Machinery Fault Diagnosis Scheme Using Redefined Dimensionless Indicators and mRMR Feature Selection

QIN HU [ID][1,2]**, XIAO-SHENG SI** [ID][1]**, AI-SONG QIN** [ID][2]**, YUN-RONG LV** [ID][2]**, AND QING-HUA ZHANG** [ID][2]

[1]Department of Automation, Rocket Force University of Engineering, Xi'an 710025, China
[2]Guangdong Provincial Key Laboratory of Petrochemical Equipment Fault Diagnosis, Guangdong University of Petrochemical Technology, Maoming 525000, China

Corresponding author: Xiao-Sheng Si (sxs09@mails.tsinghua.edu.cn)

ABSTRACT Machinery fault diagnosis methods based on dimensionless indicators have long been studied. However, traditional dimensionless indicators usually suffer a low diagnostic accuracy for mechanical components. Toward this end, an effective fault diagnosis method based on redefined dimensionless indicators (RDIs) and minimum redundancy maximum relevance (mRMR) is proposed to identify the health conditions of mechanical components. In the proposed method, the vibration signals are first processed by the variational mode decomposition, and multiple RDIs are constructed based on the decomposed signals. Subsequently, the mRMR approach is introduced to select the RDIs and several important RDIs can be obtained. Finally, the obtained RDIs are fed into a grid search support vector machine to perform fault pattern identification. To verify the superiority of the proposed method, two experimental examples for different fault types of mechanical components including rolling bearing and gearbox are conducted. The experimental results demonstrated that the RDIs as new fault features can effectively solve the deficiency of the traditional dimensionless indicator, and has stronger distinguishing ability for machinery faults. Additionally, our proposed method successfully differentiated 12 fault conditions of rolling bearings and nine fault conditions of gears with average accuracies of 97.47% and 97.12% with 11 and 5 RDIs, respectively.

INDEX TERMS Redefined dimensionless indicators, variational mode decomposition, minimum redundancy maximum relevance, grid search support vector machine, fault diagnosis.

I. INTRODUCTION

Rolling bearings and gears are the most commonly used components in modern machinery that frequently functions under harsh environmental conditions. When fault emerges in such mechanical components, it may cause catastrophic machine breakdown and significant economic loss. Therefore, fault diagnosis technology for mechanical components must be investigated to improve reliability and reduce unscheduled downtime [1]. As machinery vibration signals typically exhibit nonlinear and nonstationary characteristics,

The associate editor coordinating the review of this manuscript and approving it for publication was Li He [ID].

the health condition information distributed in machinery vibration signals is complex. Over the past decades, many signal processing techniques, such as symplectic geometry mode decomposition [2], wavelet packet decomposition [3], empirical mode decomposition [4], local mean decomposition [5], variational mode decomposition [6], empirical wavelet transform [7], etc., have been widely used to analyze the nonlinear and nonstationary signals. Among them, a self-adaptive and entirely nonrecursive signal decomposition method, called variational mode decomposition (VMD), has been proposed by Dragomiretskiy and Zosso [8]. This method can decompose a complex vibration signal into a sum of amplitude modulation and frequency modulation components, as well

as effectively preprocess nonlinear and nonstationary signals. Several studies have been conducted regarding the applications and extensions of VMD to fault diagnosis [6], [9]–[11], and VMD has been regarded as highly effective in rotating machinery fault diagnosis [12], [13]. Therefore, we utilized VMD to produce the original feature sets in this study. Based on signal processing, various types of fault features can be extracted [14]–[18]. Dimensionless indicators as time-domain features are sensitive to faults and not affected by disturbances. Therefore, they have been widely used as fault features in several studies [19]–[21].

Nevertheless, the existing types of traditional dimensionless indicators (TDIs) are limited; they are only sensitive to certain fault types. In practical engineering, an increasing number of fault types, fault severities, and fault orientations have emerged. Utilizing the limited number of dimensionless indicators to accurately diagnose more faults has become extremely difficult. Additionally, TDIs often suffer from varying amounts of aliasing in the feature space, causing a decrease in diagnosis accuracy and an increase in diagnosis uncertainty [22], [23]. To solve the abovementioned problems of TDIs, a new fault extraction method based on redefined dimensionless indicators (RDIs) is proposed in this study. Based on the RDIs, more types of dimensionless indicators and more useful fault information hidden in the vibration signal can be obtained.

After feature extraction, feature selection is typically conducted to reduce the feature dimension and obtain an appropriate subset of features. Currently, several feature selection methods have been investigated and applied in fault diagnosis. In [24], a genetic algorithm (GA)-based feature selection was applied to identify crack types for pressure vessels. In [25], two particle swarm optimization (PSO)-based feature selection algorithms were proposed. However, the GA- or PSO-based feature selection either traps easily into the local optima or has a low computational efficiency. In [26], kernel principal component analysis was used to select important features by the feature contribution rate. In [27], a locally linear embedding algorithm was used to reduce the dimensions of each row vector of a feature matrix. In [28], locality preserving projections were used to select the more sensitive low-dimensional information hidden in high-dimensional fusion feature structures. Nevertheless, these methods have failed to consider the possible redundant and irrelevant features in high-dimensional feature sets, of which redundant and irrelevant features might reduce the fault identification accuracy and increase the computation time. Among the several available feature selection methods, the max-relevance and min-redundancy (mRMR) method [29] is an effective feature selection approach, which can automatically eliminate redundant and irrelevant and features in a high-dimensional feature space and select important features according to the maximum correlation and minimum redundancy criteria. In [30], [31], the mRMR approach was employed to eliminate redundant features and select the first important features to construct new fault feature vectors

to represent fault characteristics. Additionally, the mRMR method, which affords fast calculation and strong robustness, has been validated in [32].

Based on the above analysis, an effective fault diagnosis method for mechanical components is proposed based on VMD, RDIs extraction, mRMR feature selection, and grid search support vector machine (GSSVM) fault identification. To our best knowledge, this study is the first that investigates the superiority of RDIs in machinery fault diagnosis, thereby providing a new feature extraction strategy for intelligent fault diagnosis. Two experimental examples for different fault types of mechanical components including rolling bearing and gearbox demonstrated that the RDIs can effectively solve the deficiency of the TDIs, and has stronger distinguishing ability for the fault types. Also, the experimental comparison results reinforce the superiority of mRMR in selecting the useful RDIs.

The remainder of this paper is organized as follows: section 2 presents the proposed diagnostic framework and techniques involved; section 3 describes the related theories; section 4 presents the experimental demonstration, VMD-based RDIs construction, and mRMR-based RDI selection; section 5 describes the comparative studies; section 6 concludes this paper.

II. HYBRID INTELLIGENT FAULT DIAGNOSIS METHOD

In fault diagnosis filed, feature extraction and selection are important prior to fault identification. Feature extraction is the basis of feature selection, which directly affects the fault identification accuracy. Selecting an appropriate and discriminative feature set to reflect the operating condition of the machinery is critical in fault diagnosis. Considering the increasing number of fault types, fault severities, and fault orientations of machinery components, it is necessary to construct more fault features to represent useful fault information. In this study, a new feature extraction method based on RDIs is proposed. The RDIs can extract more useful fault information hidden in the vibration signals than TDIs, which is beneficial for fault identification. However, different types of RDIs have varying amounts of distinguishing ability for fault patterns. Therefore, the distinguishing ability of each RDI must be considered. Additionally, because the fault feature set is typically high dimensional and contain irrelevant and redundant features, a high-dimensional feature set tends to deteriorate the performance of fault identification. This applies for RDIs as well. Therefore, in this study, the mRMR method was used to select RDIs possessing higher degrees of distinguishing ability and eliminate redundant and irrelevant RDIs in high-dimensional feature sets. The key idea behind the mRMR approach is to maximize the relevance between features and output class labels, and minimize the redundancy between features.

Based on the aforementioned analysis, this study proposes an intelligent fault diagnosis approach for mechanical components based on VMD, RDIs extraction, mRMR feature selection, and GSSVM fault identification. First, the

VMD method was used to process the original vibration signals of rolling bearings and gears under different operating conditions. Using VMD method, each vibration signal can be decomposed into a number of intrinsic mode functions (IMFs). Subsequently, TDIs are redefined and multiple types of RDIs are constructed. Based on the original signals and IMFs, multiple RDIs as fault features are extracted. As such, a high-dimensional feature vector comprising $(n + 1) \times m$ RDIs is built, where n represents the number of IMFs, and m the number of RDIs of the original signals. Subsequently, the mRMR method is used to select the appropriate and discriminative features from the extracted RDIs to establish a new feature set that contains the informative fault information. The mutual information between each RDI and class has been calculated, and the maximum relevance RDI with the largest mutual information is first selected. Subsequently, the mutual information between every two RDIs are calculated, and the minimum redundant RDIs are then selected. Utilizing the mRMR feature selection, several top ranked RDIs are selected as fault features. Finally, the GSSVM pattern classifier is employed to recognize 12 operating conditions of rolling bearings and nine fault conditions of gears. The GSSVM is useful in solving small samples as well as nonlinear and high-dimensional pattern recognition. The selected RDIs are fed into the GSSVM to distinguish several fault types of rolling bearings and gears. By adopting the grid search method, we can obtain the best kernel parameter γ and penalty parameter C for the support vector machine (SVM). Specifically, a set of candidates is first selected for both C and γ. Using the training sample set, each pair of C and γ is evaluated by cross validation, and the pair with the highest accuracy is determined as the optimal parameters. As such, the optimal GSSVM model can be obtained, and the testing sample set is used to validate the optimal GSSVM model. Experimental results demonstrated that RDIs as fault features can successfully reveal the fault characteristics. Furthermore, RDIs can overcome the major shortcomings of TDIs, and the selected RDIs using mRMR can improve the fault identification accuracy of rolling bearings and gears. The flowchart of the proposed fault diagnostic framework is shown in Fig. 1.

