Fault Diagnosis Based on Multi-Scale Redefined Dimensionless Indicators and Density Peak Clustering With Geodesic Distances

QIN HU\textsuperscript{1,2}, QI ZHANG\textsuperscript{2}, XIAO-SHENG SI\textsuperscript{2}, AI-SONG QIN\textsuperscript{1,} AND QING-HUA ZHANG\textsuperscript{1}

\textsuperscript{1}Guangdong Provincial Key Laboratory of Petrochemical Equipment Fault Diagnosis, Guangdong University of Petrochemical Technology, Maoming 525000, China
\textsuperscript{2}Department of Automation, Rocket Force University of Engineering, Xi’an 710025, China

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

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1306100, in part by the National Natural Science Foundation of China under Grant 61922089, Grant 61773386, Grant 61673311, Grant 61573366, Grant 61673127, and Grant 61473094, in part by the Young Elite Scientists Sponsorship Program of China Association for Science and Technology under Grant 2016QNRC001, in part by the Young Innovative Talents Program of Guangdong University under Grant 2019KQNCX088, in part by the Young Innovative Talents Program of Guangdong University of Petrochemical Technology under Grant 2016qn17, and in part by the Key Project of Natural Science Foundation of Guangdong under Grant 2018B030311054.

ABSTRACT A novel fault diagnosis method for rolling bearings based on multi-scale redefined dimensionless indicators and an unsupervised feature selection method using density peak clustering with geodesic distances is proposed in this paper. First, a new feature extraction method is proposed based on redefined dimensionless indicators and multi-scale analysis called multi-scale redefined dimensionless indicators. Then, density peak clustering with geodesic distances is utilized to automatically find the important multi-scale redefined dimensionless indicators. To the best of our knowledge, this is the first study to use density peak clustering with geodesic distances to explore unsupervised feature selection in the fault diagnosis field. Finally, the selected multi-scale redefined dimensionless indicators are fed into a quadratic discriminant analysis classifier to simultaneously identify 12 different conditions of rolling bearings. Experimental results demonstrated that the proposed method can successfully differentiate 12 localized fault types, fault severities, and fault orientations of rolling bearings.

INDEX TERMS Feature extraction, unsupervised learning, nearest neighbor searches, clustering algorithms, mechanical engineering, vibration measurement, knowledge engineering, machine learning, pattern recognition, fault diagnosis.

I. INTRODUCTION
Rolling bearings play significant roles in different types of machinery. However, these bearings are important, damage-prone components in rotating machinery. In practical engineering, bearing failures are among the most frequent causes of breakdowns in rotating machinery. For example in the case of large induction motors, bearing faults can constitute the 44\% of the total number of faults [1]. A bearing failure will cause unexpected downtime and economic loss. Consequently, there is a great demand for rolling bearing fault diagnosis techniques that can be used to make reliable maintenance decisions [2]. In general, fault diagnosis and prognosis algorithms can be categorized into model-based approaches, knowledge-based approaches and data-driven approaches [3]. In data-driven methods, features can be extracted to characterize the historical process data as a priori knowledge to a diagnostic system [4], [5]. To date, many fault feature extraction methods have been developed for vibration signal feature extraction [6]–[10]. In [11], several advanced feature extraction approaches (i.e., fast Fourier transform, singular spectrum analysis, wavelet packet transform, empirical mode decomposition and local mean decomposition), in parallel, extract informative features from the raw signals for diagnosing gear faults. In [12], in order to search the most discriminative features for bearing fault diagnosis, heterogeneous feature extraction methods based on the complex envelope spectrum, statistical time- and frequency-domain parameters, and wavelet packet analysis were proposed. In [13], to improve the ability to extract valuable information from
the measured vibration data sets, singular spectrum analysis had been employed to identify bearing fault. In [14], considering that condition-based monitoring of rotating machines requires robust features for accurate fault diagnosis, statistical spectral analysis was used to extract the fault features. Nevertheless, these methods are based on advanced signal processing methods to extract the fault features, which failed to consider the complexity and efficiency of fault diagnosis method. Dimensionless indicators are not influenced by many disturbances and are sensitive to faults. Because of these merits, they have been used as features for fault diagnosis in several studies [15]–[17]. However, the existing types of original dimensionless indicators are limited, and they usually suffer from varying amounts of aliasing in the feature space, causing a decrease in the diagnosis accuracy and an increase in the diagnosis uncertainty [18]. In view of the above mentioned problems, this paper proposes a new feature extraction method based on redefined dimensionless indicators (RDIs) and multi-scale analysis, called multi-scale redefined dimensionless indicators (MRDIs). MRDIs can reveal more informative fault information hidden in the original vibration signals in the case without using the signal processing methods. Because only the original vibration signals are used to extract the fault features, the complexity of fault diagnosis method can be largely cut down and the time consumption can be effectively decreased, which is different from the above studies.

