambda)\} \quad (1)$$

where $1 \leq i \leq N - (m - 1)\lambda$. Then, the m number of real values contained in each \mathbf{X}_i can be rearranged in an increasing order as

$$x(i + (j_1 - 1)\lambda) \leq x(i + (j_2 - 1)\lambda) \leq \dots \leq x(i + (j_m - 1)\lambda) \quad (2)$$

Therefore, any vector \mathbf{X}_i can be mapped onto a group of symbols as

$$\pi_n = (j_1, j_2, \dots, j_m) \quad (3)$$

where π_n is one of the $m!$ symbol permutations having m distinct symbols and $n = 1, 2, \dots, k$, $k \leq m!$ ($m!$ is the largest number of distinct symbols). If we suppose that $P(\pi_1), P(\pi_2), \dots, P(\pi_k)$ denote the probability distribution of each symbol sequences respectively, and $\sum_{n=1}^k P(\pi_n) = 1$.

For each permutation π_n , the relative probability distribution can be determined by:

$$P(\pi_n) = \frac{\text{Number}\{\mathbf{X}_i \text{ has type } \pi_n \mid 1 \leq i \leq N - (m - 1)\lambda\}}{N - (m - 1)\lambda} \quad (4)$$

Then, the normalized permutation entropy of order m is defined as [23]:

$$PE = -\ln(m!)^{-1} \sum_{j=1}^{m!} P(\pi_j) \ln(P(\pi_j)) \quad (5)$$

It is notable that the PE value ranges from 0 to 1. The smaller the PE value is, the more regular the time series is.

B. MPE and CMPE

The MPE algorithm was developed by Aziz and Arif [16] in 2005 based on the coarse-grained procedure and PE. Given a time series $\{x(i), i = 1, 2, 3, \dots, N\}$ with data length N and the scale factor τ , construct consecutive coarse-grained time series, $\mathbf{y}^{(\tau)}$. Each element of $\mathbf{y}^{(\tau)}$ is calculated by averaging a successively increasing number of data points in non-overlapping windows at scale factor τ as follows:

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x(i), \quad 1 \leq j \leq \frac{N}{\tau}, \tau \leq N \quad (6)$$

Then, the MPE can be obtained by calculating PE via various coarse-grained time series as

$$MPE(X, \tau, m, \lambda) = PE(\mathbf{y}^{(\tau)}, m, r) \quad (7)$$

It can be observed that the coarse-grained procedure is similar to a sub-sampling operation; hence, it may yield imprecise values because of the increasingly reduced data length in coarse-grained time series. Moreover, in the MPE, a time series is divided into equal non-overlapping fragments, as a result of which some potentially useful information hidden in adjacent data points may be neglected.

To overcome the shortcomings above mentioned, in 2013, Wu et al. proposed the concept of Composite Multiscale Entropy (CMSE) to reduce the variance of estimated entropy values at high scales [18]. Later, improved MPE methods were proposed based on CMSE [19], [20], in which the improved coarse-grained procedure combines information of multiple coarse-grained time series at one scale factor. Herein, we call those improved MPE as CMPE since the CMSE is the first work to propose the composite-based coarse-graining procedure [18]. Similar to MPE, the CMPE method mainly includes two steps. Firstly, given a time series $\{x(i), i = 1, 2, 3, \dots, N\}$, its coarse-grained forms at multiple scales can be obtained. In the CMPE, the k th coarse-grained time series for a given scale factor τ , $\mathbf{y}_k^{(\tau)} = \{y_{k,1}^{(\tau)}, y_{k,2}^{(\tau)}, y_{k,3}^{(\tau)}, y_{k,p}^{(\tau)}\}$ is defined as

$$y_{k,j}^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+k}^{j\tau+k-1} x_i, \quad 1 \leq j \leq \frac{N}{\tau}, 1 \leq k \leq \tau \quad (8)$$

then, by averaging of all the k number of CMPE values at τ scale, the CMPE value with scale factor τ can be obtained as

$$CMPE(X, \tau, m, \lambda) = \frac{1}{\tau} \sum_{k=1}^{\tau} PE(\mathbf{y}_k^{(\tau)}, m, \lambda) \quad (9)$$

where τ is the scale factor for producing coarse-grained time series, m is the embedding dimension, and λ is the time delay in the calculation of the PE measure. As shown in Fig. 2, two coarse-grained time series are separated from the original time series at the 2nd scale factor in the CMPE algorithm.

Scale 2

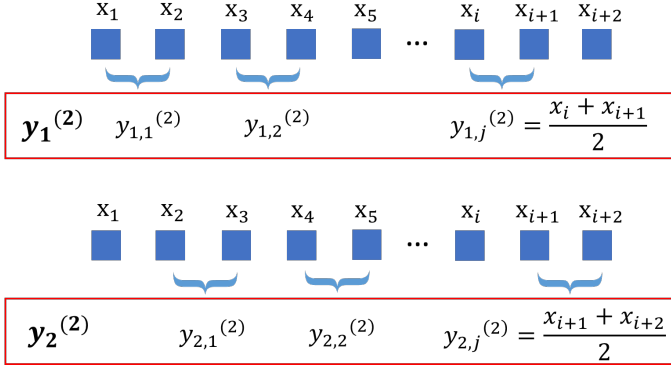


FIGURE 2. Illustration of the coarse-graining procedure at the 2nd scale factor in the CMPE algorithm where only low-frequency components are considered.

III. FINE-TO-COARSE MULTISCALE PERMUTATION ENTROPY (F2CMPE)

This section first briefly reviews the WPD analysis and then introduces the proposed F2CMPE algorithm. Based on the reconstructed wavelet coefficients using WPD, F2C signals are generated by consecutively removing high-frequency components, and then PE measure is used to estimate the complexity of the F2C signals for the calculation of F2CMPE measure.

A. Wavelet Packet Decomposition (WPD)

Wavelet analysis has been widely applied to analyze non-stationary signals and enjoys the advantages of offering good time and frequency resolutions [24]–[26]. WPD can provide both approximate and detail wavelet decomposition coefficients with same frequency bandwidths by successively using wavelet filtering operations to each decomposition level until desired frequency resolution is achieved. The principle of WPD can be briefly described as follows [27]. WPD can be implemented by means of a pair of low-pass and high-pass wavelet filters, denoted as $h(k)$ and $g(k) = (-1)^k h(1-k)$. The decomposition of a signal $x(t)$ is described as [28]:

$$\begin{cases} C_{j+1,2n} = \sum_l h(l-2k)C_{j,n} \\ C_{j+1,2n+1} = \sum_l g(l-2k)C_{j,n} \end{cases} \quad (10)$$

where $C_{j,n}$ denotes the wavelet coefficient at the j -th decomposition level, the n -th sub-band, and l is the number of the wavelet coefficients. Thereby, the input signal can be decomposed into a series of wavelet decomposition coefficients including approximation coefficients with low-frequency information and detail coefficients with high-frequency information.

To prove the superiority of WPD that could improve the efficiency of multiple-scale entropy analysis, an example was given to demonstrate the advantage of WPD when applying for producing coarse-grained and fine-grained time series. Fig. 3 presents a procedure of WPD transformation using

Level 1

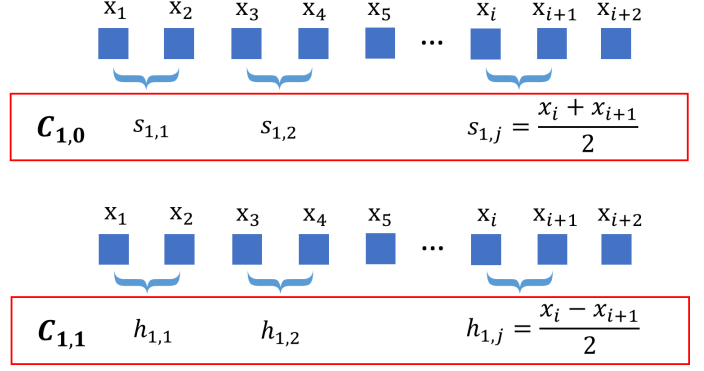


FIGURE 3. Illustration of WPD on the first level based on Haar wavelet ($C_{1,0}$ and $C_{1,1}$ are, respectively, the approximation coefficients and detail coefficients after use of low-pass and high-pass filters).

Haar wavelet on the first level; it also shows the principal distinction between the WPD analysis and composite-based coarse-grained procedure in CMPE (presented in Fig. 2) when applying to extract frequency information from an original time series. The former considers both lower and higher frequency components into account; however, the latter ignores the high-frequency information, which merely considers the low-frequency components. Therefore, WPD makes full use of information hidden in both low- and high-frequency components. Based on obtained wavelet packet coefficients, the reconstruction procedure of the wavelet transform is defined as [29]:

$$C_{j,n} = \sum_l [h(k-2l)C_{j+1,2n}] + \sum_l [g(k-2l)C_{j+1,2n+1}] \quad (11)$$

where $h(k-2l)$ and $g(k-2l)$ denote the low-pass and high-pass wavelet reconstruction filters respectively. h is related to the scaling function and g is related to the wavelet function.