III. RELATED THEORIES

A. VARIATIONAL MODE DECOMPOSITION

VMD is an adaptive and non-recursive signal decomposition method, which can decompose the original signal into narrow-band components with independent centre frequencies. In the VMD framework, the signal is decomposed into k discrete number of components, and the bandwidth of a component can be evaluated with constrained variational optimization problem, and the formulated constrained variational problem can be expressed as follows:

$$\min_{\{u_k\}\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}$$

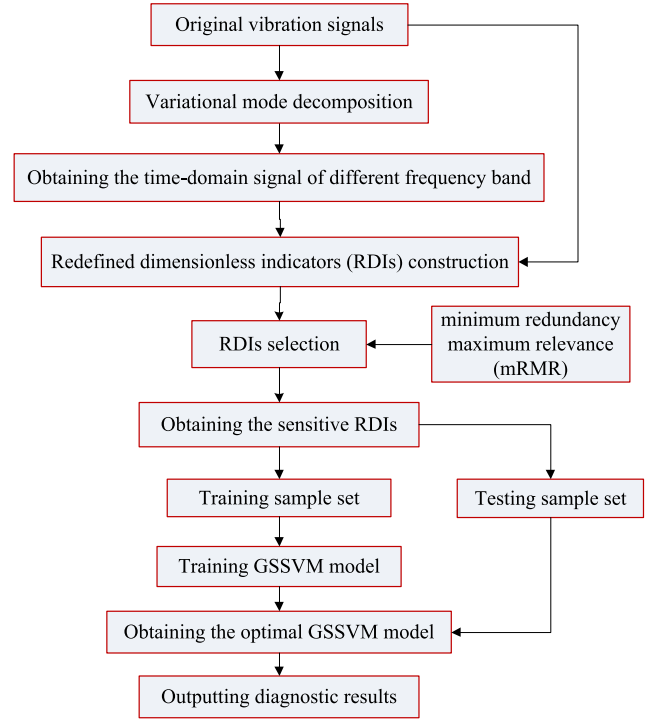$$s.t. \sum_k u_k = f \tag{1}$$

FIGURE 1. Flowchart of the proposed fault diagnosis method.

where u_k is the kth component of the signal, and the ω_k denotes center frequency of the kth component of the signal, f is the origin signal, δ is the Dirac distribution, t is the time script, and * denotes convolution. By introducing a quadratic penalty and Lagrangian multipliers, the above constrained optimization problem can be expressed as:

$$L(\{u_k\}, \{\omega_k\}, \lambda)$$

$$:= \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2$$

$$+ \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \tag{2}$$

where α denotes the balancing parameter of the data-fidelity constraint. Eq. (2) is then solved with the alternate direction method of multipliers, and then all the modes gained from solutions in spectral domain are written as:

$$u_k^{n+1}(\omega) = \frac{f(\omega) - \sum_{i \neq k} u_i(\omega) + \frac{\lambda(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{3}$$

where $f(\omega)$, $u_i(\omega)$, $\lambda(\omega)$ and $u_k^{n+1}(\omega)$ are the Fourier transforms of $y(t)$, $y_i(t)$, $\lambda(t)$ and $y_k^{n+1}(t)$. It is noted that Eq. (3) contains the Wiener filter structure. The component in time domain can be obtained from the real part of inverse Fourier transform of the filtered signal. More detailed information concerning the VMD can be found in the literature [8].

Through using the VMD method, multiple sub-signals under different frequency band are obtained. For each

sub-signal, 145 types of RDIs are calculated. The issue of how to calculate RDIs is provided in the following.

B. REDEFINED DIMENSIONLESS INDICATORS

Dimensionless indicators are better diagnosis characteristics, which do not require transformation and processing of the vibration signal. In practice, dimensionless indicators are sensitive to faults instead of working conditions, and they are not affected by the disturbance. Thus, they are employed widely in complex industrial conditions. Traditional dimensionless parameters are defined as follows [22]:

$$\zeta_x = \frac{\left[\int_{-\infty}^{+\infty} |x|^l p(x) dx\right]^{\frac{1}{l}}}{\left[\int_{-\infty}^{+\infty} |x|^m p(x) dx\right]^{\frac{1}{m}}} \quad (4)$$

where x denotes the vibration amplitude, $p(x)$ denotes the probability density function of vibration amplitude, and l and m are two different constants.

Details are given as follows:
(i) if $l = 2$, $m = 1$, we obtain the waveform indicator.
(ii) if $l \to \infty$, $m = 1$, we obtain the impulse indicator.
(iii) if $l \to \infty$, $m = \frac{1}{2}$, we obtain the margin indicator.
(iv) if $l \to \infty$, $m = 2$, we obtain the peak indicator.
(v) the kurtosis indicator is $K = \frac{\varphi}{X_{rms}^4}$. In the formula, φ is defined as $\varphi = \int_{-\infty}^{\infty} x^4 p(x) dx$. X_{rms} is defined as $X_{rms} = \sqrt{\int_{-\infty}^{\infty} x^2 p(x) dx}$.
(vi) the skewness indicator is $S = \frac{\sigma}{X_{rms}^3}$. In the formula, σ is defined as $\sigma = \int_{-\infty}^{\infty} x^3 p(x) dx$.

To date, there are existing six types of dimensionless indicators, namely waveform indicator, impulse indicator, margin indicator, peak indicator, kurtosis indicator, and skewness indicator. However, considering that the number of the fault types, fault severities and fault orientations of rolling bearings has gradually increased, more useful fault information are needed to extract from the vibration signals for identifying more fault categories. For this reason, in this study RDI is proposed to find more informative fault information hide in the vibration signals. The construction of the RDIs is based on the traditional dimensionless indicators. From the above definition of the dimensionless indicators, it can be seen that the dimensionless indicators are depended on the parameters l and m. Six types of existing dimensionless indicators are obtained when six groups l and m are determined. That is, if the value of the appropriate l and m is different, meanwhile, ensuring the numerator and the denominator of the definition formulas possessing the same dimension, more types of dimensionless indicators can be obtained. As such, we only choose different values of the l and m, a large number of new dimensional indicators, called as RDIs, can be obtained. RDIs provide a new feature extraction strategy for intelligent fault diagnosis. The definitions of the RDIs are as follows.

Definition 1: Similar to the definition of the waveform indicator, we let l and m vary from α to β, with a change

interval of λ for α and β, where α, $\beta \in [0.5, 5]$, $\lambda = 0.5$, and $\alpha \neq \beta$.

Definition 2: Similar to the definitions of the impulse indicator, margin indicator, and peak indicator, we let $l \to \infty, m = \gamma$, where the change interval is λ for $\gamma, \gamma \in [0.5, 5]$, and $\lambda = 0.5$.

Definition 3: Similar to the definitions of the kurtosis indicator and skewness indicator, we let l and m vary from ε to η, where the change interval is λ_1 for ε and the change interval is λ_2 for η. When the numerator is $\int_{-\infty}^{+\infty} x^\varepsilon p(x) dx$, the ε power of $\left[\int_{-\infty}^{+\infty} |x|^\eta p(x) dx\right]^{\frac{1}{\eta}}$ is used as the denominator, with the goal of ensuring that the numerator and denominator of the definition formula have the same dimension, where $\varepsilon \in [1, 5]$, $\eta \in [0.5, 5]$, $\lambda_1 = 1$, $\lambda_2 = 0.5$, and $\varepsilon \neq \eta$.

For the definition 1, definition 2 and definition 3, 90 types of dimensionless indicator, 10 types of dimensionless indicator and 45 types of the dimensionless indicator are obtained, respectively, thus a total of 145 types of dimensionless indicator are used as the fault features in this study. That is, the number of the RDIs is 145 (note that: the 145 RDIs contain the 6 TDIs).

Based on the aforementioned three definition types of the dimensionless indicators, new dimensionless indicators can be obtained, and we define such dimensionless indicators as redefined dimensionless indicators (RDIs), which can provide much more informative fault information regarding the operating condition of rolling bearings. The RDIs are respectively defined as follows:

$$\zeta_1 = \frac{\left[\int_{-\infty}^{+\infty} |x|^\alpha p(x) dx\right]^{\frac{1}{\alpha}}}{\left[\int_{-\infty}^{+\infty} |x|^\beta p(x) dx\right]^{\frac{1}{\beta}}} = \frac{\sqrt[\alpha]{E(|x|^\alpha)}}{\sqrt[\beta]{E(|x|^\beta)}} \quad (5)$$

$$\zeta_2 = \lim_{l \to \infty} \frac{\left[\int_{-\infty}^{+\infty} |x|^l p(x) dx\right]^{\frac{1}{l}}}{\left[\int_{-\infty}^{+\infty} |x|^\gamma p(x) dx\right]^{\frac{1}{\gamma}}} = \lim_{\varepsilon \to \infty} \frac{\sqrt[l]{E(|x|^l)}}{\sqrt[\gamma]{E(|x|^\gamma)}} \quad (6)$$

$$\zeta_3 = \frac{\int_{-\infty}^{+\infty} x^\varepsilon p(x) dx}{\left[\int_{-\infty}^{+\infty} |x|^\eta p(x) dx\right]^{\frac{\varepsilon}{\eta}}} = \frac{E(x^\varepsilon)}{\left[E(|x|^\eta)\right]^{\frac{\varepsilon}{\eta}}} \quad (7)$$

where x denotes the vibration amplitude, $p(x)$ denotes the probability density function of vibration amplitude, and α, β, γ, ε, η, λ, λ_1, λ_2 denotes different constants. When different constants are chosen, meanwhile, ensuring the numerator and the denominator of the definition formulas possessing the same dimension, more types of dimensionless indicators can be obtained. As for vibration signals, the integrals in the above formulation can be easily calculated by the samples.

