IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 September 2023, accepted 8 October 2023, date of publication 12 October 2023, date of current version 19 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3324054

RESEARCH ARTICLE

A Novel Fault Diagnosis Method Based on NEEEMD-RUSLP Feature Selection and BTLSTSVM

RONGRONG LU[1], MIAO XU[1], CHENGJIANG ZHOU[1], ZHAODONG ZHANG[1], SHANYOU HE[1], QIHUA YANG[1], MIN MAO[2], AND JINGZONG YANG[3]

[1]School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China
[2]Faculty of Information Engineering, Quzhou College of Technology, Quzhou 324000, China
[3]School of Big Data, Baoshan University, Baoshan, Yunnan 678000, China

Corresponding author: Chengjiang Zhou (chengjiangzhou@foxmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 62363036, in part by the Ph.D. Research Startup Foundation of Yunnan Normal University under Grant 01000205020503131, in part by the Fundamental Research Program of Yunnan Province under Grant 202201AU070055, in part by the Project of Educational Commission of Yunnan Province of China under Grant 2022J0131, in part by the Project of Quzhou Science and Technology Plan under Grant 2021K31 and Grant 2020014, in part by the Yunnan Fundamental Research Projects under Grant 202301AT070256, and in part by the Baoshan Xingbao Young Talent Training Project under Grant 202303.

ABSTRACT The vibration signal of rolling bearings is a nonlinear and non-stationary signal, which is affected by the working condition change and background noise, and the reliability of traditional feature extraction methods and fault identification methods is low. In order to effectively extract feature vectors and improve the accuracy and reliability of fault identification, we propose a new fault diagnosis method based on noise eliminated ensemble empirical mode decomposition and robust unsupervised feature selection with local preservation (NEEEMD-RUSLP) and binary tree least squares twin support vector machine (BTLSTSVM). Firstly, NEEEMD is introduced to suppress background noise and decompose the vibration signal into a series of intrinsic mode functions (IMF), and the wavelet packet energy entropy, packet energy coefficient, and Gini coefficient of each IMF are extracted to construct time-frequency domain features. Then, 16 time-domain features and 13 frequency-domain features of the original signal are extracted and combined with the time-frequency domain features of each IMF to construct a high-dimensional feature space. In order to reduce the feature dimension and improve the diagnostic accuracy of the model, the RUSLP feature selection method is introduced to select effective low-dimensional features from the high-dimensional features. In addition, the binary tree (BT) strategy is introduced into the LSTSVM binary classifier to construct the BTLSTSVM multi-classifier, which aims to improve the recognition accuracy of low-dimensional features. In the bearing fault diagnosis of Case Western Reserve University, the fault diagnosis accuracy obtained by the proposed method is improved by 10.67%. In the bearing fault diagnosis of the University of Ottawa, the fault diagnosis accuracy obtained by the proposed method is improved by 10%. In the fault diagnosis of check valve in the actual industrial production environment, the fault diagnosis accuracy obtained by the proposed method is improved by 22%. The results show that the proposed method can not only effectively extract and select the low-dimensional fault characteristics of the bearing, but also achieve competitive fault diagnosis accuracy. Therefore, this method can provide a new method reference for the field of fault diagnosis, and has great theoretical significance and application value.

INDEX TERMS Empirical mode decomposition, unsupervised feature selection, mixed domain, fault diagnosis, least square twin support vector machine.

I. INTRODUCTION

Bearings are important components in electric motors, generators, gearboxes, and couplings and are widely used in

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

various fields. Bearings are prone to failure due to many factors, but the vibration signal has nonlinear and non-stationary characteristics due to variable operating conditions and strong noise, which poses a serious challenge to bearing fault diagnosis. Therefore, the research on bearing fault diagnosis methods based on signal processing and machine learning has important economic significance and practical value.

In terms of feature extraction, the time-domain, frequency-domain and time-frequency domain features will change with the transition of state, but the single feature information is incomplete and cannot effectively represent the state information of mechanical parts. Therefore, it is necessary to construct high-dimensional mixed features to mine the state information and inherent characteristics of the original vibration signal. The existing feature extraction methods include domain analysis, frequency domain analysis, time-frequency domain analysis and other methods. Tao et al. [1] decomposed and reconstructed the bearing vibration signal by wavelet packet decomposition (WPD) method, extracted the energy characteristics of each subband and constructed a two-dimensional time-frequency diagram to achieve fault diagnosis. Sun et al. [2] decomposed the equal part signal into multiple IMFs through empirical mode decomposition (EMD) method, and effectively diagnosed rolling bearing faults with the improved Chebyshev distance as the feature. Motahari-Nezhad et al. [3] first extracted 60 time-domain features of bearing vibration signals, then introduced an improved distance evaluation (IDE) method for feature dimensionality reduction, and finally classified healthy and faulty bearings by the k-nearest neighbor (KNN) algorithm. Prakash Kumar et al. [4] used fault detection time-domain vibration analysis technology to explore the time-domain characteristics and information contained in vibration signals. Wu et al. [5] denoised the original signal through ensemble empirical mode decomposition (EEMD), and extracted key bearing features through a feature extraction module composed of convolutional layer, attention mechanism module, and pooling layer. Li et al. [6] took the kurtosis value of the envelope signal as the fitness function and extracted the composite fault characteristics of the bearing by parameter-optimized variational mode decomposition (VMD). Wang et al. [7] first used the grasshopper optimization algorithm (GOA) to optimize the parameters of the VMD, and then decomposed the bearing signal through the optimized VMD. Tang et al. [8] extracted time-domain features by complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), while deep frequency-domain features were extracted by fast Fourier transform (FFT), and the optimal features were input into a classifier to identify bearing faults. Under the loosely distributed cross-domain dataset, Chen et al. [9] aggregated a single local feature through the vector of locally aggregated descriptors based network (Net VLAD), and the intra-class compactness and inter-class separability of cross-domain features were improved. Jiang et al. [10] combined the time-frequency spectral amplitude modulation (TFSAM) method with the short-time Fourier transform, and more accurate and detailed amplitude information in the time-frequency domain was extracted, and the accuracy of fault identification was improved. Vashishtha and Kumar [11] optimized the parameters of time-varying filter based empirical mode decomposition (TVF-EMD) through kernel estimate for mutual information (KEMI) and amended grey wolf optimization (AGWO), and the fault diagnosis efficiency was improved. However, the above extraction method has the following problems: (1) WPD generates frequency confusion and false components at frequency band segmentation, and does not have adaptability, and the IMFs obtained by decomposition methods such as EMD and EEMD has endpoint effect and modal aliasing, and the decomposition effect of VMD depends on parameter settings. (2) The feature extraction process is affected by background noise, and most feature extraction methods do not specifically consider noise suppression, which leads to background noise components in the fault features. (3) Some artificial features are too singular, while others have redundancy, and feature selection and filtering have not been fully considered.

A high-dimensional dataset formed by multi domain features can reveal the inherent characteristics of the original signal more widely, but it also brings some redundant and negative feature information. In addition, the higher the feature dimension, the more time-consuming the classification model will be. Given this, it is necessary to perform appropriate feature selection after multi domain feature extraction. According to the availability of class label information, feature selection methods can be roughly divided into three categories: supervised, unsupervised and semi supervised. Lao et al. [12] extracted the time-domain features and multi-scale permutation entropy features of the turnout switch machine, and constructed an adaptive feature selection model to extract the key sensitive features. He et al. [13] optimized VMD through an improved sparrow search algorithm to obtain effective IMF, and extracted the energy entropy of IMF to construct a fault feature matrix, an Inverted Residual Convolutional Neural Network (IRCNN) can adaptively select sensitive features and reduce feature dimensionality. Mohd Saufi and Hassan [14] solved the problem of artificial selection of super parameters in deep learning through Laplacian Score (LS) and Long Short-Term Memory (LSTM) methods. Zheng et al. [15] effectively extracted complex nonlinear dynamic information by combining Multi-Cluster Feature Selection (MCFS) with The Gravitational Search Algorithm Optimized Support Vector. Hashemi et al. [16] had modelled the multi-label feature selection problem into a bi-objective optimization problem regarding the relevancy and redundancy degree of the features, to solve the problem of high feature dimensionality and noise interference in multi label datasets. Ma et al. [17] took the expectation, entropy, and hyper entropy of the cloud models representing uncertainty in features as spatial vectors. By fusing redundant

elements of space vectors through Interpretable and Unsupervised Dimension Reduction Method, low dimensional feature components that were conducive to fault classification are obtained. Wahid et al. [18] determined the Huber-Type Weight Function by Unsupervised Discriminative Feature Selection (UDFS) method and Mahalanobis distance, and reduced the weight of cluster observations with large distance. Guo et al. [19] used Unsupervised Feature Selection With Adaptive Structure (FSASL) method to eliminate the impact of noise, redundancy, and missing values on the original signal, improving the effectiveness of feature selection. Zeng et al. [20] eliminated the influence of noise or outlier in real fault data through Robust Unsupervised Feature Selection (RUFS) algorithm. Zhu et al. [21] combined the construction of similarity matrices with the feature selection process through the Graph Learning Unsupervised Feature Selection (GLUFS) algorithm, improving the effectiveness and superiority of the feature selection process. However, the above feature selection methods have the following problems: (1) the correlation between features is ignored, resulting in redundant features being selected and reducing the performance of the algorithm; (2) The crucial discriminative information in feature selection is ignored; (3) The effects of outlier and noise on performance are ignored.