In practice, a large number of features are usually extracted to collect all of the available information. Irrelevant, redundant, and similar information may exist in the feature set. Such information will not only swamp the effective features but also increase the model solution difficulty. This is also true for the extracted MRDIs. Feature selection method should be used to search for the most important features. In recent years, several feature selection methods have been presented in related literature [19]–[22]. In this study, density peak clustering (DPC) with geodesic distances (GD) is for the first time used to explore feature selection in the fault diagnosis field. DPC with the Euclidean distance (ED) as a clustering method has been successfully used in several studies [23]–[27]. However, if the data points lie on a complex nonlinear manifold hidden in a high-dimensional space, it is difficult for DPC with ED to obtain an accurate clustering result. To overcome this inadequacy, the GD was used to reveal the true distance between every pair of data points; this strategy was inspired by the idea of the isomap algorithm [28]. To date, only a few studies have considered the use of the DPC-GD for grouping data with arbitrary shapes or image data sets [29], [30]. Although the above studies have shown the successful use of DPC-GD in synthetic data field or image data field, DPC-GD still has the following problem that needs to be addressed: in the field of machine fault diagnosis, it may be difficult to directly cluster the correct number of sample types for sample dataset with a high-dimensional feature set. To handle this problem, we first use DPC-GD to perform fault feature clustering analysis and then employ pattern classifier to conduct sample type classification, which can address the above issue of existing studies that DPC-GD is directly used to perform sample type clustering analysis. DPC-GD for unsupervised feature selection is based on the observation that two fault features are probably related to the same group if they are close to each other. Thus, the fault features can be clustered into several groups so that fault features in the same group are similar to each other, while fault features in different clusters are dissimilar. Moreover, the DPC-GD for unsupervised feature selection was based on the assumption that the cluster centres and outliers should both be regarded as important features. The centre in each group represents one kind of fault feature and is regarded as an important fault feature. An outlier can be regarded as a cluster centre that contains only a single data point but that also represents one type of fault feature. Besides, DPC-GD can reduce the error in obtaining the distance between every pair of features. Therefore, the distances between features in the same cluster of the feature space can be decreased, whereas the distances between features from different clusters of the feature space are increased. Thus, important features can be effectively selected in feature selection process.

Based on the above analysis results, a novel fault diagnosis method is proposed based on MRDIs and unsupervised feature selection using DPC-GD. First, the definition of the original dimensionless indicators is extended, and multi-scale analysis is used to further process the RDIs, which can effectively capture more useful fault signal characteristics. Second, DPC-GD is used to enhance the unsupervised feature-clustering ability and obtain several important MRDIs. Finally, quadratic discriminant analysis (QDA) is adopted as a pattern classifier to simultaneously identify 12 different working conditions of rolling bearings. The experimental results clearly demonstrated that DPC-GD was effective at selecting the important features, and the proposed method could successfully differentiate 12 localized fault types, fault severities, and fault orientations of rolling bearings.

The main contributions of this work are as follows:

1) In this study, MRDIs were used as new fault features to find more useful fault information only from the original vibration signals, which could significantly decrease the complexity of the existing fault diagnosis method and decrease the amount of time consumed.

2) The proposed DPC-GD feature selection method was used to adaptively select several important features from a high-dimensional feature set for the first time. Compared with other feature selection methods, DPC-GD showed better feature selection performance.

The remainder of this work is organized as follows: The problem with rolling bearing fault diagnosis is described in section 2. Section 3 presents some basic definitions and theories related to this study. Section 4 concerns the applicability of the proposed method to rolling bearings based on experimental results, and the results of extensive comparative
studies are presented as examples. Finally, conclusions and further work are given in section 5.

II. PROBLEM STATEMENT

The original dimensionless indicators have been widely used as fault features in the fault diagnosis field. Nevertheless, there is a limited number of existing types of original dimensionless indicators. In practical engineering, an increasing number of fault types, fault severities, and fault orientations have emerged. Utilizing the limited number of original dimensionless indicators to accurately diagnose more faults has become extremely difficult. Thus, the first problem is developing additional fault features to effectively and accurately identify additional fault categories. In addition, for high-dimensional features, the data usually show a complex, highly curved manifold structure, which causes significant difficulties in feature selection. Thus, the second problem is selecting important fault features from a high-dimensional feature set that may not be effectively selected using other existing feature selection methods. In consideration of these two problems, a new intelligent fault-diagnosis method based on MRDIs, DPC-GD, and QDA is proposed in this paper. For the first problem, MRDIs are used to construct additional types of dimensionless indicators to reveal more useful fault information. For the second problem, DPC-GD is used as an unsupervised feature selection method to select the important features. A flowchart of the proposed method is shown in Fig. 1, and the general steps are described as follows:

Step 1: The vibration signals of a rolling bearing were measured by an acceleration sensor and divided into a series of segments for feature extraction. Rolling bearing fault signals were used to verify the effectiveness of the proposed method. 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. A total of 12 different operating conditions were tested.

Step 2: Based on the original vibration signals, $\mu$ types of RDIs were extracted as fault features. Additionally, multi-scale analysis was used to further process the RDIs, aiming to search for more useful fault information hidden in the original vibration signals. In this study, the scale factor was denoted as $\tau$. A high-dimensional feature set comprising $\mu \times \tau$ MRDIs was built. These MRDIs could measure the complexity of the nonlinearity and nonlinear vibration signals at different scales.

Step 3: An unsupervised feature-selection method based on DPC-GD was used to select several important MRDIs from the $\mu \times \tau$ MRDIs. The Floyd algorithm was used to calculate the GD, and the Gaussian kernel function was utilized to calculate the local density of each feature. A representation called a decision graph was introduced to find the MRDI cluster centres and outliers, and each of the remaining MRDIs was assigned to the same cluster as its nearest neighbour of higher density. The MRDIs possessing relatively large local density $\rho$ and distance $\delta$ were the cluster centres, and the MRDIs possessing a small local density $\rho$ and relatively large distance $\delta$ were regarded as potential outliers. An outlier was regarded as a cluster centre that contained only a single data point but represented one type of fault feature. Thus, the cluster centres and potential outliers were both regarded as important features. Using the DPC-GD feature selection method, the MRDI set could be greatly reduced, and several important MRDIs could be obtained.

Step 4: To illustrate the classification performance of the original dimensionless indicators, RDIs, and MRDIs, QDA was used to identify the 12 working conditions of the rolling bearing, and the 10-fold cross-validation approach was employed to validate the performance of the QDA. The average identification accuracy for all 10 trials was computed and used as the final identification accuracy. For the RDIs and MRDIs, DPC-GD was used to select several important features. Then, these important features were added to the QDA one by one. Thus, the relationships between the identification accuracy and different numbers of RDIs and MRDIs were obtained.