C. MAX-RELEVANCE AND MIN-REDUNDANCY

The core concept of the mRMR approach is to select a feature subset set that best characterizes the statistical property of a target classification variable, subject to the constraint that

these features are mutually as dissimilar to each other but marginally similar to the classification variable. Hitherto, several different forms of mRMR have been proposed, where "relevance" and "redundancy" were defined using factors such as mutual information, correlation, and distances. In this study, a mutual-information-based mRMR criterion is used to measure the similarity level between features and the correlation level between a feature and class.

Let $T = \{x_{ij}\}_{N \times m}$ denote a sample matrix, where x_{ij} is the *ith* sample value of the *jth* feature; N the total number of samples; and m the total number of features. The mRMR approach can select an optimal p $(p < m)$ dimension feature subspace to maximize the correlation between a feature and class, and minimize the correlation between each feature. Given two jointly discrete random variables X and Y, their mutual information is defined as follows:

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \qquad (8)$$

where $p(x, y)$ is the joint probability mass function of X and Y; $p(x)$ and $p(y)$ denote the marginal probability mass functions of X and Y, respectively.

Suppose a_i represents the individual feature and c the class. The max-relevance criterion can be expressed as

$$\begin{cases} \max D\left(S, c\right) \\ D = \dfrac{1}{|S|} \sum_{a_i \in s} I\left(a_i; c\right) \end{cases} \qquad (9)$$

where $|S|$ is the dimension of the feature space S, $I\left(a_i; c\right)$ is the mutual information between the individual feature a_i and class c and D is a vector containing the mutual information values between the features and class.

In addition, because the features selected based on the max-relevance criterion may be redundant, the min-redundancy criterion was used to minimize the redundancy between features, which is expressed as

$$\begin{cases} \min R\left(S\right) \\ R = \dfrac{1}{|S|^2} \sum_{a_i, a_j \in S} I\left(a_i; a_j\right) \end{cases} \qquad (10)$$

where $I\left(a_i; a_j\right)$ is the mutual information between features a_i and a_j, and R is a vector containing the mutual information values between the features.

Combining the max-relevance and min-redundancy criteria mentioned above, the mRMR feature selection method is formed, which can optimize D and R simultaneously.

$$\begin{cases} \max \Phi\left(D, R\right) \\ \Phi = D - R \end{cases} \qquad (11)$$

Using formula (9), the problems of the max-relevance and min-redundancy criteria can be transformed to maximize the mutual information difference $\Phi\left(D, R\right)$. Typically, the incremental search method is used to obtain the near optimal features. Suppose $m - 1$ features have been selected to establish

feature set S, and the following criteria must be satisfied to select the *m*-th feature.

$$\max \left\{ I\left(a_j; c\right) - \frac{1}{m-1} \sum_{a_i \in S} I\left(a_j; a_i\right) \right\} \qquad (12)$$

Formula (10) can guarantee that the correlation between the feature and variables achieves the maximum, and that the correlation between each feature achieves the minimum. A more detailed and comprehensive description concerning the mRMR feature selection method is available in [29].

D. GRID SEARCH SUPPORT VECTOR MACHINE

The SVM performs classification by constructing a hyperplane that optimally separates the data into two categories. We assume that a training set is given by $\{x_i, y_i\}_{i=1}^l$. The critical problem of SVM is to obtain the optimal separation hyperplane. This problem may be handled by solving the following optimization problem [33]:

$$\min_{w, b, \xi} \frac{1}{2} w^T w + C \sum_{i=1}^l \xi_i, \qquad (13)$$

subject to

$$y_i(w^T \varphi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \ i = 1, \ldots, l \qquad (14)$$

where w is normal to the hyperplane, C is a penalty parameter, which defines how important it is to avoid misclassification errors, ξ_i is a positive slack variable respectively. SVM obtains a linear separating hyperplane with maximal margin by using function φ to map the training instance to a high-dimensional space, which assists in solving the nonlinearly separable problem. A radial basis function is adopted for building function φ, which has the form:

$$K(x, x_i) = \exp\{-\gamma \parallel x - x_i \parallel^2\} \qquad (15)$$

where γ is the kernel parameter.

In the previous research, SVM was used to solve the classification problems. However, SVM still has some disadvantages owing to the questionable determination of the parameters such as penalty parameter C and kernel parameter γ. For the traditional SVM, the setting of parameters is affected by the defects caused by human subjectivity. Improper parameters directly decrease the final classification accuracy. In this study, the grid search algorithm is used to optimize the penalty parameter C and kernel parameter γ [34]. Specifically, a set of candidates is first selected for both C and γ. Using the training sample set, each pair of C and γ is evaluated by cross validation, and the pair with the highest accuracy is determined as the optimal parameters. As such, the optimal GSSVM model can be obtained.

IV. CASE STUDIES

In this section, the validity of our proposed method is assessed through two experimental examples. The first experiment was conducted from Case Western Reserve University for

TABLE 1. Detailed arrangements of the experimental datasets for fault identification.

Health conditions	Defect size (inches)	Abbreviation	Category labels	Sample number	Number of training samples	Number of testing samples
Normal condition		NB	1	50	35	15
Ball fault	0.007 mils	B-007	2	50	35	15
	0.014 mils	B-014	3	50	35	15
	0.021mils	B-021	4	50	35	15
	0.028mils	B-028	5	50	35	15
Inner race fault	0.007 mils	IR-007	6	50	35	15
	0.014 mils	IR-014	7	50	35	15
	0.021mils	IR-021	8	50	35	15
	0.028mils	IR-028	9	50	35	15
Outer race fault	0.007 mils	OR-007	10	50	35	15
	0.014 mils	OR-014	11	50	35	15
	0.021mils	OR-021	12	50	35	15

12 different operating conditions of rolling bearings, and the second experiment was conducted for nine different operating conditions of gears. For each experiment, the VMD method was first employed to decompose the original signals. The mRMR feature selection method was subsequently applied to select several sensitive RDIs and finally fed into the multiclassifier GSSVM for pattern recognition. For each experiment, comparative studies between TDIs and RDIs and for RDI selection methods were conducted.
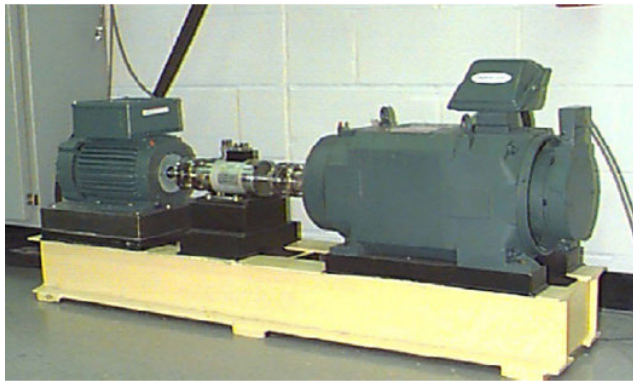
FIGURE 2. Experimental platform device.

A. CASE 1: ROLLING BEARING FAULT DIAGNOSIS

In this case, the vibration fault signals of rolling bearings of the Electrical Engineering Laboratory of Case Western Reserve University were used to validate the effectiveness of the proposed method. Fig. 2 shows the experimental platform device [35]. This case employed drive-end bearing data for a shaft speed of 1750 rpm. Three bearing fault types (ball fault, inner race fault, and outer race fault) with various levels of fault severity 0.007–0.028 mils were introduced using electro-discharge machining. A total of 12 different operating conditions were tested, which were as follows: normal condition, ball fault (0.007 mils), ball fault (0.014 mils), ball fault (0.021 mils), ball fault (0.028 mils), inner race fault (0.007 mils), inner race fault (0.014 mils), inner race fault (0.021 mils), inner race fault (0.028mils), outer race fault (0.007 mils), outer race fault (0.014 mils), and outer race fault (0.021 mils). The fault seeded at the outer race was placed at

a position equivalent to the 6:00 o'clock time configuration. For convenience, these 12 operating conditions are expressed as *N, B-007, B-014, B-021, B-028, IR-007, IR-014, IR-021, IR-028, OR-007, OR-014,* and *OR-021*, and labeled as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12, respectively.

For each operating condition, the first 120,000 points were divided into 50 sub-signals, where each sub-signal contained 2400 points. Thus, the whole data set consisted of 600 samples. The time-domain vibration waves of a single sample for the 12 operating conditions are presented in Fig. 3. The detailed arrangements of the experimental datasets for fault identification are shown in Table 1. Subsequently, VMD was applied to decompose the 600 original signals separately. The number of decomposition levels was set as six. Using the VMD method, the original signals were decomposed into several subsignals. Owing to space limitations, only the decomposition signals of the ball fault (0.007 mils) are presented in Fig. 4, where the x-axis represents the number of sampling points and the y-axis the IMF component. In this experiment, the first 35 groups of vibration data in each state were used as the training data. The remaining 15 groups of vibration data were used to verify the effectiveness of the proposed method.