In addition to the above methods, fault identification models are also key to ensuring the reliability of fault diagnosis. By using pattern recognition methods to accurately identify the type, location, or degree of faults, the recognition accuracy of the model can be improved. At present, fault identification methods include neural networks, deep learning, support vector machines, etc. Fotso et al. [22] achieved the construction of a power loss model for wind turbine bearings using Back-Propagation Neural Network (BPNN) method, and predicted the expected values through the constructed model. Ghorvi et al. [23] combined adversarial domain adaptation and local maximum mean difference (LMMD) to reduce structural differences between subdomains and global domains, and accurately identified bearing fault types through graph convolutional neural networks (GCNN). Wang et al. [24] constructed suitable Spiking Neurons through Improved Spiking Neural Network (ISNN) to compensate for information loss during forward propagation and simplify the back propagation process. Xie et al. [25] used Locally Generalized Preserving Projection (LGPP) to reduce the dimensionality of high-dimensional features, and then constructed Flexible Grey Wolf Optimizer (FGWO) to optimize the parameters of the Extreme Learning Machine (ELM), significantly improving the fault recognition rate. Gong et al. [26] extracted fault features through Variational Mode Decomposition (VMD) and Refined Composite Multiscale Bubble Entropy (RCMBE), and then used the Gorilla Troops Optimizer Optimized Kernel Extreme Learning Machine (GTO-KELM) to identify bearing faults. However, the short duration of mechanical equipment failures, small number of fault samples, and non-linear distribution make it difficult to meet the large sample requirements of neural networks and deep learning, as well as the data requirements of statistical learning methods. Support vector machine (SVM) can solve nonlinear high-dimensional space problems by a small number of samples, and has outstanding advantages for nonlinear, small-sample, high-dimensional pattern recognition and other problems, and has good learning ability and adaptability. Therefore, many researchers use SVM to identify the type and degree of mechanical faults. Yaman et al. [27] extracted features from sound signals using Mel-frequency Cepstral Coefficients (MFCC) and used SVM to classify selected features for faults. Gao et al. [28] combined isometric mapping with the Isometric Mapping And Nonhomogeneous Cuckoo Search-Least Squares Support Vector Machine (NoCuSa-LSSVM) method to monitor the operating status of rolling bearings. He et al. [29] first extracted the Composite Multiscale Weighted Permutation Entropy (CMWPE) features of bearings, and then constructed a Least Squares Support Vector Machine (LSSVM) model for bearing fault diagnosis, significantly improving recognition accuracy. In fact, only one hyperplane is constructed in the two classifiers of SVM, which is computationally complex, inefficient and sensitive to parameters, and is not suitable for complex nonlinear classification problems and large-scale data samples. Wei et al. [30] used Moth-Flame Optimization (MFO) algorithm to optimize the hyperparameters (Sigma & γ) of the LS-SVM classifier, achieving higher fault diagnosis recognition rate and algorithm robustness. Twin Support Vector Machines (TSVM) used two hyperplanes for pattern recognition, which was four times more efficient and more accurate than SVM. Dhiman et al. [31] extracted the SCADA parameters of the wind turbine gearbox and combined the Adaptive Threshold with TSVM to achieve gear fault diagnosis. Bai et al. [32] used TSVM to regenerate unlabeled samples, effectively solving the problem of difficulty in extracting fault features and a small number of fault samples. Yuan et al. [33] proposed a chemical reaction optimized TSVM model to address the issue of SVM being affected by imbalanced samples. The accuracy in transformer fault diagnosis is superior to SVM, k-nearest neighbor method, and decision tree. However, TSVM needs to solve the Quadratic Programming Problem (QPP) with inequality constraints, which affects the classification accuracy and efficiency. The Least Square Twin Support Vector Machine (LSTSVM) proposed by Arun Kumar and Gopal [34] replaced the QPP with inequality constraints in TSVM through a system of linear equations, with better computational efficiency and classification accuracy than TSVM. Ganae and Tanveer [35] first generated weights through LSTSVM, then applies the product of input features and weights to nonlinear functions to obtain enhanced features, and finally classifies based on these features through LSTSVM. Ali et al. [36] generated the optimal hyperplane by solving a pair of linear equations with LSTSVM, which improved the efficiency of the TWSVM model. Zhou et al. [37] improved the diagnostic method

by combining Multi-Scale Weighted Permutation Entropy (MWPE) and LSTSVM to improve the signal-to-noise ratio of signals and the diagnostic accuracy of bearings and one-way valves. Chen et al. [38] used the Conjugate Gradient (CG) algorithm to solve the linear equations appearing in the Lap LSTSVM model, accelerating the training process of the model. Yuan and Yang [39] reduced the impact of noise and outlier on LSTSVM model by Capped L2, P-Norm Distance Metric. Yu et al. [40] first constructed the Hessian Scatter Regularization (HSR) term, then constructed the least squares version of HSR-TSVM (HSR-LSTSVM) through HSR, and finally solved the constructed model using the Conjugate Gradient method. However, the above fault classification methods have the following problems: (1) outlier detection and membership construction of samples affect the performance of anti-outlier; (2) The objective function of the diagnostic model only considers the training error and ignores the generalization error, so the generalization performance of the model needs to be improved; (3) LSTSVM has very few applications in fault diagnosis, and as a binary classifier, it is often used for complex binary classification problems. How to apply the LSTSVM model to multi classification recognition is also a problem worth exploring.

In addition, those methods based on deep learning and methods combining feature extraction with machine learning are increasingly being used for vibration signal analysis and fault diagnosis. Othman et al. [41] proposed a custom residual deep neural network to detect passive seismic events, which not only suppresses background noise through the IIR Wiener filter, but also does not rely on real data to train a diagnostic model. Iqbal et al. [42] proposed an intelligent deep convolutional neural network method, which can train models without real data, suppress various irrelevant noises in earthquake signals, and provide a good foundation for earthquake prediction. Lao et al. [12] extracted the time-domain features and multi-scale permutation entropy features of the turnout switch machine and constructed an adaptive feature selection model (AFS) to reduce feature dimension, a light gradient boosting machine based on the improved focal loss (IFL) function improved the diagnostic model's ability to distinguish similar feature samples. He et al. [13] optimized the parameters of VMD through improved sparrow search algorithm and dispersion entropy fitness, extracted the energy entropy features of IMF, and constructed an inverted residual convolutional neural network (IRCNN) to obtain a diagnostic accuracy of 97.5%, and IRCNN itself has feature selection ability. Wei et al. [43] proposed a third order tensor model, a density-based affinity propagation tensor clustering algorithm is presented to identify different failures with unlabeled, which unsupervised characteristics show it has potential for applications on rail transit trains. Jin et al. [44] optimized the parameters of VMD through a gray wolf algorithm based on hybrid strategy, extracted multi-scale dispersion entropy and diagnosed train axle box bearing faults through the deep belief network (DBN). Yu et al. [45] constructed the multi group Resnet (MGRN) structure to extract multi-scale and multi resolution wavelet packet time-frequency feature maps from vibration time-frequency features, and then the multi kernel maximum mean diversity (MK-MMD) is employed to evaluate the distribution difference of depth features between the source and target domain. This strategy has good noise reduction performance and variable condition transfer diagnosis performance. Wen et al. [46] constructed a clustering graph convolutional network with multiple adaptive learning (c-GCN MALs) based on auto encoder (AE), AE and graph convolutional networks can extract the structural correlation of the dataset, and the setting of loss functions enhances the transfer and clustering capabilities of data and domain adversaries. It can be seen from the above that deep learning and deep transfer learning not only avoid the artificial experience in feature extraction, but also do not need to consider dimensionality reduction issues, and have gradually become popular research directions. However, for small sample problems, sample imbalance problems, and outlier problems, fault diagnosis methods based on deep learning and deep transfer learning still have some shortcomings. In addition, deep learning methods generally require a large number of training samples, and deep transfer learning performs poorly when the differences in the data domain are too large. Therefore, the fault diagnosis method based on dimensionality reduction feature extraction and recognition model is adopted in this paper.

In order to solve the above problems of feature extraction, feature selection, and recognition models, a new fault diagnosis method based on NEEEMD-RUSLP feature selection and BTLSTSVM is proposed. Firstly, a denoising integrated empirical mode decomposition NEEEMD method is constructed to suppress noise in the signal, and the original signal is decomposed into a series of IMFs; The wavelet packet energy entropy extracted from each IMF, small packet energy coefficients, and Gini coefficients are constructed as time-frequency domain features of the signal; The time-domain, frequency-domain, and time-frequency features extracted from signal samples are constructed as a mixed fault feature vector set. Then, a robust unsupervised feature selection RUSLP method with local preservation is constructed in this paper, where RUSLP obtains information by decomposing the clustering labels of the predicted data into matrices. The constraints of the orthogonal two decomposition matrices help to achieve more accurate class labels for selecting features with high discrimination ability. Finally, we use linear equations to replace the QPP with inequality constraints, and the Binary Tree (BT) strategy is introduced to build a multi classification LSTSVM model (BT-LSTSVM). BT-LSTSVM does not need any optimizer, and its computational efficiency is better than TSVM and SVM, and its classification performance is stronger. Therefore, the innovation points and main contributions of this article are as follows:

1) The time-frequency domain features based on NEEEMD are constructed in this paper. Among

them, NEEEMD is superior in signal separation and noise robustness, and can obtain richer fault feature information.

2) The feature selection method based on RUSLP and the low dimensional feature space are constructed in this paper. Among them, RUSLP selects features with high discriminative ability through more precise class labels in feature selection.
3) A multi classification model based on BT-LSTSVM is constructed in this paper. Among them, The BT-LSTSVM multi classification model achieves higher efficiency and better recognition accuracy in fault recognition.

The chapters of the text are arranged as follows:

Section II discusses the theories and methods involved in this paper. Section II-A discusses the high-dimensional feature extraction process based on the NEEEMD method, and proposes the problems and optimizations solved by the NEEEMD method. Section II-B describes the theoretical method of robust unsupervised feature selection with local preservation. Section II-C describes the theoretical method of binary tree least squares twin support vector machines. Section III discusses the implementation process of a new fault diagnosis method based on NEEEMD-RUSLP feature selection and BTLSTSVM proposed in this paper. Section IV conducts experimental verification. Section IV-A uses simulation signals to verify the effectiveness of the NEEEMD decomposition method. Section IV-B verifies the effectiveness of RUSLP dimensionality reduction method and BTLSTSVM classifier with data collected from Case Western Reserve University Experimental Platform. Section IV-C verifies the effectiveness of the RUSLP dimensionality reduction method and the BTLSTSVM classifier by using the data collected by the experimental platform of the University of Ottawa. Section IV-D verifies the validity of the RUSLP dimensionality reduction method and the BTLSTSVM classifier by using the data set generated by a check valve in a real industrial production environment. Section V provides a summary of the writing.

II. THEORY AND METHODS

A. HIGH DIMENSIONAL FEATURE EXTRACTION BASED ON NEEEMD

1) NOISE ELIMINATED ENSEMBLE EMPIRICAL MODE DECOMPOSITION

Noise Elimination Ensemble Empirical Mode Decomposition (NEEEMD) adopts a different method from CEEMD to eliminate white noise in the final stage. Instead of adding negative white noise at the primary level, it subtracts the intermediate frequency of the same white noise from the final intermediate frequency. In order to evaluate the effectiveness of the proposed feature extraction methods, relevant feature extraction methods such as EEMD, CEEMD, and Continuous Wavelet Transform (CWT) are compared. By using the original vibration signal decomposition, all output results show that NEEEMD method not only has the highest accuracy, sensitivity and robustness, but also can effectively eliminate the existence of white noise and save a lot of calculation time. The key steps of NEEEMD are as follows:

Step 1: Add the set white noise $w_i(t)$ (its length is the same as the original signal, the average value is 0, and the standard deviation is 1) and the original signal $X(t)$ to obtain $X_i(t)$.