Correspondingly, given a wavelet packet tree at the j -th decomposition level, in total a set of 2^j wavelet packet coefficients, $\{C_{j,n}, 1 \leq n \leq 2^j\}$, can be obtained where n is the order of the coefficient in the j -th decomposition level. Then, based on each coefficient vector $C_{j,n}$, a reconstructed signal $R_{j,n}$, with the same length of the original signal, can be produced by setting the all the other decomposition coefficients on level j to zero and recursively implementing the wavelet reconstruction transform in the inverse procedure until j decreases to zero [30]. Among each reconstruction procedure, the wavelet decomposition coefficient has 1/2 data points by comparing with the upper level. Finally, for each $R_{j,n}$, it has an approximative frequency range with that of $C_{j,n}$ and remains the same length with the original signal. Therefore, given the sampling frequency F_s , the frequency

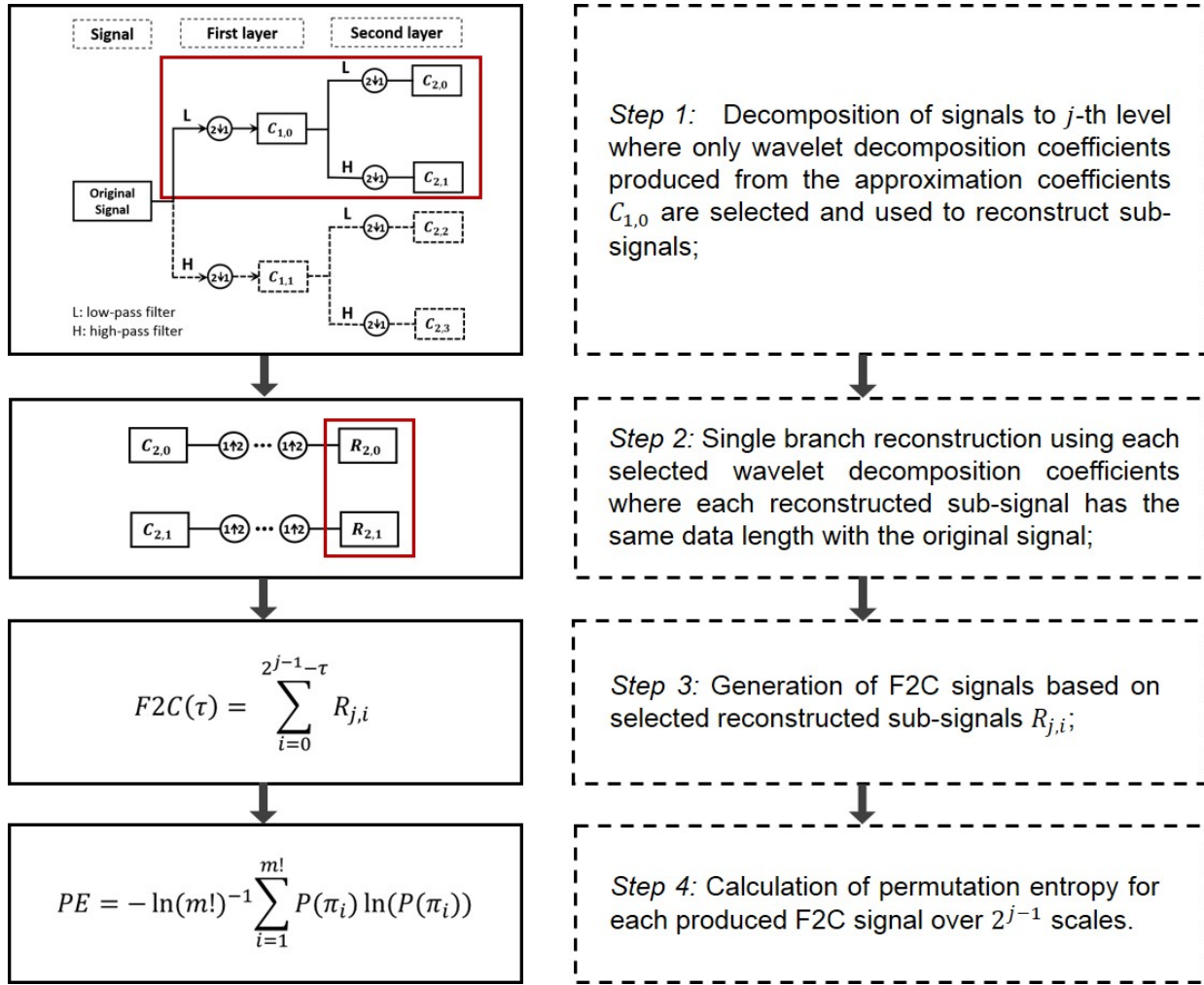


FIGURE 4. Flowchart of the proposed F2CMPE algorithm.

intervals of each $R_{j,n}$ can be approximately computed by:

$$((n-1) * 2^{-j} F_s, n * 2^{-j} F_s], \quad n = 1, 2, \dots, 2^j \quad (12)$$

Reconstructed signals equally partition the whole frequency spectrum of the signal and contain frequency information ranging from low to high. Furthermore, the reconstructed signals have the same data length as that of the original signal, which avoids the large variance caused by the decreased data length in the calculation of MPE and CMPE [7]. Thereby, the definition of F2CMPE is proposed on the basis of WPD transformation and reconstruction procedures.

B. F2CMPE Algorithm

The superiority of WPD allows decomposing non-stationary signals into wavelet coefficients with good time and frequency resolution. Also, the reconstruction procedure enables the inversion of each wavelet decomposition coefficient to a reconstructed sub-signal that remains the same length with the original signal. Owing to the advantages of WPD analysis, the F2C signals are produced by constructing reconstructed sub-signals with a fine-grained to coarse-grained manner [31].

More specifically, in the F2C procedure, when the scale factor increases, the high-frequency information is consecutively removed from previously acquired F2C signals at lower scales. Hence, given F2C signals with increasing scales, high-frequency and low-frequency information is consecutively refined and obtained from the original signal through the F2C procedure, which can contribute to appropriately characterize the dynamic changes associated with fault symptoms in vibration signals. The flowchart of our proposed F2CMPE method is illustrated in Fig. 4. The F2CMPE algorithm is described as follows:

- 1) Apply WPD to decompose an original signal to j -th decomposition level where only wavelet decomposition coefficients produced from the branch of $C_{1,0}$ on the 1-th level are selected and used. Thereby, there are 2^{j-1} sets of wavelet decomposition coefficients $\{C_{j,n}, (0 \leq n \leq 2^{j-1} - 1)\}$ are obtained and used in the next step;
- 2) Reconstruct single branch using each acquired wavelet decomposition coefficients on the j -th level, by setting the coefficients of all the other vectors on level j to zero and recursively implementing the wavelet reconstruction

transform in the inverse procedure until j decreases to zero. Thus, each reconstructed sub-signal has the same data length as the original signal. Therefore, totally 2^{j-1} reconstructed signals, $\{R_{j,n}, (0 \leq n \leq 2^{j-1} - 1)\}$, can be produced using the wavelet reconstruction procedure;

- 3) F2C procedure: construct F2C signals by consecutively removing one reconstructed signal from previously obtained F2C signals, commencing from the accumulation of all 2^{j-1} reconstructed signals. Thereby, F2C signals are produced as

$$F2C(\tau) = \sum_{i=0}^{2^{j-1}-\tau} R_{j,i}, 0 \leq i \leq 2^{j-1} - 1, 1 \leq \tau \leq 2^{j-1} \quad (13)$$

where j is the decomposition level, τ is the scale factor, and the maximum number of τ is equal to 2^{j-1} . Herein, the proposed F2C procedure refers to a process that produces sub-signals with fine-grained to coarse-grained time-frequency information refined from the original signal;

- 4) Perform PE to estimate the complexity of F2C signals over different scale factor τ , and we call these procedures as Fine-to-Coarse Multiscale Permutation Entropy (F2CMPE) analysis.

In summary, the main idea behinds the F2CMPE can be concluded as two stages: 1) generate the F2C signals representing fine-grained to coarse-grained time-frequency information obtained by consecutively removing the F2C signal having high-frequency information from previous obtained F2C signal based on the F2C procedure; 2) apply PE to estimate the irregularity and dynamic changes of the F2C signals over different scale factors. Herein, the frequency range of the F2C signals obtained from the original signal is gradually decreased since high-frequency components are consecutively removed from previously produced F2C signals. Hence, through the F2C procedure, low-frequency information finally remains in the F2C signals at high scales. With an increasing scale factor, dynamic changes hidden in lower-frequency components can be gradually characterized from the original signal. Besides, in the proposed F2CMPE algorithm, the only half frequency spectrum of the original signal is used for producing F2C signals. Indeed, similarly, in MSE, MPE, and CMPE, both of them only apply no more than half frequency spectrum of the original signal commencing from 2nd scale, because the coarse-grained procedure is similar to a sub-sampling as well as a linear smoothing method.

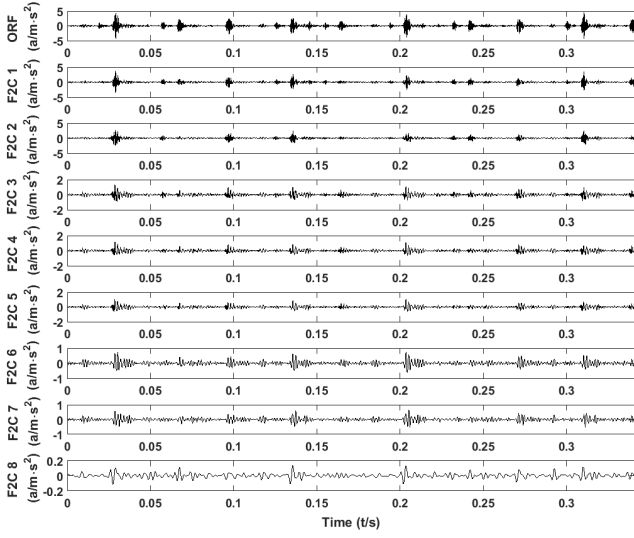
Previously published works have verified the effectiveness of the use of a half-frequency spectrum, which could provide sufficient and accurate information for discriminating different bearing conditions, such as in [6], [7], [11], [15]–[20]. Furthermore, vibration acquisition system usually has a high sampling frequency. Hence, very high-frequency components typically contain too much detail information that may be potentially considered as noises to some extent so that it has less useful information for extracting intrinsic fault symptoms. Moreover,

the emergence of incipient failures in rolling bearing components introduces impulse waves and finally results in the occurrence of coupling frequency in both lower and higher frequency components due to periodical friction and strikes between faulty and healthy components. Therefore, capturing the dynamic characteristics using the F2CMPE measure allows discriminating different bearing conditions based on the non-linear entropy analysis and pinpointing the root cause of bearing defects.