1) RDI EXTRACTION AND SELECTION

Based on the original signals and VMD decomposition signals, the values of the previously defined RDIs of each IMF and original signals were calculated. From the analysis above, we can obtain a feature set containing 145 RDIs from the original signals and 870 RDIs from the VMD decomposed signals. Therefore, 1015 dimensionless indicators were extracted as a feature subset to diagnose the 12 fault types of rolling bearings. However, although 1015 RDIs as a high-dimension feature set increase the useful information for fault diagnosis, they inevitably generate redundant information. Redundant information may reduce the accuracy and efficiency of fault diagnosis. Therefore, the mRMR approach was used to select the most discriminant features, which were highly effective for the pattern classifier when identifying fault types.

Step 1: The mutual information between each RDI and class is calculated according to formula (9). The RDI
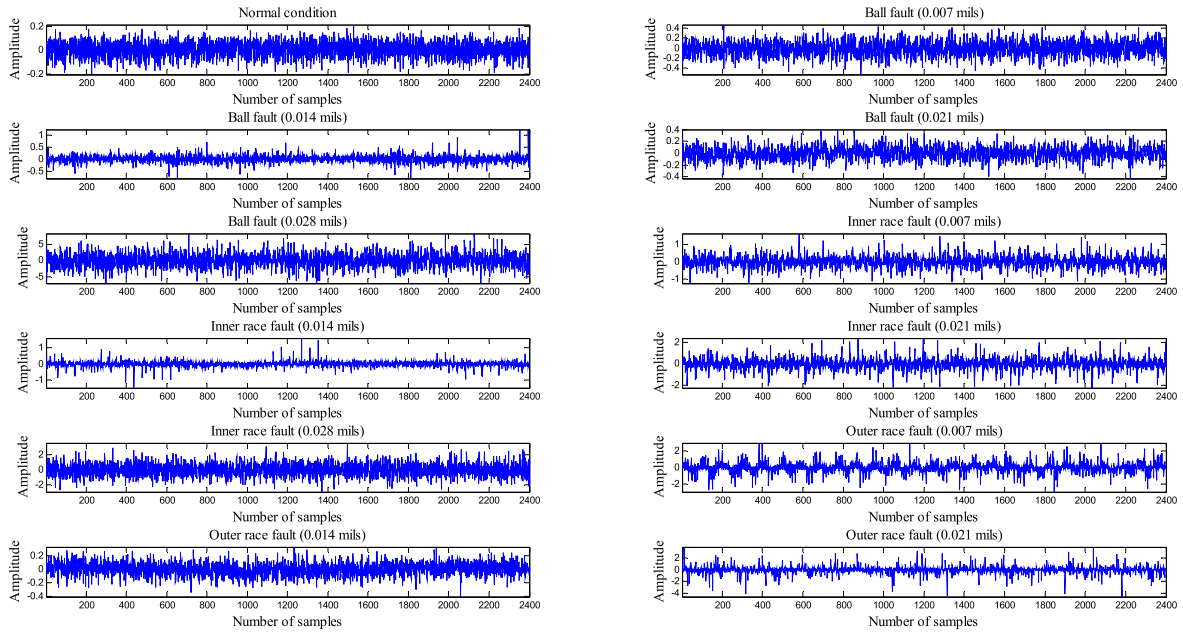
FIGURE 3. Original vibration signals of a single sample for the 12 different operating conditions of rolling bearing.
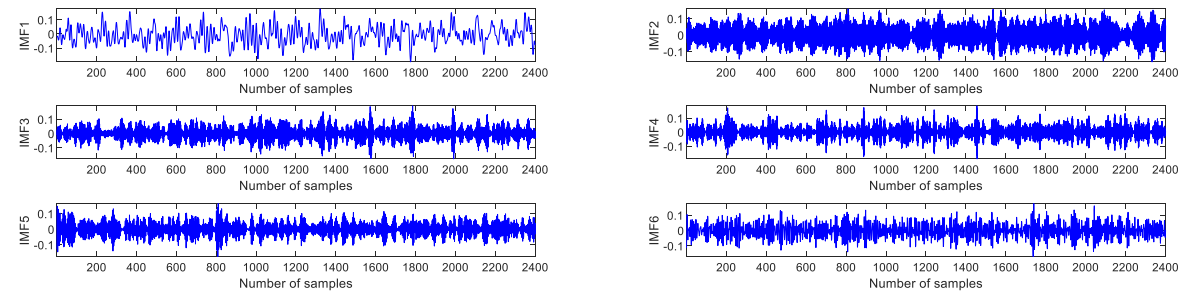

FIGURE 4. Decomposed signals under ball fault (0.007 mils).

possessing the largest mutual information value is selected as the most sensitive feature and joined into an empty set {S}, and the mutual information values of the remaining RDIs are stored in a vector D, which are then used to measure the relevance of the RDI and class.

Step 2: The mutual information values between each remaining RDI and the RDI in set {S} are calculated according to formula (10) and then stored in vector R. The mutual information value in vector R is used to measure the redundancy of the RDIs. We used the vector D to subtract the vector R and select the RDI possessing the largest difference as the sensitive feature to join into the RDI set {S}.

Step 3: Judging the number of RDIs in the RDI set {S}, if the number of RDIs is lower than the setting value K, we should return to step 2. Otherwise, if the number of RDIs is equal to K, we output the RDI set {S} containing K RDIs as the final sensitive features. The setting value K represents the number of sensitive RDIs that we must select from the original feature set. In this case, K is set as 15.

Using the mRMR method, 15 sensitive RDIs were selected from the 1015 RDIs. The detailed mutual information values of the selected 15 RDIs are shown in Table 2.

TABLE 2. Mutual information values of the selected 15 RDIs.

Serial number	1	2	3	4	5
mutual information value	1.4457	0.5991	0.7961	0.7654	0.7289
Serial number	6	7	8	9	10
mutual information value	0.6235	0.6044	0.5828	0.5788	0.5978
Serial number	11	12	13	14	15
mutual information value	0.586	0.6029	0.6027	0.5568	0.5545

2) COMPARATIVE STUDIES BETWEEN RDIs AND TDIs

To quantify the distinguishability of the TDIs and the proposed RDIs for different operating conditions of rolling bearing, the distance ratios (DR) of between-class and within-class distances for the TDIs and RDIs, respectively, were calculated. A dimensionless indicator possessing a larger DR value means that it can maximize the between-class distance and minimize the within-class distance, indicating a better performance in distinguishing different operating conditions. The DR values of the 145 RDIs are shown in Fig. 5. Among the 145 RDIs in Fig. 5, the six TDIs are marked in red, and the remaining dimensionless indicators (NDIs) are marked in blue. By analyzing the DR values of the 145 RDIs, it is observed that although there are 83 dimensionless indicators
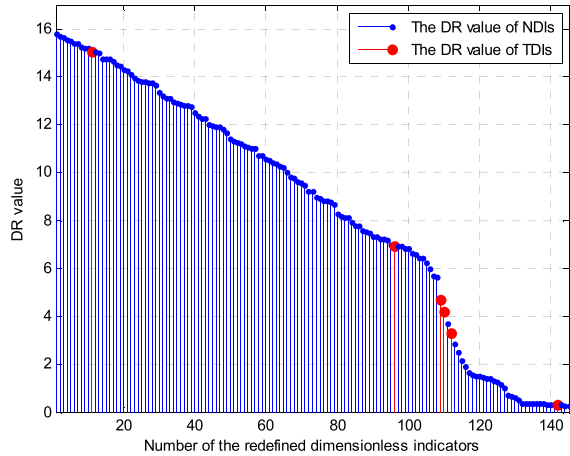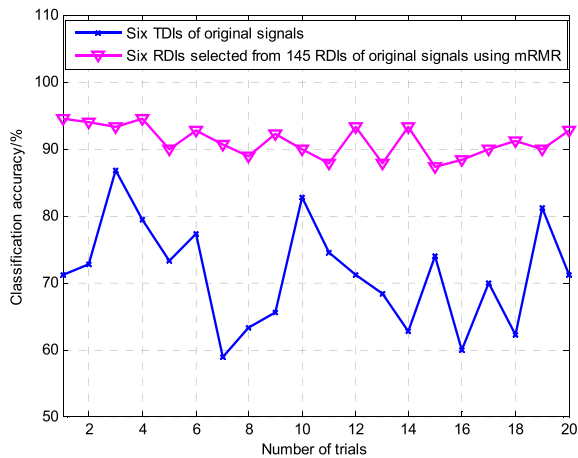
FIGURE 5. DR value of the RDIs.

FIGURE 6. Diagnosis accuracy for 20 trials.

with DR values greater than 8, there is only one TDI among them. Hence, the DR values of most NDIs were greater than those of TDIs, indicating that the RDIs have better distinguishability for different fault types than the TDIs. To further illustrate the advantages of the RDIs over the TDIs, the classification accuracy of the RDIs and TDIs was investigated under two circumstances: 1) six TDIs of the original signals; and 2) six RDIs selected from 145 RDIs of the original signals using the mRMR method. The GSSVM as a pattern classifier was used to perform the accuracy comparison of the RDIs and TDIs. For each circumstance, the diagnosis accuracy was obtained by repeating the testing procedures 20 times to reduce the effect of randomness on the results. Fig. 6 shows the classification results under the two circumstances for 20 trials. Additionally, the comparative indexes from 20 trials are listed in Table 3, which includes the maximum, minimum, mean, and standard deviation of the classification accuracy. As shown in Table 3, the maximum accuracy of 94.44%, minimum accuracy 87.22%, and average accuracy 91.11% were obtained using the RDIs, which were higher than those using the TDIs. Additionally, the standard deviation of 2.37% of the classification accuracy with the RDIs was significantly lower than that with the TDIs. The comparison results above

TABLE 3. Diagnosis accuracy comparison between TDIs and RDIs.