Step 2: Use EMD to decompose $X_i(t)$ and obtain the IMF set average value $c_j(t)$.

$$c_j(t) = \frac{1}{M}\sum_{i=1}^{M} c_{ij}(t) \tag{1}$$

Step 3: Take the input set white noise $w_i(t)$, and apply EMD to each of them.

$$W_i = \sum_{j=1}^{N} wc_{ij}(t) + wr_i(t). \tag{2}$$

Among them, $j = 1, 2, \ldots, N$, N is the number of IMF, and $wc_{ij}(t)$ is the IMF $(c_{i1}, c_{i2}, \ldots, c_{iN})$ of noise. $wr_{ij}(t)$ represents the residual of the i-th trace.

Step 4: Calculate the set average of IMF for noise.

$$wc_j(t) = \frac{1}{M}\sum_{i=1}^{M} wc_{ij}(t). \tag{3}$$

Subtract the noisy IMF from the IMF obtained by EEMD to reduce white noise.

$$IMF_j = c_j(t) - wc_j(t) \tag{4}$$

Step 5: The original signal can be obtained and make following things.

$$X(t) = \sum_{i=1}^{M} IMF_{ij}(t) + rM_j(t) - wrM_j(t). \tag{5}$$

where $wrM(t)$ is the residual of white noise. More detailed information can be found in reference [47]. The parameter selection method for white noise can be referenced in references [48] and [49], the number of ensembles in this research is 100 and the standard deviation of the white noise is 1 and the mean of it is 0.

2) HIGH DIMENSIONAL FEATURES BASED ON NEEEMD

a: TIME DOMAIN FEATURE EXTRACTION

This article uses statistical methods to extract 16 time-domain features, whose expressions are listed in Table 1. As shown in Table 1, the 10 features $TF_1 \sim TF_{10}$ are commonly referred to as dimensional statistical parameters, namely mean, standard deviation, square root amplitude, absolute mean, skewness, kurtosis, variance, maximum, minimum and peak to peak values. The six characteristics of $TF_{11} \sim TF_{16}$ are called waveform index, peak index, pulse index, edge index, skewness index and kurtosis index, also known as dimensionless statistical parameters. Among the 16 expressions given in Table 1, $x(n)$, $n = 1, 2, \ldots, N$ is a given discrete time series, and N is the number of data points for signal $x(n)$.

TABLE 1. Expression of time domain characteristic parameters [50].

Feature expression			
$TF_1 = \frac{1}{N}\sum_{n=1}^{N} x(n)$	$TF_5 = \frac{1}{N}\sum_{n=1}^{N}(x(n))^3$	$TF_9 = \min\lvert x(n)\rvert$	$TF_{13} = \frac{TF_8}{TF_4}$
$TF_2 = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}[x(n)-TF_1]^2}$	$TF_6 = \frac{1}{N}\sum_{n=1}^{N}(x(n))^4$	$TF_{10} = TF_8 - TF_9$	$TF_{14} = \frac{TF_8}{TF_3}$
$TF_3 = \left(\frac{1}{N}\sum_{n=1}^{N}\sqrt{\lvert x(n)\rvert}\right)^2$	$TF_7 = \frac{1}{N}\sum_{n=1}^{N}(x(n))^2$	$TF_{11} = \frac{TF_2}{TF_4}$	$TF_{15} = \frac{TF_5}{\left(\sqrt{TF_7}\right)^3}$
$TF_4 = \frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert$	$TF_8 = \max\lvert x(n)\rvert$	$TF_{12} = \frac{TF_8}{TF_2}$	$TF_{16} = \frac{TF_6}{(TF_7)^2}$

b: FREQUENCY DOMAIN FEATURE EXTRACTION

For a given time series $x(n)$, the text can extract 13 frequency domain features $FF_1 \sim FF_{13}$ using FFT. Table 2 describes the expressions of these features. As shown in Table 2, eigenvalues FF_1 reflect the vibration energy in the frequency domain, eigenvalues $FF_1 \sim FF_4$, FF_6, and $FF_{10} \sim FF_{13}$ represent the concentration and dispersion of the spectrum, and eigenvalues FF_5 and $FF_7 \sim FF_9$ represent the positional changes of the main frequency band. Among the thirteen expressions given in Table 2, $y(k)$ is the FFT spectrum of the given time series $x(n)$, $k = 1, 2, \ldots, K$, K is the number of spectral lines in the FFT spectrum, and f_k is the frequency value corresponding to the $k-th$ spectral line.

c: WAVELET PACKET ENERGY ENTROPY FEATURE EXTRACTION

Wavelet transform has the ability to characterize local features of signals in the time-frequency domain, but it only decomposes the low-frequency part of the signal in the next step and does not process high-frequency signals. Wavelet packet transform decomposes low-frequency and high-frequency signals, providing more complete information. The energy entropy H_{jk} of the wavelet packet is calculated using equation (6), where the letters N and J represent the signal length and the number of layers of wavelet packet decomposition, respectively, to obtain the decomposition sequence $X_{ij}(k = 0 \sim 2j-1)$. Component S_{jk} is obtained through a single reconstruction of the decomposition coefficients, and E_{jk} is set to the power of the reconstructed signal, $E_{jk} = \left|S_{jk}(i)\right|^2$, and $\varepsilon_{jk} = E_{jk}/E$, then $\sum_k \varepsilon_{jk} = 1$.

$$H_{jk} = -\sum_{i=1}^{N} \varepsilon_{jk}(i) \log \varepsilon_{jk}(i) \tag{6}$$

Obtain the energy entropy of the wavelet packet from equation (6) to form a feature vector $[H_1, H_2, \cdots, H_n]$.

d: FEATURE EXTRACTION OF WAVELET ENERGY COEFFICIENTS

The wavelet energy coefficient represents the energy distribution of signals within each frequency range, and different distributions result in different characteristics of acoustic emission sources and different damage conditions. Therefore, by comparing the changes in wavelet coefficients of each layer, the time-frequency characteristics of bearing fault expansion signals at different stages are obtained, and the corresponding relationship between wavelet energy coefficients and bearing fault expansion signal process is established. Wavelet energy coefficients can effectively represent the characteristics of bearing fault process. The extraction of wavelet energy coefficients first requires wavelet decomposition of the signal. Assuming that the original signal undergoes wavelet decomposition with K layers and K+1 frequency range components, the total energy of the signal can be represented by the energy of the wavelet coefficients of each layer, which is the wavelet energy coefficient:

$$E_f = Ea_k + \sum_{j=1}^{k} Ed_j (j = 1, 2, \cdots, k, k \in Z) \tag{7}$$

In the equation: E_f represents the total signal energy; Ea_k is the energy of the approximate wavelet coefficients for k-scale decomposition; Ed_j is the energy of the j-level detail wavelet coefficients.

e: GINI COEFFICIENT FEATURE EXTRACTION

When a rolling bearing malfunctions, the energy of its vibration signal is usually concentrated at certain positions in the

TABLE 2. Expression of frequency domain characteristic parameters [50].

Feature expression		
$FF_1 = \frac{\sum_{k=1}^{K} y(k)}{K}$	$FF_6 = \sqrt{\frac{\sum_{k=1}^{K}\left[(f_k - FF_5)^2 y(k)\right]}{K}}$	$FF_{11} = \frac{\sum_{k=1}^{K}\left[(f_k - FF_5)^3 y(k)\right]}{K(FF_6)^3}$
$FF_2 = \frac{\sum_{k=1}^{K}\left[y(k) - FF_1\right]^2}{K-1}$	$FF_7 = \sqrt{\frac{\sum_{k=1}^{K}(f_k^2 y(k))}{\sum_{k=1}^{K} y(k)}}$	$FF_{12} = \frac{\sum_{k=1}^{K}\left[(f_k - FF_5)^4 y(k)\right]}{K(FF_6)^4}$
$FF_3 = \frac{\sum_{k=1}^{K}\left[y(k) - FF_1\right]^3}{K\left(\sqrt{FF^2}\right)^3}$	$FF_8 = \sqrt{\frac{\sum_{k=1}^{K}(f_k^4 y(k))}{\sum_{k=1}^{K}(f_k^2 y(k))}}$	$FF_{13} = \frac{\sum_{k=1}^{K}\left[\sqrt{\lvert f_k - FF_5\rvert} y(k)\right]}{K\sqrt{FF_6}}$
$FF_4 = \frac{\sum_{k=1}^{K}\left[y(k) - FF_1\right]^4}{K\left(FF_2\right)^2}$	$FF_9 = \frac{\sum_{k=1}^{K}(f_k^2 y(k))}{\sqrt{\left[\sum_{k=1}^{K}(f_k^4 y(k))\right]\left[\sum_{k=1}^{K} y(k)\right]}}$	—
$FF_5 = \frac{\sum_{k=1}^{K}(f_k y(k))}{\sum_{k-1}^{K} y(k)}$	$FF_{10} = \frac{FF_6}{FF_5}$	—

signal sequence, and the Gini coefficient has good resistance to random strong pulse signals in the signal. Therefore, the Gini coefficient, which is not sensitive to random pulse signals and can measure fault characteristics, is used to characterize the development trend of rolling bearing faults and evaluate the health status. Specifically, the definition of Gini coefficient is

$$G = 1 - 2\sum_{j=1}^{L} \frac{x_j}{\|x\|}\left(\frac{L - j + 0.5}{L}\right) \tag{8}$$

Among them, $x = (x_1, x_2, \cdots, x_L)$.

In the formula: G is the value of Gini coefficient; x is the vector of the vibration acceleration signal x_j sorted in ascending order of amplitude; L is the total length of vector x; $\|x\|$ is the sum of the absolute values of all amplitudes of x.

f: STEPS FOR CONSTRUCTING HIGH-DIMENSIONAL FEATURES

Extracting fault feature information from multiple fields such as time domain, frequency domain, energy coefficient, energy entropy, and Gini coefficient can comprehensively explore the state information and inherent characteristics of the original vibration signal. Here are the steps for extracting high-dimensional features.

Step 1: Set the sampling frequency $f_s = 12000$, sample length $N = 2000$, and sample division into $[smp_1, smp_2, \ldots, smp_{60}]$, totaling 60 samples.

