IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 23, 2018, accepted December 5, 2018, date of publication December 12, 2018,
date of current version January 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2885816

Multi-Fault Rapid Diagnosis for Wind Turbine Gearbox Using Sparse Bayesian Extreme Learning Machine

JIAN-HUA ZHONG[1], JUN ZHANG[1], JIEJUNYI LIANG[2,3], AND HAIQING WANG[3]

[1]School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350108, China
[2]School of Automotive Engineering, Wuhan University of Technology, Wuhan 430070, China
[3]School of Mechanical and Mechatronics Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia

Corresponding author: Jiejunyi Liang (jiejunyiliang@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51875105 and Grant 51375013, and in part by the Youth Teacher Educational Research Fund of Fujian Provincial Education Office under Grant JAT170090.

ABSTRACT In order to reduce operation and maintenance costs, reliability, and quick response capability of multi-fault intelligent diagnosis for the wind turbine system are becoming more important. This paper proposes a rapid data-driven fault diagnostic method, which integrates data pre-processing and machine learning techniques. In terms of data pre-processing, fault features are extracted by using the proposed modified Hilbert–Huang transforms (HHT) and correlation techniques. Then, time domain analysis is conducted to make the feature more concise. A dimension vector will then be constructed by including the intrinsic mode function energy, time domain statistical features, and the maximum value of the HHT marginal spectrum. On the other hand, as the architecture and the learning algorithm of pairwise-coupled sparse Bayesian extreme learning machine (PC-SBELM) are more concise and effective, it could identify the single- and simultaneous-fault more quickly and precisely when compared with traditional identification techniques such as pairwise-coupled probabilistic neural networks (PC-PNN) and pairwise-coupled relevance vector machine (PC-RVM). In this case study, PC-SBELM is applied to build a real-time multi-fault diagnostic system. To verify the effectiveness of the proposed fault diagnostic framework, it is carried out on a real wind turbine gearbox system. The evaluation results show that the proposed framework can detect multi-fault in wind turbine gearbox much faster and more accurately than traditional identification techniques.

INDEX TERMS Wind turbine, gearbox, multi-fault diagnosis, Hilbert-Huang transform, pairwise-coupled, sparse Bayesian extreme learning machine.

I. INTRODUCTION

The development of wind energy has attracted the attention of both the academics and the industries. It is predicted that, by 2020, wind energy will account for 12% of the total electricity production [1]. Due to the growing number of wind farms and the harsh working conditions such as the load, torque, and turbulence, the wind turbine is more prone to damage than other rotating machinery. Moreover, as the wind farms and turbines are located in areas which are hard to access, the maintenance cost is always significant. To be more specific, a statistical investigation has shown that 17% of the total failures and 30% of the maintenance cost are caused by the gearbox [2]–[7]. As a result, real-time wind turbine gearbox fault diagnosis is of great importance which would help reduce the operation cost and the collateral damage caused by a power cut.

Currently, there are many existing literature focused on wind turbine fault diagnosis [1], [8]–[11] and can be generally classified into (i) model-based methods, (ii) knowledge-based methods and (iii) data-driven methods/ signal-based methods. Table 1 summarizes the pros and cons of these three kinds of methods.

It can be noticed that none of these three categories of rotating machinery fault diagnostic methods can perfectly overcome the issues that the others own. However, data-driven based approaches don't require much parameter assumption and are prior knowledge free when compared with the other two, which makes it more adaptive and accurate. As a result, the new multi-fault diagnostic framework proposed in this paper is based on the data-driven method which includes two main stages: (1) signal processing and (2) pattern recognition.

TABLE 1. Advantages and disadvantages of current rotating machinery fault diagnostic methods.

Methods	Examples	Advantages	Disadvantages
Model-based	[12-14]	High diagnostic accuracy; Require fewer data.	Exact models are difficult to obtain; Labor-demanding to determine model parameters; Machines require different models.
Knowledge -based	[15-17]	Low cost for users; Suitable for qualitative diagnosis.	Diagnostic knowledge is incomplete and vague; Strongly relies on technician's experience; Low diagnostic accuracy.
Data-driven (signal-based)	[18-24]	High diagnostic accuracy; Not require assumption or empirical estimation of physical parameters	Expensive diagnostic equipment; Diagnostic accuracy relies on input data quality and classifier performance.

Vibration is the main signal sources adopted by many researchers because of easy access. However, its accuracy and the processing cost is always downgraded by the noise interference and the relatively large datasets. To solve this problem, signal processing techniques are studied in this research for the selection of the most suitable one.

TABLE 2. Summary of different possible feature extraction methods.

Methods	Examples	Properties	Limitations
Fast Fourier transform (FFT)	[29-31]	Keep original properties well.	Not suitable for non-stationary signals.
Short time Fourier transform (STFT)	[32-34]	Suitable for non-stationary signal in the time domain.	Cannot achieve good time and frequency resolutions, resulting in information loss.
Wavelet packet transform (WPT)	[35-38]	Achieve good time and frequency resolutions at the same time.	Not a self-adaptive signal processing method in nature; Energy leaking.
Hilbert-Huang transform (HHT)	[39-42]	A self-adaptive time-frequency analysis method.	Mode mixing; Redundant information.

Table 2 lists the dominant techniques used in the fault diagnosis field, it can be seen that HHT overcomes most of the of disadvantages encountered by other signal processing methods. It mainly includes two steps, the first step is to decompose the original signal using empirical mode decomposition (EMD) into a series of IMFs which ranging from high frequency to the low. Then, the Hilbert transform is adopted to conduct spectrum analysis. Although HHT has superior performance compared with other techniques and is widely applied in fault diagnosis [25], the main disadvantage is the issue of mode mixing, as it uses EMD to decompose signal. To solve the problem of the mode mixing, an improved EMD method, ensemble empirical mode decomposition (EEMD) [26]–[28], is applied to improve the performance of EMD in Hilbert-Huang transform. As the original signal contains many redundant components and EEMD is only used to decompose the signal, not remove the noise, there are still a lot of noise components in the generated IMFs which will increase the computational cost and deteriorate the accuracy. Therefore, selecting the fault related IMF is still an imperative problem. To solve this problem, the correlation coefficient in introduced and integrated into the proposed method which can describe the correlation between each IMF and the raw signal, eliminating the redundant IMF. Moreover, to further extract representative fault features from the time and frequency domain, statistical features in the time domain and energy patterns of the processed IMF are selected. The details of the feature extraction are discussed in Section III.

For various rotating machinery fault diagnosis circumstances, the most adopted techniques are relevance vector machine (RVM) and probabilistic neural network (PNN). However, the performance of PNN based methods is superior to that of SVM based approaches when dealing with multi-label classifications [43]. Although PNN is better in terms of accuracy, the network structure construction, like the locations of neurons, the network size, and the smoothing parameters, still hinder its applications.