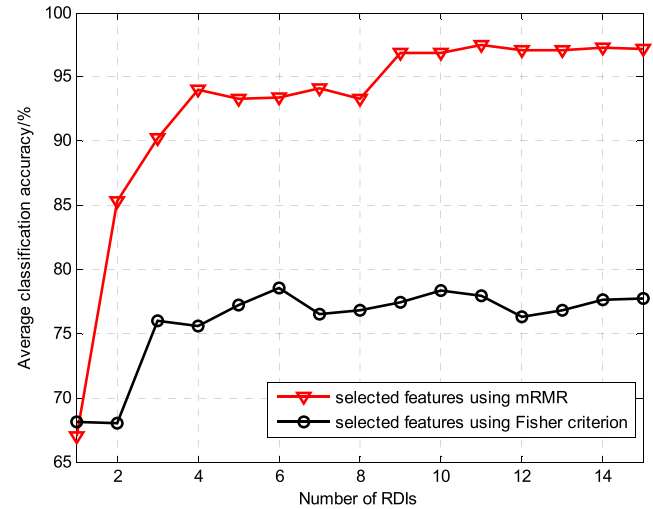| The selected features	Diagnosis accuracy/%			
	Max	Mean	Min	Std
Six TDIs of original signals	86.67	71.31	58.89	7.72
Six RDIs selected from 145 RDIs of original signals using mRMR	94.44	91.11	87.22	2.37

FIGURE 7. Average classification accuracy under different number of RDIs.

indicate that the RDIs are superior to the TDIs in terms of not only the average diagnosis accuracy, but also the stability of the diagnosis result.

3) COMPARATIVE STUDIES FOR RDI SELECTION METHOD

To verify the superiority of the mRMR feature selection method, a comparison between the Fisher criterion (FC) and the mRMR feature selection was performed to select sensitive RDIs from the 1015 RDIs. Twenty trials were conducted to reduce the effect of randomness on the results. Table 4 shows the detailed diagnosis results with the first 15 RDIs selected using the two feature selection methods in 20 trials, and the average classification accuracies using different numbers of RDIs are shown in Fig. 7. The diagnosis accuracy using the mRMR was higher than that using the FC feature selection. As shown in Fig. 7, for the 15 RDIs selected using the mRMR method, the classification accuracy presented a fluctuation change with the increasing number of RDIs at the early stage. However, when the number of input RDIs increased to 11, the average classification accuracy reached the maximum of 97.47%. Subsequently, the average classification accuracy changed slightly when the number of input RDIs continued to increase. The classification results of the first 11 RDIs corresponding to the maximum accuracy of 97.47% for 12 different operating conditions of rolling bearing are shown in Fig. 8. For the faulty conditions *B-014*, *B-028*, *IR-007*, *IR-014*, *IR-021*, *OR-007*, and *OR-021*, the fault classification rates reached 100%. For the normal condition and the faulty conditions *B-007*, *B-021*, *IR-028*, and *OR-014*, the fault classification rates were 99%, 98.67%, 88.33%, 94.33%,
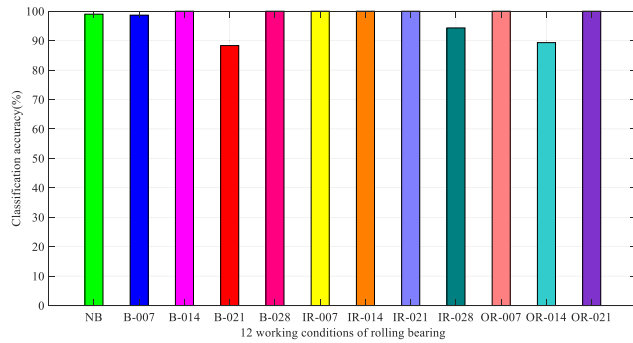
FIGURE 8. Classification results of the first 11 RDIs selected from 1015 RDIs using mRMR method.
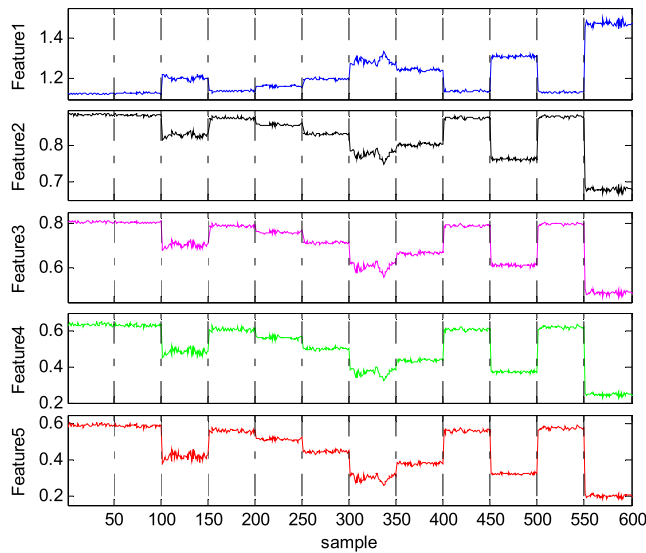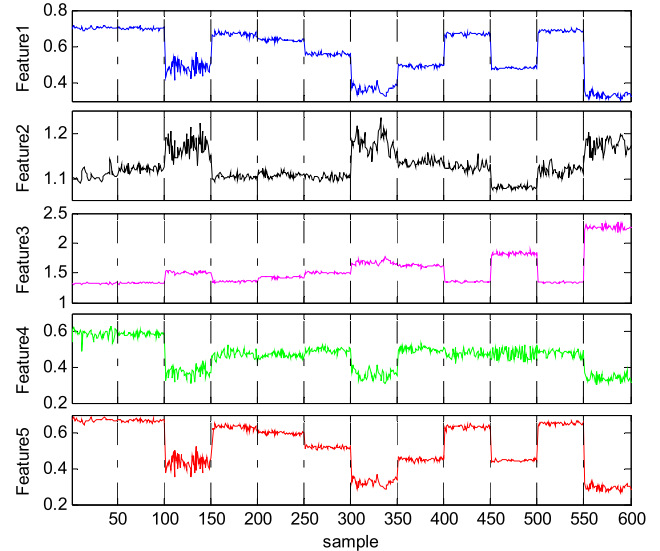

FIGURE 9. First five RDIs selected by FC method.

FIGURE 10. First five RDIs selected by mRMR method.

FIGURE 11. Gearbox system.

and 89.33%, respectively. The overall recognition rate was 97.47%. However, using the FC feature selection method, the highest average classification accuracy was only 78.56%, which was significantly lower than the 97.47% obtained by mRMR method. Additionally, the mRMR method achieved a lower standard deviation (1.14%). Hence, the accuracy rate of the proposed method was more stable compared with that of the FC method. In addition, for the proposed method, the original set of 1015 RDIs was reduced around 98.92% but still reached a high identification accuracy of 97.47%.

The classification effectiveness of the first five RDIs selected by different feature selection methods was illustrated in Figs. 9–10. As shown, the mRMR method could obtain the relevant and nonredundant RDIs to improve the fault identification accuracy. In contrast to the mRMR method, the FC method only considers the relevancy of each RDI measured with respect to the class labels, thereby causing similarity or redundant features to appear in the selected RDIs, which causes the low identification accuracy.

B. CASE 2: GEAR FAULT DIAGNOSIS

In this case, the vibration signals of the gear were collected from a benchmark two-stage gearbox [36], as shown

in Fig. 11. A 32-tooth pinion and an 80-tooth gear were installed on the first stage input shaft. The second stage comprised a 48-tooth pinion and 64-tooth gear. The input shaft speed and gear vibration signals were measured using a tachometer and accelerometer, respectively. Nine different gear conditions were tested, as follows: healthy condition, missing tooth, root crack, spalling, and chipping tip with five different levels of severity. The vibration signals were recorded with a sampling frequency of 20 KHz. For convenience, these nine operating conditions are expressed as NC, MT, RC, SP, C1, C2, C3, C4, and C5, and labeled as 1, 2, 3, 4, 5, 6, 7, 8, and 9, respectively.

For each operating condition, 104 signals were collected using the gearbox system and each signal contained 3600 sample points. Hence, 104 groups of vibration data were measured for each state, and 936 samples of the nine different operating conditions were collected for further investigations. The time-domain vibration waves of a single sample for the nine operating conditions of the gear are presented in Fig. 12. The detailed arrangements of the experimental datasets for fault identification are shown in Table 5. The VMD method was subsequently applied to decompose the 936 original
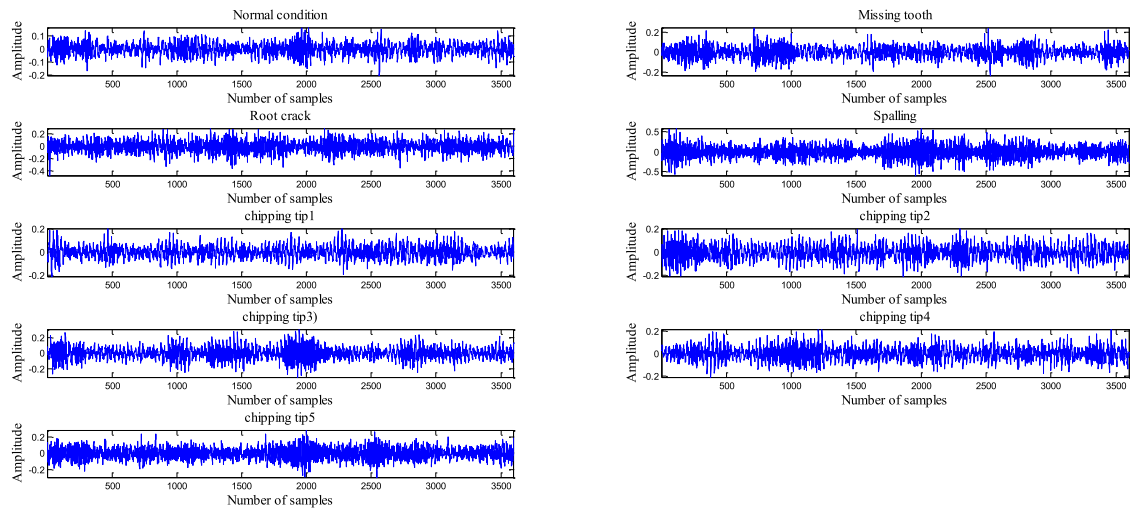
FIGURE 12. Original vibration signals of a single sample for the nine different gear operating conditions.