Step 2: Extract 16 time-domain features for each sample, namely equation (9).

$$[TF_1, TF_2, \ldots, TF_{16}] \tag{9}$$

Extract 13 frequency domain features, namely equation (10).

$$[FF_1, FF_2, \ldots, FF_{13}] \tag{10}$$

Step 3: By using the NEEEMD decomposition method, where each smp_i sample is decomposed into $[IMF_1, IMF_2, \ldots, IMF_8]$, the wavelet packet energy entropy features, small packet energy coefficient features, and Gini coefficient features of each IMF_i are extracted, namely equation (11).

$$[WPEE_i, SPEC_i, GINI_i]. \tag{11}$$

Construct the multi domain features of 8 IMF as the time-frequency domain features of each a sample, namely equation (12).

$$MDF_i = \begin{bmatrix} WPEE_i^1 & SPEC_i^1 & GINI_i^1 \\ \vdots & \vdots & \vdots \\ WPEE_i^8 & SPEC_i^8 & GINI_i^8 \end{bmatrix} \tag{12}$$

Step 4: Finally, a high-dimensional feature set GTC_{data} is constructed based on the time-domain, frequency-domain, and time-frequency features of each sample smp_i, where each feature is considered as its element. Among them, data MDF_i represents the mixed domain features of sample smp_i. The characteristic dimension of matrix GTC_{data} is $60 \times (16+13+3 \times 8)$. $TF_{i\times 16}$ is the 16-dimensional time-domain

feature vector of sample i, $FF_{i\times 13}$ is the 13-dimensional frequency-domain feature vector of sample i, $WPEE_{i\times 8}$ is the 8-dimensional wavelet packet energy entropy feature of sample i, $SPEC_{i\times 8}$ is the 8-dimensional small packet energy coefficient feature of sample i, and $GINI_{i\times 8}$ is the 8-dimensional Gini coefficient feature of sample i. Equations (13) and (14), as shown at the bottom of the page.

B. ROBUST UNSUPERVISED FEATURE SELECTION WITH LOCAL PRESERVATION

In this paper, we propose a Robust Unsupervised Feature Selection with Local Preservation (RUSLP) method that utilizes matrix decomposition while preserving local feature information. Inspired by matrix decomposition technology, the objective matrix is decomposed into two different matrices, namely the parameter matrix and the basis matrix. Applying orthogonal constraints on these two matrices to learn discriminative clustering labels for the projection data of the target matrix enhances the feature importance information in the projection matrix to select more relevant features. The key steps of RUSLP are as follows:

Requirements: data matrix $X \in \mathbb{R}^{n\times d}$, Laplacian matrix $L \in \mathbb{R}^{n\times n}$, parameters $\alpha, \lambda, \gamma, \mu$.

Step 1: Initialize: $t = 0$, $\mu = 10^{-6}$, $\in= 10^{-2}$ and W_t, G_t, F_t, H_t.

Step 2: Repeat steps 3, 4, 5, 6, 7, and 8.

Step 3: Update G_{t+1} and H_{t+1} through algorithm 1.

*Step 4:*When equation (15) satisfied,

$$Q_{ii} = \frac{1}{2\sqrt{w_i^T w_i + \in}}. \tag{15}$$

according to equation (16),

$$W = (2\alpha X^T L X + \mu X^T X + 2\lambda Q)^{-1} X^T (\mu E + \mu G F^T - Y). \tag{16}$$

updated equation (17).

$$W_{t+1} = (2\alpha X^T L X + \mu X^T X + 2\lambda Q)^{-1} X^T \times (\mu E_t + \mu G_{t+1} F_t^T - Y_t). \tag{17}$$

Step 5: When equation (18) satisfied,

$$G_t^T (X W_{t+1} - E_t + Y_t/\mu) = U_F I_m V_F^T \tag{18}$$

according to equation (19),

$$F = V_F I_{m,c} U_F \tag{19}$$

updated equation (20).

$$F_{t+1} = V_F I_{m,c} U_F^T. \tag{20}$$

Step 6: When equation (21) (22) satisfied,

$$J = XW - GF^T + Y/\mu. \tag{21}$$

$$M_{ii} = \frac{1}{2\sqrt{e_i^T e_i + \in}} \tag{22}$$

according to equation (23),

$$E = \mu(2M + \mu I)^{-1} J \tag{23}$$

updated equation (24).

$$E_{t+1} = \mu(2M + \mu I)^{-1} J \tag{24}$$

Step 7: By equation (25),

$$Y_{t+1} = Y_t + (E_{t+1} + X W_{t+1} + G_{t+1} F_{t+1}^T) \tag{25}$$

updated multiplier.

Step 8: $t = t + 1$.

*Step 9:*Check if the convergence condition meets equation (26) or equation (27).

$$\|\Delta\varepsilon(W_t, G_t, F_t, E_t, Y_t)\| \leq \in \tag{26}$$

$$t \geq T \tag{27}$$

If satisfied, ensure that the feature $\|W_t^i\|$ is selected based on the maximum value. When $i = 1, \ldots, d$, it is sorted in descending order and ends. More detailed information on RUSLP and parameter selection methods based on grid search strategies can be found in reference [51].

It is worth noting that this new feature extraction strategy has excellent noise suppression performance and periodic impact extraction performance. There are three specific reasons for this. Firstly, NEEEMD can suppress some white noise in the vibration signal by subtracting the intermediate frequency of the same white noise from the final intermediate frequency during the decomposition process. Second, during the feature extraction of wavelet packet energy entropy and packet energy coefficient, wavelet packet decomposition can suppress some noise, and the Gini coefficient

$$GTC_{data} = \begin{bmatrix} MDF_1 \\ MDF_2 \\ \vdots \\ MDF_i \end{bmatrix}. \tag{13}$$

$$GTC_{data} = \begin{bmatrix} TF_{1\times 16} & FF_{1\times 13} & WPEE_{1\times 8} & SPEC_{1\times 8} & GINI_{1\times 8} \\ TF_{2\times 16} & FF_{2\times 13} & WPEE_{2\times 8} & SPEC_{2\times 8} & GINI_{2\times 8} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ TF_{60\times 16} & FF_{60\times 13} & WPEE_{60\times 8} & SPEC_{60\times 8} & GINI_{60\times 8} \end{bmatrix}_{60\times(16+13+8+8+8)} \tag{14}$$

can extract the periodic impact characteristics contained in IMFs. Thirdly, the RUSLP feature selection strategy enhances the importance information in the projection matrix to select more relevant features. Therefore, the proposed dimensionality reduction feature extraction strategy can suppress noise in vibration signals and extract effective low dimensional fault features, and the proposed method also has great potential for application in fault feature extraction under white noise, pink noise, and salt and pepper noise backgrounds.

C. BINARY TREE LEAST SQUARES TWIN SUPPORT VECTOR MACHINE

When TSVM solves dual problems, $(G^TG)^{-1}$ and $(H^TH)^{-1}$ must exist or G^TG and H^TH must be non-singular, but not both of these conditions can be met. Arun Kumar M used a quadratic loss function to replace the hinge loss, and used a convex linear equation to replace the convex Quadratic Programming Problem (QPP) of TSVM, and proposed LSTSVM [143]. LSTSVM did not require any optimizer, has better computational efficiency than TSVM and SVM, and has stronger classification performance.

1) NONLINEAR LSTSVM

The purpose of LSTSVM is to find two nonparallel kernel hyperplanes and make one hyperplane as close to one type of sample as possible and away from the other type of sample. The target core hyperplane is as follows:

$$K\left(x^T, C^t\right) w_1 + b = 0 \text{ and } K\left(x^T, C^T\right) w_2 + b_2 = 0 \tag{28}$$

where $A \in R^{m_1 \times n}$ is m_1 positive samples, $B \in R^{m_2 \times n}$ is m_2 negative samples, $K\left(x^T, C^T\right)$ is the kernel function, $C = [A; B]$, $w_i \in R^n$ are the normal vector of the hyperplane, and b_i is the offset. The optimization problem for nonlinear LSTSVM is :

$$\begin{aligned} \min_{w_1,b_1,\xi_2} \quad & \frac{1}{2}\left\|K\left(A, C^T\right) w_1 + e_1 b_1\right\|^2 + \frac{c_1}{2}\xi_2^T \xi_2 \\ s.t. \quad & -\left(K(B, C^T)w_1 + e_2 b_1\right) + \xi_2 = e_2 \end{aligned} \tag{29}$$

$$\begin{aligned} \min_{w_2,b_2,\xi_1} \quad & \frac{1}{2}\left\|K\left(B, C^T\right) w_2 + e_2 b_2\right\|^2 + \frac{c_2}{2}\xi_1^T \xi_1 \\ s.t. \quad & -\left(K(A, C^T)w_2 + e_1 b_2\right) + \xi_1 = e_1 \end{aligned} \tag{30}$$

Among them, c_1, c_2 are penalty parameters, ξ_1 and ξ_2 are relaxation variables, e_1 and e_2 are vectors with each element being 1. By introducing equality constraints into the objective function, the following optimization problems without constrained optimization can be obtained:

$$\begin{aligned} \min_{w_1,b_1} \frac{1}{2} & \left\|K(A, C^T)w_1 + e_1 b_1\right\|^2 \\ & + \frac{c_1}{2}(e_2 + K(B, C^T)w_1 + e_2 b_1)^T \\ & \times (e_2 + K(B, C^T)w_1 + e_2 b_1) \end{aligned} \tag{31}$$

$$\begin{aligned} \min_{w_2,b_2} \frac{1}{2} & \left\|K(B, C^T)w_2 + e_2 b_2\right\|^2 \\ & + \frac{c_2}{2}(e_1 - K(A, C^T)w_2 - e_1 b_2)^T \\ & \times (e_1 - K(A, C^T)w_2 - e_1 b_2) \end{aligned} \tag{32}$$

Order $KerE = [K(A, C^T)e_1] \in R^{m_1 \times (m+1)}$, $KerF = [K(B, C^T)e_2] \in R^{m_2 \times (m+1)}$. The above equation calculates the partial derivatives of w_1 and b_1, w_2 and b_2 respectively, and the solutions of Quadratic programming problems (29) and (30) can be obtained by constructing linear equations.

$$v_+ = -(KerF^T \cdot KerF + (1/c_1)KerE^T \cdot KerE)^{-1} \cdot KerF^T e_2 \tag{33}$$

$$v_- = -(KerE^T \cdot KerE + (1/c_2)KerF^T \cdot KerF)^{-1} \cdot KerE^T e_1 \tag{34}$$

Order $KerE = \left[K(A, C^T)e_1\right] \in R^{m_1 \times (m+1)}$, $KerF = \left[K(B, C^T)e_2\right] \in R^{m_2 \times (m+1)}$, in the linear case, the solutions for OPP (31) and (32) are obtained:

$$(KerE^T \cdot KerE + c_1 KerF^T \cdot KerF) \cdot v_+ = -c_1 KerF^T \cdot e_2 \tag{35}$$

$$(KerF^T \cdot KerF + c_2 KerE^T \cdot KerE) \cdot v_- = -c_2 KerE^T \cdot e_1 \tag{36}$$

where $v_+ = [w_1; b_1] \in R^{m+1}$, $v_- = [w_2; b_2] \in R^{m+1}$, Simplify (35) and (36) to obtain:

$$v_+ = -(KerF^T \cdot KerF + (1/c_1)KerE^T \cdot KerE)^{-1} \cdot KerF^T e_2 \tag{37}$$

$$v_- = -(KerE^T \cdot KerE + (1/c_2)KerF^T \cdot KerF)^{-1} \cdot KerE^T e_1 \tag{38}$$

Nonlinear LSTSVM requires solving the inverse of a matrix with size $(l + 1) \times (l + 1)$ twice. Then, using the Sherman-Morr-rison-Woodbury (SMW) formula, the nonlinear LSTSVM can be solved by the inverse of three matrices with dimensions less than $(l + 1) \times (l + 1)$.