The reason is that all training datasets will be included in the pattern layer and the redundant datasets would considerably increase the size of the network structure. Two serious problems will come along with the large network which is the higher computational cost and the over-sensitiveness. The required computation will grow proportionally to the scale of the network when dealing with an unknown pattern and the over-sensitiveness would result in poor generalization performance. Moreover, the tuning of the smoothing parameters will also have a great influence on the performance and it is data dependent. Relevance vector machine (RVM), also known as a sparse Bayesian learning approach of SVM [44], is another machine learning method capable of predicting the posterior probability of the class. A study [45] already showed that RVM outperformed SVM in terms of diagnostic accuracy for fault diagnosis of low-speed bearings. For engine fault diagnosis, a study [46] also demonstrated that RVM was more robust than PNN for probabilistic classification. One obstacle of RVM is that the complexity of Hessian matrix computing attains $O(N^2)$, with N being the number of training data. This degrades the performance of RVM when dealing with large datasets.

Sparse Bayesian extreme learning machine (SBELM) is a variant of extreme learning machine (ELM) developed in [47] that combines the advantages of ELM and RVM. It mainly

has three advantages which are the sparsity of the weights, the low computational cost, and the probabilistic prediction ability. It is demonstrated in [47] that compared with the N in RVM, the L in SBELM is much smaller which is usually around 50. This makes the Hessian matrix much simpler and the execution time of SBELM much shorter when dealing with large-scale classification problems. Moreover, the number of hidden nodes and the hyperparameters have much less influence on its performance. This is a benefit to deal with the real-time monitoring engineering problem which requires the classifier has a quick response capability to analyze the mass acquired signal data. From the structural point of view, ELM is more suitable for multi-class problems as it has multi-input and output structure. As a result, due to the superiority of SBELM, it is chosen to be the fundamental structure in this study. However, as SBELM is designed only for binary classification problems, it is modified by integrating the pairwise coupling strategy which will be discussed in Section II.

The organization of this paper is as follows. The proposed framework and the corresponding techniques are demonstrated in Section II. Section III shows the experimental setup and the data pre-processing. In the fourth section, a detailed discussion will be conducted on the results and comparisons will be given. Conclusions will be drawn in the last section.

II. PROPOSED FRAMEWORK AND RELATED TECHNIQUES

Figure 1 shows the proposed real-time wind turbine gearbox fault diagnosis framework. There are mainly three sub-modules: (1) an online monitoring system; (2) multi-fault diagnostic model; and (3) fault identification system.

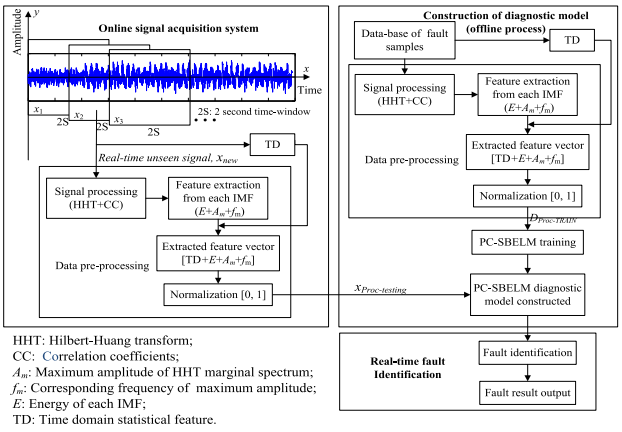


FIGURE 1. Proposed real-time fault diagnostic system for wind turbine gearbox.

In order to reduce the redundant information and improve the overall diagnosis accuracy, data pre-processing will be firstly conducted. A series of IMF will be generated by decomposing the acquired vibration signal using HHT, then CC algorithm selects useful IMF component to eliminate the redundant IMF. After signal decomposition, E, A_{m}, f_{m}, and TD indicators are extracted from time and frequency domain to construct a fault feature vector. In this proposed framework, fault samples in the database firstly go through the procedure of data pre-processing to construct the offline diagnostic model. Based on the online signal acquisition system, the real-time unseen signal, x_{new}, is acquired, and then it goes through the procedure of data pre-processing. Finally, the trained diagnostic model is applied to identify the real-time unseen signal.

A. HIBERT-HUANG TRANSFORM

The proposed modified HHT integrates EEMD and the Hilbert transform. In the rest of the paper, HHT will be used to represent the modified HHT integrating EEMD. For EEMD, the white noise of finite amplitude will be added into the original signal to strengthen the energy of the original components. As the average sum of the white noise is around zero, the true IMF is defined as the ensemble mean of each IMF set. In [27] and [48], the range of the standard deviation of EEMD is set to be between 0.1 to 0.4. The details of the EEMD algorithm can be described as:

1. A white noise series $\sigma n(t)$ is added to the investigated signal $x(t)$:
$$x_1(t) = x(t) + \sigma n(t) \tag{1}$$
2. With the EMD algorithm, $x_1(t)$ is decomposed into some IMF and a residual signal:
$$x_1(t) = \sum_{j=1}^{n} c_{1j}(t) + r_{1n}(t) \tag{2}$$
3. For each value of white noise until from j to n, repeat calculate Step 1 and 2:
$$x_i(t) = \sum_{i=1}^{n} c_{ij}(t) + r_{in}(t) \tag{3}$$
4. Calculate the mean for each IMF:
$$c_i(t) = \frac{1}{n} \sum_{i=1}^{n} c_{ij}(t) \tag{4}$$

After the EEMD decomposed, the Hilbert transform is applied to calculate each IMF. The Hilbert spectrum $H(\omega, t)$ is then calculated as the following:

$$H(\omega, t) = \mathrm{Re} \sum_{i=1}^{n} a_i(t) e^{i \int \omega_j(t) dt} \tag{5}$$

where $\omega_j(t)$ is instantaneous frequency, $a_j(t)$ is the amplitude of the signal. According to the Hilbert spectrum $H(\omega, t)$, the marginal spectrum $h(\omega)$ of Hilbert-Huang transform can be defined:

$$h(\omega) = \int_0^T H(\omega, t) dt \tag{6}$$

where T is the length of the signal, $h(\omega)$ reflects the amplitude changing with frequency in the entire frequency range, and $H(\omega, t)$ reveals important characteristics of the signal in the time-frequency domain.