TABLE 4. Detailed diagnosis results of using different feature selection methods.

Number of RDIs	mRMR method classification accuracy/%				FC method classification accuracy/%			
	Max	Mean	Min	Std	Max	Mean	Min	Std
1	73.33	67.00	62.78	2.68	75	68.17	61.67	3.89
2	90	85.36	79.44	2.63	71.67	67.97	62.78	2.49
3	95	90.22	87.22	2.33	81.11	76.03	72.78	2.09
4	96.67	94.00	91.67	1.44	79.44	75.64	68.89	2.13
5	96.67	92.17	88.89	2.01	81.67	77.25	67.78	3.05
6	96.11	93.42	90.00	1.87	**83.33**	**78.56**	**75.00**	**2.22**
7	97.22	94.14	90.56	1.56	82.22	76.53	71.67	2.50
8	97.22	93.25	87.78	2.54	80.00	76.86	71.11	2.16
9	98.89	96.86	94.44	1.20	82.22	77.39	73.33	3.04
10	98.89	96.92	95.56	0.85	83.33	78.36	72.78	2.20
11	**99.44**	**97.47**	**95.56**	**1.14**	82.78	77.94	72.78	2.70
12	98.89	97.08	93.89	1.30	81.11	76.33	71.11	2.60
13	98.89	97.11	93.89	1.39	82.22	76.81	70.56	3.12
14	99.44	97.33	96.11	0.91	82.78	77.67	73.33	2.49
15	99.44	97.14	94.44	1.17	81.11	77.72	73.33	2.43

signals separately. The number of the decomposition level was set as six. Using the VMD method, the original signals were decomposed into several subsignals. Owing to space limitations, only the decomposition signals of the missing tooth gear are presented in Fig. 13, where the x-axis represents the number of sampling points and the y-axis the IMF component. In this experiment, the first 70 groups of vibration data in each state were used as the training data. The remaining 34 groups of the vibration data were used to verify the effectiveness of the proposed method.

1) RDI EXTRACTION AND SELECTION
Based on the original signals and the VMD signals, the values of the previously defined RDIs of each IMF and original signals were calculated. From the analysis above, a feature set containing 145 RDIs from the original signals and 870 RDIs from the VMD decomposed signals can be obtained. Therefore, 1015 dimensionless indicators were extracted as a

feature subset to diagnose the nine gear fault types. Although the 1015 RDIs as a high-dimension feature set increased the amount of useful information for fault diagnosis, some redundant information was generated inevitably, which may reduce the accuracy and efficiency of fault diagnosis. Therefore, the mRMR approach was used to select the most discriminant features from the 1015 RDIs. Using the mRMR method, 10 sensitive RDIs were selected from the 1015 RDIs. The detailed mutual information values of the selected 10 RDIs are shown in Table 6.

2) COMPARATIVE STUDIES BETWEEN TDIs AND RDIs
To quantify the distinguishability of the TDIs and the proposed RDIs for different gear operating conditions, the DRs of between-class and within-class distances for the TDIs and RDIs, respectively, were calculated. A dimensionless indicator possessing a larger DR value means that it can maximize the between-class distance and minimize the within-class distance, indicating a better performance in distinguishing different operating conditions. The DR values of the 145 RDIs are shown in Fig. 14. Among the 145 RDIs in Fig. 5, six TDIs are marked in red, and the NDIs are marked in blue. Analyzing the DR values of the 145 RDIs, it is clear that the DR values of most NDIs are greater than that of the TDIs, indicating that the RDIs have better distinguishability for different fault types than the TDIs. To further illustrate the advantages of the RDIs over the TDIs, the classification accuracy of the RDIs and TDIs was investigated under two circumstances: 1) six TDIs of the original signals; and 2) six RDIs selected from 145 RDIs of the original signals using the mRMR method. The GSSVM as a pattern classifier was used to compare the accuracy of the RDIs and TDIs. For each circumstance, the diagnosis accuracy was obtained by repeating the testing procedures 20 times to reduce the effect of randomness on the results. Fig. 15 shows the classification results under the two circumstances for 20 trials. Additionally, the comparative

TABLE 5. Detailed arrangements of the experimental datasets for fault identification.

Health conditions	Abbreviation	Category labels	Sample number	Number of training samples	Number of testing samples
Normal condition	NC	1	104	70	34
Missing tooth	MT	2	104	70	34
Root crack	RC	3	104	70	34
Spalling	SP	4	104	70	34
Chipping tip with five different levels of severity	C1	5	104	70	34
	C2	6	104	70	34
	C3	7	104	70	34
	C4	8	104	70	34
	C5	9	104	70	34
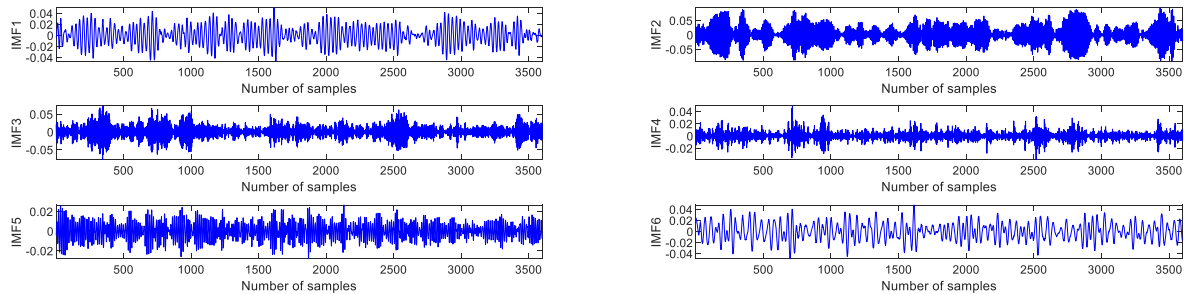
FIGURE 13. Decomposed signals under missing tooth.

FIGURE 14. DR value of the RDIs.

TABLE 6. Mutual information values of the selected 10 RDIs.

Serial number	1	2	3	4	5
mutual information value	1.0652	0.6324	0.6656	0.6831	0.6365
Serial number	6	7	8	9	10
mutual information value	0.5566	0.5687	0.5467	0.5523	0.5672

TABLE 7. Diagnosis accuracy comparison between TDIs and RDIs.

The selected features	Diagnosis accuracy/%			
	Max	Mean	Min	Std
Six TDIs of original signals	73.86	67.29	61.76	3.03
Six RDIs selected from the 145 RDIs of original signals using mRMR	77.78	72.01	67.65	3.14

TABLE 8. Detailed diagnosis results from using different feature selection methods.

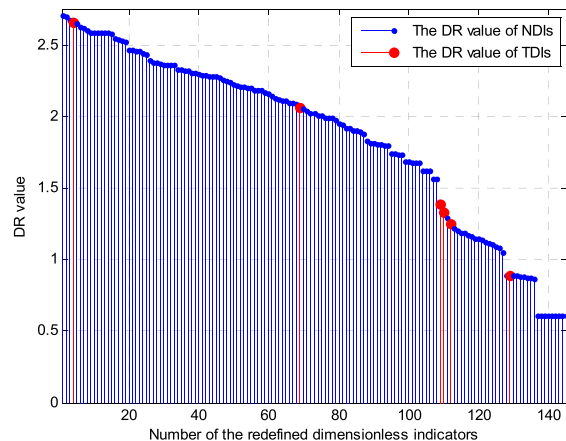| Number of features	The proposed method				Fisher criterion			
	classification accuracy/%				classification accuracy/%			
	Max	Mean	Min	Std	Max	Mean	Min	Std
1	65.03	61.88	56.86	2.02	64.38	60.90	57.84	1.93
2	87.58	84.43	82.03	1.55	73.2	66.72	62.09	2.68
3	93.79	90.77	87.58	1.52	75.82	72.52	67.65	2.24
4	95.75	93.43	91.50	1.02	78.76	73.86	68.63	2.43
5	**98.69**	**97.12**	**94.77**	**1.09**	77.78	73.96	68.63	2.14
6	98.69	96.59	94.12	1.22	82.03	77.16	69.28	3.16
7	98.04	96.65	94.12	1.08	84.31	80.34	76.8	2.23
8	98.37	96.54	94.12	1.10	84.64	80.20	75.49	2.36
9	98.37	96.46	93.79	1.27	83.66	80.65	75.49	2.01
10	98.04	96.24	93.79	1.23	**84.31**	**81.27**	**77.45**	**1.79**

indexes in the 20 trials are listed in Table 7, which includes the maximum, minimum, mean, and standard deviation of the classification accuracy. It can be observed that the maximum accuracy of 77.78%, minimum accuracy 67.65%, and average accuracy 72.01% were obtained using the RDIs, which were higher than those obtained using the TDIs. Additionally, a 3.14% standard deviation in classification accuracy with the RDIs is approximately equal to a 3.03% standard deviation in classification accuracy with the TDIs. The comparison results indicate that the RDIs are superior to the TDIs in terms of the average diagnosis accuracy.