1) When $m_1 < m_2$, rewrite (37) and (38) after SMW as follows:

$$\begin{aligned} v_+ = & -(Y - Y \cdot KerE^T(c_1 \cdot I + KerE \cdot Y \cdot KerE^T)^{-1} \\ & KerE \cdot Y)KerF^T e_2 \end{aligned} \tag{39}$$

$$\begin{aligned} v_- = & -c_2(Y - Y \cdot KerE^T(I/c_2 + KerE \cdot Y \cdot KerE^T)^{-1} \\ & KerE \cdot Y)KerE^T e_1 \end{aligned} \tag{40}$$

where $Y = (KerF^T \cdot KerF)^{-1}$. Using TSVM as a reference to introduce the regularization term $\varepsilon I, \varepsilon > 0$ in Y to solve the possible ill-posed inverse problem $KerF^T \cdot KerF$.

$$Y = \frac{1}{\varepsilon}(I - KerF^T(\varepsilon I + KerF \cdot KerF^T)^{-1}KerF). \tag{41}$$

2) When $m_2 < m_1$, rewrite (37) and (38) after SMW as follows:

$$\begin{aligned} v_+ = & -c_1(Z - Z \cdot KerF^T(I/c_1 + KerF \cdot Z \cdot KerF^T)^{-1} \\ & KerF \cdot Z)KerF^T e_2 \end{aligned} \tag{42}$$

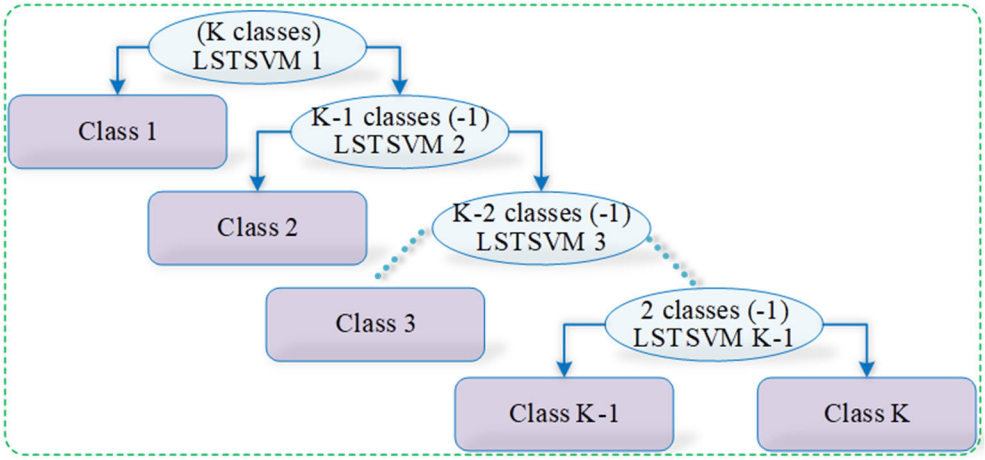


FIGURE 1. BT based multi-classification LSTSVM model BTLSTSVM.

$$v_- = (Z - Z \cdot KerF^T(c_2 \cdot I + KerF \cdot Z \cdot KerF^T)^{-1} KerF \cdot Z)KerF^T e_1 \quad (43)$$

$$Z = \frac{1}{\varepsilon}(I - KerE^T(\varepsilon I + KerE \cdot KerE^T)^{-1}KerE). \quad (44)$$

where $v_+ = [w_1; b_1] \in R^{m+1}$, $v_- = [w_2; b_2] \in R^{m+1}$. Once w_1 and b_1, w_2 and b_2 are obtained, two nonparallel hyperplanes (28) can be solved. By the following decision formula (45), the new data point $x \in R^n$ is divided into a positive class W_1 or a negative class W_2, where $|\cdot|$ represents the distance from the data point to the core hyperplane.

$$k = \arg\min_{k=1,2}\{\left|K(x, C^T)w_1 + b_1\right|, \left|K\left(x, C^T\right)w_2 + b_2\right|\}$$
$$x \in W_k \quad (45)$$

The steps of non-linear LSTSVM are as follows:

Step 1: Input the positive sample $A \in R^{m_1 \times n}$ and negative sample $B \in R^{m_2 \times n}$, and select the appropriate kernel function K;

Step 2: Define $KerE = [K(A, C^T)e_1] \in R^{m_1 \times (m+1)}$ and $KerF = [K(B, C^T)e_2] \in R^{m_2 \times (m+1)}$;

Step 3: Select the penalty parameter c_1, c_2 by verification;

Step 4: If $m_1 < m_2$, the parameters w_1, b_1 and w_2, b_2 of the two hyperplanes are determined by (39) and (40), otherwise w_1, b_1 and w_2, b_2 are determined by (42) and (43);

Step 5: Calculate the distances $\left|K(x, C^T)w_1 + b_1\right|$ and $\left|K(x, C^T)w_2 + b_2\right|$ from the new sample $x \in R^n$ to the two hyperplanes and assign this sample to the positive or negative class by (45).

2) MULTI-CLASSIFICATION LSTSVM

To obtain better fault diagnosis results, we extend non-linear LSTSVM to multi-classification LSTSVM through binary tree (BT) strategy according to different fault diagnosis requirements and use it for fault diagnosis of mechanical equipment. If the training set is:

$$T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_L, y_L)\} \quad (46)$$

where $x_i \in R^n, i = 1, 2, \ldots, L$ is the data point of n dimensional real space R^n, $y_i \in \{1, 2, \ldots, K\}$ is the category label, and L is the number of samples. In this section, LSTSVM is extended to a multi-classification model through the binary tree (BT), as shown in Figure 1.

3) BTLSTSVM

Assuming that the total number of categories is K, BTLSTSVM constructs $K - 1$ binary classifiers and the i classifier identifies Class i ($+1$ class) data and Class $[i + 1, K]$ (-1 class) data. Class i data is $X_i \in R^{l_i \times n}$, and Class $[i + 1, K]$ data is $Y \in R^{(l-l_i') \times n}$, where l_i' is the number of samples with class labels less than or equal to i, that is, $Y_i = [(X_{i+1})^T, (X_{i+2})^T, \ldots, (X_K)^T]^T$. Therefore, the farther away the classifier is from the root node, the fewer negative class samples it has, and the higher its efficiency.

Let the i hyperplane be $f_i = (w_i x) + b_i = 0$, the $K - 1$ nonparallel hyperplanes of non-linear BTLSTSVM are solved according to the above training process, and the category is determined according to the distance. The classifier trained by the BT model is farther away from the root node, the smaller the number of negative samples and the higher the efficiency. BTLSTSVM combines the advantages of the high computational efficiency of BT and the high classification accuracy of LSTSVM, and BT trains fewer classifiers and does not have the problem of inseparable regions. The parameter settings of multi classification model such as BTLSTSVM can be found in the paper [52], and the experimental results also show that the parameter settings have little effect on the recognition accuracy of multi-classification LSTSVM.

III. A NEW FAULT DIAGNOSIS METHOD BASED ON NEEEMD-RUSLP FEATURE SELECTION AND BTLSTSVM

We extract mixed fault features based on NEEEMD and perform more compact feature selection through RUSLP, and low-dimensional and effective feature space is obtained. To accurately identify different fault types and fault degrees

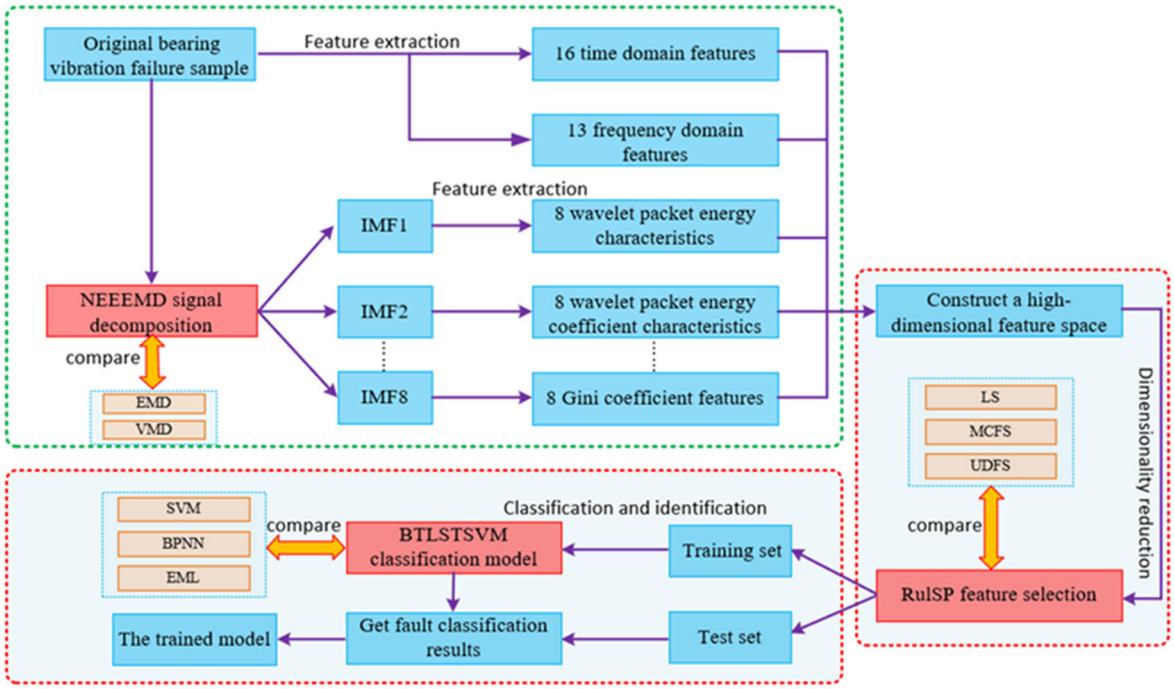


FIGURE 2. Experimental flow chart.

