IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 27, 2019, accepted January 11, 2020, date of publication January 17, 2020, date of current version January 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2967537

Bearing Fault Feature Selection Method Based on Weighted Multidimensional Feature Fusion

YAZHOU LI[1], WEI DAI[2], AND WEIFANG ZHANG[2]
[1]School of Energy and Power Engineering, Beihang University, Beijing 100191, China
[2]School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China

Corresponding author: Wei Dai (dw@buaa.edu.cn)

This work was supported by the National Natural Science Foundation of China (No. 51705015), and the Technical foundation program (No. JSZL2017601C002 and JCKY2018203C005) from the Ministry of Industry and Information Technology of China.

ABSTRACT Rolling bearing is one of the most critical components in rotating machinery, so in order to efficiently select features, reduce feature dimensions and improve the correctness of fault diagnosis, a feature selection and fusion method based on weighted multi-dimensional feature fusion is proposed. Firstly, features are extracted from different domains to constitute the original high-dimensional feature set. Considering the large number of invalid and redundant features contained in such original feature set, a feature selection process that combines with support vector machine (SVM) single feature evaluation, correlation analysis and principal component analysis-weighted load evaluation (PCA-WLE) is put forward in this paper for selecting sensitive features. The selected features are weighted and fused according to their sensitivity so as to further weaken the interference of low important features. Finally, this process is applied to the data provided by the Case Western Reserve University Bearing Data Center and Xi'an Jiaotong University School of Mechanical Engineering, respectively, and the fault is diagnosed by using the particle swarm optimization-support vector machine (PSO-SVM). The results show that this method can accurately identify different fault categories and degrees of bearing, which is superior and practical than single-domain fault diagnosis with higher recognition ability.

INDEX TERMS Features selection, feature weighting, sensitive features, fault diagnosis.

I. INTRODUCTION

Rotating machinery is a very essential power unit for industrial applications and is widely used in various production and processing fields [1]. As a key component of the transmission of power in a rotating machine, the running state of the rolling bearing is directly related to the performance state of the mechanical equipment [2], [3]. Due to the harsh working environment and often at full load, the rolling bearings are extremely easy to wear out and accumulate to form faults. Once the fault occurs, it may cause a series of impacts on the enterprise, such as production equipment shutdown, economic benefit damage and casualties [4]. According to statistics, due to the damage of the bearing, the rotating mechanical equipment can not operate normally, accounting for about 40% [5]. Therefore, monitoring the bearing status, discovering and eliminating potential faults in time, and maintaining the safe operation of the equipment are of great significance [6].

Generally, fault diagnosis can be divided into three types [7]–[9], namely, analytical model-based method, qualitative empirical knowledge based method, and data driven based method. The analytical model-based method is based on the mathematical model of the known diagnostic object, and the information of the measured object is processed according to a certain mathematical method. Using this method requires having enough sensors, understanding the process mechanism structure, and a more accurate quantitative mathematical model. At present, the method mainly includes three methods: method based on parameter estimation [10], method based on state estimation and method based on equivalent space, and all of them have been studied in depth. However, due to the fact that it is difficult to obtain an accurate mathematical model of the research object in practice, the scope and effect of the method are greatly limited [11].

The associate editor coordinating the review of this manuscript and approving it for publication was Yu Wang.

The qualitative empirical knowledge based method mainly depends on the accumulated experience gained during the operation of the system. According to the incomplete prior experience, the operating state of the equipment is described and a qualitative model is established. The next state of the equipment is predicted by reasoning. This kind of fault diagnosis method includes singed directed graph [12], fault tree [13], expert system [14] and so on. However, the diagnostic ability of knowledge-based fault diagnosis methods depends only on the historical experience of experts or field workers. With the acceleration of industrial upgrading and the deepening of relevant professional knowledge, the empirical knowledge often exceeds the range that can be grasped by ordinary workers, making it difficult to carry out. operating. This method is especially not suitable for large industrial systems. Moreover, the above two methods are more suitable for systems with fewer input, output and state variables, and are less practical for multi-sensor and mass acquisition data systems.

Data-driven fault diagnosis methods include: (1) statistical-based methods; (2) signal-based methods; and (3) artificial intelligence-based methods. With the rapid development of data mining, computer technology and artificial intelligence [15], data-driven fault diagnosis methods have increasingly shown their strong applicability, and often use a combination of three methods. Based on the redundant second generation wavelet packet transform (RSGWPT), Liu *et al.* [16] extracted 56 features of the vibration signal and input support vector machine (SVM) for fault identification; Tian *et al.* [17] selected permutation entropy (PE) as the fault feature, and proposed a manifold-based dynamic time warping method for fault diagnosis; Li *et al.* [18] selected 1634 features and classified the bearing faults using the method of fuzzy C-means with a variable focal point (FCMFP); In [19], composite multiscale fuzzy entropy (CMFE) was selected as the feature to train the ensemble support vector machine (ESVM) for fault diagnosis of the rolling element bearings; In [20], the energy entropy of the intrinsic mode function (IMFs) of the bearing vibration signal is extracted, and combined with probabilistic neural network (PNN) and simplified fuzzy adaptive resonance theory map (SFAM) for online bearing fault diagnosis; In [21], the hierarchical symbol dynamic entropy (HSDE) is used as a sensitive feature input binary tree support vector machine (BT-SVM) to effectively identify the fault of the bearing. Most of these tasks use statistic and signal analysis to extract the features of vibration signals and to diagnose faults based on artificial intelligence. Some documents also use deep learning methods to automatically extract fault features for diagnosis. All kinds of them have greatly advanced the fault diagnosis research of bearings.

The data-driven fault diagnosis process can be divided into four steps: signal processing, feature extraction, features reduction, and patterns recognition [22], [23], and the first three are the foundation of the fourth step. Features reduction includes feature selection and feature dimension reduction. Compared with features reduction and patterns recognition, there are relatively few studies on features reduction. On the one hand, the increasing feature extraction method leads to an increase in the feature vector dimension, but not all fault features have an effect on bearing fault diagnosis. The increase of invalid features is likely to cause the diagnosis process to be more complicated and the accuracy of the diagnosis results to be reduced [24]. On the other hand, different types of features have different applicability in different types of bearing failures or different stages of bearing operation [25]. Therefore, features should be simplified after feature extraction is completed, and the optimal features for maintaining the intrinsic information about the faults should be retained under the condition of reducing the number of features as much as possible, so as to effectively and efficiently diagnose the faults of bearings. Liao *et al.* [26] selected two different clustering analysis methods to classify the bearing data, and used the correlation analysis method to reduce the dimensionality of the data; Yang *et al.* [27] extracted the fault features in the vibration signal by means of ensemble empirical mode decomposition (EEMD), and reduced them by using principal component analysis (PCA); In [28], correlation, monotonicity and robustness were selected as the evaluation indicators of the features. Using these indicators, the residual life trend of the bearing was well displayed, and the remaining service life of the bearing was effectively predicted; In [29], an adaptive feature selection technique was proposed. This technique can be used to remove redundant features and reduce the amount of computation for pattern recognition; In [3], the Hilbert time-time (HTT) transform was combined with principal component analysis to extract and reduced the bearing fault features. At present, some researchers have studied the selection and dimension reduction of bearing fault features, but there are still some deficiencies in these research work. For one thing, many articles only consider single-fault features, such as time domain statistics or frequency domain statistics, which cannot reflect fault information more comprehensively, and the comprehensiveness of features is poor; For another, the existing methods of dimensionality reduction mostly use a single method such as Linear Discriminant Analysis (LDA) and PCA, which cannot reflect the difference between samples. Moreover, these methods use mathematical means to reprocess the data. The new features are obtained by combining a plurality of original features, and the physical information cannot be directly represented to guide the subsequent equipment processes. Therefore, the selection and dimension reduction of fault features should be further explored in order to adaptively select the optative sensitive features.

The rest of this paper is organized as follows. In section II, a basic theories of the SVM, PCA, correlation analysis and multi-dimensional feature extraction techniques is outlined. In section III, the specific steps of the proposed feature selection method is described in detail, and the system framework

of the method is given. Experimental verifications of actual data are conducted in Sections 4 and conclusion and recommendations for future work are summarized in section V.

II. BASIC THEORIES

A. MULTIDIMENSIONAL FEATURE EXTRACTION

1) TIME-DOMAIN FEATURE

As the simplest and most direct signal analysis method, time-domain analysis is currently applied to most rolling bearing online monitoring systems. Generally, it performs signal analysis by calculating the simple statistical characteristic quantity of the signal, and then selects appropriate feature parameters to accurately classify different types of faults. The statistical parameters are mainly divided into two categories according to the presence or absence of the dimension [30]. The first category is the dimensional statistical parameters, including maximum value, minimum value, mean value, root mean square (RMS) value, peak-to-peak value, and standard deviation. The other category is dimensionless statistical parameters, including skewness [31], kurtosis factor [32], peak factor, form factor, pulse factor, and margin factor.

Using dimensional statistical parameters to describe the bearing state can reflect part of the fault information, such as RMS of vibration signal, which can directly reflect the vibration intensity of the bearing and is an important evaluation index. However, the dimensionless statistical characteristic values are not only related to the type, size and state of the bearing, but also in connection with the changes in external motion parameters (e.g. speed, pressure, load, etc.), and for different working conditions, there will be large variation in feature values so that it is unable to draw an unified conclusion. The dimensionless statistical parameter is insensitive to changes in external parameters, i.e., independent of the bearing's motion conditions, so it is an ideal monitoring parameter in machine condition. For example, the kurtosis, peak factor, and pulse factor can be adopted for detecting the impact component in the vibration signal. The dimensionless characteristic parameter shall be zero-averarized, which is, removing the mean from the original data and leaving only the dynamic part.

It can be seen that the information reflected by the time domain parameters is limited, and the information that different eigenvalues can display is also different. Since the wear-type failure of the bearing is usually reflected in the high amplitude level of the vibration signal, RMS and the peak value can be used to determine the degree of wear [33]. The RMS of the vibration signal increases with the wear of the bearing. However, although RMS can reflect the surface roughness caused by the manufacturing quality or wear of the bearing working surface, it has certain limitations on the failures such as partial peeling, scratches, indentations and pits on the bearing components. The pulse shape of these discrete faults has a high peak, and for such an impact fault [34], the peak factor is more representative than RMS. When the peak factor is relatively small, it can reflect that the bearing has poor lubrication. The stability of such features is poor, and sometimes the value decreases as the degree of failure increases. In general, time-domain based fault feature extraction is still in a relatively early stage.

2) FREQUENCY-DOMAIN FEATURE

The working principle of bearing determines that the corresponding fault frequency component will be inevitably generated in the frequency-domain when the bearing breaks down. Therefore, from the perspective of the frequency domain, fault extraction of bearings is theoretically feasible.

The Fast Fourier Transform (FFT) can obtain the distribution of the frequency components of the signal in the spectrogram, and can provide more intuitive information content than the time domain waveform. A state spectrum analysis of the rolling bearing can help monitor its operating state or find the location of vibration source. E.g., the bearing rotational frequency and amplitude as well as phase of the main frequency components such as higher harmonics can be obtained through spectrum analysis, which provides an effective analysis method for judging the location, type and severity of the bearing fault; in the process of inspecting bearing turntable, the operating condition and deterioration degree of the bearing can be judged by comparing the amplitude variation under the same frequency component and the presence of a new frequency. The frequency domain-based analysis method can smooth the non-stationary components in the signal and reflect the frequency information in the signal, but it also has the limitations of being unable to reflect the change of the signal frequency with time, so it is not suitable for analyzing non-stationary signals.