B. SPARSE BAYESIAN EXTREME LEARNING MACHINE

The SBELM method employs the Bayesian learning (SBL) algorithm for ELM. The SBL is a benefit to overcome the drawbacks in ELM [47], [49]. For binary classification problem, SBELM treats it as an independent Bernoulli event with the probability expressed as:

$$P(\boldsymbol{t}|\,\boldsymbol{\beta},\mathbf{H})=\sum_{i=1}^{N}\sigma(\boldsymbol{h}_i\boldsymbol{\beta})^{t_i}[1-\sigma(\boldsymbol{h}_i\boldsymbol{\beta})]^{1-t_i} \tag{7}$$

where $\sigma(.)$ is sigmoid function and $\mathbf{h}_j = (\mathrm{g}(\boldsymbol{\omega}_1\boldsymbol{x}_j + b_1), \ldots, \mathrm{g}(\boldsymbol{\omega}_m\boldsymbol{x}_j + b_m))$, where $j = 1,\ldots,N$. Applying the Laplace approximation approach, automatic relevance determination (ARD) approximates a Gauss for the marginal likelihood. The $\ln\{p(\mathbf{t}|\,\boldsymbol{\beta},\mathbf{H})p(\boldsymbol{\beta}|\,\boldsymbol{a})\}$ can be defined as follows:

$$\begin{aligned}&\ln\{p(\mathbf{t}|\,\boldsymbol{\beta},\mathbf{H})p(\boldsymbol{\beta}|\,\boldsymbol{a})\}\\&=\sum_{i=1}^{N}\{t_i\ln y_i+(1-t_i)\ln(1-y_i)\}-\frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{A}\boldsymbol{\beta}+const\end{aligned} \tag{8}$$

where a is the hyperparameter of ARD, A is a diagonal matrix and $y_i = \sigma(\boldsymbol{h}_i\boldsymbol{\beta})$. The Laplace's mode $\hat{\beta}$ can be calculated using the Newton-Raphson method - iterative reweighted least-squares algorithm (IRLS).

$$\beta_{new}=(\mathbf{H}^T\mathbf{B}\mathbf{H}+\mathbf{A})^{-1}\mathbf{H}^T\mathbf{B}\hat{\boldsymbol{t}} \tag{9}$$

where $\hat{\boldsymbol{t}} = \mathbf{H}\beta_{old} + \mathbf{B}^{-1}(\mathbf{t}-y)$, $\mathbf{B} = \mathrm{diag}(\beta_1, \beta_2, .., \beta_N)$.

Setting the differential of $\mathcal{L}(\boldsymbol{a}) = \mathrm{Log}(P(\mathbf{t}|\,\boldsymbol{a},\mathbf{H}))$ with respect to $\boldsymbol{a}$ to zero, it yields

$$\begin{aligned}\frac{\partial\mathcal{L}(\mathbf{a})}{\partial a_i}&=\frac{1}{2}\left(\frac{1}{a_i}-\Sigma_{ii}-\hat{\beta}_i^2\right)=0\\a_i^{new}&=\frac{1-a_i\Sigma_{ii}}{\hat{\beta}_i^2}\end{aligned} \tag{10}$$

Through the iterations through Equation (9) and (10), the operation continues to maximize the marginal likelihood function until reaching the convergence criteria. The final probability distribution $\mathrm{P}(t\,|\,\mathbf{f},\hat{\beta})$, where $\mathbf{f}$ is an unseen input case or vector, is predicted by using sparse weight based on $y(\mathbf{h}\hat{\beta}) = \mathbf{h}\hat{\beta}$ and $\sigma[y(\mathbf{h}\hat{\beta})] = (1+e^{-y(\mathbf{h}\hat{\beta})})^{-1}$.

C. PAIRWISE COUPLING STRATEGY (PCS)

The one-versus-all strategy is simple and easy to deal with multi-class classification problems. However, it does not consider the pairwise correlation between different problems which causes larger indecisive region than pairwise coupling strategy (using one-versus-one) as shown in Figure 2. In the pairwise coupling strategy, a group of classifiers $\boldsymbol{l}_{\mathrm{class}} = [\boldsymbol{C}_1, \boldsymbol{C}_2, \ldots, \boldsymbol{C}_d]$ for d-label classification problem is constructed. Each $\boldsymbol{C}_i = [C_{i1}, \ldots C_{ij}, \ldots, C_{id}]$ is composed of a set of d-1 different pairwise classifiers $C_{ij}, i \neq j$. Since C_{ij} and C_{ji} are complementary, there are totally $d(d-1)/2$ classifiers in $\boldsymbol{l}_{\mathrm{class}}$ as shown in Figure 3. To solve the multi-class classification as well as probabilistic output problem, pairwise coupling strategy is adopted for the aforementioned probabilistic classifiers PNN, RVM, and SBELM, which is noted PC-PNN, PC-RVM, and PC-SBELM respectively.

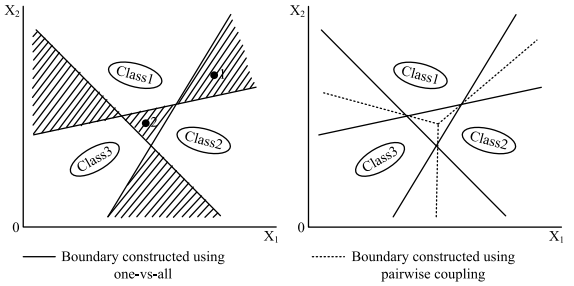


FIGURE 2. One-vs-all (left) contains shaded regions and pairwise coupling (right).

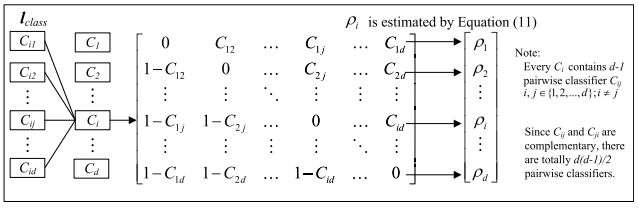


FIGURE 3. Probabilistic output using PCS.

The strategy combines all the outputs of every pair of classifier C_{ij} to re-estimate the overall probability ρ_i where $i = 1$ to d. In the case study, a simple pairwise coupling strategy for simultaneous-fault diagnosis is proposed. Every ρ_i is calculated as

$$\rho_i = C_i(\boldsymbol{x}) = \frac{\sum_{j=1:i\neq j}^{d} C_{ij}(\boldsymbol{x})}{d} \tag{11}$$

III. CASE STUDY AND EXPERIMENTAL SETUP

In order to demonstrate the effectiveness of the proposed system, experiments on the wind turbine simulator were carried out. The required raw datasets are acquired from the sensor mounted on the turbine gearbox. Details of the experiment design and the signal acquisition are discussed.