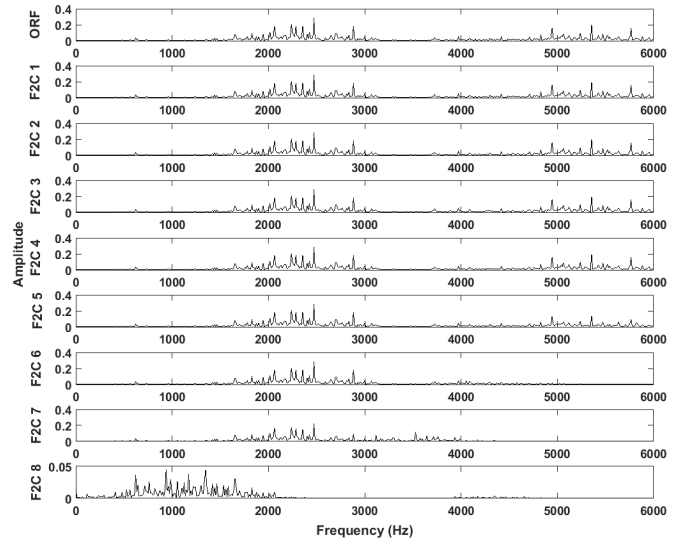
Comparatively, the use of the coarse-grained procedure has some shortcomings in MPE and CMPE. Firstly, the fundamental idea behind MPE is essentially a linear smoothing procedure. As a result, both MPE and CMPE lack of high-frequency information obtained from the original time series. That is, useful high-frequency components are proactively discarded, which may not provide relatively stable and precise entropy values. Furthermore, the data length of the coarse-grained time series decreases over increasing scale factors. But, the calculation of PE greatly depends on the data length. For instance, short time series may inevitably yield imprecise estimation of entropy values [6]. The proposed F2CMPE method could alleviate those problems.

Theoretically, the proposed F2CMPE method not only provides low- and high-frequency information but also alleviates the influence of short time series for PE calculation. Thanks to the WPD analysis, it can give fine-grained resolution in time and frequency domains. By using orthogonal wavelet kernels, it allows generating wavelet coefficients by applying low- and high-pass filters rather than the use of the simple linear filter. Therefore, both low- and high-frequency information can be maintained in the F2C signals. Furthermore, by reconstructing coefficients to signals with the same data length as the original time series, it could avoid imprecise PE measure due to the decreased data length, especially when the length of original data is already too short. Given an example, a vibration signal of rolling bearing with Outer Race Fault (ORF) is analyzed by using the F2CMPE algorithm, and 8 F2C signals are obtained when a four-level WPD tree is used. The F2C signals and their corresponding frequency spectrums are presented in Fig. 5.

One can see from Fig. 5 (a) that the F2C signals with increasing scale factors are becoming more and more smooth and flat because high-frequency information representing detail changes has been consecutively removed from former F2C signals according to the theory of the F2C procedure. Therefore, only low-frequency information remains in the F2C signals at very high scales. Furthermore, Fig. 5 (b) indicates that the frequency range of each F2C signal is in line with the concept of the F2C procedure. In particular, both low- and high-frequency information is extracted from the original signal and exists in F2C signals at small scales. With an increasing scale, low-frequency information mainly composes the frequency spectrum because of the use of reconstruction signals produced by coefficients after the low-pass filter. Besides, the frequency of the F2C signal at the first scale factor is very similar to that of the original ORF signal, which indicates that the F2C signals can maintain significant



(a) Time domain waveforms of ORF and F2C signals



(b) Frequency spectrum of ORF and F2C signals

FIGURE 5. Time domain waveforms and frequency spectrum of the original ORF signal and F2C signals using the F2CMPE. Herein, ORF stands for the original bearing vibration signal with outer race fault, and the left waveforms are eight F2C signals.

frequency information in the original signal to the most extent. Therefore, the F2CMPE method can be used to characterize the complexity and dynamic changes for extracting fault symptoms hidden in bearing vibration signals.

C. Parameter Selection of F2CMPE

The appropriate use of mother wavelet function and decomposition level can significantly improve the performance of generating F2C signals based on signal decomposition and reconstruction in WPD analysis. More specifically, desired frequency resolution in F2C signals can be achieved by using a suitable wavelet kernel and decomposition level, where the latter determines the frequency band in each F2C signal. Therefore, the selection of wavelet and decomposition level is vital for achieving optimum performance using the F2CMPE algorithm, especially for generating the F2C signals appropriately. In wavelet analysis, the performance of a mother wavelet is based on two major factors, namely the support size and the number of vanishing moments. More specifically, a mother wavelet containing a large number of vanishing moments and small support size can locate valuable information from the original signal with less redundant information [32]. Among various mother wavelets, the Daubechies and Symlet family of wavelets are well-known for their orthogonality and efficiency in filter implementation for the Mallat fast algorithm, which are considered as desired wavelet functions in this study. Besides, the Relative Wavelet Energy (RWE) method has been applied to compare and select the appropriate mother wavelet [33], [34]. The RWE can provide information regarding relative energy distribution in transformed signals, which is also suitable for measuring the energy ratio in the F2C signals. An appropriate mother wavelet is supposed to

extract the most significant amount of energy because of the occurrence of failures which introduce large magnitude in a few wavelet coefficients. Hence, the wavelet kernel having the highest RWE value is considered as the optimum mother wavelet for generating F2C signals. Herein, the use of RWE method is described as follows. Given a sets of F2C signals $\{F2C_\tau, \tau = 1, 2, \dots, 2^{j-1}\}$, the energy of each F2C signal can be obtained by:

$$E(\tau) = \sum_{i=1}^N |F2C_{i,\tau}|^2, 1 \leq i \leq N, 1 \leq \tau \leq 2^{j-1} \quad (14)$$

where i is the index of the data point in each F2C signal, N is the data length of the F2C signal at the scale factor τ . Then, the total energy of F2C signals obtained in the j -th decomposition level can be obtained as

$$E_{sum} = \sum_{i=\tau}^{2^j} E(\tau), 1 \leq \tau \leq 2^{j-1} \quad (15)$$

Finally, the normalized value represents the relative energy of each F2C signal among overall F2C signals:

$$RWE(\tau) = \frac{E(\tau)}{E_{sum}}, 1 \leq \tau \leq 2^{j-1} \quad (16)$$

where $\sum_{\tau=1}^{2^{j-1}} RWE(\tau) = 1$, and the energy probability distribution $RWE(\tau)$ is considered as a time-scale density. Besides, the variance of one indicator quantifies to what extent the indicator varies and fluctuates. Normally, the high variance index also means that there are extra dynamic changes and possibly additional information existed in this indicator. Hence, in this study, the variance of RWE is also applied

TABLE I. Description of wavelet functions and their maximum RWE and average variance values (RWE: Relative Wavelet Energy).

Wavelet name	Maximum Relative Wavelet Energy	Variance
Daubechies2	0.110795	0.030126
Daubechies4	0.113272	0.030434
Daubechies6	0.113340	0.030224
Daubechies8	0.112933	0.030026
symlet2	0.110795	0.030126
symlet4	0.113234	0.030539
symlet6	0.113246	0.030303
symlet8	0.112933	0.030026

to evaluate optional mother wavelet functions. The larger variance value of REW is, the greater possibility of extracting useful information associated with fault symptoms from non-stationary signals [35]. In this study, the RWE values and their corresponding variance are both applied to evaluate four Daubechies (“db2”, “db4”, “db6”, “db8”) and four Symlet (“sym2”, “sym4”, “sym6”, “sym8”) wavelets respectively to select the optimum one for generating F2C signals. Additionally, the wavelet decomposition level determines the range of sub-frequency band in wavelet coefficients as well as the reconstructed signals. The larger decomposition level, the higher frequency resolution in each sub-band can be obtained. Nevertheless, a very high decomposition level will require more computational time and computing resources. By taking the effectiveness of signal decomposition and reconstruction based on WPD into consideration, a six-level wavelet tree is used (the decomposition level $j = 6$) in the F2CMPE algorithm. Subsequently, 32 F2C signals and a set of F2CMPE features over 32 scale factors are produced correspondingly.

To select the appropriate mother wavelet, the vibration signals of rolling bearing with ten conditions are randomly chosen from Case Western Reserve University (CWRU) Data Center [36]. Eight number of different mother wavelet kernels are then applied to construct F2C signals based on the F2CMPE algorithm. In this study, the fifth-decomposition level is used, and 32 sets of F2C signals are therefore obtained correspondingly. This experiment is operated 100 times, and the average maximum RWE values and their average variances are presented in Table I. It can be seen that “db4” and “db6” wavelet functions outperform the rest. Besides, the two indicators (namely RWE and its variance) of “db4” and “db6” wavelets are very similar, and “db4” is finally selected as the desired mother wavelet in this study for the F2CMPE feature extraction.

Additionally, the calculation of PE also greatly affects the effectiveness of the F2CMPE feature extraction. To provide reliable PE measurements, the selection of the embedding

dimension m and the time delay λ are necessary. Practically, when $m < 4$ it cannot detect the dynamic change of the mechanical vibration signals. Besides, when $m > 8$, not only the reconstruction of phase space will homogenize vibration signals but also the calculation of PE is time-consuming; hence, it cannot truly reflect the small varying range. According to literatures [6], [37], it was recommended to select $m = 4 - 7$. Regarding the use of time delay, when $\lambda > 5$, it cannot detect a slight change in the time series. Comparatively, the effect of time delay λ has small effects on the calculation of PE [15], especially when $\lambda \leq 4$. Moreover, a very short time series cannot produce prominent statistical significance on PE values. Therefore, in this study, $m = 5$ and $\lambda = 3$ are specified for the calculation of PE values from the F2C signals by the F2CMPE measure as well as in the CMPE measure for a comparative study, and the data length of time series is set to $N = 4,096$.

IV. THE PROPOSED BEARING DIAGNOSIS METHOD

This section describes the proposed method for fault diagnosis of rolling bearing based on the proposed F2CMPE, LS, and SVM methods, the procedure of which is presented in Fig. 1.