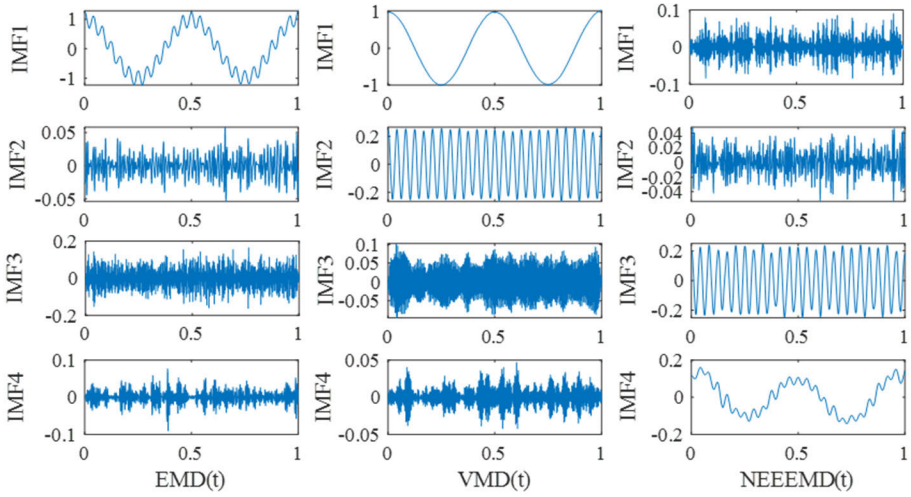


FIGURE 3. Signal decomposition comparison diagram.

of bearings, we construct a BTLSTSVM multi-classification model. The detailed steps are as follows:

1) Firstly, the vibration signals of each state of the rolling bearing are collected through the acceleration sensor (one set of data each from Case Western Reserve University and the University of Ottawa); Then, the vibration signals of each state of the rolling bearing are divided into 60 non-overlapping samples (each sample size is 2000) with the same length.
2) 16 time-domain features and 13 frequency-domain features of 60 original vibration signal samples are extracted respectively.
3) The signal decomposition method based on NEEEMD is constructed to extract 8 IMFs of the original vibration signal; Then, the wavelet packet energy entropy, packet energy coefficient and Gini coefficient extracted from each IMF are constructed as the time-frequency domain features of the signal.
4) The time-domain features, frequency-domain features, and time-frequency domain features extracted from the signal samples are constructed as a mixed fault feature vector set, where the maximum dimension of the feature vector set is $D = 53$, and the size of the feature matrix for each state is 60·53.

TABLE 3. Feature extraction time.

decomposition	EMD	VMD	NEEEMD
time(s)	47.72	8133.54	114.68

5) The RULSP algorithm is constructed to select the optimal low-dimensional features from the fault feature vector set, and the optimal low-dimensional training sample dataset and low-dimensional test sample dataset can be obtained.
6) 600 low-dimensional feature samples are divided into 300 training samples and 300 test samples, and the training set is input into the BTLSTSVM classifier to train the model. The trained BTLSTSVM model is used for fault pattern recognition of the test samples.

To verify the effectiveness of the new fault diagnosis method based on NEEEMD-RUSLP feature selection and BTLSTSVM, the experiment is organized by the flow chart shown in Figure 2.

1) To verify the effectiveness of NEEEMD in fault feature extraction, we simultaneously extract EMD and VMD and compare the fault diagnosis results with NEEEMD.
2) To verify the effectiveness of the RUSLP dimensionality reduction method, we compare the fault diagnosis results of LS, UDFS, MCFS and RUSLP.
3) To verify the effectiveness of the multi-classification BTLSTSVM model in fault identification, we compare the fault diagnosis results of SVM, BBN, ELM and BTLSTSVM.

IV. EXPERIMENTAL VERIFICATION

A. SIMULATION SIGNAL VERIFICATION

In the feature extraction stage, the vibration frequency of the simulation signal is used to carry out experiments, and the following bearing simulation signal is constructed.

$$x(t) = y_0 e^{-2\pi f_n \zeta t} \sin(2\pi f_n \sqrt{1-\zeta^2} t) \tag{47}$$

Among them, the natural frequency $f_n = 3000$ Hz, displacement constant $y_0 = 2.5$, damping coefficient $\zeta = 0.1$, impact failure period T = 0.00625 s (characteristic frequency $f_{chthe} =$ 0.00625 s = 160 Hz), sampling frequency $f_s = 12KHz$, and number of data points $N = 4096$ of the bearing. Meanwhile, 5dB of white noise is added to the fault simulation signal. The parameters based on the NEEEMD method are $num\_IMF = 8$, $NR = 10$, $NstdMax = 0.2$. In order to verify the performance of NEEEMD in characterizing bearing fault features, we also extract the fault features of EMD and VMD from each signal sample for comparative analysis.

The original signal is decomposed into four IMFs by EMD, VMD and NEEEMD, and the results are shown in Figure 3. The signal decomposed by EMD does not contain strong periodicity and contains messy noise components, that is, the performance of EMD filtering noise is poor. In the signal decomposed by VMD, IMF1∼IMF2 have good periodicity and smooth trend, but the noise component in IMF3 still exists, that is, VMD cannot filter the noise component in the signal well. In the signals decomposed by NEEEMD, IMF2∼IMF3 contain obvious periodic characteristics, and the noise component in IMF1 is suppressed, which can better eliminate the influence of the noise proposed in this paper on fault extraction. The three signal decomposition methods used in this paper and the feature extraction time are shown in Table 3. EMD consumes the least time but the signal decomposition performance is not high, and the performance of VMD signal decomposition is general and consumes up to 70 times that of the NEEEMD method. NEEEMD not only effectively suppresses noise and extracts vibration shocks from signals, but also consumes relatively less time. Therefore, this paper constructs time-frequency domain features through NEEEMD to ensure the accuracy and efficiency of feature extraction.

B. CASE WESTERN RESERVE BEARING FAULT DIAGNOSIS EXPERIMENT

1) INTRODUCTION TO THE EXPERIMENTAL PLATFORM

The experimental platform for rolling bearings at Case Western Reserve University is shown in Figure 4.

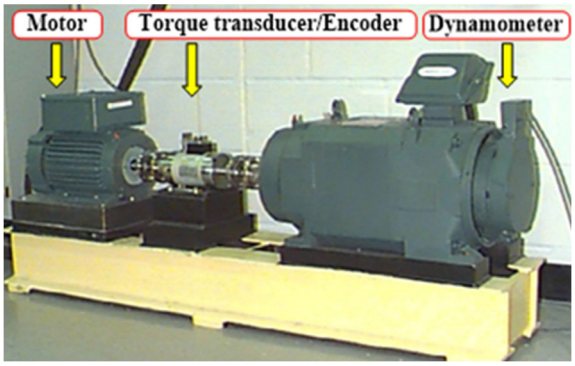


FIGURE 4. Case western reserve university rolling bearing experimental platform [37].

The rolling bearing type is 6205-2RSJEMSKF, the motor speed is 1797rpm (conversion frequency is 1797/60 Hz = 29.53 Hz), the sampling frequency of the system is 12 KHz, and the experimental data length is 2048. The parameters are shown in Table 4.

The vibration signal of the fan end bearing at the motor speed of 1797 r/min is used as the experimental data. Each set of data has 120,000 sampling points, and 2000 sampling points are taken for each sample. Therefore, there are

TABLE 4. Structural parameters of skf 6205 bearing [37].

bearing designation	Roller diameter	Rolling elements	Pitch diameter	contact angle
SKF 6205	7.938 mm	9	39 mm	0

60 sample data for each group of faults and 10 sets of fault data for a total of 600 samples. Among them, 1∼60 are normal data; 61∼120, 121∼180, 181∼240 are inner ring fault data; 241∼300, 301∼360, 361∼420 are outer ring fault data; 421∼480, 481∼540, 541∼600 are rolling element fault data. The data description is shown in Table 5. The three fault diameters of the inner ring, outer ring and rolling element are 0.007 inches, 0.014 inches and 0.021 inches, respectively. The sample labels for the 10 types of feature vectors are 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10.

The time-domain waveform of the vibration signal is shown in Figure 5, and it can be seen that the signals IR07, IR21, OR07, OR14 and OR21 have obvious periodic impacts, and the amplitude difference between the signals is obvious. The impact characteristics and cycles of the remaining signals are not obvious, so it is necessary to diagnose the bearing fault type and fault degree through feature extraction and fault diagnosis model.

2) FAULT FEATURE EXTRACTION

The bearing vibration signal at the motor speed of Case Western Reserve University at 1797 r/min is used as experimental data, and the data description refers to the experimental introduction in this section. Firstly, the vibration signal of each state is divided into 60 non-overlapping samples with a length of 2000. Then, the time-domain and frequency-domain characteristics of each original bearing vibration signal are extracted. Secondly, in order to construct the time-frequency domain characteristics of the vibration signal, the wavelet packet energy entropy, packet energy coefficient and Gini coefficient are extracted from the IMF obtained by NEEEMD decomposition. Finally, the extracted time-domain, frequency-domain, and time-frequency domain features are constructed into a fault feature vector set. The obtained high-dimensional feature matrix size is 600·53, and the maximum dimension of is $D = 53$. The distribution of feature samples under each dimension D is shown in Figure 6 (a) and (b).

The distribution of feature samples in partial dimension D is shown in Figure 6(b). In Figure 6(a), the mixed features of different fault types have differences in dimension D, but some features of the same fault type deviate from most of the features of that type, that is, there are outliers. As shown in Figure 6(b), when dimension D=2, the mixed features of 10 bearing fault states are separable to a certain extent, but the edges of the mixed features of various fault states are fuzzy. When dimension D = 20, the mixed features of various fault states overlap, and the calculated feature values are clustered below. When dimension D=38, the mixed characteristics of various fault states are chaotic and scattered. When dimension D=46, the mixed features of various fault states overlap, and the calculated eigenvalues are clustered above. Therefore, the mixed features obtained by the original extraction are difficult to find the dividing line for dividing sample categories under most dimensions, resulting in low sample separability. Therefore, this paper proposes RUSLP to sort and select hybrid features, filter out features with small correlation to reduce feature redundancy, and improve fault diagnosis reliability.

3) FAULT FEATURE SELECTION

In order to verify the influence of RUSLP dimensionality reduction method on the diagnostic performance of bearing faults, using the NEEEMD decomposition method and BTLSTSVM multi-classification, we compare the accuracy of RUSLP dimensionality reduction method with LS, MCFS and UDFS methods under the same mixed high-dimensional fault feature vector set and the same parameters. The parameters of BTLSTSVM multi-classification are set to OptPara.c1=0.25, OptPara.c2=4, OptParaKerfPara.pars=4. The results are shown in Figure 7.