A. WIND TURBINE SIMULATOR AND SAMPLE DATA ACQUISITION

In this research, Spectra Quest's wind turbine simulator is employed for conducting fundamental research in wind turbine energy production and vibration diagnostics. The system mainly consists of the blade, yaw system, rotor, planetary gearbox, generator, and tower. The planetary gearbox is directly connected to the input shaft through the torque sensor via couplings as shown in Figure 4. The input shaft drives the planet carrier of the planetary system and the sun gear provides the output.

Due to the complexity of the whole wind turbine system and the faults residing in the gearbox accounts for most of the failures, this paper focuses on the fault detection of

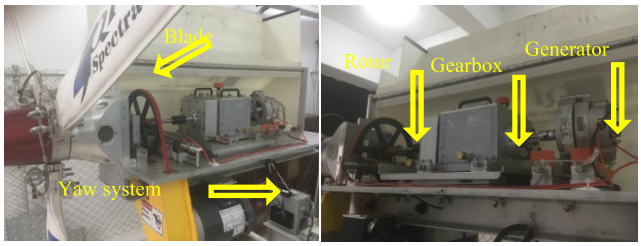


FIGURE 4. Spectra Quest's wind turbine simulator.

TABLE 3. Sample single-faults and possible simultaneous-fault.

Case no.	Condition	Label	Fault Description
C1	Normal	Label={1}	Normal
C2		Label={2}	Looseness
C3		Label={3}	Mechanical misalignment
C4	Single-fault	Label={4}	Tooth broken
C5		Label={5}	Chipped tooth
C6		Label={6}	Wear of gear tooth
C7		Label={7}	Inner race fault
C8		Label={8}	Outer race fault
C9		Label={9}	Ball fault
C10	Simultaneou	Label={4,9}	Tooth broken & Ball fault
C11	s-fault	Label={5,8}	Chipped tooth & Outer race fault

FIGURE 5. Tooth broken (C4); Chipped tooth (C5); Wear of gear tooth (C6); Inner race fault (C7); Outer race fault (C8); Ball fault (C9).

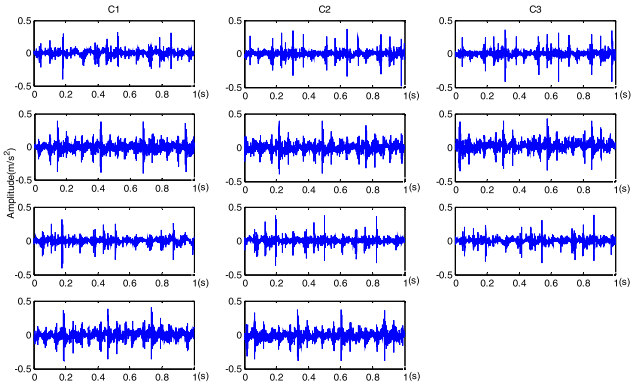


FIGURE 6. Example of faulty signals.

the gearbox. The test rig shown in Figure 4 has the ability to simulate many common faults cases, such as looseness, inner race fault, and gear crack. Table 3 and Figure 5 present a total of eleven cases including one normal case, eight single-faults, and two simultaneous-fault cases. Samples of these signals are illustrated in Figure 6. Note that the combination of single-faults is not randomly generated. According to practical experiences, a machine system cannot be operated under too many faults appearing concurrently; therefore, simultaneous-faults are an experimental selection.

TABLE 4. Division of the sample dataset into different subsets.

	Type of dataset	Single-fault (1800)	Simultaneous-fault (200)
Raw sample data	Validation dataset	D_{vaild_1} (630)	D_{vaild_S} (110)
	Training dataset	D_{train_1} (900)	
	Test dataset	D_{test_1} (270)	D_{test_S} (90)
Feature extraction	Validation dataset	$D_{proc\_vaild\_1}$ (630)	$D_{proc\_vaild\_S}$ (110)
	Training dataset	$D_{proc\_train\_1}$ (900)	
	Test dataset	$D_{proc\_test\_1}$ (270)	$D_{proc\_test\_S}$ (90)

To make the experiment more representatives, two electric loads under the constant rotational speed (1800rpm) are adopted. For each load, each single fault is repeated one hundred times, and fifty for each simultaneous fault. The sampling frequency is set to 4096 Hz and the recording window is two seconds. It should be noted that in order to avoid missing details of the fault information, the sampling frequency is set higher than the twice of the meshing frequency. As a result, for each sample dataset, there will be 8092 sampling points. The acquired 2000 sample sets are shown in Table 4 which includes 1 normal case, 8 single cases and 2 simultaneous faults ($1 \times 100 \times 2$ load $+ 8 \times 100 \times 2$load $+ 2 \times 50 \times 2$ load). In order to make the experiments more typical, the acquired datasets are randomly divided into three subsets which are for training, validation and testing, respectively. The training sets are used to train the classifiers, the validation sets are used for the tuning of the parameters and the testing sets are for the evaluation of the performance. It should be pointed out that the classifiers are only trained by the single faults as the symptoms of the fault in simultaneous faults would be identified due to the proposed feature extraction method.

B. SIGNAL PROCESSING

In this case study, the correlation coefficient approach [39] is applied to select the proper number of IMFs. It is obtained by:

$$Coe_{x(t)',I_i(t)} = \frac{\sum_{i=1}^{M}(x(t)' - \overline{x})(I_i(t) - \overline{I_i})}{\sqrt{\sum_{i=1}^{M}(x(t)' - \overline{x})^2}\sqrt{\sum_{i=1}^{M}(I_i(t) - \overline{I_i})^2}} \quad (12)$$

where M is the number of IMFs, $\overline{x}$ and $\overline{I_i}$ is the mean values of the $x(t)$' and $I_i(t)$ respectively. A large $Coe_{x(t)',I_i(t)}$ value means a high correlation between $I_i(t)$ and $x(t)$', and also implies that $I_i(t)$ contains more fault information. A sample of signal C6 is presented in Figure 7, the correlation coefficient of IMF I_7 and I_8 is obviously smaller than the others. Thus, in this case study, I_7 and I_8 are ignored to extract the energy pattern.

To extract the effective fault features, six-dimensional energy feature vector $e = [E_1, E_2, \ldots, E_6]$ is calculated as

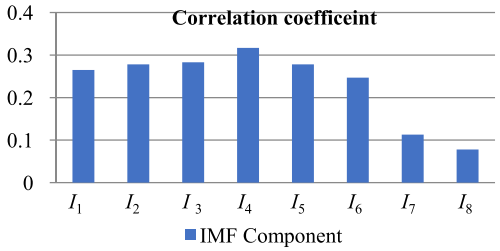


FIGURE 7. Correlation coefficients of each IMF component for the signal of C6.

TABLE 5. Definition of common statistical features for time-domain signals (TD).