Commonly used spectral methods include Fourier transform, cepstrum analysis, refinement spectrum analysis [35], order tracking spectrum analysis, etc. Different analysis methods focus on different directions. For example, in addition to identifying and separating the periodic components in the signal, cepstrum analysis can also effectively extract the fault information in the signal when there is an unrecognizable multi-cluster modulation sideband in the bearing fault signal [36]. The order tracking spectrum analysis is available for extracting bearing fault features under variable speed. By establishing the corresponding relationship between the rotation speed and frequency in speed-up and speed-down stages, it analyzes by converting the time-domain signal into angle signal [37]. In addition, some statistical indicators [38], such as center frequency (CF), root mean square frequency(RMSF) and standard deviation frequency(STDF), also have good discrimination ability for bearing faults. CF and the RMSF can describe the position change of the main spectrum of the power spectrum, and STDF can describe the degree of dispersion of the spectral energy [39].

In summary, both time-domain features and frequency-domain features are a representation of the overall signal. Or completely in the time-domain, or completely in the frequency-domain, it is impossible to characterize when and how the signal will change at a certain frequency component.

The information expressed by this method is not comprehensive and requires a joint distribution of time and frequency to characterize the signal.

3) ENERGY FEATURE

Since the measured vibration signal contains not only the operating condition information related to the bearing itself, but also a large amount of information about other rotating parts and structures in the unit equipment, of which the latter belong to background noise compared to the former [40]. Background noise is usually so large that the slight bearing fault information will be submerged and difficult for extraction. Thus, it is hard to accurately assess the working condition of bearing through the conventional time-domain and frequency-domain methods [41], [42]. Therefore, the method of time-frequency analysis based on Wigner-Ville Distribution (WVD) [43], Wavelet Transform (WT) [44] and Empirical Mode Decomposition (EMD) [45] has been widely used in recent years.

The time-frequency analysis can characterize the variation of the signal spectral components over time, and finally characterize the distribution of signal strength or energy simultaneously in time and frequency. As soon as the rolling bearing breaks down, the energy of the fault feature band corresponding to the vibration signal will be significantly increased, so that the fault type and the fault location can be determined by judging the characteristic frequency band in the wavelet decomposition result that includes fault information. Therefore, the deep information of the fault type can be reflected by decomposing the signal via the time-frequency method and extracting the energy characteristics in different frequency bands.

4) ENTROPY FEATURE

Entropy is a measure of information uncertainty [46]. The entropy of different frequency bands can be used to measure the uncertainty of signal distribution state and signal complexity, so it can quantitatively describe the information contained in the signal. According to the overall average characteristics of the signal source, entropy can manifest the complexity of system internal information, so the essential information of the bearing fault can be extracted based on the effective entropy value. Commonly used entropy features are Shannon entropy, index entropy [47], and permutation entropy [48]. For a discrete random variable X with a sample space of $[x_1, x_2, \ldots, x_n]$, the Shannon entropy is:

$$H(X) = E\left(\log_2 \frac{1}{p(x_i)}\right) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \quad (1)$$

where $p(x_i)$ represents the probability of the sample. Index entropy can avoid the case where the logarithm of Shannon entropy is prone to undefined and zero values. Its definition is as follows:

$$H_{EXP} = -\sum_{i=1}^{n} p_i e^{(1-p_i)} \quad (2)$$

The feature extraction according to entropy theory is applicable to the environment with high signal-to-noise ratio, but when the effective signal is completely submerged by noise, a large overlap will be triggered between different signal entropies, making it difficult to accurately distinguish features.

One of the cores of comprehensive diagnosis and prediction of bearing development fault state is the extraction of signal fault features. It is particularly crucial to select features that can accurately represent the fault category to improve the accuracy of the diagnosis results. From the above analysis, we can know that the fault information displayed by different categories of features is not the same, so it is necessary to establish a high-dimensional feature set that can represent the fault state of the bearing to a large extent. In this paper, based on the different characteristics of the system, the time domain, frequency domain statistical parameters, wavelet packet decomposition energy and entropy composition feature set are extracted for subsequent operations.

B. SUPPORT VECTOR MACHINE

The support Vector Machine (SVM) [49], [50] is a classification method based on the principle of structural risk minimization proposed by *Vapnik et al.*. The main purpose of the SVM is to not only correctly classify the various sample points, but also to maximize the spacing between the them, that is, to maximize the minimum distance between the optimally divided hyperplane and all training sample points. The principle can be described as follows:

Given an original data sample set:

$$\left\{(x_i, y_i) \,\middle|\, x_i \in \mathrm{R}^d, y_i \in \{-1, +1\}, i = 1, 2, \ldots, n\right\}. \quad (3)$$

where n is the number of training data samples, and x_i is the input of the model; d represents the dimension of the training sample; y_i is the sample category; -1 and 1 are category labels.

For the linearly separable case, the separation plane equation is $w \cdot x + b = 0$. The sample (x_i, y_i) needs to satisfy:

$$y_i [(w \cdot x_i) + b] \geq 1, \quad i = 1, 2, \ldots, n \quad (4)$$

where w is the plane normal vector and b is the constant term.

The distance between the nearest sampling point and the separation plane is $1/\|w\|$. Therefore, the maximum spacing of $1/\|w\|$ can be equivalent to the minimum value of $\|w\|^2$. The separation line determined by w is the optimal separation line, and the sample points on the separation line $w \cdot x + b = \pm 1$ are called support vectors.

The Lagrange optimization method is adopted to convert it into its dual problem, namely, the maximization function:

$$\max W(\alpha) = \sum_{I=1}^{N} \alpha_i - \frac{1}{2} \alpha_i \alpha_j y_i y_j \left(x_i \cdot x_j\right) \quad (5)$$

where, α_i is Lagrange multiplier, and $\alpha_i \geq 0$, $i = 1, \ldots, n$. It actually aims to find the optimal solution of quadratic function with constraints, and the sample corresponds to the

non-zero α_i in the solution is support vector, so that the optimal classification function can be obtained in this way:

$$f(x) = \mathrm{sgn}\left\{(w \cdot x) + b^*\right\} = \mathrm{sgn}\left\{\sum_{i=1}^{n} \alpha_i^* y_i (x_i \cdot x) + b^*\right\} \tag{6}$$

where, α_i^* is optimal Lagrange factor and b^* is classification threshold, which are the parameters for determining optimal hyper-plane partition. The positive or negative function indicates the class attributes.

Regarding the linear inseparable case, the slack variable ξ_i is introduced, so as to convert the problem of looking for hyper-plane into quadratic programming problem:

$$\begin{cases} \phi(\omega) = \min \dfrac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N} \varepsilon_i \\ \text{s.t.} y_i\left[(w \cdot x_i) + b\right] \geq 1 - \xi_i, \xi_i \geq 0, \quad i = 1, \ldots n \end{cases} \tag{7}$$

where, ξ_i is the positive slack variable that allows misclassification, representing the deviation amount of corresponding data point x_i from the hyper-plane. C is penalty factor, indicating the degree of punishment for misclassification. It is used for control the weight between looking for the hyper-plane with maximal spacing in the objective function and guaranteeing the minimum deviation amount at the data point.

For the nonlinear separable case, the low-dimensional input space can be mapped into the high-dimensional feature space by introducing the kernel function, so as to realize the linear classification after nonlinear classification transformation. In this case, the classification function becomes:

$$f(x) = \mathrm{sgn}\left\{\sum_{i=1}^{n} \alpha_i^* y_i K(x_i, x) + b^*\right\} \tag{8}$$

where $K(x_i \cdot x)$ is the kernel function.

Replacing the inner product of the original space with a kernel function is the key to SVM. Common kernel functions [51] are as follow: 1) Linear kernel $K(x, y) = x \cdot y$; 2) Polynomial kernel $K(x, y) = [(x \cdot y) + 1]^q$; 3) Radial basis function (RBF) kernel $K(x, y) = \exp\left(\|x - y\|^2 / 2\sigma^2\right)$; 4) Sigmoid kernel $K(x, y) = \tanh(\alpha(x \cdot y) + b)$.

The fault diagnosis of rolling bearings is usually a multi-class identification task. In view of the better classification ability of SVM for nonlinear and small training samples, it is still widely used in the field of machine fault diagnosis. Liu *et al.* [52] used the SVM to verify the superiority of the method by merging the Minim Entropy Deconvolution (MED) with the hierarchical fuzzy entropy; Wan *et al.* [53] combined the objective wavelet transform (EWT) with multi-scale entropy to obtain new features, and input SVM to improve the fault diagnosis efficiency of the bearing; In [54], a novel rolling bearing fault diagnosis strategy was proposed based on Improved multi-scale permutation entropy (IMPE), Laplacian score (LS) and Least squares support vector machine-Quantum behaved particle swarm optimization (QPSO-LSSVM); In [55], Zhu et al. proposed a multi-scale global fuzzy entropy (MGFE) feature extraction method, and introduced multiple class feature selection (MCFS) method to filter features, and finally input SVM for fault diagnosis.

However, most literature on bearing diagnostics uses SVM for pattern recognition only in the final step of its algorithm. In this paper, the SVM is directly introduced into the feature selection part to diagnose a single feature. According to its diagnostic rate, it is judged whether the feature has strong correlation with the bearing fault information, and the invalid information is eliminated. By this method, features capable of expressing fault information in the feature set can be extracted to a greater extent. The method can perform a screening of the original feature set.

C. CORRELATION ANALYSIS

In the field of statistical signal processing research, correlation analysis has been the focus of scholars. The study of correlation is a method that uses the relevant two sets of variables to reflect the overall relevance. The Pearson correlation coefficient represents the degree of linear correlation between the two sets of variables [56]. The Pearson correlation coefficient R can be expressed as a formula:

$$R = \frac{\mathrm{cov}(A, B)}{\sigma_A \sigma_B} = \frac{\sum_{i=1}^{N}\left(A_i - \overline{A}\right)\left(B_i - \overline{B}\right)}{\sqrt{\sum_{i=1}^{N}\left(A_i - \overline{A}\right)^2 \sum_{i=1}^{N}\left(B_i - \overline{B}\right)^2}} \tag{9}$$

where A and B represent two sets of features of equal length. N is the number of samples in the variable; A_i and B_i are the i^{th} measurements of variables A and B; $\overline{A}$ and $\overline{B}$ are the average of variables A and B, respectively.

The correlation coefficient R ranges from -1 to $+1$. When the value is 0, there is no linear correlation between the two features. If the value is at $[-1, 0)$, it indicates that the two features are negatively correlated; if the value is at $(0, +1]$, the two features are positively correlated. The closer the absolute value of the correlation coefficient R is to 1, the higher the degree of correlation between the two features, indicating that the duplicate information of the two features is larger; When the absolute value of R is 1, the information represented by the two features can be replaced with each other. Therefore, the larger the absolute value of R, the lower the significance of the corresponding feature [57]: 1) $|R| \geq 0.8$, highly correlated; 2) $0.5 \leq |R| < 0.8$, moderate correlation; 3) $0.3 \leq |R| < 0.5$, low correlation; 4) $|R| < 0.3$, weak correlation, which can be regarded as nonlinear correlation.

In this part, the Pearson correlation coefficient method is used to select the selected features again. The highly correlated features of each type of feature are selected, and only one of the main features is taken as a sensitive feature. It is considered that the physical information expressed by the remaining features is basically the same as this sensitive feature, which is a redundant feature and is excluded to achieve the purpose of the second screening.

D. PRINCIPAL COMPONENT ANALYSIS AND WEIGHTED LOAD EVALUATION

1) PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) is a commonly used data processing and analysis method. Its purpose is to reduce the data to eliminate overlapping information in the coexistence of many information [3], [58], [59].