A. Laplacian Score (LS) for Feature Selection

The obtained F2CMPE features over all scales could be applied to discriminate bearing conditions for fault diagnosis purpose. Nevertheless, not all entropy measures are directly associated with incipient bearing defects. Furthermore, a vast number of feature vectors can be time-consuming due to the increase of computational complexity because of the high dimensions. Hence, it is necessary to select the most critical features to avoid the dimension disaster and enhance the fault identification efficiency. In this paper, LS is applied to choose salient features and construct the optimum feature vectors by reducing feature dimension. It has been verified for its efficiency in fault diagnosis of rolling bearing, such as in [15]. LS is a feature selection method used to evaluate the importance of features according to their power of locality preserving based on Laplacian Eigenmaps and Locality Preserving Projection [38].

Given m data samples and each sample has n features, suppose L_r represent the Laplacian score of the r -th feature, and $f_{r,i}$ represent the i -th sample of the r -th feature ($r = 1, 2, \dots, n$ and $i = 1, 2, \dots, m$). The main calculation of LS algorithm can be described as follows:

- (1) Construct a nearest neighbor graph \mathbf{G} with m nodes, where the i -th node corresponds to the i -th data point x_i . Then an edge is put between i -th and j -th node, if x_i and x_j are “close”, which is defined as x_i is among k nearest neighbors of x_j , or x_j is among k nearest neighbors of x_i where k is practically set to 5. When the label information is available, one can put an edge between two nodes sharing the same label.
- (2) Define weight matrix \mathbf{S} of the models as

$$S_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{t}} & \text{if nodes } i \text{ and } j \text{ are connected} \\ 0 & \text{otherwise.} \end{cases} \quad (17)$$

where t is a suitable constant.

(3) For the r -th feature, suppose \mathbf{f}_r as

$$\mathbf{f}_r = [f_{r1}, f_{r2}, \dots, f_{rm}]^T, \mathbf{D} = \text{diag}(\mathbf{S}\mathbf{I}), \quad (18)$$

$$\mathbf{I} = [1, 1, \dots, 1]^T, \mathbf{L} = \mathbf{D} - \mathbf{S}.$$

where \mathbf{I} is the unit vector with dimension m and the matrix \mathbf{L} is often called graph Laplacian. Let

$$\tilde{\mathbf{f}}_r = \mathbf{f}_r - \frac{\mathbf{f}_r^T \mathbf{D} \mathbf{I}}{\mathbf{I}^T \mathbf{D} \mathbf{I}} \mathbf{I} \quad (19)$$

(4) Compute the Laplacian Score of the r -th feature as follows:

$$L_r = \frac{\sum_{ij} (f_{ri} - f_{rj})^2 S_{ij}}{\text{Var}(\mathbf{f}_r)} = \frac{\tilde{\mathbf{f}}_r^T \mathbf{L} \tilde{\mathbf{f}}_r}{\tilde{\mathbf{f}}_r^T \mathbf{D} \tilde{\mathbf{f}}_r} \quad (20)$$

where $\text{Var}(\mathbf{f}_r)$ is used to estimate variance of the r -th feature. The larger S_{ij} is, the smaller $(f_{ri} - f_{rj})$ is; therefore the more important the feature is, the smaller LS value of the feature obtains. In this study, the use of LS helps to select salient entropy features, which are then applied to construct fault feature vectors according to their LSs from low to high. Specifically, entropy features with lower scores are considered as critical features for representing fault symptoms extracted from different vibration signals of rolling bearing [15], [38]. After that, the optimum feature vectors constructed using F2CMPE features and LS can be fed into multi-class SVM for fault identification, which is used to classify health conditions of rolling bearing.

B. The proposed rolling bearing fault diagnosis method

The procedure of the proposed rolling bearing fault diagnosis method is described as follows:

- 1) Vibration signals are collected from rolling bearings with different conditions using an acceleration sensor with a specified sampling frequency.
- 2) The F2CMPE algorithm is then applied to extract fault features from measured vibration signals. In this study, a six-level wavelet tree is used in the F2MPE, and 32 F2CMPE features are produced. In the calculation of PE, the embedding dimension m is set to 5 and time delay λ is set to 3 respectively.
- 3) The initial fault feature vectors constructed using 32 F2CMPE features are then divided into training and testing data sets respectively. After that, LS is applied to refine and rank training and testing elements of features according to their LS values from low to high.
- 4) The first several salient features in each feature vector after LS analysis are selected to construct new training and testing fault feature vectors respectively.
- 5) The obtained optimum feature vectors are then fed into multi-class SVM for fault identification and classification.

V. COMPARATIVE ANALYSIS OF F2CMPE AND CMPE BASED ON SYNTHETIC SIGNALS

In this study, simulation analysis is first carried out to investigate the performance of the proposed F2CMPE and conventional CMPE for analyzing synthetic signals containing different noise levels and SNRs.

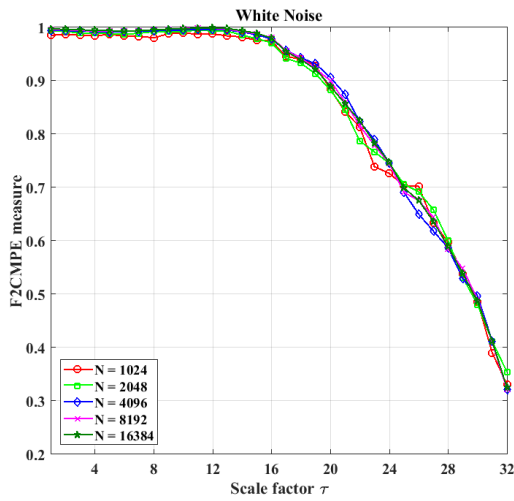
A. Analysis of Gaussian White Noise and 1/f Noise

Gaussian white noises and 1/f noises with different lengths ($N = 1024, 2048, 4096, 8192, \text{ and } 16384$) are considered in this study. Both of the F2CMPE and CMPE are applied to estimate entropy values from noise signals, and the numerical results are presented in Fig. 6. From Figs 6 (a) and (b), it can be noted that, with an increasing scale factor τ , the F2CMPE features obtained from the Gaussian white noise and 1/f noise signals gradually decrease from roughly 0.97 to 0.32. This phenomenon is in line with the theory of the F2CMPE algorithm, because both low and high-frequency information is extracted from the F2C signals at low scales. The waveforms of these F2C signals rapidly change and fluctuate over time than those of the smooth signals at high scales. Therefore, it potentially provides more permutation sequences and obtains large PE values from those F2C signals with low scales. Afterward, PE values start to descend with an increasing scale since the detail information is continuously removed from previously obtained F2C signals; thus, waveforms of F2C signals at low scales are becoming more and more smooth and regular. As a result, the higher the scale is, the smaller the PE value is.

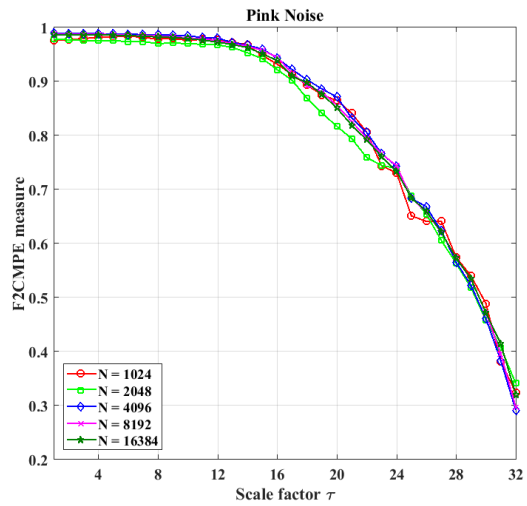
Additionally, in contrast with CMPE values obtained from white and 1/f noise signals, one can note that F2CMPE feature values show more consistency and stability with increasing scales. Comparatively, the CMPE features, however, relatively fluctuate over scale factors, especially when time series with decreased data length are analyzed. Indeed, the calculation of PE depends on the data length of the time series. Small data length can yield imprecise PE values and give an incorrect estimation of the dynamic changes of time series. From the comparative study, the simulation results indicate that the F2CMPE could provide consistent and stable values for estimating the complexity of time series compared to that of the CMPE.

B. Analysis of Signals with different SNRs

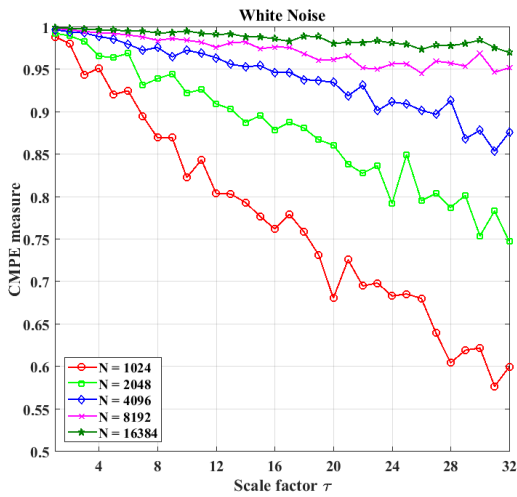
Mixed signals are analyzed to verify the distinguishability and consistence of the proposed F2CMPE measure. Gaussian white noise is added in a sinusoidal signal with a frequency 30 Hz, and the signals with SNRs: 10, 15, 20, 25, and 30 are considered in this analysis. Then, the F2CMPE and CMPE features of mixed signals are calculated, the results of which are presented in Fig. 7. From Fig. 7, it shows that, for both F2CMPE and CMPE features, entropy values of signals with small SNRs are greater than that of signals with large SNRs, because more white noises are added in the signals with a decreasing SNR. This indicates that F2CMPE and CMPE verify the case. Moreover, Fig. 7 (a) shows that the F2CMPE



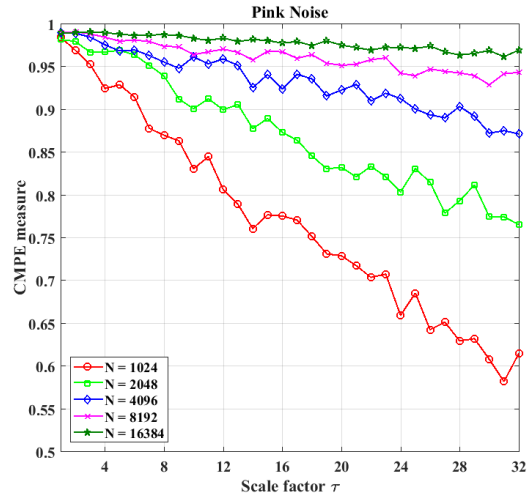
(a) F2CMPE measurements of Gaussian white noise