Step 5: To further demonstrate the superiority of DPC-GD, it was compared with other feature selection methods. For each feature selection method, the detailed feature selection results and classification results of the MRDIs corresponding to the maximum accuracy for normal and eleven faulty working conditions for a rolling bearing were given.
III. RELATED THEORIES

A. MULTI-SCALE REDEFINED DIMENSIONLESS INDICATORS

Dimensionless indicators are good diagnosis characteristics that do not require transformation and processing of the vibration signal. In practice, dimensionless indicators are sensitive to faults rather than working conditions, and they are not affected by disturbance. Traditional dimensionless parameters (TDIs) are defined as follows [18]:

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

(1)

where \( x \) denotes the vibration amplitude and \( p(x) \) denotes the probability density function of the vibration amplitude. Details are given as follows:

(i) If \( l = 2, m = 1 \), we obtain the waveform indicator.
(ii) If \( l \rightarrow \infty, m = 1 \), we obtain the impulse indicator.
(iii) If \( l \rightarrow \infty, m = 1/2 \), we obtain the margin indicator.
(iv) If \( l \rightarrow \infty, m = 2 \), we obtain the peak indicator.
(v) The kurtosis indicator is \( K = \frac{\int_{-\infty}^{+\infty} x^4 p(x) dx}{\int_{-\infty}^{+\infty} x^2 p(x) dx} \). In this formula, \( X_{rms} \) is defined as \( X_{rms} = \sqrt{\int_{-\infty}^{+\infty} x^2 p(x) dx} \).
(vi) The skewness indicator is \( S = \frac{\int_{-\infty}^{+\infty} x^3 p(x) dx}{\int_{-\infty}^{+\infty} x^2 p(x) dx} \).

To date, there are six types of dimensionless indicators, namely, the waveform indicator, impulse indicator, margin indicator, peak indicator, kurtosis indicator, and skewness indicator. However, because the number of fault types, fault severity, and fault orientations of rolling bearings has gradually increased, it is necessary to extract more useful fault information from the vibration signals to identify additional fault categories. Therefore, RDIs were proposed to find more informative fault information hidden in the vibration signals.

The construction of the RDIs was based on TDIs. From the previously provided definition of the dimensionless indicators, it can be seen that they depend on parameters \( l \) and \( m \). Six types of existing dimensionless indicators are obtained when six groups of \( l \) and \( m \) values are determined. Additional types of dimensionless indicators can be obtained if the appropriate values for \( l \) and \( m \) are changed, while ensuring that the numerators and denominators of the defining formulas have the same dimension. Thus, a large number of new dimensional indicators, called RDIs, can be obtained simply by selecting different values for \( l \) and \( m \). In addition, because of its good performance in fault feature extraction, multi-scale analysis can be used to further process the RDIs. The goal is to search for more informative fault information hidden in the vibration signals. The definitions of the RDIs are as follows [31]:

**Definition 1:** Similar to the definition of the waveform indicator, we let \( l \) and \( m \) vary from \( \alpha \) to \( \beta \), with a change interval of \( \lambda \) for \( \alpha \) and \( \beta \), where \( \alpha, \beta \in [0.5, 5] \), \( \lambda = 0.5 \), and \( \alpha \neq \beta \).

**Definition 2:** Similar to the definitions of the impulse indicator, margin indicator, and peak indicator, we let \( l \rightarrow \infty, m = \gamma \), where the change interval is \( \lambda \) 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 \( \varepsilon \) to \( \eta \), where the change interval is \( \lambda_1 \) for \( \varepsilon \) and the change interval is \( \lambda_2 \) for \( \eta \). When the numerator is \( \int_{-\infty}^{+\infty} x^\varepsilon p(x) dx \), the \( \varepsilon \) power of \( \left( \int_{-\infty}^{+\infty} x^\gamma p(x) dx \right)^\frac{1}{2} \) 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 definition 1, definition 2, and definition 3, 90 types of dimensionless indicators, 10 types of dimensionless indicators, and 45 types of dimensionless indicators are obtained, respectively. Thus, a total of 145 types of dimensionless indicators were used as the fault features in this study. In other words, the number of RDIs was 145.

In addition, in order to further improve the diagnosis ability of the RDIs, multi-scale analysis is used to transform the original signal into different scales of data through the coarse-graining process. Given a one-dimensional discrete time series, \( S = \{x_1, x_2, \ldots, x_n\} \), first, we divide the original time series into non-overlapping windows of length \( \tau \). Second, consider the \( \tau \)-length vectors \( S_\tau(j) = \{x_{(j-1)\tau+1}, x_{(j-1)\tau+2}, \ldots, x_{j\tau}\} \), \( 1 \leq j \leq \frac{n}{\tau} \). The consecutive coarse-grained time series sets \( \{y^{(1)}_{(\tau)}\} \) are reconstructed according to the following equation:

\[ y^{(\tau)}_j = \frac{1}{\tau} \sum_{(j-1)\tau+1}^{j\tau} x_i, \quad 1 \leq j \leq \frac{n}{\tau}, \]  

(2)

where \( \tau \) is also called the scale factor; for scale \( \tau \), the time series \( \{y^{(\tau)}_j\} \) is \( \{y^{(1)}_1, y^{(1)}_2, \ldots, y^{(1)}_{v}\} \) and \( v \) is the quotient of \( n/\tau \). In particular, for scale one, the time series \( \{y^{(1)}_1, y^{(1)}_2, \ldots, y^{(1)}_{v}\} \) is simply the original time series \( \{y^{(1)}_1, y^{(1)}_2, \ldots, y^{(1)}_{v}\} \). Finally, we calculate the RDI value for each coarse-grained time series plotted as a function of the scale factor.