PCA maps high-dimensional data space to low-dimensional space by orthogonal transform, which recombines many original features with certain correlation into a group of relatively uncorrelated integrated features. This not only retains the main information of the original variables, but the new features are not related to each other. From the mathematical point of view, the m-dimensional feature is mapped to the $k(k<m)$ dimension, and the obtained k-dimensional feature is the principal component feature extracted from the original data feature. This k-dimensional principal component feature already contains most of the information. These new variables are irrelevant and are arranged in descending order of variance [60]. The specific analysis steps are as follows:

(1) Original data standardization: For the evaluation objects in a group n, there are m features: $X_1, X_2, \ldots X_m$; then the j-th indicator value of the i-th evaluation object is marked as x_{ij}, so the sample space matrix of the evaluation object can be obtained:

$$X = (X_1, X_2, \ldots, X_m) = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix} \quad (10)$$

where, $X_i = (x_{1i}, x_{2i} \ldots, x_{ni}), i = 1, 2, \ldots, m$. The various indicators x_{ij} are standardized:

$$\tilde{x}_{ij} = \frac{x_{ij} - \mu_j}{S_j}, \quad i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m. \quad (11)$$

where, $\mu_j = \frac{1}{n}\sum_{i=1}^{n} a_{ij}$, $S_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(a_{ij} - \mu_j\right)^2}$; μ_j and S_j are the sample mean value and standard deviation of the j-th features, respectively; then the corresponding $\tilde{X}_j = \frac{X_j - \mu_j}{S_j}$ is standardized characteristic variable.

(2) Correlation coefficient matrix: the correlation coefficient of the standardized feature is $r_{ij} = \frac{\sum_{k=1}^{n} \tilde{x}_{ki} \cdot \tilde{x}_{kj}}{n-1}, i, j = 1, 2, \ldots, m$. The correlation coefficient matrix is composed as $R = \left(r_{ij}\right)_{m*n}$, and $r_{ii} = 1, r_{ij} = r_{ji}$.

(3) Computerization of eigenvalues and eigenvectors: according to the correlation coefficient matrix, the eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_m \geq 0$ can be obtained from big to small. $\alpha_j = \left(\alpha_{1j}, \alpha_{2j}, \ldots, \alpha_{mj}\right)^{\mathrm{T}}$ represents the eigenvector corresponding to the i-th eigenvalue λ_i. The eigenvectors $\beta_1, \beta_2, \ldots, \beta_m$ can be obtained after orthogonalization and unitization on this basis, where $\beta_j = \left(\beta_{1j}, \beta_{2j}, \ldots, \beta_{mj}\right)^{\mathrm{T}}$.

(4) Selection of important principal components: Let the principal component as $F_1, F_2, \ldots, F_m$. The contribution rate and cumulative contribution rate of principal components are mainly calculated based on the previously computerized eigenvalues. The contribution of each principal component corresponds to the eigenvalues λ_i of the original feature, and then the contribution rate of the j-th principal component is:

$$C_j = \frac{\lambda_j}{\sum_{i=1}^{m} \lambda_j} * 100\% \quad (12)$$

Since the variance of each principal component is decreasing, the amount of information contained is also decreasing. Therefore, in the actual analysis, it is generally not to select m principal components, but to select the first k principal components (A_k reaches 85%-90%) according to the cumulative contribution rate of each principal component. The contribution rate here refers to the proportion of the variance of a principal component to the total variance, that is, the proportion of a certain eigenvalue to the total eigenvalues:

$$A_k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{m} \lambda_i} * 100\% \quad (13)$$

The greater the variance contribution rate, the stronger the ability of the selected principal components to reflect comprehensive information.

2) WEIGHTED LOAD EVALUATION

In practical applications, after selecting the important principal components, we must also pay attention to the interpretation of the actual meaning of the principal components. Since the new principal component is obtained by orthogonal transformation of the original features, each principal component reflects the comprehensive information of multiple original variables. Therefore, it is difficult to obtain the original fault information directly from the principal component. To this end, the article uses the load analysis method to obtain the load factor matrix of the k principal components and score the original features.

(1) Load factor matrix:

It can be learned from the principle of the principal component analysis method that each principal component can be obtained by linear combination $X_1, X_2, \ldots, X_m$:

$$\begin{aligned} F_1 &= \alpha_{11} \times X_1 + \alpha_{21} \times X_2 + \cdots + \alpha_{m1} \times X_m \\ F_2 &= \alpha_{12} \times X_1 + \alpha_{22} \times X_2 + \cdots + \alpha_{m2} \times X_m \\ &\vdots \\ F_m &= \alpha_{1m} \times X_1 + \alpha_{2m} \times X_2 + \cdots + \alpha_{mm} \times X_m \end{aligned} \quad (14)$$

Each principal component F_i in (14) corresponds to i-th eigenvalue λ_i:

$$F_i = X\alpha_i\,(i = 1, 2, \ldots, m)\,. \quad (15)$$

According to the correlation matrix theorem, α_i satisfies the Equation (16):

$$\sum_{1}^{m} \alpha_i \alpha_i^{\mathrm{T}} = 1 \quad (16)$$

Combining (15) and (16), we get:

$$X = \sum_{1}^{m} F_i \alpha_i^{\mathrm{T}} \quad (17)$$

As mentioned above, the k principal components $F_1, F_2, \ldots, F_k$ $(k < m)$ is obtained according to the original features $X_1, X_2, \ldots, X_m$, and this meets $COV\left(F_i, F_j\right) = 0$, namely, F_i and F_j are not correlated; The variance $D\left(F_i\right)$ is greater, so the first k principal components can stand for the majority of information in original features with lowered dimensionalities. The linear equations of the first k principal components can be derived:

$$F_i = \alpha_{1i} \times X_1 + \alpha_{2i} \times X_2 + \cdots + \alpha_{ki} \times X_k, \quad i = 1, 2, \ldots, k. \tag{18}$$

Then the original features can be expressed:

$$\hat{X} = \sum_{1}^{m} F_i \alpha_i^{\mathrm{T}} \tag{19}$$

where, the combination coefficient $\alpha_i = (\alpha_{1i}, \alpha_{2i}, \ldots, \alpha_{ki})^{\mathrm{T}}$ of each principal component is the load factor matrix corresponding to the eigenvalue of the original feature.

(2) Original feature evaluation

The feature vector coefficients of the principal components can be calculated based on the obtained principal component feature values, contribution rates, and load factors, as (20) displays [61].

$$f_{ij} = \frac{c_{ij}}{\sqrt{\lambda_i}}, \quad i = 1, 2, \ldots k, \; j = 1, 2, \ldots m. \tag{20}$$

where f_{ij} is the eigenvector coefficient of the main component, α_{ij} is the component load of each feature under the principal component, and λ_i is the eigenvalue of the corresponding principal component.

The weight ω_i corresponding to each principal component can be obtained from the corresponding variance contribution rate. A mathematical model of the principal component composite score can be obtained by linearly summing all the principal components.

$$F = \sum_{1}^{k} \omega_i \times F_i, \quad i = 1, \ldots, k. \tag{21}$$

The weighted load score v of the original feature can be expressed:

$$v = \sum_{1}^{k} \omega_i \times \alpha_{ij}, \quad i = 1, \ldots, k, \; j = 1, \ldots m. \tag{22}$$

III. METHOD AND SYSTEM FRAMEWORK

The information represented by a single feature is limited, and does not fully reflect the fault information of the bearing signal. Extracting multiple features can more accurately determine the fault category. Therefore, it is necessary to construct multi-features with different dimensions such as statistical parameters, energy and various entropies, and to use the difference complementarity between different features to construct a more comprehensive high-dimensional feature set that expresses fault type information. However, if the feature concentration dimension is too high, it will inevitably be doped with some invalid or redundant features, and may lead to ''dimension disaster'', which will increase the calculation amount and reduce the prediction efficiency. Therefore, it is necessary to reduce the dimension of the feature set as much as possible while ensuring the integrity of the information to obtain the best feature vector in accordance with the processing background. Depending on the best sensitive characteristics, bearing faults can be diagnosed and the operation of the equipment can be further guided.

A. CONSTRUCTION OF FAULT FEATURE SET

According to the multi-dimensional feature extraction method mentioned in section 2.1, the feature set Q_1 of the bearing under a certain machining condition can be constructed. The feature set contains four types of features, namely four sub-feature sets $\{T_1, F_1, E_1, S_1\}$, which represent time-domain features, frequency domain features, energy features and information entropy features, respectively, where $Q_1 = T_1 + F_1 + E_1 + S_1$.

12 commonly used time-domain parameters $t_1 \sim t_{12}$ are selected to form a time-domain feature set T_1, including: mean, RMS, absolute mean, amplitude of RMS, peak-to-peak value, peak factor, standard deviation, kurtosis factor, form factor, pulse factor, margin factor and skewness factor. For the frequency-domain parameters, due to the working principle of the bearing, the corresponding fault frequency component will be generated when the bearing fails. The change of each frequency component in the signal will cause corresponding changes in the power spectrum. By describing the variation of the main frequency band in the power spectrum, the frequency-domain feature variation of the bearing signal can be well described. The frequency domain feature set consists of CF, RMSF, and STDF: $P_1 = \{p_1, p_2, p_3\}$. The relevant calculation equation is shown in Table 1.

The energy of each frequency component in the signal contains a wealth of fault information. The article decomposes the original signal by means of wavelet packet decomposition. By conducting i-th layer of wavelet packet decomposition on original signal X, a wavelet packet decomposition sequence $S_{i,j}(j = 1, 2, \ldots, 2^i)$ can be obtained. The secondary energy type is used to indicate the reconstructed signal corresponding to each frequency band; then the energy spectrum [62] of the j-th frequency band of i-th layer of wavelet packet decomposition is:

$$E_{i,j}(l) = \left|x_{i,j}(l)\right|^2 \tag{23}$$

wherein, $x_{i,j}(l)$ is the discrete point amplitude of the reconstructed signal, j is the frequency band serial number of the i-th layer after decomposition, l is the sampling point serial number$(l = 1, 2, \ldots n)$, n is the total number of signal sampling points. Then the wavelet packet energy spectrum of each frequency band can be obtained: $E_i = \left[E_{i,1}, E_{i,2}, \cdots, E_{i,2^i}\right]^T$. The total signal energy E_T at certain time window is equal to the sum of the energy of each component. This constitutes an energy feature set: $E_1 = \left\{e_1, e_2, \ldots e_{2^i}, E_T\right\}$.

TABLE 1. Time-domain and frequency-domain feature parameters.