(b) F2CMPE measurements of 1/f noise

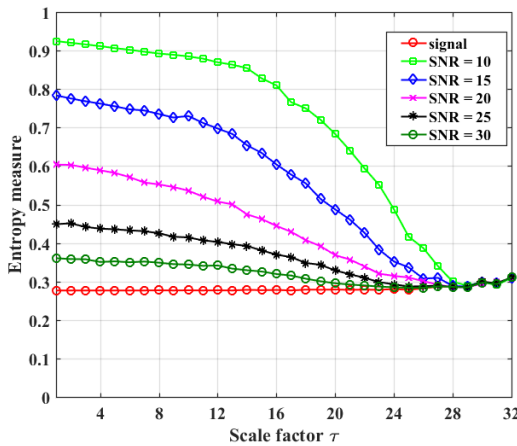


(c) CMPE measurements of Gaussian white noise

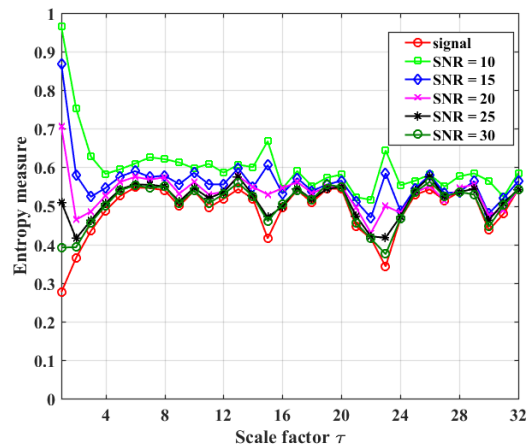


(d) CMPE measurements of 1/f noise

FIGURE 6. F2CMPE and CMPE features of Gaussian white noise and 1/f noise with different data length.



(a) F2CMPE measurements over 32 scales



(b) CMPE measurements over 32 scales

FIGURE 7. F2CMPE and CMPE features of synthetic signals with different SNRs.

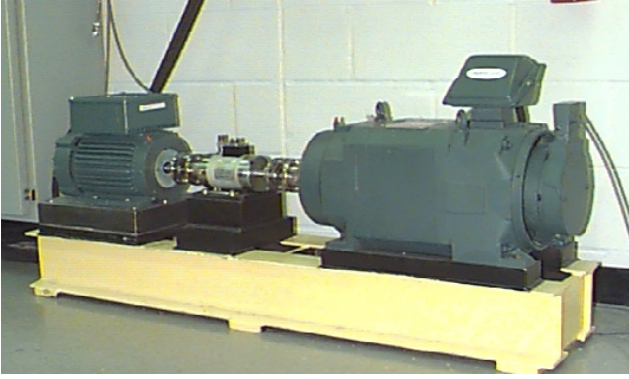


FIGURE 8. CWRU bearing testbed [36].

features of the original signal keep steady over 32 scales, and other signals with white noises gradually decrease until scales are very high. This is because high-frequency information, namely potential detail dynamic changes, is consecutively removed from F2C signals with increasing scales. Especially, F2CMPE features extracted from signals at very high scales cannot be discriminated from each other as their waveforms are very smooth at very high scales. Comparatively, Fig. 7 (b) shows that the CMPE features relatively fluctuate over 32 scales, and most CMPE feature curves mix with each other so that it is comparatively more difficult to distinguish between those signals.

Therefore, the analysis results indicate that the F2CMPE outperforms the CMPE in discriminating different conditions of complex signals, which can efficiently characterize the complexity of signals with different white noises, by extracting low and high-frequency information from the F2C signals. Besides, the F2CMPE feature values have more consistency and robustness than that of the CMPE when different white noises are added to the original signal. Especially, the curves of F2CMPE measure, in signals with small noises (e.g., SNR = 25 or 30), have the similar trend with that of the original sinusoidal signal, which indicates that the F2CMPE measure is robust to small noises and can effectively represent intrinsic complexity characteristics in non-stationary signals.

VI. EXPERIMENTAL DATA ANALYSIS

This section presents an experimental study to investigate the performance of the F2CMPE for fault diagnosis of rolling bearing by analyzing vibration signals. The experiment setup is introduced, then results and discussions are presented and described respectively.

A. Experiment Setup

Experimental validation is carried out to verify the feasibility of the proposed method for bearing fault diagnosis. Bearing data sets are provided by the CWRU Bearing Data Center [36]. In the experiments, bearing data, having ten classes of states and three kinds of fault severity levels, were collected from the drive end channel. This experiment system is shown in Fig. 8, which includes a 3HP motor, a torque

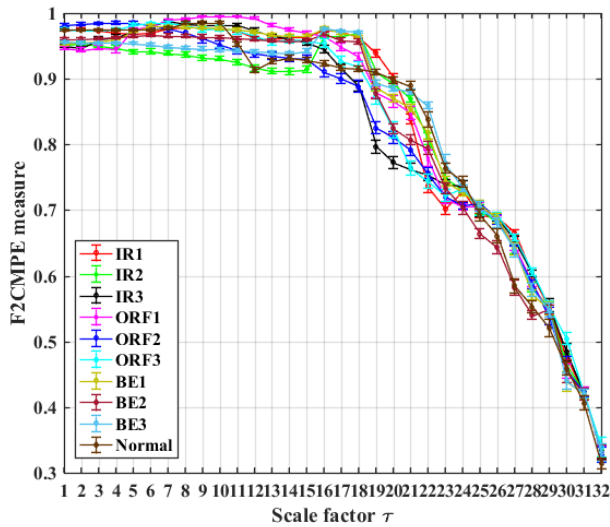
TABLE II. Description of each bearing condition and its class label.

States	Fault diameter	Class label	Number of training samples	Number of testing samples
IR1	0.1778	1	14	15
IR2	0.3556	2	14	15
IR3	0.5334	3	14	15
OR1	0.1778	4	14	15
OR2	0.3556	5	14	15
OR3	0.5334	6	14	15
BE1	0.1778	7	14	15
BE2	0.3556	8	14	15
BE3	0.5334	9	14	15
Norm	0	10	14	15

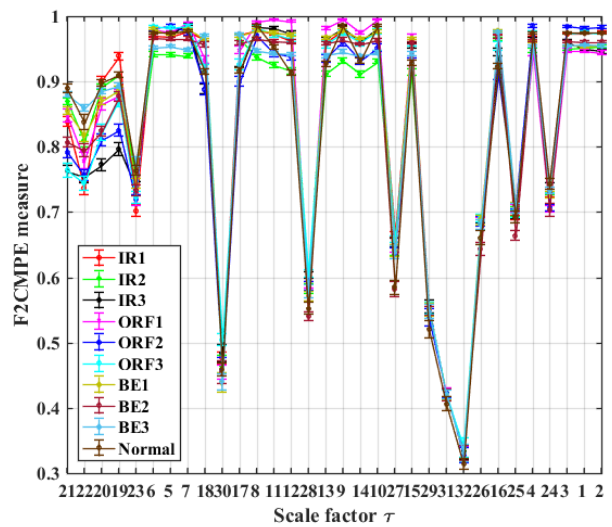
transducer, a dynamometer, control electronics, and the 6205-2RS JEM SKF deep groove ball bearing. Single point failures were introduced into SKF bearings using electro-discharge machining with local fault diameters of 0.1778 mm, 0.3556 mm, and 0.5334 mm and fault depth of 0.2794 mm. In this experiment, vibration data of rolling bearing include ten conditions, i.e., the normal condition (Norm), and the damages on the inner race (IR), the outer race (OR) at 6 o'clock, and the ball element (BE) respectively. In this study, for all fault conditions, three defect sizes of point fault were considered 0.1778 mm, 0.3556 mm, and 0.5334mm under 1730 r/min with Load 3 HP, and 12 kHz sampling rate. These ten conditions are respectively labeled as Norm, IR1, IR2, IR3, OR1, OR2, OR3, BE1, BE2, and BE3. Then, these vibration signals were split into a set of non-overlapping segments with a specified data length ($N = 4,096$). The detail specification of each rolling bearing state is presented in Table II.

B. Experimental Results and Discussion

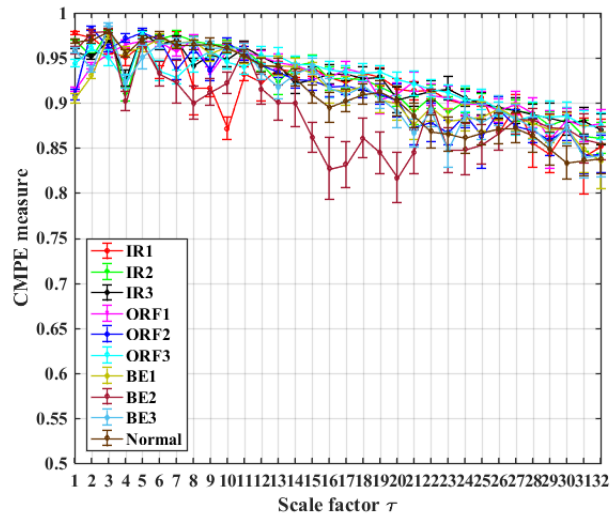
In this case, the performance of the proposed F2CMPE for rolling bearing fault diagnosis is validated and compared with that of the conventional CMPE, where LS and SVM are used as standard feature selection and fault classification methods to justify the comparison results. Firstly, after the data acquisition, there are in total 29 samples collected for each condition, and 290 samples are therefore obtained for ten conditions of rolling bearing. Each sample is a time series with 4,096 data points. Herein, 14 samples in each state are selected as training samples, and the left 15 samples are chosen as testing samples. Then, the F2CMPE and CMPE are applied to estimate the complexity of the vibration signals and produce 32 features to construct feature vectors for each sample. Therefore, for all 29 samples in each condition, the mean and standard deviation values of the F2CMPE and CMPE features representing each state are acquired, the results of which are shown in Figs 9 (a) and (c) respectively. One can find that both F2CMPE and CMPE feature values gradually decrease with an increasing