3) COMPARATIVE STUDIES FOR RDI SELECTION METHOD
To verify the superiority of the mRMR feature selection method, a comparison between the FC and the mRMR feature

selection was performed to select sensitive RDIs from the 1015 RDIs. Twenty trials were conducted to reduce the effect of randomness on the results. Table 8 shows the detailed diagnosis results with the first 10 RDIs selected by the two feature selection methods above in 20 trials, and the average
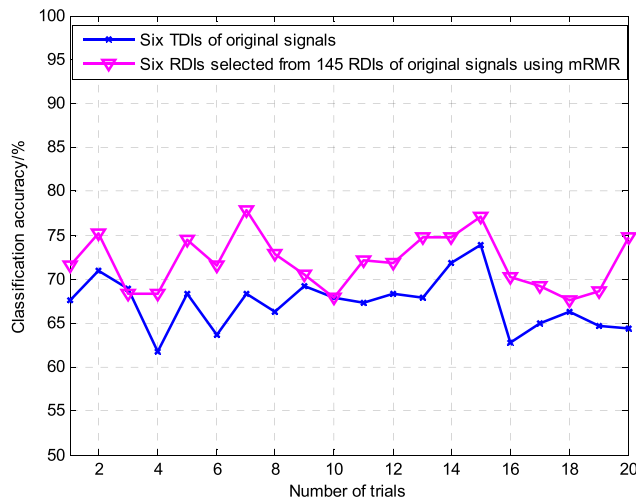
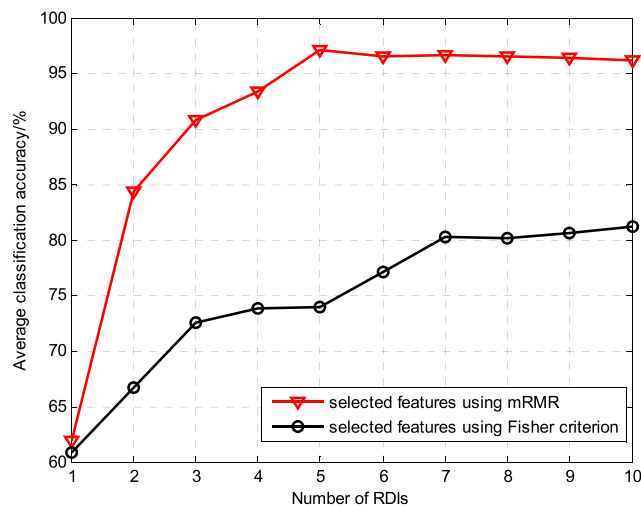**FIGURE 15.** Diagnosis accuracy for 20 trials.

FIGURE 16. Average classification accuracy under different number of RDIs.

classification accuracies using different numbers of RDIs are shown in Fig. 16. The diagnosis accuracy using the mRMR was higher than that of using the FC feature selection. As shown in Fig. 16, for the 10 RDIs selected by the mRMR method, the classification accuracy exhibited a fluctuation change with the increasing number of RDIs at the early stage. However, when the number of input RDIs increased to five, the average classification accuracy was the maximum at 97.12%. Subsequently, the average classification accuracy changed slightly when the number of input RDIs continued to increase. The classification results of the first five RDIs corresponding to the maximum accuracy of 97.12% for nine different gear operating conditions are shown in Fig. 17. For the normal condition and the faulty conditions C3 and C5, the fault classification rates reached 100%. For the faulty conditions MT, RC, SP, C1, C2, and C4, the fault classification rates were 96.03%, 96.62%, 97.65%, 99.41%, 92.35%, and 91.91%, respectively. The overall recognition rate was 97.12%. However, using the FC feature selection
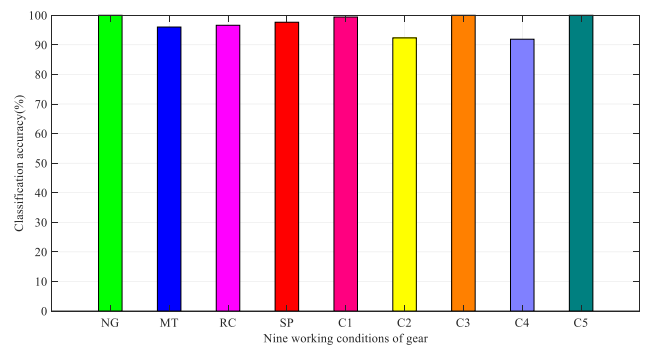

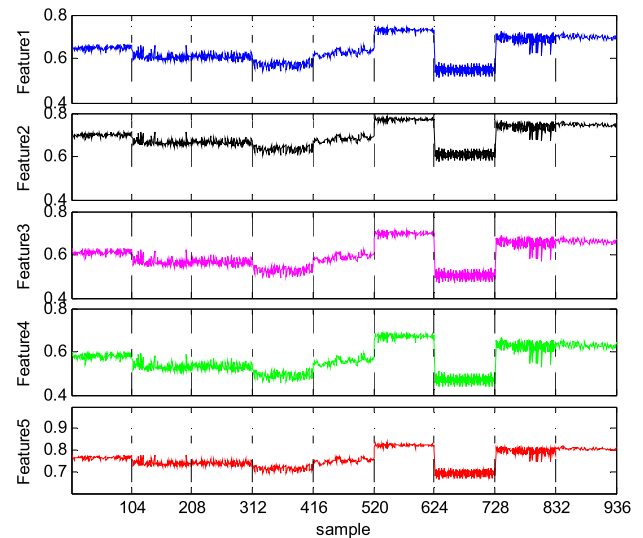**FIGURE 17.** Classification results of the first 11 RDIs selected from 1015 RDIs using mRMR method.

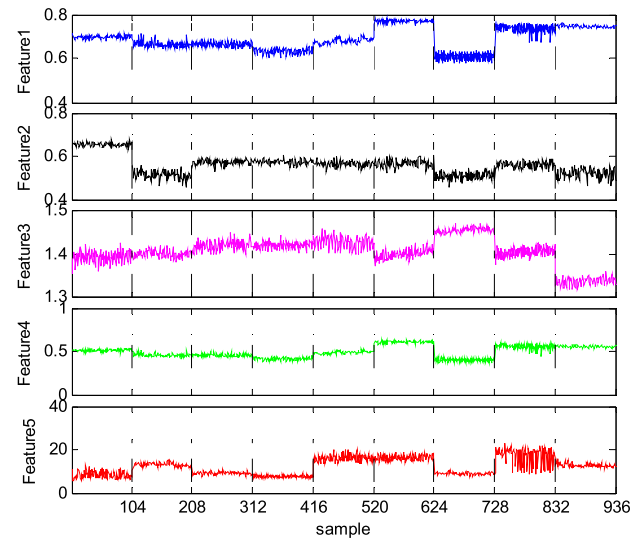**FIGURE 18.** First five RDIs selected using FC method.

FIGURE 19. First five RDIs selected using mRMR method.

method, the highest average classification accuracy was only 81.27%, which was significantly lower than the 97.12% obtained by the mRMR method. Additionally, the mRMR method achieved the smaller standard deviation (1.09%).

Hence, the accuracy rate of the proposed method was more stable compared with that of the FC method. In addition, for the proposed method, the original set of 1015 RDIs was reduced around 99.01% while reaching a high identification accuracy of 97.12%.

The classification effectiveness of the first five RDIs selected by different feature selection methods was illustrated in Figs. 18–19. As shown, the mRMR method could obtain relevant and nonredundant RDIs to improve the fault identification accuracy. In contrast to the mRMR approach, the FC method only considers the relevancy of each RDI measured with respect to the class labels, thereby causing similarity or redundant features to appear in the selected RDIs, which causes the low identification accuracy.

V. CONCLUDING REMARKS

With the increasing number of fault types, fault severities, and fault orientations of machinery components in industrial applications, it is necessary to construct more fault features to represent the useful fault information hidden in the vibration signals. Hence, this study provides a new fault feature extraction strategy for the fault diagnosis of machinery components based on RDIs. The mRMR approach was introduced to select sensitive RDIs with higher distinguishability as the input of the GSSVM. Using the mRMR approach, the mutual information of each RDI was calculated, and a set of maximally relevant and minimally redundant RDIs was obtained. Two experiments based on 12 different operating conditions of rolling bearings and nine different operation conditions of gearboxes were performed to verify the effectiveness of the proposed method. Comparative experiments demonstrated that the RDIs performed better than the TDIs and could improve the diagnosis accuracy. Experimental results showed that our proposed method could successfully differentiate 12 and nine faulty conditions of rolling bearings and gears, respectively, with the average accuracy of 97.47% with 11 RDIs and 97.12% with 5 RDIs. Although our proposed method could yield a satisfactory diagnosis accuracy, human subjectivity still existed in the parameter settings of RDI definitions. The effectiveness of RDIs using different parameters is therefore unknown. Further studies will be performed to test the effectiveness of the proposed method with different parameter settings in RDI definitions.

REFERENCES

[1] Y. Lei, Z. Qiao, X. Xu, J. Lin, and S. Niu, "An underdamped stochastic resonance method with stable-state matching for incipient fault diagnosis of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 94, pp. 148–164, Sep. 2017.