It can be seen that the fault diagnosis accuracy without the dimension reduction method can reach 95% in $8 \leq x \leq 19$ interval. However, as the feature dimension gradually increases, the fault diagnosis accuracy continues to decrease to 10%. In addition to the LS method, other dimensionality reduction methods can also maintain high fault diagnosis accuracy when the dimension is increasing. Therefore, it is necessary to select an effective dimensionality reduction method to select more compact low-dimensional fault features for high-dimensional fault feature vector sets. For the model constructed in this paper, except for LS, the accuracy of fault diagnosis can reach ≥ 95% when the other dimension reduction methods are in $11 \leq x \leq 39$ interval. Table 6 shows the highest accuracy obtained by each dimensionality reduction method in Figure 7, and the corresponding feature dimension x. When using the LS dimensionality reduction method, as the feature dimension continues to increase, the fault diagnosis accuracy decreases from 63% to 10%, indicating that the LS dimensionality reduction method is not suitable for the model constructed in this paper. However, when $x = 31$, the fault diagnosis accuracy of 96% can be obtained by using the MCFS dimension reduction method. When $x = 26$, the fault diagnosis accuracy of 96.67% can be obtained by using the UDFS dimension reduction method.

TABLE 5. Bearing vibration signal description [37].

Category	Size	Abbreviation	Label
Normal	0.000 inches	Normal	1
Failure of inner ring	0.007 inches	IR07	2
	0.014 inches	IR14	3
	0.021 inches	IR21	4
Failure of outer ring	0.007 inches	OR07	5
	0.014 inches	OR14	6
	0.021 inches	OR21	7
Failure of roller	0.007 inches	B07	8
	0.014 inches	B14	9
	0.021 inches	B21	10

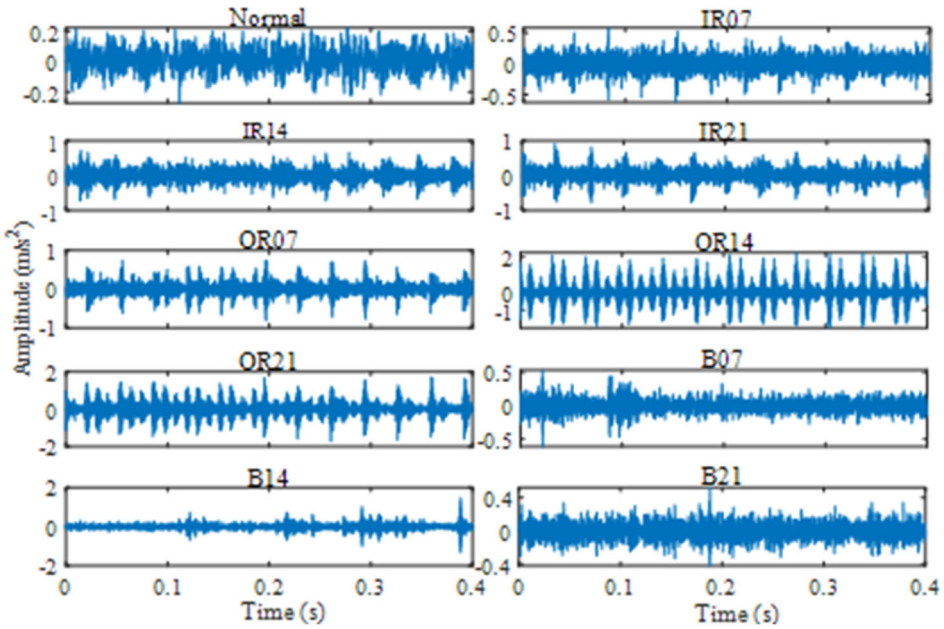


FIGURE 5. Time domain waveform of vibration signal [37].

When $x = 39$, the fault diagnosis accuracy of 97% can be obtained by using the RUSLP dimension reduction method. Therefore, for the model constructed in this paper, the use of the RUSLP dimensionality reduction method can improve the fault diagnosis accuracy of bearings.

When the LS, MCFS, UDFS, and RUSLP methods extract the first 1 dimensional, first 31 dimensional, first 26 dimensional, and first 39 dimensional features, respectively, the resulting confusion matrix is shown in Figure 8. If the LS dimensionality reduction method is used, there is a false diagnosis between the seven states of IR07, IR14, IR21, OR07, OR14, B07 and B14, and the fault categories are difficult to be accurately distinguished. If the MCFS dimensionality reduction method is used, 3% of B14 samples will be wrongly diagnosed as IR14 status. If the UDFS dimensionality reduction method is used, 3% of B14 samples will be wrongly diagnosed as IR14 state. If the RUSLP dimensionality reduction method is used, samples of

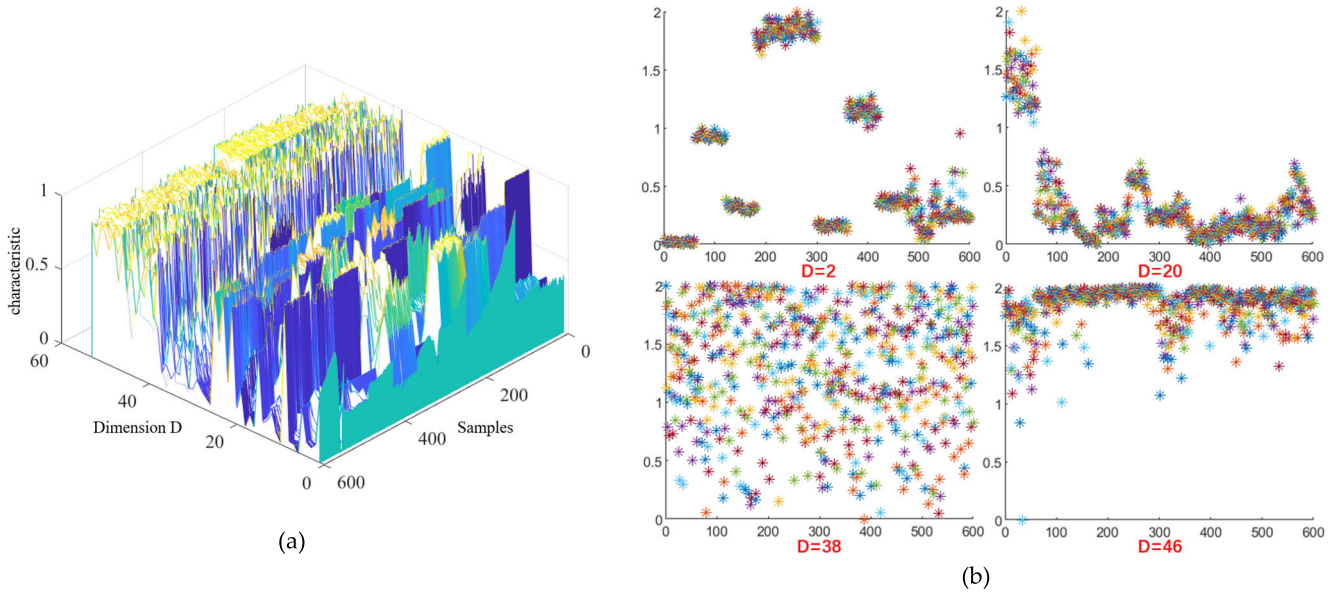


FIGURE 6. Distribution of characteristic samples for Casey Reservoir bearings with different dimensions D.

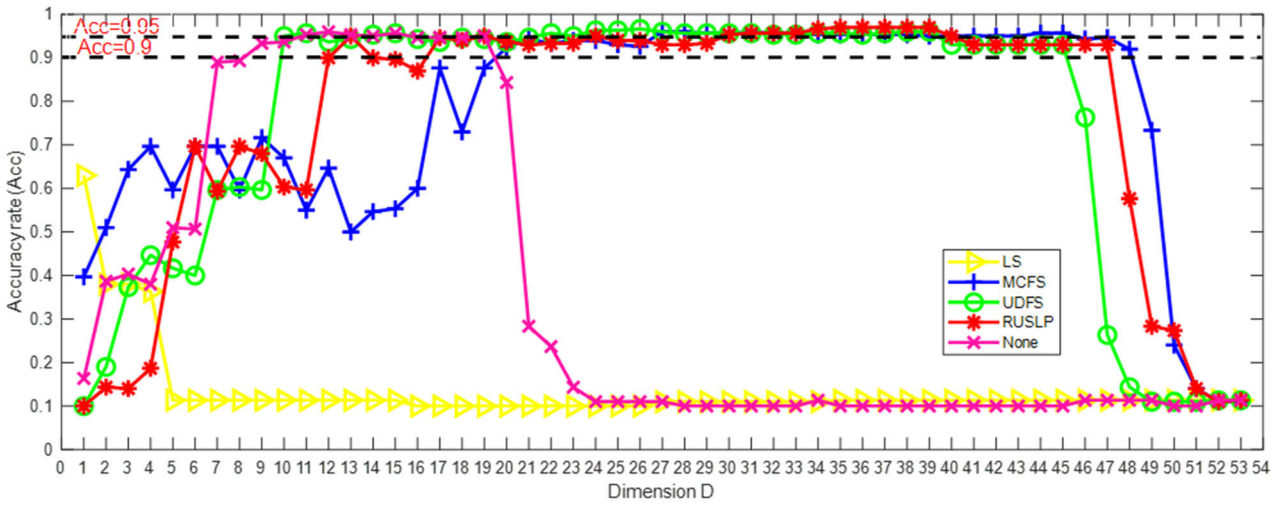


FIGURE 7. Fault diagnosis accuracy under different dimensionality reduction methods.

10 different states can be correctly diagnosed. The experimental results show that the RUSLP dimensionality reduction method can effectively improve the accuracy of bearing fault diagnosis.

4) FAULT IDENTIFICATION

To verify the effect of the BT-LSTSVM multi-classifier method on the bearing fault diagnosis performance, we use the NEEEMD decomposition method and the BT-LSTSVM multi-classifier under the condition of the same mixed high-dimensional fault feature vector set and the same parameters. The accuracy of BT-LSTSVM multi-classifier is compared with SVM, BPNN and ELM methods. The parameters of the RUSLP-based method are set as X= data matrix, G= encoding matrix, F= orthogonal basis matrix, and W= feature selection matrix.