Dimensional feature	Equation	Dimensionless feature	Equation
Mean	$t_1=\frac{1}{N_S}\sum_{i=1}^{N_S}x_i$	Skewness factor	$t_7=(\sum_{i=1}^{N_S}(x_i-T_1)^3)/T_6^3N_S$
Root mean square (RMS)	$t_2=\sqrt{\frac{1}{N_S}\sum_{i=1}^{N_S}(x_i)^2}$	Kurtosis factor	$t_8=(\sum_{i=1}^{N_S}(x_i-T_1)^4)/T_6^4N_S$
Absolute mean	$t_3=\frac{1}{N_S}\sum_{i=1}^{N_S}\lvert x_i\rvert$	Form factor	$t_9=t_2/t_3$
Amplitude of RMS	$t_4=\left[\frac{1}{N_S}\sum_{i=1}^{N_S}\sqrt{\lvert x_i\rvert}\right]^2$	Peak factor	$t_{10}=\max(x_i)/t_2$
Peak-to-peak	$t_5=\max(x_i)-\min(x_i)$	Margin factor	$t_{11}=\max(x_i)/t_4$
Standard deviation	$t_6=\sqrt{\frac{1}{N_s}\sum_{i=1}^{N_s}(x_i-T_1)^2}$	Pulse factor	$t_{12}=\max(x_i)/t_3$
Center frequency (CF)	$p_1=(\sum_{j=1}^{K}f_j\times s_j)/\sum_{j=1}^{K}s_j$	Standard deviation frequency (STDF)	$p_3=\sqrt{(\sum_{j=1}^{K}(f_j-p_1)^2\times s_j)/\sum_{j=1}^{K}s_j}$
Root mean square frequency (RMSF)	$p_2=\sqrt{(\sum_{j=1}^{K}f_j^2\times s_j)/\sum_{j=1}^{K}s_j}$		

Here x_i is a signal series of a dataset for i= *1,2, ... ,*N_S; N_S is the number of data points; s_j is a spectrum for j = *1,2, ... ,*K,; K is the number of spectrum lines; f_j is the frequency value of the j-th spectrum line.

Let $p_j = E_{i,j}/E$ and $\sum p_j = 1$, then the corresponding wavelet energy spectrum entropy can be given according to the measurement of information entropy, that is:

$$S = -\sum_{j=1}^{2^i} p_j \log_2 p_j \tag{24}$$

This constitute an entropy feature set: $S_1 = \{s_1, s_2, \ldots s_{2^i}, S_T\}$.

B. FEATURE SELECTION AND WEIGHTING FUSION

1) FEATURE SELECTION

First, the single-variable feature selection is conducted through SVM, and the respective diagnostic rate φ can be initially obtained by taking each feature in the feature set as the input of the SVM classifier. The features with $\varphi \geq 50\%$ are considered as in close correlation with the bearing fault information, so they are retained; the features with $\varphi < 50\%$ are regarded as invalid features and eliminated. Thereby a screened feature set Q_2 is obtained.

Then, the corresponding similarity γ is obtained based on the correlation analysis of the four sub-feature sets in Q_2. The group of features with $\gamma \geq 85\%$ is considered to have greater similarity in the contained bearing fault information, so only the features with the highest diagnostic rate screened in last round of screening are retained as the main features, and the remaining features are regarded as redundant and removed. In this way, the second round of screening is completed and the feature set Q_3 is obtained.

Directing at a certain sub-feature set with the feature number greater than 5, it is further screened by means of PCA-WLE method. Firstly, the PCA is used, and only k principal components with contribution degree greater than 90% and in close correlation with bearing fault information are selected. Secondly, the original features are then scored according to the principal components. The load matrix of these principal components is calculated, and the principal components are weighted and fused in accordance with their contribution rate so as to obtain the score ranking of the original features. The three features with the highest sum score in the principal component are selected through the weighted calculation method of load evaluation, so that the four new sub-feature sets (T_4, P_4, E_4 and S_4) are formed, which constitute the total feature set Q_4. Finally, the feature selection process is completed.

2) FEATURE WEIGHTING FUSION

The weight of a feature is deemed as an evaluation of the feature importance. The sensitivity of the evaluated features to failure is determined by assigning the figures between [0, 1] to them. According to the feature selection process, a set of features that are sensitive to fault information has been obtained. However, the correlation analysis shows that the fault information reflected by different sensitive features is distinguished, so the weighted fusion on the selected sensitive features is necessary to obtain more accurate and reliable data analysis results.

The corresponding diagnosis success rate R_i is obtained via putting the obtained new sub-feature set Q_4 into the SVM

classifier for diagnosis; the weight W_i of the sub-feature set is obtained according to R_i, which, to some extent, can represent the ability of the fault information in diagnosing the bearing [63]. The corresponding calculation formula is:

$$W_i = \frac{R_i}{\sum_{i=1}^{M} R_i}, \quad i = 1, 2, \ldots, M. \tag{25}$$

where, M is the number of features.

Before the weighted fusion of features, it is necessary to standardize the features to prevent from flushing out the features with smaller data values by those with greater data values, so as to avoid affecting the calculation results due to different dimensions. The feature value q_i of the i-th feature is normalized according to (26).

$$q_i' = \frac{q_i - \min(q_i)}{\max(q_i) - \min(q_i)}, \quad i = 1, 2, \ldots, M. \tag{26}$$

A feature that can describe the fault information to the greatest extent can be obtained through multiplying the feature q_i in feature set Q_4 by its corresponding weight W_i and then summing them up. The sum of the weights is 1. The new fusion feature obtained from this linear weighted combination is calculated as follows:

$$M = q_1 \bullet W_1 + q_2 \bullet W_2 + \ldots + q_i \bullet W_i \times \begin{cases} W_i > 0 \\ \sum W = 1_i \end{cases} \quad i = 1,\ 2,\ \ldots,\ M. \tag{27}$$

Depending on the weight W_i, the sensitivity of the selected feature to the fault can be obtained. Combined with SVM, the fault type and severity of the experimental data are identified to verify the effectiveness of the method.

C. SYSTEM BLOCK DIAGRAM

The implementation flow of the feature screening model proposed in this paper is shown in Fig.1. Taking the fault diagnosis of the bearing as the goal, the process is divided into five steps: signal processing, feature extraction, feature selection, feature weighting fusion and patterns recognition.

Step 1: Collecting the vibration signal in the bearing operation and decomposing the collected signal into different frequency bands.

Step 2: Feature extraction is performed on the original signal and each frequency band signal, and its time-domain, frequency-domain, energy and entropy features are obtained and constitute the original high dimensional feature set.

Step 3: The original features are sequentially subjected to three feature selection processes: SVM single feature selection, correlation analysis and PCA-WLE. The invalid feature and redundant feature in the feature set are eliminated to obtain the low-dimensional sensitive feature set.

Step 4: The corresponding diagnosis rates are obtained by inputting the features in low-dimensional feature set into the SVM, respectively, and the corresponding weight is obtained based on the diagnosis rate. A fusion feature that can describe the fault information to the greatest extent can be obtained through multiplying each standardized feature by its corresponding weight and then summing them up.

Step 5: The fusion feature is used as an input for pattern recognition so as to train the fault classifier, and the obtained weights can be further reflected in the feature extraction process to guide the model to perform feature extraction according to a certain weight.

IV. EXPERIMENTS AND ANALYSIS RESULTS

A. CASE1

1) DATA DESCRIPTION

In order to verify the feature selection method proposed in this paper, the rolling bearing fault signal provided by the laboratory of Case Western Reserve University(CWRU) [64] is taken as an example for testing. The bearing parameters used in the test are shown in Table 2. The entire test stand consists of a three-phase asynchronous motor (left), a torque encoder (center), a dynamometer (right) and associated vibration acceleration sensors, as shown in Fig. 2.

The single-point-fault was introduced to the test bearings using electro-discharge machining with fault positions of inner raceway, outer raceway and rolling element. Fault diameters include 0.1778mm, 0.3556mm and 0.5334mm (fault severity: mild fault, moderate fault and severe fault). With normal bearing data, bearing data can be divided into 10 types for each condition. The motor no-load speed is 1797r/min.

The vibration signals at the driving end is recorded by the acceleration vibration sensor with a sampling frequency of 12 kHz under different motor loads of 0-3 horsepower(motor speeds of 1730 to 1797 rpm). The data under the three load conditions form three data sets **A**, **B** and **C** respectively. As can be learned from the motor speed and the sensor sampling frequency, about 400 data points are collected in one rotation of the bearing. Therefore, in order to ensure that the length of a single sample can completely and accurately reflect that data distribution of the bearing vibration signals in this state, the first 120000 points of the raw data in each sample are taken, and every 1200 data points are regarded as a small sample length, so that each raw data can produce 100 samples. Let the first 70 groups be used for establishing the sample knowledge base and the last 30 groups be the validation samples to test the method validity. Detailed information of the bearing vibration data set is shown in Table 3.

When the fault diameter is 0.5334mm, the samples with the bearing condition of 1hp in the normal state and different fault states are extracted, and the respective vibration signals are shown in Fig. 3.

2) ANALYSIS RESULTS

According to the system flow shown in Fig. 1, the extracted original signal $X\{x_1, x_2, \ldots, x_n\}$ is processed first. The ''db5'' wavelet is selected to decompose the vibration signal into four layers, and the characteristic signals 16 frequency

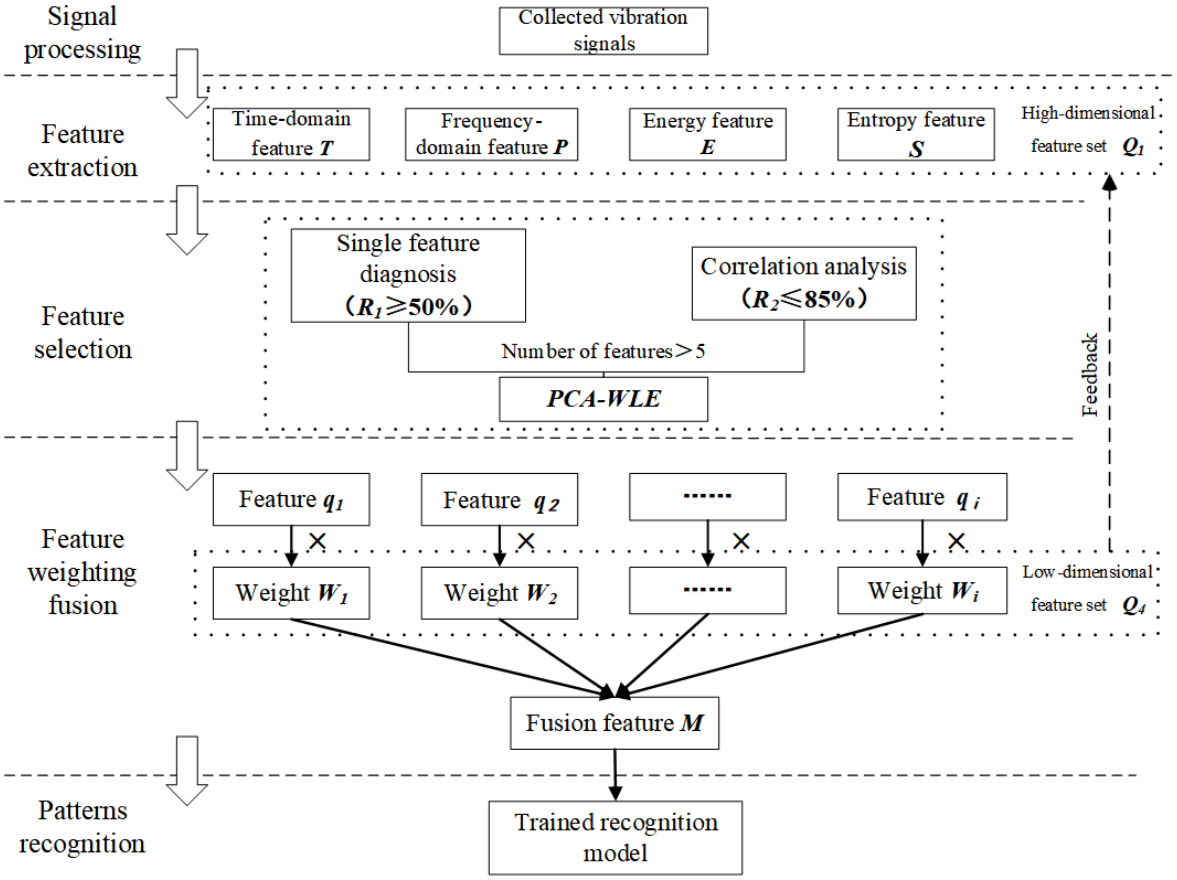

FIGURE 1. Implementation of feature selection model.