Based on the aforementioned three definitions of the dimensionless indicators and multi-scale analysis, new dimensionless indicators can be obtained, and we define these dimensionless indicators as the multi-scale redefined dimensionless indicators (MRDIs), which can provide much more informative fault information regarding the operating condition of rolling bearings. The MRDIs are defined as follows:

\[ \xi_1 = \frac{\int_{-\infty}^{+\infty} |y^{(\tau)}_j|^\alpha p(y^{(\tau)}_j) dy^{(\tau)}_j}{\int_{-\infty}^{+\infty} |y^{(\tau)}_j|^\beta p(y^{(\tau)}_j) dy^{(\tau)}_j} = \frac{\sqrt[\alpha]{E(|y^{(\tau)}_j|^\alpha)}}{\sqrt[\beta]{E(|y^{(\tau)}_j|^\beta)}} \]  

(3)
The DPC method utilizes two important quantities: one is 
the amplitude. 

distance 

\[ d \leq \sum_{i,j \neq i} \exp \left( -\frac{d^2_{ij}}{d_c^2} \right) \] 

where the cut-off distance \( d_c \) is an adjustable parameter.

The process for selecting \( d_c \) is the only part of Formulas 6 and 7 that has not been described; it is defined as:

\[ d_c = d_{[M \times P]}, \] 

where \( M = \binom{N}{2} \) and \( d_{[M \times P]} \in D = \{ d_{ij} \}_{i,j=1}^N \). \( D \) is the set of all the distances between each pair of points in the data set, which are sorted in ascending order. \( p \) is a percentage.

The computation of \( \delta_i \) is quite simple, because its definition is the nearest separation distance between the samples of higher density and data point \( i \), as follows:

\[ \delta_i = \min_{j, d_{ij} > \rho_i} d_{ij}. \] 

For the data point that has the highest local density, we simply let

\[ \delta_i = \max_{j \neq i} \delta_j. \] 

Based on the local density \( \rho_i \) and separation distance \( \delta_i \), a representation called the decision graph is introduced to help one find the cluster centres and the probable outliers. In the decision graph, a data point possessing relatively large \( \rho_i \) and \( \delta_i \) values implies that the data point may be the centre of a cluster. When \( \rho_i \) and \( \delta_i \) are both small, the data point lies near the edge of the cluster. When \( \rho_i \) is large and \( \delta_i \) is small, it implies that the data point is close to but not the centre of the cluster. When \( \rho_i \) is small and \( \delta_i \) is large, on the other hand, the point is far from the entire data set, indicating that the data point is probably an outlier.

From Equation (6), it can easily be observed that the distance \( d_{ij} \) between \( x_i \) and \( x_j \) has a direct impact on the values of local density \( \rho_i \) and separation distance \( \delta_i \). Thus, this paper uses Floyd’s algorithm [32] to calculate the GD. The GD is the unique shortest path connecting two features in a high-dimensional fault feature space. The GD is specifically calculated as follows: First, the neighbourhood for each point is calculated. For example, the neighbourhood of a point may be the \( k \) nearest points. If data point \( j \) is one of the \( k \) nearest neighbours of data point \( i \), then \( G(i, j) = 1 \) and \( G(j, i) = 1 \); otherwise, \( G(i, j) = 0 \) and \( G(j, i) = 0 \). The value of \( G(i, j) \) indicates whether data point \( j \) is a nearest neighbour of data point \( i \). In this way, a neighbourhood graph can be constructed. Next, the GD \( d_G(x_i, x_j) \) is initialized to the ED between data point \( i \) and data point \( j \)...

\[ d_G(x_i, x_j) = \begin{cases} d(x_i, x_j), & \text{if } G(i, j) = 1 \\ \infty, & \text{otherwise} \end{cases} \] 

where \( d(x_i, x_j) \) represents the ED between data point \( x_i \) and data point \( x_j \). Finally, the GD between two data points is calculated as follows: For each value of \( q = 1, 2, \ldots, N \) in turn, replace all entries \( d_G(x_i, x_j) \) by

\[ d_G(x_i, x_j) = \min \{ d_G(x_i, x_j), d_G(x_i, x_q) + d_G(x_q, x_j) \} \] 

The procedure of the DPC-GD algorithm for adaptive feature selection is presented in Table 1.

IV. EXPERIMENTAL RESULTS AND COMPARISONS

A. EXPERIMENTAL DATA ANALYSIS

The effectiveness of the proposed method was validated using the measured vibration fault signals of a rolling bearing provided by the Electrical Engineering Laboratory of Case Western Reserve University [33]. This study 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 (0.028 mils),
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 that of the
hour hand of a clock at 6:00. For convenience, these 12 oper-
ating conditions are expressed as \( N, B-007, B-014, B-021,
B-028, IR-007, IR-014, IR-021, IR-028, OR-007, OR-014,
and OR-021 \), respectively. For each operating condi-
tion, 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 classified
effectiveness of the six original dimensionless indicators are
illustrated in Fig. 2, which shows the sample number on the
vertical axis. Sample numbers 1–50, 51–100, 101–150,
151–200, 201–250, 251–300, 301–350, 351–400, 401–450,
451–500, 501–550, and 551–600 represent the conditions \( N,
B-007, B-014, B-021, B-028, IR-007, IR-014, IR-021, IR-028,
OR-007, OR-014, \) and \( OR-021 \), respectively. The vibration
signals for the 12 different operating condi-
tions were processed using multi-scale analysis, and multiple
course-grained time series were obtained. In the implementa-
tion process of the multi-scale analysis, the scale factor was \( \tau = 1, 2, 3, \ldots, 6 \), the selection of which was based on [34], [35]. Based on each coarse-grained time series, the values of the RDIs were calculated under different scale
factors. From the above analysis, it was possible to obtain a
feature set containing 870 MRDIs from the original vibration
signals. The specific order of the 870 MRDIs is as follows:
the first 145 MRDIs are for the original signals at scale one
(Note: when the scale factor is one, the 145 MRDIs are simply the
145 RDIs). The next 145 MRDIs are for the original signals at scale two. In this way, the last 145 MRDIs are for the original signals at scale six. To qualitatively show the difference between the RDIs and TDIs, four RDIs, namely, the 1st RDI, 9th RDI, 55th RDI, and 108th RDI, are illustrated in Fig. 3. In addition, in order to quantitatively illustrate how