[2] H. Pan, Y. Yang, X. Li, J. Zheng, and J. Cheng, "Symplectic geometry mode decomposition and its application to rotating machinery compound fault diagnosis," *Mech. Syst. Signal Process.*, vol. 114, pp. 189–211, Jan. 2019.

[3] Z. Zhang, Y. Wang, and K. Wang, "Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network," *J. Intell. Manuf.*, vol. 24, no. 6, pp. 1213–1227, May 2012.

[4] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 35, nos. 1–2, pp. 108–126, Feb. 2013.

[5] J. Yu and J. Lv, "Weak fault feature extraction of rolling bearings using local mean decomposition-based multilayer hybrid denoising," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 12, pp. 3148–3159, Dec. 2017.

[6] M. Zhang, Z. Jiang, and K. Feng, "Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump," *Mech. Syst. Signal Process.*, vol. 93, pp. 460–493, Sep. 2017.

[7] 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. 35042–35056, 2018.

[8] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.

[9] X. Wang, W. Fan, X. Li, and L. Wang, "Weak degradation characteristics analysis of UAV motors based on laplacian eigenmaps and variational mode decomposition," *Sensors*, vol. 19, no. 3, p. 524, Jan. 2019.

[10] Y. Li, G. Cheng, C. Liu, and X. Chen, "Study on planetary gear fault diagnosis based on variational mode decomposition and deep neural networks," *Measurement*, vol. 130, pp. 94–104, Dec. 2018.

[11] G. Li, G. Tang, G. Luo, and H. Wang, "Underdetermined blind separation of bearing faults in hyperplane space with variational mode decomposition," *Mech. Syst. Signal Process.*, vol. 120, pp. 83–97, Apr. 2019.

[12] Y. Li, G. Li, Y. Wei, B. Liu, and X. Liang, "Health condition identification of planetary gearboxes based on variational mode decomposition and generalized composite multi-scale symbolic dynamic entropy," *ISA Trans.*, vol. 81, pp. 329–341, Oct. 2018.

[13] F. Li, R. Li, L. Tian, L. Chen, and J. Liu, "Data-driven time-frequency analysis method based on variational mode decomposition and its application to gear fault diagnosis in variable working conditions," *Mech. Syst. Signal Process.*, vol. 116, pp. 462–479, Feb. 2019.

[14] Y. Lei, N. Li, J. Lin, and Z. He, "Two new features for condition monitoring and fault diagnosis of planetary gearboxes," *J. Vibrat. Control*, vol. 21, no. 4, pp. 755–764, Jul. 2013.

[15] H. Zhao, M. Sun, W. Deng, and X. Yang, "A new feature extraction method based on EEMD and multi-scale fuzzy entropy for motor bearing," *Entropy*, vol. 19, no. 1, p. 14, Dec. 2016.

[16] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016.

[17] H. Zhao, H. Liu, J. Xu, and W. Deng, "Performance prediction using high-order differential mathematical morphology gradient spectrum entropy and extreme learning machine," *IEEE Trans. Instrum. Meas.*, to be published, doi: 10.1109/TIM.2019.2948414.

[18] Q. Hu, X.-S. Si, Q.-H. Zhang, and A.-S. Qin, "A rotating machinery fault diagnosis method based on multi-scale dimensionless indicators and random forests," *Mech. Syst. Signal Process.*, vol. 139, May 2020, Art. no. 106609.

[19] J. Xiong, Q. Zhang, J. Wan, L. Liang, P. Cheng, and Q. Liang, "Data fusion method based on mutual dimensionless," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 2, pp. 506–517, Apr. 2018.

[20] J. Xiong, Q. Liang, J. Wan, Q. Zhang, X. Chen, and R. Ma, "The order statistics correlation coefficient and PPMCC fuse non-dimension in fault diagnosis of rotating petrochemical unit," *IEEE Sensors J.*, vol. 18, no. 11, pp. 4704–4714, Jun. 2018.

[21] A. Qin, Q. Hu, Y. Lv, and Q. Zhang, "Concurrent fault diagnosis based on Bayesian discriminating analysis and time series analysis with dimensionless parameters," *IEEE Sensors J.*, vol. 19, no. 6, pp. 2254–2265, Mar. 2019.

[22] J. Xiong, Q. Zhang, G. Sun, X. Zhu, M. Liu, and Z. Li, "An information fusion fault diagnosis method based on dimensionless indicators with static discounting factor and KNN," *IEEE Sensors J.*, vol. 16, no. 7, pp. 2060–2069, Apr. 2016.

[23] Z. Wang, Q. Zhang, J. Xiong, M. Xiao, G. Sun, and J. He, "Fault diagnosis of a rolling bearing using wavelet packet denoising and random forests," *IEEE Sensors J.*, vol. 17, no. 17, pp. 5581–5588, Sep. 2017.

[24] M. Islam, M. Sohaib, J. Kim, and J.-M. Kim, "Crack classification of a pressure vessel using feature selection and deep learning methods," *Sensors*, vol. 18, no. 12, p. 4379, Dec. 2018.

[25] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimization for feature selection in classification: A multi-objective approach," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1656–1671, Dec. 2013.

[26] Q. Hu, A. Qin, Q. Zhang, J. He, and G. Sun, "Fault diagnosis based on weighted extreme learning machine with wavelet packet decomposition and KPCA," *IEEE Sensors J.*, vol. 18, no. 20, pp. 8472–8483, Oct. 2018.

[27] S. Kang, D. Ma, Y. Wang, C. Lan, Q. Chen, and V. I. Mikulovich, "Method of assessing the state of a rolling bearing based on the relative compensation distance of multiple-domain features and locally linear embedding," *Mech. Syst. Signal Process.*, vol. 86, pp. 40–57, Mar. 2017.

[28] X. Ding, Q. He, and N. Luo, "A fusion feature and its improvement based on locality preserving projections for rolling element bearing fault classification," *J. Sound Vib.*, vol. 335, pp. 367–383, Jan. 2015.

[29] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[30] Y. Li, Y. Yang, G. Li, M. Xu, and W. Huang, "A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mRMR feature selection," *Mech. Syst. Signal Process.*, vol. 91, pp. 295–312, Jul. 2017.

[31] X. Jin, E. W. M. Ma, L. L. Cheng, and M. Pecht, "Health monitoring of cooling fans based on mahalanobis distance with mRMR feature selection," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 8, pp. 2222–2229, Aug. 2012.

[32] X. Yan and M. Jia, "Intelligent fault diagnosis of rotating machinery using improved multiscale dispersion entropy and mRMR feature selection," *Knowl.-Based Syst.*, vol. 163, pp. 450–471, Jan. 2019.

[33] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mech. Signal Process.*, vol. 21, no. 6, pp. 2560–2574, Aug. 2007.

[34] L. Qing-yi, H. Zhou, A. P. Lin, K. Z. Qiu, and K. F. Cen, "Prediction of ash fusion temperature based on grid search and support vector machine," *J. Zhejiang Univ.*, vol. 45, no. 12, pp. 2181–2187, 2011.

[35] (Jul. 20, 2012). *Case Western Reserve University Bearings Data Set.* [Online]. Available: http://csegroups.case.edu/bearingdatacenter/pages/download-data-file

[36] P. Cao, S. Zhang, and J. Tang, "Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning," *IEEE Access*, vol. 6, pp. 26241–26253, 2018.

XIAO-SHENG SI received the B.Eng., M.Eng., and Ph.D. degrees in control engineering from the Department of Automation, Xi'an Institute of High-Technology, Xi'an, China, in 2006, 2009, and 2014, respectively. He is currently an Associate Professor with the Rocket Force University of Engineering and also a Postdoctoral Researcher with Xi'an Jiaotong University, Xi'an. He has authored or coauthored more than 50 articles in several journals, including *European Journal of Operational Research*, the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, the IEEE TRANSACTIONS ON RELIABILITY, the IEEE TRANSACTIONS ON FUZZY SYSTEMS, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS PART A, the IEEE TRANSACTION ON AUTOMATION SCIENCE AND ENGINEERING, *Reliability Engineering and System Safety*, and *Mechanical Systems and Signal Processing*. His research interests include evidence theory, expert systems, prognostics and health management, reliability estimation, predictive maintenance, and lifetime estimation. He received the Excellent Doctoral Dissertation Award, in 2015, from the Chinese Association of Automation. He is currently an Associate Editor of IEEE ACCESS. He is an active reviewer for a number of international journals.

AI-SONG QIN received the B.Sc. degree in automation and the M.Sc. degree in control theory and control engineering from the Taiyuan University of Technology, China, in 2010 and 2013, respectively. Her research areas include remaining useful life prediction and fault diagnosis of rotating machinery.

YUN-RONG LV received the M.Sc. degree in chemical engineering from the South China University of Technology, China, in 2001, and the Ph.D. degree in chemical machinery from Zhejiang University, China, in 2018. His research areas include corrosion and protection technique of mechanical equipment and fault diagnosis of rotating machinery.

QIN HU received the master's degree in control theory and control engineering from the Guangdong University of Technology, China, in 2013. He is currently pursuing the Ph.D. degree with the Department of Automation, Rocket Force University of Engineering, China. His research fields are fault diagnosis of rotating machinery, signal processing, and pattern recognition.

QING-HUA ZHANG received the M.Sc. degree in industrial automation and the Ph.D. degree in control theory and control engineering from the South China University of Technology, China, in 1995 and 2004, respectively. His research interests include condition monitoring and fault diagnosis of rotating machinery, and applications of intelligent algorithms.

• • •