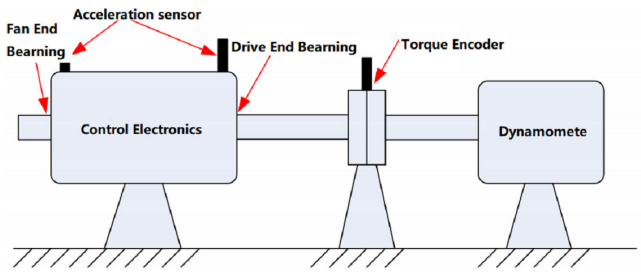

FIGURE 2. Experimental test stand [3], [64].

TABLE 2. Tested bearing parameters.

Bearing Type	Inside Diameter	Outside Diameter	Thickness	Ball Diameter	Pitch Diameter	Number of the roller	Contact Angle
6205-2R JEM SKF	25.001 mm	51.999 mm	15.001 mm	8.031 mm	39.040 mm	9	0°

bands at the fourth layer from the low frequency to the high frequency are obtained. The wavelet packet coefficients are reconstructed to obtain the reconstructed signal. By calculating the total band energy e_i $(i = 1, 2, \ldots, 16)$ of each node in

TABLE 3. Detailed information of the bearing vibration data set.

Class	Fault type	Fault diameter(mm)	Degree of failure	Sample data length	Speed (r/min)	Number of samples (Data set A/B/C)
1	Normal bearing	0	Normal	1200	1772(1hp) /1750(2hp) /1730(3hp)	100
2	Inner race fault	0.1778	mild	1200		100
3		0.3556	moderate	1200		100
4		0.5334	severe	1200		100
5	Ball fault	0.1778	mild	1200		100
6		0.3556	moderate	1200		100
7		0.5334	severe	1200		100
8	Outer race fault	0.1778	mild	1200		100
9		0.3556	moderate	1200		100
10		0.5334	severe	1200		100

the fourth layer and the total energy E_Z ($E_Z = \sum_{i=1}^{16} e_i$) of the fourth layer, the energy feature set $E_1 = \{e_1, e_2, \ldots, e_{16}, E_Z\}$ can be formed. According to the energy ratio $p_i = e_i/E_Z$ of different frequency bands, the information entropy characteristic of signal can be obtained, and the information entropy feature set $S = \{s_1, s_2, \ldots, s_{16}, S_Z\}$ can be formed. The time domain feature set and the frequency domain feature set

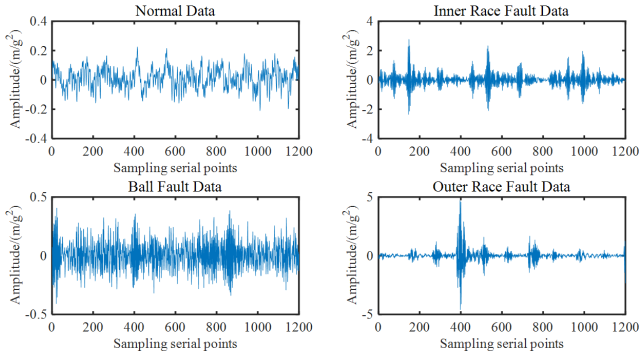

FIGURE 3. Vibration signal of bearing in different states (fault diameter is 0.5334mm).

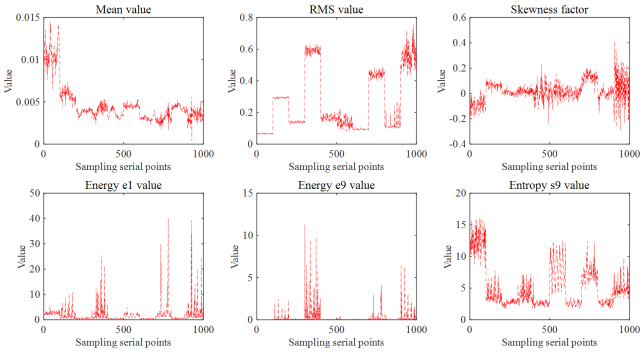

FIGURE 4. Trends in the characteristics of training samples.

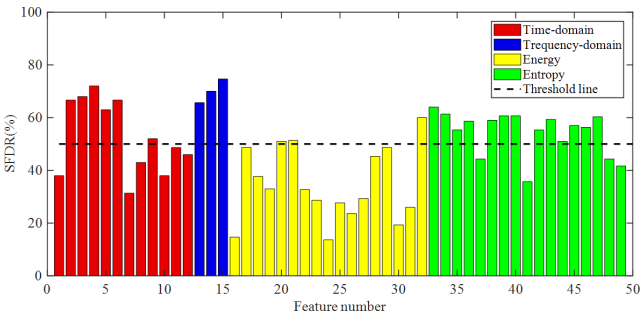

FIGURE 5. The SFDR of 49 features of training samples.

are $T = \{t_1, t_2, \ldots t_{12}\}$ and $P = \{p_1, p_2, p_3\}$, respectively. The fault feature set Q_1 contains 49 fault features, each of which has different distinguishing characteristics and variation degrees. In this paper, the features of mean value, RMS, skewness, e_1, e_9 and s_9 are used as examples, which can be shown in Fig. 4. Therefore, the degree of correlation between features and fault information is different, which requires a further screening to extract the sensitive features that are more suitable for fault diagnosis.

The obtained original feature set is input into the SVM for single feature selection, and the single feature diagnostic rates(SFDR) of different features can be respectively obtained, as shown in Fig. 5. Among them, the 1-12 (red), 13-15 (blue), 16-32 (yellow) and 33-49 (green) represent time-domain, frequency-domain, wavelet energy and wavelet

TABLE 4. Single fault feature correlation rate.

Category	Selected features	Correlation rate				Independent features
Time-domain	T_2 (RMS)	T_3	T_4	T_5	T_6	T_9
		0.977	0.943	0.922	0.999	
Frequency-domain	P_3 (STDF)	P_2		P_1		/
		0.949		0.981		
Energy	E_Z	e_6		e_5		/
		0.915		0.904		
Entropy	s_1	s_{10}				s_3
		0.976				
	s_2	s_4	s_{11}	s_{12}	s_{15}	s_6
		0.948	0.969	0.888	0.957	s_7
	s_{13}	s_{14}				s_8
		0.878				

TABLE 5. PCA analysis results of remaining entropy features.

Principal component	Eigenvalue	contribution rate %	Cumulative contribution rate %
F_1	3.0484	43.55	43.55
F_2	1.9343	27.63	71.18
F_3	1.0586	15.12	86.30
F_4	0.5572	7.96	94.26
F_5	0.2861	4.09	***98.35***
F_6	0.0699	1.00	99.35
F_7	0.0454	0.65	100

entropy features, respectively. And the horizontal axis represents the number of features. Based on the screening threshold of 50%, it is considered that the feature with a diagnosis rate lower than 50% is an invalid feature because of being in weak correlation with the intrinsic information of the bearing fault. According to Fig. 5, only the features with a diagnosis rate greater than 50% are retained, which are 25 features in total.

By conducing correlation analysis on the remaining parts of the four types of features and introducing (9), the linear correlation degree between any of the two features can be obtained. The results are shown in Table 4.

When multiple feature correlations are greater than 85%, only one feature with the highest diagnostic rate is retained, which refers to the gray portion in the table. It can be observed from the table that there are still 7 remaining features based on entropy and only 4 remaining features of three types. In order to further refine the remaining effective features and reduce the amount of computation, the remaining information entropy features are processed by the PCA-WLE method. The seven principal components and the corresponding cumulative contribution rates are obtained, as shown in Table 5.

The (first four) principal components with a cumulative contribution rate of about 90% are retained, and their load

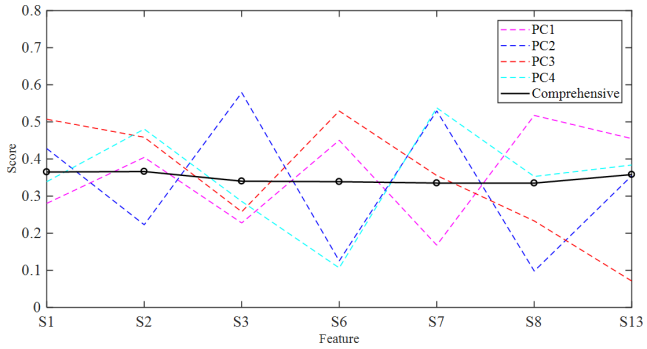

FIGURE 6. The principal component and comprehensive score.

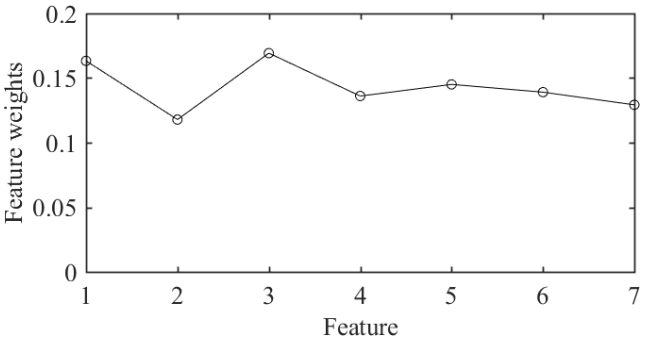

FIGURE 7. The Feature weights of sensitive features.

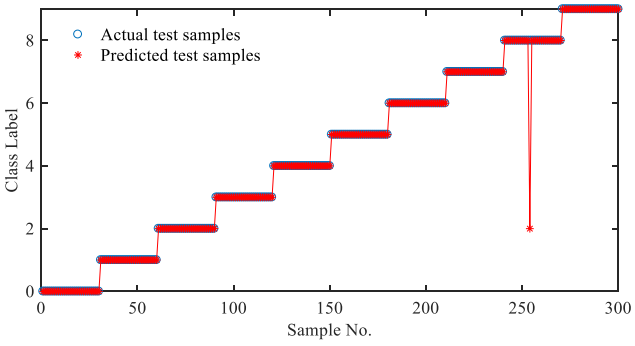

FIGURE 8. Bearing fault diagnosis result based on weighted fusion feature.

factor matrixes (TABLE 5) are calculated. Taking principal component F_1 as an example, its linear expression is:

$$F_1 = 0.4895 \times S_1 + 0.7061 \times S_2 + 0.3975 \times S_3 + 0.7866 \times S_6 - 0.2930 \times S_7 + 0.9036 \times S_8 + 0.7944 \times S_{13}. \quad (28)$$

By introducing the obtained principal component eigenvalues and load factors into (20), the eigenvector coefficients of the principal components can be obtained.

The weight ω_i for each principal component can be obtained from the corresponding variance contribution rate, which are 0.4620, 0.2931, 0.1604 and 0.0844, respectively. By linearly summing all the principal components, the weighted load scores of the original features can be obtained (Table 5). Fig. 6 is a graph of each principal component and comprehensive score. It can be seen that the comprehensive score considers the difference of the four principal components and eliminates the problem of large fluctuation between different principal components.

$$F = 0.4620 \times F_1 + 0.2931 \times F_2 + 0.1604 \times F_3 + 0.0844 \times F_4 \quad (29)$$

According to the score, the top three features are: S_2, S_1 and S_{13}. By this means, the specific fault feature set of this case can be formed:

$$A = \{T_2, T_9, P_3, E_Z, S_1, S_2, S_{13}\}$$

By inputting the remaining features into the SVM again, the diagnostic result rate R_i can be obtained. According to (27), the corresponding weights are obtained as 0.1633, 0.1179, 0.1693, 0.1361, 0.1451, 0.1391 and 0.1293 which respectively represent the proportion of the features in the fault feature set. Fig. 7 shows the weights of 7 features after screening.