Time-domain feature	Equation	Time-domain feature	Equation
1.Mean	$x_m = \frac{1}{N}\sum_{i=1}^{N} x_i$	4. Peak	$x_{pk} = \max\|x_i\|$
2.Standard deviation	$x_{std} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - x_m)^2}{N-1}}$	5. Kurtosis	$x_{kur} = \frac{\sum_{i=1}^{N}(x_i - x_m)^4}{(N-1)x_{std}^4}$
3.Root mean square	$x_{rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$	6.Impulse factor	$IF = \frac{x_{pk}}{\frac{1}{N}\sum_{i=1}^{N}\|x_i\|}$

x_i represents a signal series for i =1,2,…, N, where N is the number of data points of a raw signal.

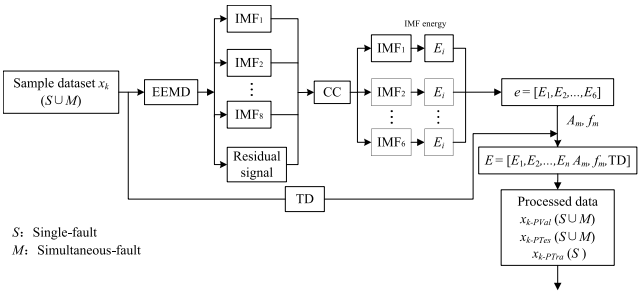


FIGURE 8. Flowchart of the proposed feature extraction approach.

follows:

$$E_i = \sum_{j=1}^{n} [(j \cdot \Delta t) \cdot |I_i(j \cdot \Delta t)|^2] \tag{13}$$

where Δt is the time interval and j is the index of data points. Considering the HHT marginal spectra of different fault conditions display various corresponding frequency f_m and maximum amplitude A_m, they are added into the extracted energy feature vector e. Besides, to enrich the fault information, six statistical features of the time-domain (TD) are further extracted as shown in Table 5. The procedure of data processing is demonstrated in Figure 8. In this case study, the extracted feature vector is finally extended to an eleven-dimensional vector as $E = [E_1, E_2, \ldots, E_6, A_m, f_m, \mathrm{TD}]$.

IV. EVALUATION OF THE PROPOSED FRAMEWORK

In the real-time fault diagnosis of WTGS, two main tasks need to be accomplished including dealing with large vibration signal from the online monitoring system and detecting various faults in the WTGS. The proposed framework can detect whether there are some faults in the samples data quickly and correctly.

A. EFFECTIVENESS OF THE SIGNAL PROCESSING TECHNIQUES

To reveal the superiority of the proposed modified HHT approach, conventional wavelet package transform with principal component analysis (WPT+PCA), and empirical mode decomposition with singular value decomposition (EMD+SVD) are adopted as the comparisons. For the wavelet-based method, the mother wavelets should be firstly addressed. Daubechies wavelet with level 4 decomposition is adopted in this study due to its effectiveness after carrying out many trails to decompose and reconstruct the signal. Then PCA was employed to reduce the dimension of the reconstructed signal of vibration from 4096 inputs to 40 respectively.

TABLE 6. Evaluation of different feature extraction techniques.

Feature extraction	Classifier	Accuracies for Test Case (%)		
		Sing-fault	Simultaneous -fault	Overall fault
None	PNN	53.35	47.26	51.43
	RVM	54.46	47.63	52.27
	SBELM	54.32	47.58	51.68
WPT+PCA	PNN	89.57	81.06	85.12
	RVM	90.45	81.78	85.67
	SBELM	90.30	81.69	85.43
WPT+E	PNN	92.36	83.31	88.46
	RVM	93.62	84.64	89.25
	SBELM	93.86	85.00	90.12
EMD+SVD	PNN	88.67	80.74	84.68
	RVM	89.74	81.24	85.23
	SBELM	89.46	81.16	85.03
EMD+E	PNN	92.70	83.68	89.74
	RVM	93.74	84.98	90.46
	SBELM	94.02	85.22	90.78
HHT+SVD	PNN	89.84	81.63	85.62
	RVM	91.48	83.04	87.76
	SBELM	91.36	82.86	87.53
HHT+E	PNN	93.44	84.43	90.68
	RVM	94.63	85.48	91.46
	SBELM	95.12	86.25	92.23

The results of different signal pre-processing and feature extraction methods are shown in Table 6. Compared with the none processed approach, the proposed method using SBLEM has improved the overall efficiency by 40.55%. When the proposed method are conducted with conventional PNN and RVM, the overall accuracies are also improved by 39.25% and 39.19, respectively. It should be noted that the classifier is only trained by single-fault features and the testing dataset is from the simultaneous-faults. It also can be seen that the proposed HHT+E signal processing method is superior to other approaches independently of the classifiers. The reason is that the modified HHT incorporates EEMD which could decompose the original signal both according to the frequencies and the signals, giving it the capability

of dealing with both non-linear and non-stationary signal. Moreover, the mode mixing problem residing in the conventional HHT is also overcome. The comparisons demonstrate the effectiveness of the proposed modified HHT+E feature extraction method.

B. EFFECTIVENESS OF PAIRWISE COUPLING STRATEGY

To verify the effectiveness of the PCS, the extracted features are used to train pairwise-coupled classifiers, namely PC-SBELM, PC-PNN, and PC-RVM. According to the PCS, there are totally $11(11\text{-}1)/2 = 55$ SBELM classifiers trained based on 11 fault types in the proposed WTGS. For those classification algorithms, hidden node h, spread s, and width w^* are important hyper-parameters for the PC-SBELM, PC-PNN, and PC-RVM respectively. The hyper-parameter of hidden node h is default set to close the number of trained classifiers 55. Meanwhile, by using the fivefold cross-validation method, s, and w^* were set to be 0.4, 0.62 respectively. With the proposed signal processing techniques, the classification results for pairwise-coupling strategy classifiers are shown in Table 7.

TABLE 7. Comparison results of different classifiers using pairwise coupling strategy.

Feature extraction	Classifier	Accuracies for Test Case (%)		
		Single-fault	Simultaneous-fault	Overall fault
HHT+E	PNN	93.44	84.43	90.68
	RVM	94.63	85.48	91.46
	SBELM	95.12	86.25	92.23
	PC-PNN	94.62	85.43	91.82
	PC-RVM	95.72	86.52	92.87
	PC-SBELM	96.36	87.21	93.92

It can be seen that the fault diagnosis accuracy of the proposed PC-SBELM in overall faults using the feature extraction method (HHT+E) outperforms the SBELM by 1.69%, which demonstrate the proposed PC-SBELM superiority. The classifier PC-PNN and PC-RVM also give the improvement of 1.14% and 1.41% when compared with PNN and RVM respectively. The improvement comes from the proposed one-versus-one strategy as it considers different feature groups as individuals not just classifying them as two groups. This technique avoids the large undecided areas in the one-versus-all strategy and giving all feature vectors the corresponding labels. The results demonstrate that the proposed pairwise-coupling strategy could improve the classification accuracy for both single and simultaneous faults.