| TABLE 1. Algorithm of DPC-GD for feature selection. |
|-----------------------------------------------|
| Input: \( X = \{ x_i \}_{i=1}^{n}, x_i \in \mathbb{R}^m \) (data), \( d_c \) (cut-off distance), \( k \) (number of nearest neighbours) |
| Output: The label vector of the cluster index and outlier index |
| Step 1: calculate the ED between every pair of features in the high-
dimensional feature set according to Formula (6). |
| Step 2: determine the number of nearest neighbours and construct a neighbourhood graph. |
| Step 3: calculate the geodesic distance matrix that contains all pairwise-geodesic distances between every pair of fault features according to Formulas (11)–(12). |
| Step 4: calculate the local density for each feature according to Formula (7). |
| Step 5: calculate the separation distance between each feature and other features with higher density according to Formula (9). |
| Step 6: plot the decision graph and select cluster centres and outliers as the final fault features. |

the MRDIs and RDIs are different from the TDIs, the first
145 MRDIs (note that the first 145 MRDIs are the 145 RDIs)
and 6 TDIs are taken as an example. Distance ratios (DRs)
of between-class and within-class distances for the TDIs
and MRDIs, respectively, were calculated and arranged in
decreasing order. The DR values of the 145 MRDIs are shown
in Fig. 4. Among the 145 MRDIs, the six TDIs are marked in
red, and the remaining dimensionless indicators are marked
in blue. If a dimensionless indicator has a large DR value,
It can maximize the between-class distance and minimize the
within-class distance, indicating a good performance in dis-
tinguishing different operating conditions. By analysing the
DR values of the 145 MRDIs, it can be concluded that the DR
values of most MRDIs were greater than those of the TDIs,
indicating that the MRDIs have better distinguishability for
different fault types than the TDIs.

B. COMPARISONS FOR TDIS, RDIS, AND MRDIS
This subsection compares the classification performances of
the TDIs, RDIs, and MRDIs based on QDA, which is a
powerful statistical multivariate pattern-recognition method.
A detailed introduction to QDA can be found in [36]. In this
study, the 10-fold cross validation method was used to train
and test the QDA pattern classifier. The sample dataset was
randomly divided into 10 subsets, one of which was used as
the testing set, and the other 9 of which were combined
to form the training set. Then, the average identification
accuracy for all 10 trials was computed and used as the final
identification accuracy.

For the 870 MRDIs, DPC-GD was used to adaptively
select the important features. To evaluate the selection results,
the selected important features were used as the input of QDA. In DPC-GD, the parameter \( k \) for the \( k \) nearest neigh-
bours was set at five, and the cut-off distance \( d_c \) was set
at 2%. In addition, it is worth pointing out that in the cal-
culation process for the GD between each pair of features
according to Formula (11), there may be an infinite distance
between two data points in the distance matrix, imposing
several infinite values on the delta matrix. In this study,
the maximum distance, second only to the infinite distance
value, was used to replace the infinite value in the distance
matrix. Thus, the delta matrix of the DPC did not contain
an infinite value, which was beneficial for constructing a
decision graph. The decision graph generated by DPC-GD
is shown in Fig. 5, which is a plot of the separation distance
as a function of the local density for each MRDI. It can be
seen that three MRDIs, identified as centres, are located in
the upper right part of the decision graph, and six MRDIs,
identified as outliers, are in the upper left part of the decision
graph. It is worth noting that the outliers overlap in the
decision graph, causing it to seem that only four MRDIs
are selected. The actual number of selected MRDIs is nine.
These nine MRDIs are the 1st, 2nd, 10th, 19th, 101st, 137th,
248th, 475th, and 632nd MRDIs. Among these, the 1st, 2nd,
10th, 19th, 101st, and 137th MRDIs are outliers, and the
248th, 475th, and 632nd MRDIs are centres. To evaluate the
FIGURE 2. Classified effectiveness of the original dimensionless indicators: (a) the waveform indicator; (b) the peak indicator; (c) the impulse indicator; (d) the margin indicator; (e) the kurtosis indicator; (f) the skewness indicator.

classification effectiveness of the 9 selected MRDIs, QDA is used as the pattern classifier to identify the 12 bearing working conditions. For these 9 MRDIs, the identification accuracy is 59.17%, 76.83%, 76.17%, 76.33%, 88.33%, 91.00%, 97.67%, 97.00% and 97.17%, respectively, when they are fed into DQA one by one. The maximum accuracy is 97.67% when the number of selected MRDIs reaches 7. The classification results of the MRDIs corresponding to the maximum accuracy of 97.67% are shown in Fig. 6. It can be observed that for the normal condition and the faulty conditions B-014, B-028, IR-014, OR-007 and OR-021, the fault classification rates reach 100%. For the faulty conditions B-007, B-021, IR-028 and OR-014, the fault classification rate is 94%. For the faulty conditions IR-007 and IR-021, the fault classification rate is 98%. The overall recognition rate is 97.67%.