The training set is input into the SVM classifier to train the classification model, and the model is optimized by Particle Swarm optimization (PSO) to obtain the optimal parameters c and g of the SVM. The normalized test set features, after multiplying by the corresponding weights, are input into the SVM classifier to identify the fault category, and the result is shown in Fig. 8.

With the purpose of verifying the validity and applicability of the proposed method of selecting fault features, the four different types of screened features are compared with the weighted fusion features obtained by this method. Table 7 presents the comparison results.

It is obvious that the selection feature with weighted fusion in this paper has a higher diagnostic accuracy than the single feature. And the diagnostic rate of selection feature with weighted fusion is higher than that of selection feature without weighted fusion (from 95.67% to 99.61%).

B. CASE2

1) DATA DESCRIPTION

In this case, the test data about bearing lifetime in acceleration provided by Xi'an Jiaotong University School of Mechanical Engineering is taken as the input of the model. The accelerated life testing stand consists of an alternating current (AC) induction motor, a motor speed controller, a support shaft, two support bearings (heavy duty roller bearings) and a hydraulic loading system [65], as shown in Fig. 9. The bearing parameters used in the test are shown in Table 8. Each set of data contains the whole process data of the rolling bearing ranging from normal operation to severe failure (10 times of the maximum amplitude in normal operation) under this operating condition.

Based on the 25.6kHz of sampling frequency of the acceleration sensor and 1min of sampling interval, in this test, about 700 data points are collected during a complete rotation

TABLE 6. Features score and rank.

Feature	F_1		F_2		F_3		F_4		Score	Rank
	α	f	α	f	α	f	α	f		
S_3	0.3975	0.2277	0.8051	0.5789	0.2653	0.2579	-0.2133	-0.2857	0.3403	4
S_2	0.7061	0.4044	-0.3098	-0.2228	0.4718	0.4586	0.3588	0.4807	0.3663	**1**
S_7	-0.2930	-0.1678	0.7367	0.5297	0.3658	0.3555	0.4015	0.5379	0.3352	6
S_{13}	0.7944	0.4550	0.4936	0.3549	0.0733	0.0712	-0.2866	-0.3839	0.3581	**3**
S_8	0.9036	0.5175	-0.1363	-0.0980	-0.2398	-0.2331	0.2634	0.3529	0.3350	7
S_6	0.7866	0.4505	0.1744	0.1254	-0.5447	-0.5294	0.0796	0.1066	0.3388	5
S_1	0.4895	0.2804	-0.5956	-0.4282	0.5218	0.5072	-0.2529	-0.3388	0.3650	**2**

TABLE 7. Bearing fault diagnosis results of different types of feature sets.

Accuracy (%)	Time-domain feature	Frequency-domain feature	Energy feature	Entropy feature	Selection feature without weighted fusion	Selection feature with weighted fusion	Parameters of SVM
Training set	88.00	73.86	61.67	92.00	97.43	99.71	c=1; g=12.25
Testing set	87.00	74.67	60.00	89.67	95.67	**99.61**	

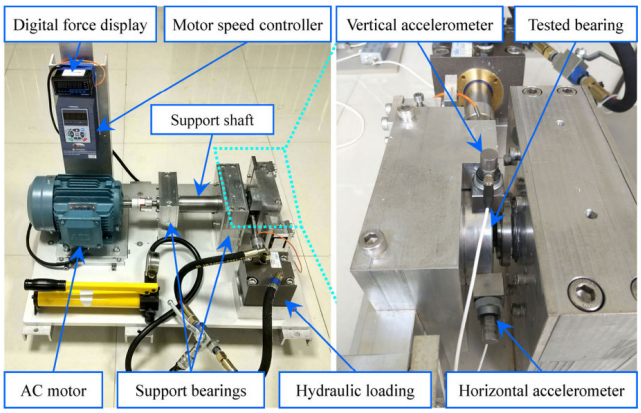

FIGURE 9. Experimental test stand [65].

of the bearing. In order to ensure that the length of a single sample can completely express the signal data distribution, the first 32200 points in each small sample are taken. With a sample length of 700 points, each original sample can generate 46 small sample sets. Let the first 30 groups be used for establishing the sample knowledge base and the last 16 groups be the validation samples to test the method validity.

The data under the operation condition at 2250r/min of rotation rate and 11kN of radial load is selected for verification. Detailed information of the bearing vibration data set is shown in Table 9. The vibration signals, according to their amplitude changes, can be divided into four types: normal data, mild damage, moderate damage and severe damage.

TABLE 8. Tested bearing parameters.

Bearing Type	Inside Diameter	Outside Diameter	Bearing Mean Diameter	Ball Diameter	Number of the roller	Contact Angle
LDK UER204	29.30 mm	39.80 mm	34.55 mm	7.92 mm	8	0°

TABLE 9. Detailed information of the bearing vibration data set.

Operating condition	Class	Number of files	Bearing lifetime	Sample data length	Number of samples	Fault element
2250r/min 11kN	1	491	8h11min	700	46	Inner race
	2	533	8h53min	700	46	Cage
	3	42	42min	700	46	Outer race

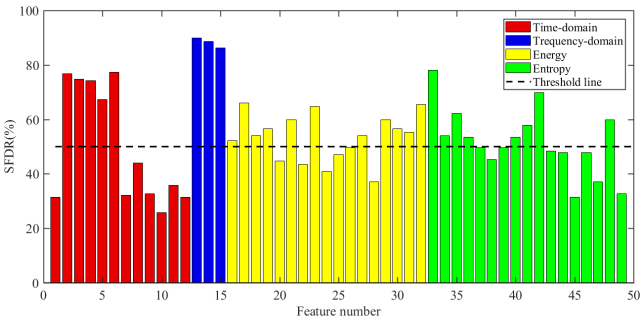

FIGURE 10. The SFDR of 49 features of training samples.

2) ANALYSIS RESULTS

Similarly, the "db5" wavelet is applied to decompose the vibration signal into 4 layers and extract the features from original vibration signal. In this way, 49 fault features can also be obtained. By inputting the obtained original feature set into SVM for single feature screening, 26 fault features with a diagnosis rate greater than 50% are finally obtained, as shown in Fig. 10.

As can be learned from Table 10, the summary of correlation analysis on the 26 remaining features, by removing

TABLE 10. Single fault feature correlation rate.

Category	Selected features	Correlation rate		Independent features
Time-domain	T_2 (RMS)	T_3	T_4	/
		0.996	0.992	
		T_5	T_6	
		0.934	0.999	
Frequency-domain	P_1 (CF)	P_2	P_3	/
		0.979	0.982	
Energy	e_6	e_2	e_3	e_1
		0.874	0.865	
	e_8	e_4		e_{12}
		0.852		
	e_{14}	e_{15}		E_Z
		0.881		
Entropy	s_1	s_{10}		s_2 / s_3 / s_4 / s_8 / s_9 / s_{16}
		0.890		

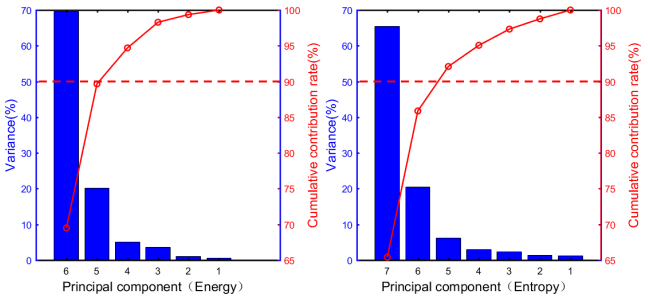

FIGURE 11. PCA analysis results of remaining entropy features.

the redundant features through correlation analysis, one time domain feature (RMS), one frequency domain feature (CF), six energy features, and seven entropy features can be obtained. Therefore, the PCA-WLE method shall be adopted to separately reduce the dimensionality of the energy features and entropy features.

The principal component analysis results of the two types of features are separately shown in Fig. 11. Taking energy features for example, the principal components (first three) with cumulative contribution rate greater than 90% are retained, and their load factor matrix as well as principal component eigenvector coefficients (Table 11) are calculated. Taking the principal component F_1 as an example, its linear expression is:

$$F_1 = 0.3661 \times e_1 + 0.4601 \times e_6 + 0.4102 \times e_8 + 0.2890 \times e_{12} + 4170 \times e_{14} + 0.4781 \times Ez \quad (30)$$

The weight of each principal component can be obtained via the corresponding variance contribution rate, which is 0.7341, 0.2121, and 0.0532, respectively. A weighted load score of the original features can be obtained by linearly summing all the main components (Table 11). Fig. 12 presents

TABLE 11. Features score and rank.

Feature	F_1		F_2		F_3		Score	Rank
	α	f	α	f	α	f		
e_1	0.7477	0.3661	0.5500	0.5004	0.3520	0.6400	0.4092	1
e_6	0.9398	0.4601	-0.2274	-0.2069	-0.1247	-0.2268	0.2817	4
e_8	0.8378	0.4102	-0.3866	-0.3517	0.0877	0.1595	0.2348	5
e_{12}	0.5902	0.2890	0.7156	0.6510	-0.3549	-0.6453	0.3162	**3**
e_{14}	0.8516	0.4170	-0.4325	-0.3934	-0.1341	-0.2439	0.2095	6
E_z	0.9765	0.4781	0.0741	0.0674	0.1068	0.1941	0.3756	**2**

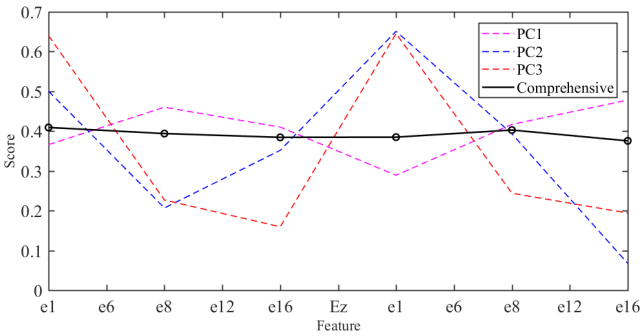

FIGURE 12. The principal component and comprehensive score.

each principal component and the integrate score.

$$F = 0.7341 \times F_1 + 0.2121 \times F_2 + 0.0532 \times F_3 \quad (31)$$

By processing the entropy features in the same way, the feature set needed by the bearing under this operation condition can finally be obtained:

$$A = \{T_2, P_3, e_1, e_6, e_{14}, S_1, S_8, S_{16}\}$$

The fault diagnosis rates corresponding to the above 8 sensitive features are: 0.7673, 0.8994, 0.5220, 0.5975, 0.5975, 0.7799, 0.5346 and 0.5975 By introducing these into formula (27), the corresponding weights can be obtained, which are 0.1449, 0.1698, 0.0986, 0.1128, 0.1128, 0.1473, 0.1009 and 0.1128. Fig. 13 shows the weights of 7 features after screening.

By inputting the normalized test set to the trained PSO-SVM model after multiplying by the corresponding weights, the type of failure can be identified, and the results are shown in Fig. 14. According to Table 12, which shows the diagnosis results of different features, unweighted fusion features and the weighted fusion features proposed in this article, the sensitive features based on weighted fusion have a higher diagnosis rate than the single-domain features and unweighted fusion features.