As we known, simultaneous-fault is difficult to identify its single-fault quantity and type because of the different single-fault signals intercoupling which display the complex shape. Another contribution of this paper is that the proposed PC-SBELM probabilistic classifier can effectively identify the single-fault quantity and type as shown in Figure 9. In Figure 9(a), 35 testing samples of simultaneous-fault C10 are classified into 32 tooth broken (C4) and 29 ball fault (C9); moreover, in Figure 9(b), 55 testing samples

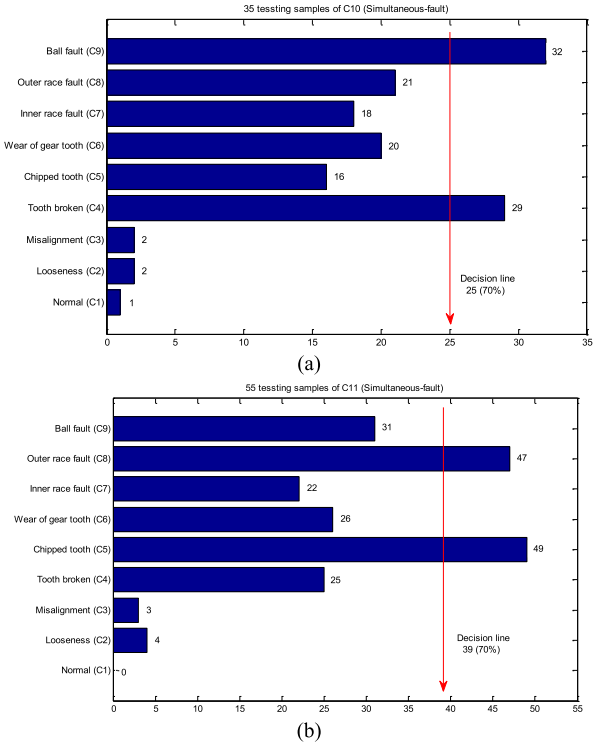


FIGURE 9. The classification results of simultaneous-fault (C10 and C11) testing rest. (a) 35 testing samples of simultaneous-fault C10. (b) 55 testing samples of simultaneous-fault C11.

of simultaneous-fault C11 are classified into 47 chipped tooth (C5) and 49 outer race fault (C8) which is belong to single-fault. In this case study, the decision line uses the default setting 70% which will be further optimized in future work, that is, surpassing the decision line is defined to the fault. It can be seen that simultaneous-fault C10 is combined with single-fault C4 and C9, and C11 is combined with C5 and C8. The average accuracy of the proposed PC-SBELM probabilistic classifier reaches 87.21%, which is a competitive result in simultaneous-fault identification.

C. RESULTS AND DISCUSSION

In order to further demonstrate the superiority of the proposed method in terms of the computational efficiency, Table 8 gives the processing time of different methods besides the corresponding accuracy. The identification time in the table is defined as the duration from the receiving the signal from the sensor to the generating the fault detection alert, which means the sensor delay is not considered.

TABLE 8. Processing time of PC-SBELM with PC-PNN and PC-RVM.

	Time for test case (s)		
	Single-fault	Simultaneous-fault	Overall fault
PC-PNN	0.68±0.8	0.50±1.0	0.59±0.9
PC-RVM	0.45±0.6	0.32±0.8	0.38±0.6
PC-SBELM	0.18±0.4	0.12±0.4	0.14±0.3

Extraction method based on the proposed HHT+E

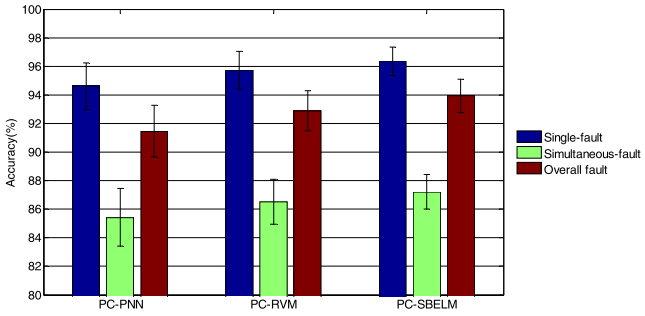


FIGURE 10. Diagnosis results using different classifiers trained by extracted feature.

Figure 10 also shows that the fault detection accuracy of PC-SBELM (93.92%) on overall faults is slightly higher than that of PC-PNN (91.46%) and PC-RVM (92.87%) by 2.46% and 1.05% respectively. In addition to fault detection accuracy, the fault identification time is also presented in Table 8. It demonstrates that the fault identification time of PC-PNN, PC-RVM, and PC-SBELM only take 0.59s, 0.38s, and 0.14s respectively. In the implementation of WTGS fault diagnosis, it is required to continuously work all the time which means that no data could be processed offline as the size of the dataset would grow beyond the computation capability. As a result, the processing time is of great importance. In Table 8, it can be seen that, although the accuracy of the proposed PC-SBELM is only improved slightly, the processing time is considerably improved by 0.45s and 0.24s compared with PC-PNN and PC-RVM, respectively. The reason is that the size and training time of PC-PNN is heavily dependent of the size of the input dataset, and for PC-RVM, it is that the complexity of Hessian matrix computing attains $O(N^2)$ is also related to the number of training data, which deteriorates the ability to deal with multi-class classifications. Compared with these two approaches, the $O(L^2)$ for the proposed PC-SBELM is relatively stable where $L \approx 50$. It can be concluded that the fast training time and the sparsity of the proposed method would present desirable performance when dealing with large-scale dataset. Moreover, the implementation of the proposed method is also much easier as it is insensitive to the hyperparameter and the number of hidden nodes. Therefore, the proposed method could have better performance in terms of accuracy and calculation efficiency.

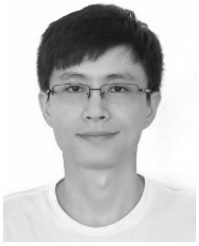
V. CONCLUSION

This paper proposes a new real-time fault diagnostic system for a wind turbine gearbox system. In the proposed system, the data pre-processing approaches employ the modified Hilbert-Huang transform adopting the EEMD to eliminate the mode mixing problem. Moreover, energy pattern calculation is also proposed to extract the fault features in conjunction with time-domain statistical analysis which could further reduce the size of the features to improve calculate efficiency. To satisfy the purpose of rapid recognition in the real-time monitoring system, the improved probabilistic classifier PC-SBELM is proposed. Although PC-SBELM only produces slightly higher diagnostic accuracy (93.92%) than PC-PNN and PC-RVM in this case study, it can generate a smaller classification model and takes less execution time (0.14s). In short, PC-SBELM is superior to PNN, RVM, PC-PNN, and PC-RVM in terms of calculation efficiency and at the same time could guarantee the fault detention accuracy. The experiment results also demonstrate the simultaneous-fault detection ability (87.21%) with the classifier only trained by single-fault datasets. As the proposed fault diagnostic framework is generic, it could be applied to the other applications of condition monitoring in which the fault identification time and simultaneous-fault diagnosis are critical.