For the 145 RDIs, the decision graph generated by DPC-GD is shown in Fig. 7. It can be seen that three RDIs identified as centres are located in the upper right part of the decision graph, and one RDI identified as an outlier is in the upper left part of the decision graph. The number of selected RDIs is four. These four selected RDIs are the 52nd, 60th, 104th, and 137th RDIs. The identification accuracy of these four RDIs are 64.83%, 75.17%, 90.67%, and 93.17%, respectively, when they are fed into DQA one by one. The classification results of the RDIs corresponding to the maximum accuracy of 93.17% are shown in Fig. 8. It can be seen that for the faulty conditions B-014, B-028, IR-007, IR-021 and OR-007, the fault classification rates reach 100%. For the normal condition and the faulty condition IR-014, the fault classification rate is 98%. For the faulty conditions B-007, B-021, IR-028, OR-014 and OR-021, the fault classification rate is 98%.
FIGURE 3. Classified effectiveness of a portion of the RDIs: (a) the 1st RDI ($\alpha = 0.5, \beta = 1$); (b) the 9th RDI ($\alpha = 0.5, \beta = 5$); (c) the 55th RDI ($\alpha = 3.5, \beta = 0.5$); (d) the 108th RDI ($\epsilon = 1, \eta = 4$).

rates are 84%, 78%, 96%, 74% and 90%, respectively. The overall recognition rate is 93.17%.

For the six TDIs, the identification accuracy is 72.67%, 79.83%, 78.00%, 78.17%, 80.67%, and 86.67%, respectively, when they are fed into DQA one by one. The classification results for the TDIs corresponding to the maximum accuracy of 86.67% are shown in Fig. 9. It can be observed that for the faulty conditions B-028, IR-007, IR-014, OR-007 and OR-021, the fault classification rates reach 100%. For the normal condition and the faulty conditions B-007, B-014, B-021, IR-021, IR-028, and OR-014, the fault classification rates are 84%, 66%, 96%, 52%, 98%, 70% and 74%, respectively. The overall recognition rate is 86.67%.

Based on the above experimental results, it can be seen that compared to the TDIs and RDIs, the MRDIs can improve the fault diagnosis accuracy by 11% and 4.5%, respectively. This demonstrates that the classification effectiveness of the MRDIs is better than that of the RDIs and TDIs.
In addition, DPC-GD is directly used to conduct clustering for the 600 samples. The sample clustering results based on the MRDIs and RDIs are illustrated in Fig. 10, where it can be observed that only six clusters and three clusters are obtained, respectively. However, the actual number of sample clusters was 12. The experimental results show that DPC-GD fails to cluster the correct number of fault types for the sample data set in a high-dimensional feature space.

C. COMPARATIVE STUDY OF ORIGINAL DPC AND DPC-GD BASED ON MRDIs

For the original DPC, several tests were performed with different values for the cut-off distance \(d_c = 0.1\%, 0.5\%, 1\%, 1.5\%, 2\%, 2.5\%\). The decision graphs generated by the original DPC with six different cut-off distances are shown in Fig. 11. It can be seen that the number of selected MRDIs is always two. These two selected MRDIs for the six group experiments are the 137th and 396th MRDIs, 137th and 396th MRDIs, 137th and 466th MRDIs, 137th and 548th MRDIs, 137th and 548th MRDIs, and 137th and 548th MRDIs. Among the six group experiments, the maximum identification accuracy is 59.67% when the cut-off distance \(d_c\) is 0.1% or 0.5%. This low accuracy was caused by the fact that the ED could not truly measure the distances between the pairs of MRDIs. The intrinsic geometry of the complex nonlinear structure was ignored, leading to an inaccurate
local density for each MRDI. The classification results corresponding to the maximum accuracy of 59.67% are shown in Fig. 12. It can be seen that for the normal condition only, the recognition rate is 100%. For the faulty conditions B-007, B-014, B-021, B-028, IR-007, IR-014, IR-021, IR-028, OR-007, OR-014 and OR-021, the fault classification rates are 58%, 52%, 60%, 38%, 52%, 64%, 92%, 50%, 22% and 74%, respectively. The overall recognition rate is 59.67%.

For DPC-GD, several tests were performed with different values of parameter $k$ for the $k$ nearest neighbours $(k = 5, 6, 7)$ and different cut-off distance values $(d_c = 0.1\%, 0.5\%, 1\%, 1.5\%, 2\%, 2.5\%)$. The decision graphs generated by DPC-GD with different values of $d_c$ at $k = 5$ are shown in Fig. 13. It can be seen that the number of selected MRDIs is always nine. The detailed MRDI selection results and identification accuracy when using DPC-GD are presented in Table 2. The identification accuracy is slightly lower only when the cut-off distance is 0.5%. The maximum identification accuracy is 98.17% when the number of nearest neighbours and the cut-off distance are set at 5 and 0.1%, respectively. The classification results corresponding to the maximum accuracy of 98.17% are shown in Fig. 14. It can be seen that for the normal condition and the faulty conditions B-014, B-028, IR-014, IR-021, OR-007 and OR-021, the fault classification rates reach 100%. For the faulty conditions B-007 and B-021, the fault classification rate is 94%. For the faulty condition IR-007, the fault classification rate is 98%. For the faulty conditions IR-028 and OR-014, the fault classification rate is 96%. The overall recognition rate is 98.17%.