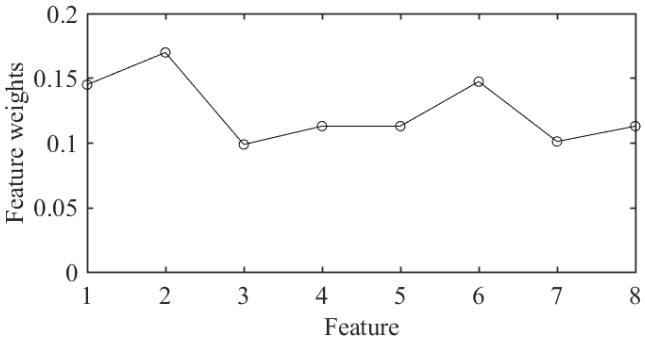

FIGURE 13. The Feature weights of sensitive features.

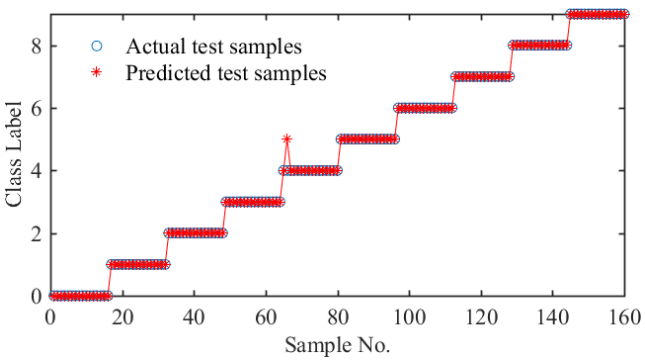

FIGURE 14. Bearing fault diagnosis result based on weighted fusion feature.

TABLE 12. Bearing fault diagnosis results of different types of feature sets.

Accuracy (%)	Time-domain feature	Frequency-domain feature	Energy feature	Entropy feature	Selection feature without weighted fusion	Selection feature with weighted fusion
Training set	73.67	86.67	73.67	88.67	98.67	98.33(295/300)
Testing set	76.10	88.05	78.61	89.31	98.75	**99.38(159/160)**

C. DISCUSSION

As can be observed from Table 7, the diagnosis results of features in different domains are various, of which the diagnosis results of time-domain features and entropy features are significantly better than that of frequency-domain features and energy features. This is because different features have their own importance degree in the fault diagnosis, and sensitive features in close relationship to fault information have higher fault identification ability. Moreover, considering the differentiated number of features included in different domains, although some feature sets contain many features, most of which are not related to the fault or in high overlap and will generate strong interference with the later fault diagnosis.

Compared to the feature extraction in a single domain, the features extracted from multiple domains have higher diagnosis rate. This shows that the multi-domain features

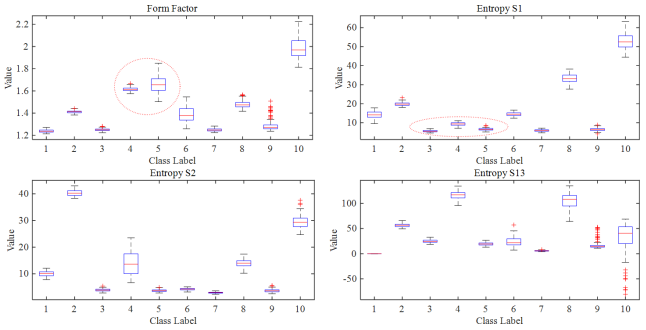

FIGURE 15. Nassi–Shneiderman diagram(N-S) of some features.

can effectively utilize the difference and complementarity between the features, so as to discover the fault information hidden in the signal more accurately and comprehensively. Compared to the features without weighted fusion, the diagnostic rate after weighted fusion of multi-domain features is also improved. This means that weighting can further highlight the importance of sensitive features and expand their role in the classification process. Meanwhile, feature weighting also makes the clustering center closer to the dense region of corresponding clustering, thus highlighting the clustering performance [38].

Although PCA is applied to process the features in this paper, the displayed results are not directly adopted since the principal component cannot directly represent the physical meaning of the fault. Instead, the principal components are further processed to find the sensitive features with highest grade. At the same time, SVM is also used in the process of feature selection, which can provide the indicators that reflect feature differentiation through single feature extraction.

In the section of Case1, the RMS, form factor, STDF, total energy of wavelet packet, Wavelet package entropy s_1, s_2 and s_{13} are eventually extracted as the final sensitive features to constitute the feature set, but this only shows that these features are closely related to this bearing failure. Different rotating machines and operating environments have different effects on vibration signals, so although the method proposed in this paper has been successfully applied to bearings, it is still available for the feature extraction and selection in other machines.

The method proposed in this paper still has room for improvement. First of all, considering the weak physical information contained in some of the features, the features that can be applied to fault diagnosis are not enumerated or applied in detail, which makes the fault feature set unable to fully describe the bearing fault information. Secondly, the SVM classifier, which is used in this paper with better diagnosis results, can explain the superiority of this feature selection method. However, the diagnostic results can be further improved by optimizing the parameters of the SVM or via selecting a more advanced diagnostic algorithm. Finally, the proposed method classifies the categories and degrees

of the bearing fault at one time instead of considering that different fault levels of various fault types may correspond to the same feature value. As shown in Fig. 15, the value of form factor features of severe fault in inner ring highly overlaps that of minor fault in rolling element, which can cause misjudgment in the fault diagnosis and affect the final diagnosis result. Therefore, the respective wear degree shall be further differentiated after identifying the fault type. But the proposed feature selection method is still applicable to this identification plan.

V. CONCLUSION

In this paper, a bearing fault feature selection method based on weighted Multidimensional fusion is proposed. In addition to taking into account the importance of different features during fault diagnosis, this method makes up for the inability of the principal component in explaining the physical meaning when reducing the dimensions based on traditional PCA algorithm. Meanwhile, by means of developing an original high-dimensional feature set via extracting feature parameters from different domains, a feature selection process algorithm that combines with SVM single feature evaluation, correlation analysis and PCA-WLE is put forward in this paper for selecting sensitive features, and the weighted fusion is conducted according to the correlation between selected features and fault information. In the end, the fusion features are input into the PSO-SVM classifier for fault diagnosis, which proves that this method has strong identification ability.

The results of experiment on test data from Case Western Reserve University bearing data center and Xi'an Jiaotong University School of Mechanical Engineering indicate that the features selected by this proposed method can accurately identify and classify the different fault categories and fault severity of bearing. Therefore, it is applicable to the feature selection of various bearings and rotating machineries with great application potential.

AUTHOR CONTRIBUTIONS

Wei Dai and Weifang Zhang conceived and they were experience at assembly process; Yazhou Li built up the model and did the simulations; Wei Dai and Yazhou Li contributed to the writing and editing of the manuscript. Wei Dai checked manuscript and provided some suggestions for revision.

REFERENCES

[1] 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, doi: 10.1016/j.ymssp.2012.09.015.

[2] S. Adamczak, K. Stępień, and M. Wrzochal, "Comparative study of measurement systems used to evaluate vibrations of rolling bearings," *Procedia Eng.*, vol. 192, pp. 971–975, 2017, doi: 10.1016/j.proeng.2017.06.167.

[3] B. Pang, G. Tang, T. Tian, and C. Zhou, "Rolling bearing fault diagnosis based on an improved HTT transform," *Sensors*, vol. 18, no. 4, p. 1203, Apr. 2018, doi: 10.3390/s18041203.

[4] X. Xue and J. Zhou, "A hybrid fault diagnosis approach based on mixed-domain state features for rotating machinery," *ISA Trans.*, vol. 66, pp. 284–295, Jan. 2017, doi: 10.1016/j.isatra.2016.10.014.

[5] 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, doi: 10.1109/jsen.2017.2726011.

[6] Y. Wei, Y. Li, M. Xu, and W. Huang, "A review of early fault diagnosis approaches and their applications in rotating machinery," *Entropy*, vol. 21, no. 4, p. 409, Apr. 2019, doi: 10.3390/e21040409.

[7] J. Peng, L. Fan, W. Xiao, and J. Tang, "Anomaly monitoring method for key components of satellite," *Sci. World J.*, vol. 2014, pp. 1–14, Jan. 2014, doi: 10.1155/2014/104052.

[8] J. Xiong, C. Li, J. Cen, Q. Liang, and Y. Cai, "Fault diagnosis method based on improved evidence reasoning," *Math. Problems Eng.*, vol. 2019, pp. 1–9, Mar. 2019, doi: 10.1155/2019/7491605.

[9] H. Li and D. Xiao, "Fault diagnosis of Tennessee Eastman process using signal geometry matching technique," *EURASIP J. Adv. Signal Process.*, vol. 2011, no. 1, p. 83, Oct. 2011, doi: 10.1186/1687-6180-2011-83.

[10] F. Bagheri, H. Khaloozaded, and K. Abbaszadeh, "Stator fault detection in induction machines by parameter estimation, using adaptive Kalman filter," in *Proc. Medit. Conf. Control Autom.*, Jun. 2007, pp. 1–6, doi: 10.1109/med.2007.4433953.

[11] H. Li and D. Y. Xiao, "Survey on data driven fault diagnosis methods," *Control Decis.*, vol. 26, no. 1, pp. 1–9 and 16, 2011.

[12] X. Ma and D. Li, "A hybrid fault diagnosis method based on fuzzy signed directed graph and neighborhood rough set," in *Proc. 6th Data Driven Control Learn. Syst. (DDCLS)*, May 2017, pp. 253–258, doi: 10.1109/ddcls.2017.8068078.

[13] Y. Wang, X. Li, J. Ma, and S. Li, "Fault diagnosis of power transformer based on fault-tree analysis (FTA)," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 64, May 2017, Art. no. 012099, doi: 10.1088/1755-1315/64/1/012099.

[14] Q. Yao, J. Wang, and G. Zhang, "A fault diagnosis expert system based on aircraft parameters," in *Proc. 12th Web Inf. Syst. Appl. Conf. (WISA)*, Sep. 2015, pp. 314–317, doi: 10.1109/wisa.2015.21.

[15] S. Yuanyuan, G. Lili, and W. Yongming, "Artificial intelligence and learning techniques in intelligent fault diagnosis," in *Proc. 4th Int. Conf. Comput. Sci. Netw. Technol. (ICCSNT)*, Dec. 2015.

[16] Z. Liu, W. Guo, J. Hu, and W. Ma, "A hybrid intelligent multi-fault detection method for rotating machinery based on RSGWPT, KPCA and Twin SVM," *ISA Trans.*, vol. 66, pp. 249–261, Jan. 2017.

[17] Y. Tian, Z. Wang, and C. Lu, "Self-adaptive bearing fault diagnosis based on permutation entropy and manifold-based dynamic time warping," *Mech. Syst. Signal Process.*, vol. 114, pp. 658–673, Jan. 2019.

[18] C. Li, J. V. De Oliveira, M. Cerrada, F. Pacheco, D. Cabrera, V. Sanchez, and G. Zurita, "Observer-biased bearing condition monitoring: From fault detection to multi-fault classification," *Eng. Appl. Artif. Intell.*, vol. 50, pp. 287–301, Apr. 2016.

[19] J. Zheng, H. Pan, and J. Cheng, "Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines," *Mech. Syst. Signal Process.*, vol. 85, pp. 746–759, Feb. 2017.

[20] J. Ben Ali, L. Saidi, A. Mouelhi, B. Chebel-Morello, and F. Fnaiech, "Linear feature selection and classification using PNN and SFAM neural networks for a nearly online diagnosis of bearing naturally progressing degradations," *Eng. Appl. Artif. Intell.*, vol. 42, pp. 67–81, Jun. 2015.