REFERENCES

[1] A. Zhou, D. Yu, and W. Zhang, "A research on intelligent fault diagnosis of wind turbines based on ontology and FMECA," *Adv. Eng. Inform.*, vol. 29, no. 1, pp. 115–125, 2015.
[2] R. W. Hyers, J. G. McGowan, K. L. Sullivan, J. F. Manwell, and B. C. Syrett, "Condition monitoring and prognosis of utility scale wind turbines," *Energy Mater.*, vol. 1, no. 3, pp. 187–203, Sep. 2006.
[3] B. Lu, Y. Li, X. Wu, and Z. Yang, "A review of recent advances in wind turbine condition monitoring and fault diagnosis," in *Proc. Power Electron. Mach. Wind Appl. (PEMWA)*, Jun. 2009, pp. 1–7.
[4] Q. Li and S. Y. Liang, "Degradation trend prognostics for rolling bearing using improved R/S statistic model and fractional Brownian motion approach," *IEEE Access*, vol. 6, pp. 21103–21114, 2018.
[5] V. Fernandez-Cavero, D. Morinigo-Sotelo, O. Duque-Perez, and J. Pons-Llinares, "A comparison of techniques for fault detection in inverter-fed induction motors in transient regime," *IEEE Access*, vol. 5, pp. 8048–8063, 2017.
[6] F. Cheng, J. Wang, L. Qu, and W. Qiao, "Rotor-current-based fault diagnosis for DFIG wind turbine drivetrain gearboxes using frequency analysis and a deep classifier," *IEEE Trans. Ind. Appl.*, vol. 54, no. 2, pp. 1062–1071, Mar./Apr. 2017.
[7] J. Wang, F. Cheng, W. Qiao, and L. Qu, "Multiscale filtering reconstruction for wind turbine gearbox fault diagnosis under varying-speed and noisy conditions," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4268–4278, May 2018.
[8] S. Simani *et al.*, "Data-driven modelling of a wind turbine benchmark for fault diagnosis application," *WIT Trans. Eng. Sci.*, vol. 1, no. 7, pp. 1–12, 2012.
[9] S. Asgari and A. Yazdizadeh, "Robust model-based fault diagnosis of mechanical drive train in V47/660 kW wind turbine," *Energy Syst.*, vol. 9, no. 4, pp. 921–952, 2017.
[10] M. S. Li, D. Yu, Z. Chen, K. Xiahou, T. Ji, and Q. H. Wu, "A data-driven residual-based method for fault diagnosis and isolation in wind turbines," *IEEE Trans. Sustain. Energy*, to be published.
[11] F. Cheng, Y. Peng, L. Qu, and W. Qiao, "Current-based fault detection and identification for wind turbine drivetrain gearboxes," *IEEE Trans. Ind. Appl.*, vol. 53, no. 2, pp. 878–887, Mar./Apr. 2017.
[12] P. Pennacchi, N. Bachschmid, A. Vania, G. A. Zanetta, and L. Gregori, "Use of modal representation for the supporting structure in model-based fault identification of large rotating machinery: Part 1—Theoretical remarks," *Mech. Syst. Signal Process.*, vol. 20, no. 3, pp. 662–681, 2006.
[13] R. Isermann, "Model-based fault-detection and diagnosis—Status and applications," *Annu. Rev. Control*, vol. 29, no. 1, pp. 71–85, 2005.
[14] P. Pennacchi, N. Bachschmid, and A. Vania, "A model-based identification method of transverse cracks in rotating shafts suitable for industrial machines," *Mech. Syst. Signal Process.*, vol. 20, no. 8, pp. 2112–2147, 2006.
[15] M. B. Çelik and R. Bayir, "Fault detection in internal combustion engines using fuzzy logic," *Proc. Inst. Mech. Eng., D, J. Automobile Eng.*, vol. 221, no. 5, pp. 579–587, 2007.
[16] B.-S. Yang, D.-S. Lim, and A. C. C. Tan, "VIBEX: An expert system for vibration fault diagnosis of rotating machinery using decision tree and decision table," *Expert Syst. Appl.*, vol. 28, no. 4, pp. 735–742, 2005.
[17] H. L. Gelgele and K. Wang, "An expert system for engine fault diagnosis: development and application," *J. Intell. Manuf.*, vol. 9, no. 6, pp. 539–545, 1998.