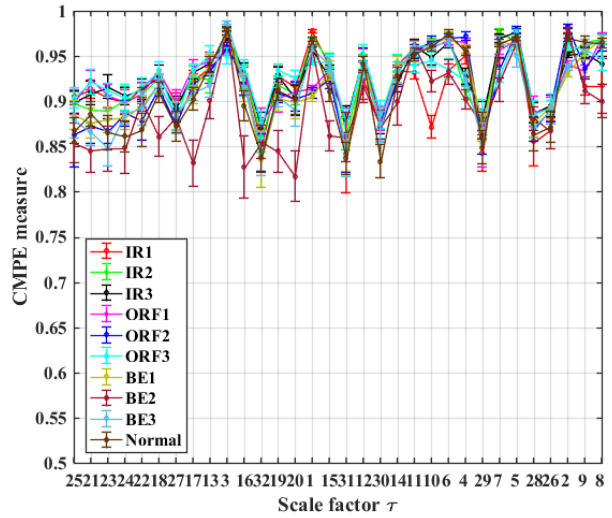
(a) Original F2CMPE features of vibration signals



(b) Ranked F2CMPE features of vibration signals after LS analysis



(c) Original CMPE features of vibration signals



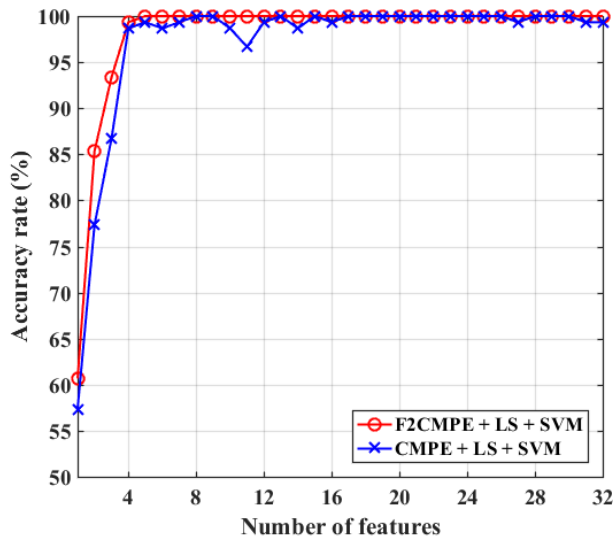
(d) Ranked CMPE features of vibration signals after LS analysis

FIGURE 9. Original and ranked F2CMPE and CMPE features of rolling bearing vibration signals over 32 scales based on LS analysis.

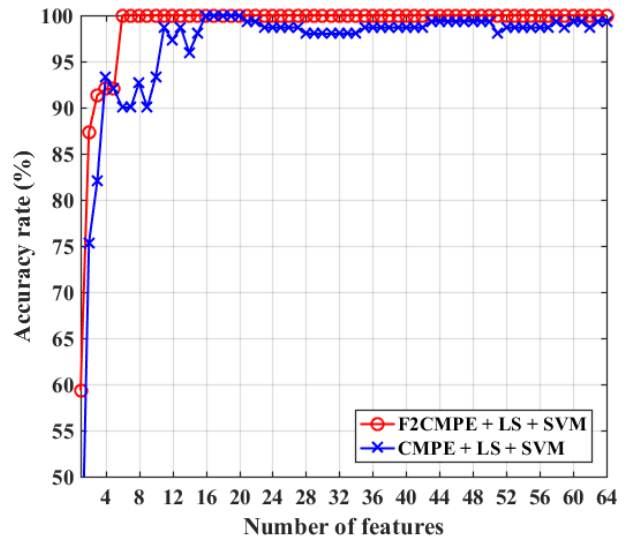
scale factor since signals are becoming smooth and regular at high scales. From Fig. 9 (a), it shows that the F2CMPE measures representing ten conditions of rolling bearing can be relatively distinguished by observing waveforms, especially from medium scales between 16 - 25. This indicates that significant differences can be extracted and identified in ten conditions using the F2CMPE. From Fig. 9 (c), it was found that CMPE features representing ten bearing conditions are very disordered and can be hardly discriminated by observing waveforms directly. Besides, Figs 9 (a) and (c) indicate that the standard deviation of the F2CMPE features at each scale is comparatively smaller than those of the CMPE features, which indicates that the calculation of F2CMPE is more consistent and reliable. It indicates that the F2CMPE may effectively extract the dynamic characteristics associated with fault symp-

toms hidden in different conditions of rolling bearing with different fault states.

After obtaining the F2CMPE and CMPE features from vibration signals, LS is then used to select and rank salient features for constructing all training and testing feature vectors, the results of which are presented in Figs 9 (b) and (d). Fig. 9 (b) shows that the first five ranked F2CMPE features greatly distinguish different bearing conditions to the most extent, the indexes of which are in line with those scales in Fig. 9 (a), which can distinguish differences between various bearing conditions. Comparatively, the first five CMPE features lie in the 21 - 25 scales as shown in Fig. 9 (b). Herein, the importance of entropy features selected using LS refers to the discriminatory capability of differentiating and distinguish between various bearing conditions. Besides,

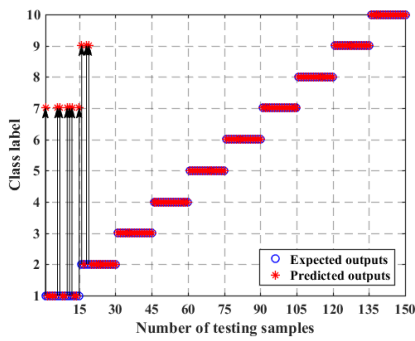


(a) Accuracy rate using increasing number of features from 1 to 32

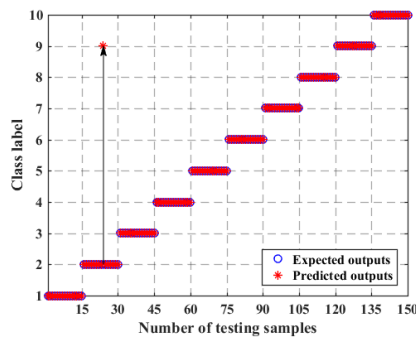


(b) Accuracy rate using increasing number of features from 1 to 64

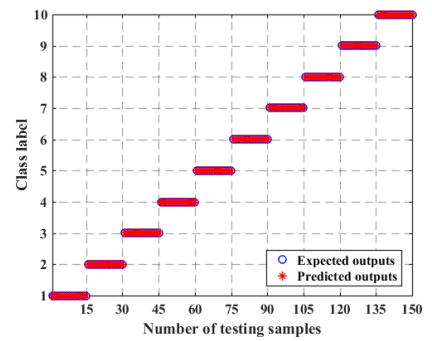
FIGURE 10. Comparative classification accuracy rate using the proposed F2CMPE and conventional CMPE methods based on LS and SVM over (a) 32 and (b) 64 number of entropy features respectively.



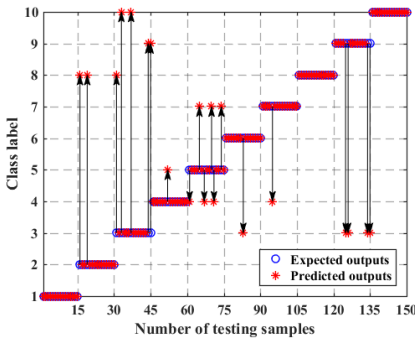
(a) Outputs by using the first 3 F2CMPEs



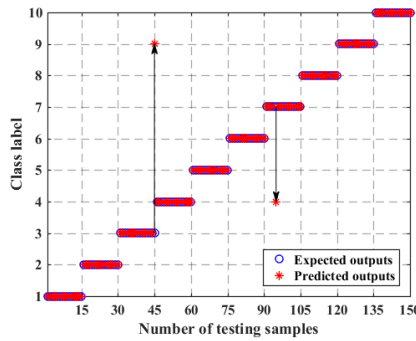
(b) Outputs by using the first 4 F2CMPEs



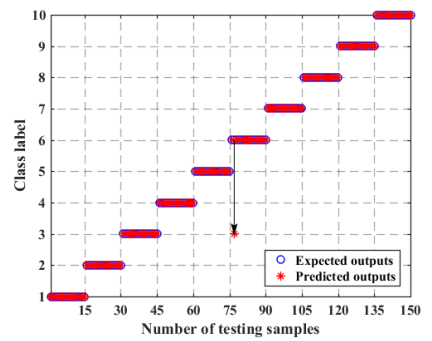
(c) Outputs by using the first 5 F2CMPEs



(d) Outputs by using the first 3 CMPEs



(e) Outputs by using the first 4 CMPEs



(f) Outputs by using the first 5 CMPEs

FIGURE 11. Misclassified outputs of testing samples using the first 3, 4 and 5 F2CMPE and CMPE features.

Fig.s 9 (b) and (d) indicate that, for both F2CMPE and CMPE features, entropy measurements at very low scales can hardly contribute to the improvement of discriminative performance since the redundant detail changes in high-frequency information may not accurately reflect the intrinsic characteristics in signals. Hence, after LS analysis, ranked entropy features in

front orders are applied to construct fault feature vectors for bearing fault classification using SVM classifier [41], [42].

Aiming at verifying the effectiveness of the proposed method for bearing fault classification, training and testing feature vectors obtained from signals are used. At first, the increasing number of features, ranging from 1 to 32, are

TABLE III. Comparative study between the proposed method and related published literature for rolling bearing fault diagnosis.