In addition, several tests were performed using DPC-GD with different cut-off distance values when the parameter $k$ was fixed at six and seven. The decision graphs are shown in Fig. 15 and Fig. 18, respectively, and the detailed MRDI selection results and identification accuracy are presented in Tables 3-4. It can be observed that when the parameter $k$ is fixed at six, the identification accuracy only shows a slight change. The maximum identification accuracy is 98.17% when the cut-off distance is 0.5% or 1.5%. The classification results corresponding to the maximum accuracy of 98.17% are shown in Figs. 16–17. When the number of nearest neighbours is set at seven, the identification accuracy only shows a slight change as well. The maximum identification accuracy is 98.33% when the cut-off distance $d_c$ is 1.5%. The classification results corresponding to the maximum accuracy of 98.33% are shown in Fig. 19.
Based on the above experimental results, it can be concluded that compared with the original DPC, DPC-GD can enhance the unsupervised feature selection ability and obtain better feature selection results. DPC-GD can truly reveal the distance between every pair of features, as well as accurately construct a decision graph to select several important features to improve identification accuracy. The proposed DPC-GD for unsupervised feature selection can achieve satisfactory diagnosis results.

**D. COMPARATIVE STUDY OF OTHER FEATURE SELECTION METHODS**

Furthermore, in order to verify the superiority of the proposed method, other existing feature selection methods were used to conduct a comparison. This comparative study primarily considered four feature selection methods, including local learning-based unsupervised clustering for feature selection (LLC) [37], the Laplacian score for feature selection (LSFS) [38], unsupervised discriminative feature selection (UDFS) [39], and the ordinal locality approach for unsupervised feature selection (OLFS) [40]. In addition,
we briefly investigate the parameter selection issue for these four feature selection methods without considering the selection of the optimal parameters, whose studies are beyond the scope of this study. For simplicity, we only reported the maximum identification accuracy and the corresponding detailed feature selection result for all compared methods. For LLC, the number of clusters $c_1$, the size of the mutual neighbors $k_1$ and the trade-off parameter $\beta_1$ were searched in: $c_1 \in \{6, 8, 12\}$, $k_1 \in \{5, 10, 20\}$ and $\beta_1 \in \{0.1, 1, 10\}$, respectively. For LSFS, no parameter selection was needed. For UDFS, the number of clusters $c_2$, the number of nearest neighbors $k_2$, the regularization parameters $\alpha_2$ and $\beta_2$ were searched in: $c_2 \in \{6, 12\}$, $k_2 \in \{5, 10\}$, $\alpha_2 \in \{10^{-6}, 10^{-4}, 10^{-2}, \cdots, 10^0\}$, $\beta_2 \in \{10^{-6}, 10^{-4}, 10^{-2}, \cdots, 10^0\}$. After a detailed simulation analysis, for LLC, the maximum identification accuracy that we achieved was 94.33% with the combination $c_1 = 12$, $k_1 = 20$, $\beta_1 = 0.1$. For UDFS, the maximum identification accuracy that we achieved was 91.5% with $\gamma_1 = 10^{-8}$. For OLFS, the maximum identification accuracy that we achieved was 91.83% with the combination $c_2 = 6$, $\alpha_2 = 10^4$, $\beta_2 = 10$, $k_2 = 5$. For each of the above-mentioned methods, the first 30 MRDIs were selected as the QDA inputs for fault identification analysis.

The classification accuracy results of these four methods are illustrated in Fig. 19. It can be observed that with an increase in the number of MRDIs, the classification accuracies of the LLC, LSFS, UDFS, and OLFS methods also improved. However, when the number of input features increased to 11, 11, 18, and 26, the classification accuracies reached their highest values of 94.33%, 92.17%, 91.50%, and 91.83%, respectively. Then, the classification accuracy gradually decreased as the number of input MRDIs continued to increase.
TABLE 5. Detailed feature selection results and identification accuracy of the four feature selection methods.

| Feature selection method | Sequence number of the selected 30 MRDIs | Corresponding number of MRDIs for obtaining the maximum accuracy | Maximum identification accuracy |
|--------------------------|----------------------------------------|---------------------------------------------------------------|---------------------------------|
| LLC                      | 283, 418, 273, 139, 708, 129, 844, 718, 428, 563, 138, 853, 572, 128, 863, 282, 717, 427, 137, 862, 573, 140, 119, 284, 429, 130, 274, 574, 141, 699 | The first 11 MRDIs | 94.33% |
| LSFS                     | 137, 862, 128, 138, 283, 717, 853, 273, 427, 129, 308, 418, 139, 283, 572, 274, 863, 563, 428, 130, 718, 419, 284, 140, 573, 854, 709, 275, 564, 429 | The first 11 MRDIs | 92.17% |
| UDFS                     | 859, 860, 74, 850, 386, 533, 387, 56, 858, 531, 65, 675, 66, 532, 857, 75, 388, 676, 385, 849, 861, 851, 534, 83, 530, 674, 856, 390, 678, 133 | The first 18 MRDIs | 91.50% |
| OLFs                     | 6, 5, 2, 137, 3, 835, 854, 1, 855, 4, 771, 241, 816, 572, 282, 818, 644, 709, 82, 418, 798, 690, 635, 91, 824, 94, 372, 572, 93, 653 | The first 26 MRDIs | 91.83% |

Compared with the classification results of the MRDIs based on DPC-GD, it can be seen that the identification accuracy of the proposed method is higher than that of the compared methods. Specifically, the proposed method achieved a classification accuracy of 98.33% with six MRDIs, which was better than the values for LLC with 11 MRDIs (94.33%), LSFS with 11 MRDIs (92.17%), UDFS with 18 MRDIs (91.50%), and OLFs with 26 MRDIs (91.83%). These comparative results demonstrate that the proposed method could successfully differentiate 12 localized fault types of rolling bearings and is superior to the aforementioned methods.