[21] Y. Li, Y. Yang, X. Wang, B. Liu, and X. Liang, "Early fault diagnosis of rolling bearings based on hierarchical symbol dynamic entropy and binary tree support vector machine," *J. Sound Vibrat.*, vol. 428, pp. 72–86, Aug. 2018.

[22] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, Jun. 2015.

[23] Z. Gao, C. Cecati, and S. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part II: Fault diagnosis with knowledge-based and hybrid/active approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3768–3774, Jun. 2015.

[24] M. Kang, J. Kim, J.-M. Kim, A. C. C. Tan, E. Y. Kim, and B.-K. Choi, "Reliable fault diagnosis for low-speed bearings using individually trained support vector machines with kernel discriminative feature analysis," *IEEE Trans. Power Electron.*, vol. 30, no. 5, pp. 2786–2797, May. 2015.
[25] R. Li, P. Sopon, and D. He, "Fault features extraction for bearing prognostics," *J. Intell. Manuf.*, vol. 23, no. 2, pp. 313–321, Apr. 2012.
[26] Y. Yang, Y. Liao, G. Meng, and J. Lee, "A hybrid feature selection scheme for unsupervised learning and its application in bearing fault diagnosis," *Expert Syst. Appl.*, vol. 38, no. 9, pp. 11311–11320, Sep. 2011.
[27] C. Yang and T. Wu, "Diagnostics of gear deterioration using EEMD approach and PCA process," *Measurement*, vol. 61, pp. 75–87, Feb. 2015.
[28] B. Zhang, L. Zhang, and J. Xu, "Degradation feature selection for remaining useful life prediction of rolling element bearings," *Qual. Rel. Engng. Int.*, vol. 32, no. 2, pp. 547–554, Mar. 2016.
[29] Z. Wei, Y. Wang, S. He, and J. Bao, "A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection," *Knowl.-Based Syst.*, vol. 116, pp. 1–12, Jan. 2017.
[30] B. Samanta and K. Al-Balushi, "Artificial neural network based fault diagnostics of rolling element bearings using time-domain features," *Mech. Syst. Signal Process.*, vol. 17, no. 2, pp. 317–328, Mar. 2003.
[31] V. Sugumaran, V. Muralidharan, and K. Ramachandran, "Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 930–942, Feb. 2007.
[32] L. Yuan, Y. He, J. Huang, and Y. Sun, "A new neural-network-based fault diagnosis approach for analog circuits by using kurtosis and entropy as a preprocessor," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 3, pp. 586–595, Mar. 2010, doi: 10.1109/tim.2009.2025068.
[33] M. S. Ballal, Z. J. Khan, H. M. Suryawanshi, and R. L. Sonolikar, "Adaptive neural fuzzy inference system for the detection of inter-turn insulation and bearing wear faults in induction motor," *IEEE Trans. Ind. Electron.*, vol. 54, no. 1, pp. 250–258, Feb. 2007.
[34] L. Zhen, H. Zhengjia, Z. Yanyang, and C. Xuefeng, "Bearing condition monitoring based on shock pulse method and improved redundant lifting scheme," *Math. Comput. Simul.*, vol. 79, no. 3, pp. 318–338, Dec. 2008, doi: 10.1016/j.matcom.2007.12.004.
[35] A. Bellini, A. Yazidi, F. Filippetti, C. Rossi, and G.-A. Capolino, "High frequency resolution techniques for rotor fault detection of induction machines," *IEEE Trans. Ind. Electron.*, vol. 55, no. 12, pp. 4200–4209, Dec. 2008.
[36] P. M. Baggenstoss and F. Kurth, "Comparing shift-autocorrelation with cepstrum for detection of burst pulses in impulsive noise," *J. Acoust. Soc. Amer.*, vol. 136, no. 4, pp. 1574–1582, Oct. 2014.
[37] Y. Guo and K. K. Tan, "Order-crossing removal in Gabor order tracking by independent component analysis," *J. Sound Vibrat.*, vol. 325, nos. 1–2, pp. 471–488, Aug. 2009.
[38] Y. Lei, Z. He, Y. Zi, and X. Chen, "New clustering algorithm-based fault diagnosis using compensation distance evaluation technique," *Mech. Syst. Signal Process.*, vol. 22, no. 2, pp. 419–435, Feb. 2008.
[39] J. Pei, S. Zhang, M. Qi, and G. Wan, "A new method for fault diagnosis of fluid end in drilling pump," *Acta Petrolei Sinica*, vol. 30, no. 4, pp. 617–620, 2009.
[40] R. Tiwari, V. K. Gupta, and P. Kankar, "Bearing fault diagnosis based on multi-scale permutation entropy and adaptive neuro fuzzy classifier," *J. Vibrat. Control*, vol. 21, no. 3, pp. 461–467, Feb. 2015.
[41] W. Sun, G. An Yang, Q. Chen, A. Palazoglu, and K. Feng, "Fault diagnosis of rolling bearing based on wavelet transform and envelope spectrum correlation," *J. Vibrat. Control*, vol. 19, no. 6, pp. 924–941, Apr. 2013.
[42] Y. Ying, J. Li, P. Chai, Y. Chen, and J. Pang, "Study on rolling bearing fault diagnosis based on multi-dimensional feature extraction," *J. Shanghai Univ. Electr. Power*, vol. 34, no. 5, pp. 413–421, 2018.
[43] J. Rosero, L. Romeral, J. Ortega, and E. Rosero, "Short-circuit detection by means of empirical mode decomposition and Wigner–Ville distribution for PMSM running under dynamic condition," *IEEE Trans. Ind. Electron.*, vol. 56, no. 11, pp. 4534–4547, Nov. 2009.
[44] R. Kumar and M. Singh, "Outer race defect width measurement in taper roller bearing using discrete wavelet transform of vibration signal," *Measurement*, vol. 46, no. 1, pp. 537–545, Jan. 2013.
[45] S. Mohanty, K. K. Gupta, and K. S. Raju, "Hurst based vibro-acoustic feature extraction of bearing using EMD and VMD," *Measurement*, vol. 117, pp. 200–220, Mar. 2018.
[46] J. Li and J. Guo, "A new feature extraction algorithm based on entropy cloud characteristics of communication signals," *Math. Problems Eng.*, vol. 2015, pp. 1–8, Jun. 2015.
[47] L. Huang, M. Wang, and L. Wu, "Research on change detection approach using PSO algorithm and multiple thresholds exponential entropy in remote sensing images," *Eng. Surveying Mapping*, vol. 27, no. 7, pp. 1–5, 2018.
[48] J. Zheng, H. Pan, S. Yang, and J. Cheng, "Generalized composite multiscale permutation entropy and Laplacian score based rolling bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 99, pp. 229–243, Jan. 2018.
[49] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 1995.
[50] Y. Li, W. Dai, X. Wu, and Y. Kan, "Surface quality evaluation based on roughness prediction model," in *Proc. Int. Conf. Inf. Technol. Electr. Eng. (ICITEE)*, 2018, doi: 10.1145/3148453.3306271.
[51] Z. Zhang, Y. Zhang, and Q. Liu, "Fault diagnosis on bearing by support vector machine and wavelet analysis," *Machinery Des. Manuf.*, vol. 313, no. 3, pp. 204–207, 2017.
[52] Y. Liu and S. Liu, "Application of MED and hierarchical fuzzy entropy to rolling bearing fault diagnosis," *Machinery Des. Manuf.*, no. 11, pp. 49–52 and 56, 2018.
[53] S. Wan, L. Dou, R. Liu, and X. Zhang, "Fault diagnosis for high voltage circuit breakers based on EWT and multi-scale entropy," *J. Vib., Meas. Diagnosis*, vol. 38, no. 4, pp. 672–678 and 867, 2018.
[54] Y. Li, W. Zhang, Q. Xiong, D. Luo, G. Mei, and T. Zhang, "A rolling bearing fault diagnosis strategy based on improved multiscale permutation entropy and least squares SVM," *J. Mech. Sci. Technol.*, vol. 31, no. 6, pp. 2711–2722, Jun. 2017, doi: 10.1007/s12206-017-0514-5.
[55] K. Zhu, L. Chen, and X. Hu, "Rolling element bearing fault diagnosis based on multi-scale global fuzzy entropy, multiple class feature selection and support vector machine," *Trans. Inst. Meas. Control*, vol. 41, no. 14, pp. 4013–4022, Oct. 2019, doi: 10.1177/0142331219844555.
[56] X. Xiao, Q. He, Z. Li, A. O. Antoce, and X. Zhang, "Improving traceability and transparency of table grapes cold chain logistics by integrating WSN and correlation analysis," *Food Control*, vol. 73, pp. 1556–1563, Mar. 2017.
[57] H. Zhao, D. Zhang, S. Huang, S. Mo, and H. Wei, "Analysis on the relation between cloud-to-ground lightning density and lightning trip rate in Hainan province based on pearson correlation coefficient," *High Voltage App.*, vol. 55, no. 8, pp. 186–192, 2019.
[58] G. Chen, J. Chen, Y. Zi, J. Pan, and W. Han, "An unsupervised feature extraction method for nonlinear deterioration process of complex equipment under multi dimensional no-label signals," *Sens. Actuators A, Phys.*, vol. 269, pp. 464–473, Jan. 2018.
[59] C. Wang and J. Cai, "Research on fault diagnosis of rolling bearing based on empirical mode decomposition and principal component analysis," *Acta Metrologica Sinica*, vol. 40, no. 6, pp. 1077–1082, 2019.
[60] X. Meng, C. Feng, and S. Gao, "Research on consistency of turbine blade temperature distribution based on principal component analysis," *J. Harbin Univ. Commerce (Natural Sci. Ed.)*, vol. 35, no. 4, pp. 451–457, 2019.
[61] Y. Ma, "Principal component analysis of quality indexes of different varieties of actinidia arguta," *Sci. Technol. Food Ind.*, vol. 40, no. 5, pp. 233–238, 2019.
[62] X. J. Zeng, X. L. Zhang, M. A. Hong-Jiang, and L. Li, "Traveling wave fault location method for power grids based on wavelet packet energy spectra," *High Voltage Eng.*, vol. 34, no. 11, pp. 2311–2316, 2008.
[63] L. Zhai, "Research on image classification based on weighted multi-feature fusion and SVM," M.S. thesis, School Comput., Central China Normal Univ., Wuhan, China, 2016.
[64] *Bearing Data Center*. Accessed: Oct. 18, 2019. [Online]. Available: https://csegroups.case.edu/bearingdatacenter/home
[65] B. Wang, Y. Lei, N. Li, and N. Li, "A hybrid prognostics approach for estimating remaining useful life of rolling element bearings," *IEEE Trans. Rel.*, to be published.

YAZHOU LI was born in Hebi, Henan, China, in 1996. He received the B.E. degree in energy and power engineering from the Beijing University of Civil Engineering and Architecture, in 2018. He is currently pursuing the M.S. degree with the School of Energy and Power Engineering, Beihang University.

His research interests include condition monitoring, fault diagnosis, reliability assessment, the evaluation and analysis based on big data.

WEI DAI was born in Datong, Shanxi, China. He received the B.S. degree with the School of Mechanical Engineering and Automation, Beihang University, and the Ph.D. degree in mechanical engineering from the Beihang University and University of Bath, in 2011.

He is currently an Associate Professor with Beihang University. His research interests include reliability manufacturing theory and technology, and the evaluation and analysis based on big data.

WEIFANG ZHANG received the B.S. and Ph.D. degrees from the School of materials science and Engineering, Harbin Institute of Technology, Harbin, China, in 1996 and 1999, respectively. He is currently a Professor with the School of Reliability and Systems Engineering, Beihang University. His main research interests include material and structural damage, structural safety and health monitoring, and life prediction.