References	Year	Defects considered	Signal Processing method	Characteristic features	Classifier used	Number of classified states	Maximum classification accuracy (%)	Feature selection method
Wu et al. [37]	2012	ORF, IRF, and BF ¹	N/A	MPE of original signals	SVM	4	100%	N/A
Zhao et al. [39]	2015	IRF, ORF, and BF	WPD	MPE of wavelet coefficients	Hidden Markov model	4	100%	N/A
Zhang et al. [40]	2015	IRF, ORF, and BF	EEMD	PE of intrinsic mode functions decomposed by EEMD	SVM	2, 3, 4, 11	100%	N/A
Li et al. [20]	2017	IRF, ORF, and BF	N/A	Improved MPE of original signals	SVM	4	100%	LS
Zheng et al. [15]	2018	IRF, ORF, and BF	N/A	GCMPE of original signals	SVM	6	100%	LS
Present work	2018	IRF, ORF, and BF	F2C procedure based on WPD analysis	F2CMPE based on constructed F2C signals	SVM	10	100%	LS

¹ IRF: Inner Race Fault; ORF: Outer Race Fault; BF: Ball element Fault; EEMD: Ensemble Empirical Mode Decomposition; GCMPE: Generalized Composite Multiscale Permutation Entropy.

used to construct feature vectors. Given k number of features, the feature vectors of training samples can be obtained with dimension $140 \times k$, and the feature vectors of testing samples are $150 \times k$. Later, the different number of feature vectors are respectively fed into the SVM classifier for performance comparison. That is, 32 groups of experiments are carried out using feature vectors that consist of a different number of F2CMPE and CMPE features in each vector varying from 1 to 32. In our study, the RBF kernel is applied in SVM, and a grid search method is applied to locate the optimum cost parameter c and the width parameter g in the training stage. Herein, c and g are respectively set between 2^{-10} to 2^{10} . A 10-fold cross-validation method is used for the validation of the proposed bearing fault diagnosis approach. In a k -fold cross-validation method, the data sets are divided into k subsets, and the holdout method is repeated k times. After that, the average error for all k trials can be obtained. The fault identification accuracy rate based on the F2CMPE and CMPE using an increasing number of features are presented in Fig. 10 (a). As can be seen, the classification accuracy rate continuously increases with the use of an increasing number of entropy features. Comparatively, the proposed method based on the F2CMPE can achieve 100% accuracy rate using only the first five features in contrast with that based on the CMPE for fault identification. Moreover, the proposed method can continuously maintain high and stable identification accuracy for bearing diagnosis under ten conditions.

The indexes of the first eight features using the F2CMPE method with 32 scales are 21, 22, 20, 19, 24, 23, 6, and 5 according to the LSs from low to high. The misclassified testing samples are shown in Fig. 11 (a) - (c) in which the first 3, 4, and 5 selected F2CMPE features are respectively applied.

Besides, the proposed method can achieve 99% accuracy rate with only the first four features. Hence, the analysis results indicate that the proposed method using the F2CMPE, LS and SVM can efficiently identify bearing faults under ten conditions. Comparatively, the method based on the CMPE, LS and SVM cannot continuously offer a stable accuracy rate, when an increasing number of CMPE features are applied, the values of which fluctuate between 86.67% and 100%. The misclassified testing samples using the first 3, 4, and 5 selected CMPE features are presented in Fig. 11 (d) - (f) respectively.

Additionally, the effectiveness of the F2CMPE and CMPE features over 64 scales are investigated using the different number of selected features using LS and SVM. The results of the classification accuracy rate are shown in Fig. 10 (b), which presents that the F2CMPE accuracy curve gradually increases from 60% to 100% with the use of an increasing number of features. Besides, it achieves a 100% accuracy rate when the first eight features are used in this case. In contrast, the CMPE accuracy curve initially rises to 93% and then goes up and down when less than 16 CMPE features are applied, after which it achieves a 100% accuracy rate. Nevertheless, it cannot continuously guarantee a high and stable accuracy rate which also fluctuates between 98.67% and 100%. Generally, with the use of an increasing number of features, both F2CMPE and CMPE can achieve a reasonable classification accuracy. Evidently, the diagnosis method based on the F2CMPE can provide a more reliable and stable fault classification accuracy than that based on the CMPE, when ten conditions of rolling bearings are considered. To sum up, experimental validation has demonstrated that the proposed method based on the F2CMPE, LS and SVM can efficiently discriminate different states of rolling bearing and identify

bearing failures by offering reliable and stable bearing fault pattern classification accuracy.

Finally, to illustrate the potential effectiveness of the proposed methodology for bearing fault diagnosis, a comparison study between the presented work and published literature is summarized in Table III. The comparative items include the year of publication, defects considered, signal processing method, characteristic features, classifier used, number of classified states, maximum classification accuracy and feature selection method.

From Table III, one can note that PE has been widely studied under a multiple-scale framework for rolling bearing fault diagnosis. Though some methodologies can achieve 100% classification accuracy using the different number of bearing conditions, their improved methods are mostly developed based on the definition of MPE. However, as has been discussed in this study, both MPE and CMPE neglect the high-frequency components; thereby, temporal time series of decreased data length may not yield entropy estimation appropriately. In this study, we are attempting to overcome these limitations by proposed the F2CMPE measure. Furthermore, it can be seen that LS and SVM methods have been widely applied, respectively, for fault characterization and classification in the field of fault diagnosis. In addition, it should be noticed that our proposed method is different from the method presented in Zhao et al. [39], in which average MPE values of wavelet coefficients are directly obtained and used as feature vectors based on wavelet coefficients. In contrast, in this paper, the F2C procedure is proposed to construct F2C signals which contain varying low- and high-frequency information based on WPD transformation and reconstruction analysis. Besides, the LS feature selection method is applied to investigate the performance of using a different number of entropy features. Furthermore, experimental results have demonstrated that the proposed method can achieve a 100% identification accuracy rate even when ten conditions of rolling bearing are applied for experimental validation.

VII. CONCLUSIONS

In this paper, a new rolling bearing fault diagnosis method is proposed based on the F2CMPE, LS, and SVM. Focusing on non-linear and non-stationary characteristics of bearing vibration signals, a new entropy measure, named F2CMPE, is presented to estimate the complexity and dynamic changes of time series. The selection of parameters in the calculation of the F2CMPE is introduced and discussed (i.e., wavelet functions and decomposition level in WPD analysis). A comparative performance study was carried out to investigate the F2CMPE and CMPE features for analyzing synthetic signals. Results indicated that the F2CMPE method could yield more consistent and robust entropy values compared with those of the CMPE method. Meanwhile, the effectiveness of the proposed bearing fault diagnosis method is verified by analyzing rolling bearing vibration signals, which represent different fault states and fault severity levels. Experimental results have demonstrated the efficacy of the proposed method

and the superiority of the newly introduced F2CMPE measure for fault detection and diagnosis of rolling bearing. For future work, further study will be on the investigation of the proposed bearing fault diagnosis method when applying for analyzing vibration and acoustic signals measured from a real industrial-scale rotary machine. Meanwhile, the focus will be on developing self-adaptive approaches using an improved multiple-scale entropy measure for detecting performance deterioration in rotating machinery.

ACKNOWLEDGEMENT

This work is partially supported by International and Hong Kong, Macao & Taiwan collaborative innovation platform and major international cooperation projects of colleges in Guangdong Province (No.2015KGJHZ026), and The Natural Science Foundation of Guangdong Province (No.2016A030307029).

REFERENCES

- [1] D. Wang, K. Tsui, and Q. Miao, "Prognostics and health management: A review of vibration based bearing and gear health indicators," *IEEE Access*, vol. 6, pp. 665–676, 2018.
- [2] C. Li, D. Cabrera, J. de Oliveira, R. Sanchez, M. Cerrada, and G. Zurita, "Extracting repetitive transients for rotating machinery diagnosis using multiscale clustered grey infogram," *Mechanical Systems and Signal Processing*, vol. 76-77, pp. 157 – 173, 2016.
- [3] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mechanical Systems and Signal Processing*, vol. 104, pp. 799–834, 2018.
- [4] G. Jombo, Y. Zhang, J. D. Griffiths, and T. Latimer, "Automated gas turbine sensor fault diagnostics," in *ASME Turbo Expo 2018: Turbomachinery Technical Conference and Exposition*. American Society of Mechanical Engineers, 2018.
- [5] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3757–3767, 2015.
- [6] R. Yan, Y. Liu, and R. X. Gao, "Permutation entropy: a nonlinear statistical measure for status characterization of rotary machines," *Mechanical Systems and Signal Processing*, vol. 29, pp. 474–484, 2012.
- [7] A. Humeau-Heurtier, "The multiscale entropy algorithm and its variants: a review," *Entropy*, vol. 17, no. 5, pp. 3110–3123, 2015.
- [8] Y. Zhang, Z. Yang, H. Lu, X. Zhou, P. Phillips, Q. Liu, and S. Wang, "Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation," *IEEE Access*, vol. 4, pp. 8375–8385, 2016.
- [9] G. Jiang, H. He, J. Yan, and P. Xie, "Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox," *IEEE Transactions on Industrial Electronics*, 2018.
- [10] W. Deng, S. Zhang, H. Zhao, and X. Yang, "A novel fault diagnosis method based on integrating empirical wavelet transform and fuzzy entropy for motor bearing," *IEEE Access*, vol. 6, pp. 35 042–35 056, 2018.
- [11] Z. Huo, Y. Zhang, and L. Shu, "A short survey on fault diagnosis of rotating machinery using entropy techniques," in *International Conference on Industrial Networks and Intelligent Systems*. Springer, 2017, pp. 279–284.
- [12] R. Yan and R. X. Gao, "Approximate entropy as a diagnostic tool for machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 824–839, 2007.
- [13] W. Chen, J. Zhuang, W. Yu, and Z. Wang, "Measuring complexity using FuzzyEn, ApEn, and SampEn," *Medical Engineering and Physics*, vol. 31, no. 1, pp. 61–68, 2009.
- [14] C. Yi, Y. Lv, M. Ge, H. Xiao, and X. Yu, "Tensor singular spectrum decomposition algorithm based on permutation entropy for rolling bearing fault diagnosis," *Entropy*, vol. 19, no. 4, p. 139, 2017.