[18] S. Yin, G. Wang, and H. R. Karimi, "Data-driven design of robust fault detection system for wind turbines," *Mechatronics*, vol. 24, no. 4, pp. 298–306, 2014.
[19] Q. Hu, Z. He, Z. Zhang, and Y. Zi, "Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 688–705, Feb. 2007.
[20] J. Sanz, R. Perera, and C. Huerta, "Fault diagnosis of rotating machinery based on auto-associative neural networks and wavelet transforms," *J. Sound Vib.*, vol. 302, nos. 4–5, pp. 981–999, 2007.
[21] D. Yang, Y. Liu, S. Li, X. Li, and L. Ma, "Gear fault diagnosis based on support vector machine optimized by artificial bee colony algorithm," *Mech. Mach. Theory*, vol. 90, pp. 219–229, Aug. 2015.
[22] D. Antory, "Application of a data-driven monitoring technique to diagnose air leaks in an automotive diesel engine: A case study," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 795–808, 2007.
[23] J. Hou and B. Xiao, "A data-driven clustering approach for fault diagnosis," *IEEE Access*, vol. 5, pp. 26512–26520, 2017.
[24] J. Yu and Y. He, "Planetary gearbox fault diagnosis based on data-driven valued characteristic multigranulation model with incomplete diagnostic information," *J. Sound Vib.*, vol. 429, pp. 63–77, Sep. 2018.
[25] J. Hu and S. Yang, "Research on HHT-based fault diagnosis method for rotating machinery," *Power Eng.*, vol. 24, no. 6, pp. 845–851, 2004.
[26] X. Wang, C. Liu, F. Bi, X. Bi, and K. Shao, "Fault diagnosis of diesel engine based on adaptive wavelet packets and EEMD-fractal dimension," *Mech. Syst. Signal Process.*, vol. 41, nos. 1–2, pp. 581–597, 2013.
[27] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, 2008.
[28] B. Liu, S. Riemenschneider, and Y. Xu, "Gearbox fault diagnosis using empirical mode decomposition and Hilbert spectrum," *Mech. Syst. Signal Process.*, vol. 20, no. 3, pp. 718–734, Apr. 2006.
[29] Z. K. Peng and F. L. Chu, "Application of the wavelet transform in machine condition monitoring and fault diagnostics: A review with bibliography," *Mech. Syst. Signal Process.*, vol. 18, no. 2, pp. 199–221, Mar. 2004.
[30] N. Saravanan, V. N. S. K. Siddabattuni, and K. I. Ramachandran, "Fault diagnosis of spur bevel gear box using artificial neural network (ANN), and proximal support vector machine (PSVM)," *Appl. Soft Comput.*, vol. 10, no. 1, pp. 344–360, 2010.
[31] G. T. Zheng and A. Y. T. Leung, "Internal combustion engine noise analysis with time-frequency distribution," *J. Eng. Gas Turbines Power*, vol. 124, no. 3, pp. 645–649, 2002.
[32] W. Bartelmus and R. Zimroz, "Vibration condition monitoring of planetary gearbox under varying external load," *Mech. Syst. Signal Process.*, vol. 23, no. 1, pp. 246–257, 2009.
[33] Z. Li, X. Yan, Z. Guo, Y. Zhang, C. Yuan, and Z. Peng, "Condition monitoring and fault diagnosis for marine diesel engines using information fusion techniques," *Elektron. Elektrotech.*, vol. 123, no. 7, pp. 109–112, 2012.
[34] J. D. Bronzino, "Biomedical signals: Origin and dynamic characteristics; frequency-domain analysis," in *Medical Devices and Systems*. Boca Raton, FL, USA: CRC Press, 2006, pp. 27–48.
[35] Z. Li, X. Yan, Z. Tian, C. Yuan, Z. Peng, and L. Li, "Blind vibration component separation and nonlinear feature extraction applied to the nonstationary vibration signals for the gearbox multi-fault diagnosis," *Measurement*, vol. 46, no. 1, pp. 259–271, 2013.
[36] J.-D. Wu and J.-M. Kuo, "An automotive generator fault diagnosis system using discrete wavelet transform and artificial neural network," *Expert Syst. Appl.*, vol. 36, no. 6, pp. 9776–9783, 2009.
[37] B. S. Anami and V. B. Pagi, "Localisation of multiple faults in motorcycles based on the wavelet packet analysis of the produced sounds," *IET Intell. Transport Syst.*, vol. 7, no. 3, pp. 296–304, Sep. 2013.
[38] S. J. Loutridis, "Damage detection in gear systems using empirical mode decomposition," *Eng. Struct.*, vol. 26, no. 12, pp. 1833–1841, 2004.
[39] Y. S. Wang, Q. H. Ma, Q. Zhu, X. T. Liu, and L. H. Zhao, "An intelligent approach for engine fault diagnosis based on Hilbert–Huang transform and support vector machine," *Appl. Acoust.*, vol. 75, pp. 1–9, Jan. 2013.
[40] Z. K. Peng, P. W. Tse, and F. L. Chu, "A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing," *Mech. Syst. Signal Process.*, vol. 19, no. 5, pp. 974–988, 2005.
[41] E. Elbouchikhi, V. Choqueuse, Y. Amirat, M. El Hachemi Benbouzid, and S. Turri, "An Efficient Hilbert–Huang transform-based bearing faults detection in induction machines," *IEEE Trans. Energy Convers.*, vol. 32, no. 2, pp. 401–413, Jun. 2017.
[42] D. Yu, Y. Yang, and J. Cheng, "Application of time–frequency entropy method based on Hilbert–Huang transform to gear fault diagnosis," *Measurement*, vol. 40, nos. 9–10, pp. 823–830, 2007.
[43] P. M. Ciarelli, E. Oliveira, C. Badue, and A. F. De Souza, "Multi-label text categorization using a probabilistic neural network," *Int. J. Comput. Inf. Syst. Ind. Manage. Appl.*, vol. 1, pp. 133–144, 2009.
[44] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.
[45] A. Widodo *et al.*, "Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 7252–7261, 2009.
[46] C.-M. Vong, P.-K. Wong, W.-F. Ip, and C.-C. Chiu, "Simultaneous-fault diagnosis of automotive engine ignition systems using prior domain knowledge and relevance vector machine," *Math. Problems Eng.*, vol. 2012, Dec. 2012, Art. no. 974862.
[47] J. Luo, C.-M. Vong, and P.-K. Wong, "Sparse Bayesian extreme learning machine for multi-classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 4, pp. 836–843, Apr. 2014.
[48] Z. P. Feng *et al.*, "Fault diagnosis for wind turbine planetary gearboxes via demodulation analysis based on ensemble empirical mode decomposition and energy separation," *Renew. Energy*, vol. 47, pp. 112–126, Nov. 2012.
[49] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.



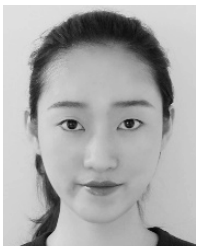
JIAN-HUA ZHONG received the M.S. and Ph.D. degrees in electromechanical engineering from the University of Macau, in 2011 and 2016, respectively.

He is currently an Assistant Professor with the School of Mechanical Engineering and Automation, Fuzhou University, China. He has published over 20 papers in refereed journals and conference proceedings. His research interests include rotating machinery condition monitoring, vehicle dynamics and control, signal processing, pattern recognition, and fault diagnosis using machine learning methods.

JUN ZHANG received the M.S. and Ph.D. degrees from the School of Mechanical Engineering, Tianjin University, China, in 2004 and 2007, respectively.

He is currently a Professor with the School of Mechanical Engineering and Automation, Fuzhou University, China. His current research interests include mechanical dynamics, robotics, signal processing, and rotating machinery fault diagnosis.

JIEJUNYI LIANG received the B.S. degree in mechanical design, manufacturing and automation from the Huazhong University of Science and Technology, Wuhan, China, in 2012, the M.S. degree in electromechanical engineering from the University of Macau, Macau, China, in 2015, and the Ph.D. degree in mechanical engineering from the School of Mechanical and Mechatronic Engineering, University of Technology Sydney, in 2018. He is currently a Research Fellow with the School of Automotive Engineering, Wuhan University of Technology, China.



HAIQING WANG received the B.S. degree in metal material engineering from Liaoning Shihua University, China, in 2015, and the M.S. degree in mechanical engineering from the School of Mechanical and Mechatronic Engineering, University of Technology Sydney, in 2018, where she is currently a Research Assistant.