E. COMPARATIVE STUDY FOR PATTERN CLASSIFIERS BASED ON THE SELECTED MRDIs

In addition, the performance of the QDA classifier was compared with that of linear discriminant analysis (LDA), a quadratic support vector machine (QSVM), and bagged trees (BT). The nine MRDIs selected by DPC-GD, in which the parameter $k$ for the $k$ nearest neighbours was set at five and the cut-off distance $d_c$ was set at 2%, were used as the input of the LDA, QSVM and BT. The 10-fold cross validation method was used to train and test the above pattern classifiers. The average identification accuracy for all 10 trials was computed and used as the final identification accuracy. The maximum classification accuracies of the LDA, QSVM and BT are 89.5% with 9 MRDIs, 93.8% with 8 MRDIs and 91.7% with 9 MRDIs, respectively. These comparative results demonstrate that, using the nine selected MRDIs, the classification accuracy of the QDA pattern classifier is higher than that of the LDA, QSVM and BT.

V. CONCLUSION

This paper proposed a novel hybrid intelligent fault diagnosis method based on MRDIs and DPC-GD to identify 12 different working conditions of a rolling bearing. The proposed method includes three steps for intelligent fault diagnosis: fault feature extraction by MRDIs, unsupervised feature selection using DPC-GD, and fault pattern recognition using QDA. The advantage of the MRDIs is that additional types of dimensionless indicators can be constructed to deal with the increasing number of fault types, fault severities, and fault orientations, and more useful fault information hidden in the original vibration signals can be found. The unsupervised feature selection method DPC-GD plays an important role in improving classification accuracy. Compared with other feature selection methods, the proposed method tended to achieve better diagnosis results with fewer selected MRDIs. However, the effectiveness of DPC-GD is sensitive to two parameters: the cut-off distance and the number of nearest neighbours. Thus, a method to select the optimal parameters when performing feature selection to improve identification accuracy needs further study.

REFERENCES
Q. Hu, X.-S. Si, Q.-H. Zhang, and A.-S. Qin, “A rotating machinery 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.

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

A. Malhi and R.-X. Gao, “PCA-based feature selection scheme for machine defect classification,” IEEE Trans. Instrum. Meas., vol. 53, no. 6, pp. 1517–1525, Dec. 2004.

Y. 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.

J. Biet, “Rotor faults diagnosis using feature selection and nearest neighbors rule: Application to a turbogenerator,” IEEE Trans. Ind. Electron., vol. 60, no. 9, pp. 4063–4073, Sep. 2013.

F. Pacheco, M. Cerrada, R.-V. Sánchez, D. Cabrera, C. Li, and J. V. de Oliveira, “Attribute clustering using rough set theory for feature selection in fault severity classification of rotating machinery,” Expert Syst. Appl., vol. 71, pp. 69–86, Apr. 2017.

A. Rodríguez and A. Laio, “Clustering by fast search and find of density peaks,” Science, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.

J. Hou and A. Zhang, “Enhancing density peak clustering via density normalization,” IEEE Trans. Ind. Informat., vol. 16, no. 4, pp. 2477–2485, Apr. 2020.

D. Yu, G. Liu, M. Guo, X. Liu, and S. Yao, “Density peaks clustering based on weighted local density sequence and nearest neighbor assignment,” IEEE Access, vol. 7, pp. 34301–34317, 2019.

M. Du, S. Ding, and H. Jia, “Study on density peaks clustering based on K-nearest neighbors and principal component analysis,” Knowl.-Based Syst., vol. 99, pp. 135–145, May 2016.

Y. Wang, Z. Wei, and J. Yang, “Feature trend extraction and adaptive density peaks search for intelligent fault diagnosis of machines,” IEEE Trans. Ind. Informat., vol. 15, no. 1, pp. 105–115, Jan. 2019.

J. B. Tenenbaum, “A global geometric framework for nonlinear dimensionality reduction,” Science, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

M. Du, S. Ding, X. Xu, and Y. Xue, “Density peaks clustering using geodesic distances,” Int. J. Mach. Learn. Cybern., vol. 9, no. 8, pp. 1335–1349, Aug. 2018.

Z. Li and Y. Tang, “Comparative density peaks clustering,” Expert Syst. Appl., vol. 95, pp. 236–247, Apr. 2018.

Q. Hu, X.-S. Si, A.-S. Qin, Y.-R. Lv, and Q.-H. Zhang, “Machinery fault diagnosis scheme using redefined dimensionless indicators and mRMR feature selection,” IEEE Access, vol. 8, pp. 40313–40326, 2020.

R. W. Floyd, “Algorithm 97: Shortest path,” Commun. ACM, vol. 5, no. 6, pp. 345, Jun. 1962.

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

M. Costa, A. L. Goldberger, and C.-K. Peng, “Multiscale entropy analysis of biological signals,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 71, no. 2, Feb. 2005, Art. no. 021906.

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.

T. W. Anderson, An Introduction to Multivariate Statistical Analysis (Wiley Series in Probability and Statistics). Stanford, CA, USA: Stanford Univ., Dept. Statist., 2003, pp. 260–261.
AI-SONG QIN received the B.Sc. degree in automation from the Taiyuan University of Technology, China, in 2010, and the M.Sc. degree in control theory and control engineering from the Taiyuan University of Technology, China, in 2013. She is currently with the Guangdong Provincial Key Laboratory of Petrochemical Equipment Fault Diagnosis, Guangdong University of Petrochemical Technology. Her research areas include remaining useful life prediction and fault diagnosis of rotating machinery.

QING-HUA ZHANG received the M.Sc. degree in industrial automation from the South China University of Technology, China, in 1995, and the Ph.D. degree in control theory and control engineering from the South China University of Technology, China, in 2004. He is currently with the Guangdong Provincial Key Laboratory of Petrochemical Equipment Fault Diagnosis, Guangdong University of Petrochemical Technology. His research interests include: condition monitoring and fault diagnosis of rotating machinery, and applications of intelligent algorithms.