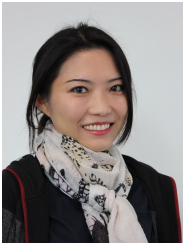
- [15] J. Zheng, H. Pan, S. Yang, and J. Cheng, "Generalized composite multiscale permutation entropy and Laplacian score based rolling bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 99, pp. 229–243, 2018.
- [16] W. Aziz and M. Arif, "Multiscale permutation entropy of physiological time series," in *9th International Multitopic Conference, IEEE INMIC 2005*. IEEE, 2005, pp. 1–6.
- [17] M. Costa, A. Goldberger, and C. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical review letters*, vol. 89, no. 6, p. 068102, 2002.
- [18] S. Wu, C. Wu, S. Lin, C. Wang, and K. Lee, "Time series analysis using composite multiscale entropy," *Entropy*, vol. 15, no. 3, pp. 1069–1084, 2013.
- [19] H. Azami and J. Escudero, "Improved multiscale permutation entropy for biomedical signal analysis: Interpretation and application to electroencephalogram recordings," *Biomedical Signal Processing and Control*, vol. 23, pp. 28–41, 2015.
- [20] Y. Li, W. Zhang, Q. Xiong, D. Luo, G. Mei, and T. Zhang, "A rolling bearing fault diagnosis strategy based on improved multiscale permutation entropy and least squares SVM," *Journal of Mechanical Science and Technology*, vol. 31, no. 6, pp. 2711–2722, 2017.
- [21] C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series," *Physical review letters*, vol. 88, no. 17, p. 174102, 2002.
- [22] M. Zanin, L. Zunino, O. A. Rosso, and D. Papo, "Permutation entropy and its main biomedical and econophysics applications: a review," *Entropy*, vol. 14, no. 8, pp. 1553–1577, 2012.
- [23] Y. Cao, W.-w. Tung, J. Gao, V. A. Protopopescu, and L. M. Hively, "Detecting dynamical changes in time series using the permutation entropy," *Physical Review E*, vol. 70, no. 4, p. 046217, 2004.
- [24] S. Wang, M. Yang, S. Du, J. Yang, B. Liu, J. M. Gorriz, J. Ramírez, T.-F. Yuan, and Y. Zhang, "Wavelet entropy and directed acyclic graph support vector machine for detection of patients with unilateral hearing loss in MRI scanning," *Frontiers in computational neuroscience*, vol. 10, p. 106, 2016.
- [25] Z. Huo, Y. Zhang, P. Francq, L. Shu, and J. Huang, "Incipient fault diagnosis of roller bearing using optimized wavelet transform based multi-speed vibration signatures," *IEEE Access*, vol. 5, pp. 19442–19456, 2017.
- [26] Y. Zhang, Z. Dong, S. Wang, G. Ji, and J. Yang, "Preclinical diagnosis of magnetic resonance (MR) brain images via discrete wavelet packet transform with Tsallis entropy and generalized eigenvalue proximal support vector machine (GEP-SVM)," *Entropy*, vol. 17, no. 4, pp. 1795–1813, 2015.
- [27] T. Ogden, *Essential wavelets for statistical applications and data analysis*. Springer Science & Business Media, 2012.
- [28] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: A review with applications," *Signal processing*, vol. 96, pp. 1–15, 2014.
- [29] G. Zhang, T. Yi, T. Zhang, and L. Cao, "A multiscale noise tuning stochastic resonance for fault diagnosis in rolling element bearings," *Chinese Journal of Physics*, vol. 56, no. 1, pp. 145–157, 2018.
- [30] L. Law, J. Kim, W. Liew, and S. Lee, "An approach based on wavelet packet decomposition and Hilbert-Huang transform (WPD-HHT) for spindle bearings condition monitoring," *Mechanical Systems and Signal Processing*, vol. 33, pp. 197–211, 2012.
- [31] Z. Huo, Y. Zhang, and L. Shu, "Fine-to-coarse multiscale permutation entropy for rolling bearing fault diagnosis." IEEE Conference on International Wireless Communications & Mobile Computing Conference (IWCMC 2018), 2018.
- [32] S. Mallat, *A wavelet tour of signal processing*. Elsevier, 1999.
- [33] R. Yan, "Base wavelet selection criteria for non-stationary vibration analysis in bearing health diagnosis," *Electronic Doctoral Dissertations for UMass Amherst*, Jan. 2007.
- [34] P. Kankar, S. Sharma, and S. Harsha, "Fault diagnosis of ball bearings using continuous wavelet transform," *Applied Soft Computing*, vol. 11, no. 2, pp. 2300–2312, 2011.
- [35] J. Rafiee and P. Tse, "Use of autocorrelation of wavelet coefficients for fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 23, no. 5, pp. 1554–1572, 2009.
- [36] "Case Western Reserve University Bearing Data Center," (Date last accessed 28-Dec.-2018). [Online]. Available: <http://csegroups.case.edu/bearingdatacenter/home>
- [37] S. Wu, P. Wu, C. Wu, J. Ding, and C. Wang, "Bearing fault diagnosis based on multiscale permutation entropy and support vector machine," *Entropy*, vol. 14, no. 8, pp. 1343–1356, 2012.
- [38] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," *Advances in Neural Information Processing Systems*, pp. 507–514, 2005.
- [39] L. Zhao, L. Wang, and R. Yan, "Rolling bearing fault diagnosis based on wavelet packet decomposition and multi-scale permutation entropy," *Entropy*, vol. 17, no. 9, pp. 6447–6461, 2015.
- [40] X. Zhang, Y. Liang, J. Zhou *et al.*, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164–179, 2015.
- [41] "LIBSVM Matlab Toolbox," (Date last accessed 28-Dec.-2018). [Online]. Available: <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [42] Y. Zhang, S. Wang, and G. Ji, "A comprehensive survey on particle swarm optimization algorithm and its applications," *Mathematical Problems in Engineering*, vol. 2015, 2015.

ABBREVIATIONS

λ	Time delay
τ	Scale factor
$C_{j,n}$	Wavelet decomposition coefficient
$g(k)$	High-pass filter
$h(k)$	Low-pass filter
j	Decomposition level
m	Embedding dimension
N	Data length of time series
$R_{j,n}$	Reconstructed sub-signal using each wavelet decomposition coefficient
ApEn	Approximate Entropy
CMPE	Composite Multiscale Permutation Entropy
CMSE	Composite Multiscale Entropy
EEMD	Ensemble Empirical Mode Decomposition
F2C	Fine-to-Coarse
F2CMPE	Fine-to-Coarse Multiscale Permutation Entropy
FuzzyEn	Fuzzy Entropy
GCMPE	Generalized Composite Multiscale Permutation Entropy
LS	Laplacian Score
MPE	Multiscale Permutation Entropy
MSE	Multiscale Entropy
PE	Permutation Entropy
RWE	Relative Wavelet Energy
SampEn	Sample Entropy
SNRs	Signal-to-Noise Ratios
SVM	Support Vector Machine
WPD	Wavelet Packet Decomposition



Zhiqiang Huo is currently working towards his Ph.D. degree at University of Lincoln, Lincoln, UK. He received his Ms. and BS. from China University of Geosciences Beijing, China in 2016 and 2013 respectively. His research interests lie in the field of fault diagnosis of industrial systems, wireless sensor networks, and participatory sensing. He had been awarded the INISCOM 2017 Best Paper Award. He has served as Co-Chairs in international conferences/workshops, such as AINIS 2015 and 2016, CollaborateCom 2017.



Yu Zhang is currently a Senior Lecturer in the School of Engineering, University of Lincoln, Lincoln, U.K. She obtained her BSc degree from School of Aerospace Engineering and Applied mechanics, Tongji University, Shanghai, China, in 2004. She has finished her MSc degree and PhD degree from the Department of Civil Engineering, University of Nottingham, Nottingham, U.K. in 2005 and 2011 separately. Her research interests include Equipment Fault Detection and Diagnosis, Grey-box System Modelling, and development of Data Analysis and

Machine Learning algorithms for industrial applications.



Lei Shu (M'07-SM'15) is a Distinguished Professor in Nanjing Agricultural University, China and a Lincoln Professor in University of Lincoln, UK. He is the director of NAU-Lincoln Joint Research Center of Intelligent Engineering. He is a member of EU Academy of Sciences. He received the B.Sc. degree in Computer Science from South Central University for Nationalities, China, in 2002 and the M.Sc. degree in Computer Engineering from Kyung Hee University, Korea, in 2005 and the Ph.D. degree in Digital Enterprise Research Institute, from National

University of Ireland, Galway, Ireland, in 2010. Until March 2012, he was a Specially Assigned Researcher in Department of Multimedia Engineering, Graduate School of Information Science and Technology, Osaka University, Japan. His main research field is Wireless Sensor Networks, Internet of Things. He has published over 380 papers in related conferences, journals, and books in the area of sensor networks. Currently, his H-index is 42 and i10-index is 153 in Google Scholar Citation. He had been awarded Globecom 2010, ICC 2013, ComManTel 2014, WICON 2016, and SigTelCom 2017 Best Paper Awards, 2017 and 2018 IEEE Systems Journal Best Paper Awards, the outstanding associate editor award of 2017 IEEE Access, 2014 Top Level Talents in "Sailing Plan" of Guangdong Province, China, 2015 Outstanding Young Professor of Guangdong Province, China. He has been serving as associate editors for IEEE Transactions on Industrial Informatics, IEEE Communications Magazine, IEEE Network Magazine, IEEE Systems Journal, IEEE Access, IEEE/CAA Journal of Automatic Sinica, Sensors, etc. He served as more than 50 various Co-Chair for international conferences/workshops, e.g., IWCMC, ICC, ISCC, ICNC, Chinacom, especially Symposium Co-Chair for IWCMC 2012, ICC 2012, General Co-Chair for Chinacom 2014, Qshine 2015, Collaboratecom 2017, DependSys 2018, SCI 2019, TPC Chair for InisCom 2015, NCCA 2015, WICON 2016, NCCA 2016, Chinacom 2017, InisCom 2017, WMNC 2017, NCCA 2018; TPC members of more than 150 conferences, e.g., ICDCS, DCOSS, MASS, ICC, Globecom, ICCCN, WCNC, ISCC.



Michael Gallimore is currently the Head of School of Engineering, University of Lincoln, Lincoln, U.K. He received the BEng(Hons) degree in Mechanical and Computer Aided Engineering in 2006 from Sheffield Hallam University, Sheffield, UK, and the PhD degree in Engineering in 2016 from the University of Lincoln. His research interests include intelligent diagnostics and prognostics, signal processing, optimization, and biomedical engineering.