The results are shown in Figure 9. It can be seen that, except for the SVM and BPNN methods, both the ELM and BT-LSTSVM classification methods can achieve the accuracy of $\geq$ 95% in the interval $3 \leq x \leq 19$. Table 7 shows the highest accuracy obtained by each classification method in Figure 9 and the corresponding feature dimension x. SVM extracts the first 12 dimensional features, and the accuracy can be 87%. BPNN extracts the first 25 dimensional features with an accuracy of 94%. ELM extracts the first three-dimensional features with an accuracy of 95.33%. BT-LSTSVM extracts the first 19 dimensional features with an accuracy of 97.67%. Therefore, the fault diagnosis

TABLE 6. The highest accuracy and corresponding feature dimension x obtained by different dimensionality reduction methods.

reduced-dimension model	LS	MCFS	UDFS	RUSLP
Pre-X-dimensional features	1	31	26	39
Highest precision	0.63%	0.96	0.9667	0.97

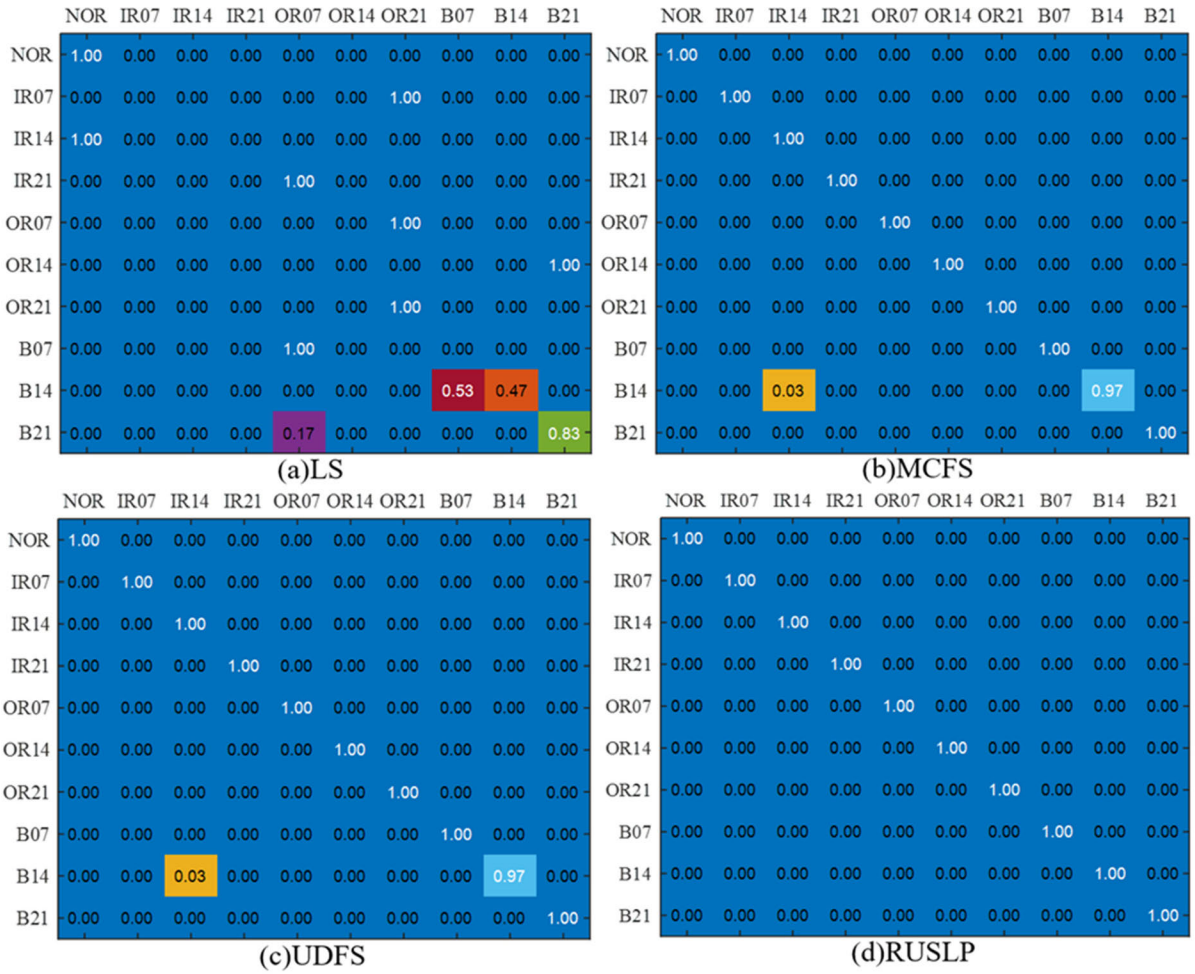


FIGURE 8. Confusion matrix Obtained by different dimension reduction methods taking the dimension with the highest accuracy.

accuracy is improved by 10.67% under BT-LSTSVM dimension reduction method.

When the SVM, BPNN, ELM, and BT-LSTSVM methods extract the first 12 dimensions, the first 25 dimensions, the first 3 dimensions, and the first 19 dimensions respectively, the confusion matrix obtained is shown in Figure 10. If the SVM recognition model is used, 10% of B14 samples will be incorrectly diagnosed as IR14, 10% of B14 samples will be incorrectly diagnosed as B07, and 3% of B14 samples will be incorrectly diagnosed as B21. 3% of B21 samples will be incorrectly diagnosed as IR07, 3% of B21 samples will be incorrectly diagnosed as IR14, and 3% of B21 samples will be incorrectly diagnosed as B07. If the BPNN is used to identify the model, there is a misdiagnosis between IR14, OR14, OR21, B07, B14 and B21. If the ELM recognition model is used, 3% of B14 samples will be incorrectly diagnosed as IR14,20% of B14 samples will be incorrectly diagnosed as OR14, 3% of B14 samples will be incorrectly diagnosed as B07, and 3% of B14 samples will be incorrectly diagnosed as B21. 20% of B21 samples will be incorrectly diagnosed as OR14, 7% of B21 samples will be incorrectly diagnosed as B07, and 7% of B21 samples will be incorrectly diagnosed as B14. If the BT-LSTSVM multi-classifier is used, only 3% of B07 samples are incorrectly diagnosed as OR14.

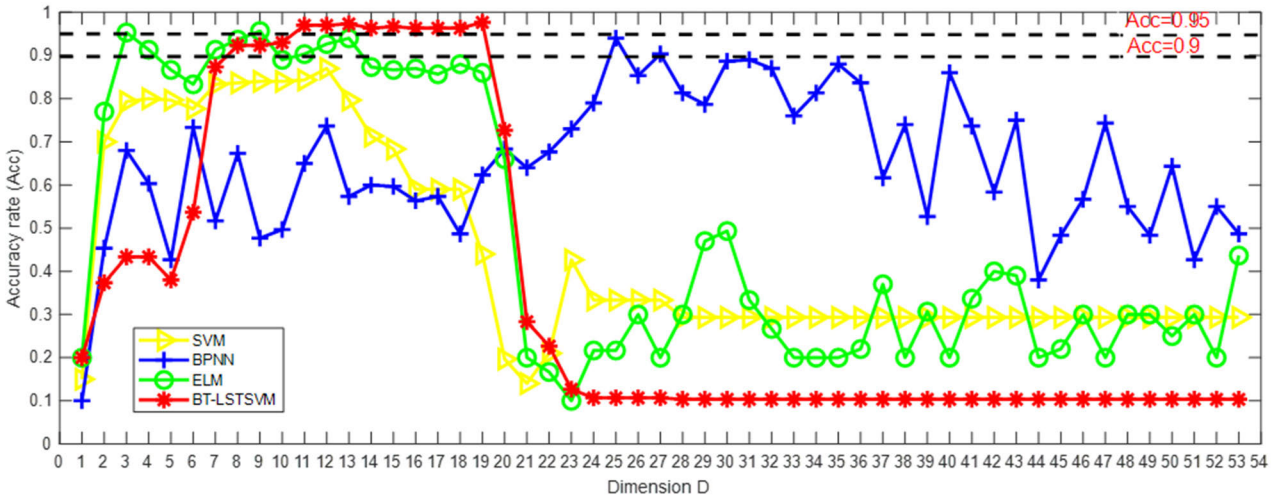


FIGURE 9. Fault diagnosis accuracy of different classification methods.

The experimental results show that the BT-LSTSVM multi-classifier can effectively improve the accuracy of bearing fault diagnosis.

C. BEARING FAULT DIAGNOSIS EXPERIMENT OF UNIVERSITY OF OTTAWA

1) INTRODUCTION TO THE EXPERIMENTAL PLATFORM

Ottawa data is carried out on the Spectra Quest Mechanical Failure Simulator (MFS-PK5M). The experimental device is shown in Figure 11.

The simulator is a data set of bearing vibration collected under the condition of time varying speed. The health status of this bearing includes health, failure and inner and outer seat defects. The running speed conditions of the data set include increasing speed, decreasing speed, increasing then decreasing speed, and decreasing then increasing speed (the data set used in this paper is running at increasing speed). The bearing is driven by a motor and the speed is controlled by an AC drive. The experimental platform is equipped with two ER16K ball bearings to support the shaft, the healthy one on the left and the experimental one on the right, which are replaced by bearings with different health conditions. The accelerometer (ICP accelerometer, model 623C01) is placed on the housing of the test bearing to collect vibration data. In addition, an incremental encoder (EPC model 775) was installed to measure shaft speed. Bearing faults can be detected and diagnosed by observing the fault characteristic frequency (FCF) in the frequency domain. For each type of failure, it has a specific FCF, which is proportional to the operating rotation frequency, and the coefficient is determined by the bearing structure parameters. The structural parameters of the bearings used in the experiment are shown in Table 8. According to the bearing parameters, the FCF coefficient of the bearing inner ring is 5.43, that is, BPFI=5.43. Similarly, the FCF of the bearing outer ring, the ball passing frequency (BPFO) of the outer ring = 3.57.

The number of datasets is shown in Table 9. The data are sampled at 200000 Hz, and the sampling duration is 10s. For each dataset, there are two experimental settings: bearing health condition and variable speed condition. The health condition of bearings includes health, faults with inner ring defects, and faults with outer ring defects. The operating speed condition is to increase the speed. Therefore, there are three different scenarios. In order to ensure the authenticity of the data, three experiments are collected in each experimental environment, and the results are in a total of nine data sets.

2) FAULT FEATURE EXTRACTION

The bearing vibration signal when the operating speed of the Ottawa motor is increasing is taken as the experimental data, and the data description is referred to the introduction of the test bench in this section. Firstly, the vibration signal of each state is divided into 60 non-overlapping samples with a length of 2000. Then, the time-domain and frequency-domain features of each original bearing vibration signal are extracted. Secondly, to construct the time-frequency domain features of the vibration signal, the wavelet packet energy entropy, packet energy coefficient and Gini coefficient are extracted from the IMF obtained by NEEEMD decomposition. Finally, the extracted time-domain, frequency-domain and time-frequency domain features are constructed into a set of fault feature vectors. The dimension of the high-dimensional feature matrix is 180·53, and the maximum dimension is D=53. The distribution of feature samples under each dimension D is shown in Figure 12 (a) and (b).

In Figure 12 (a), the mixed features of different fault types are different in dimension D, but some features of the same fault type deviate from most features of this type, that is, there are outliers. As shown in Figure 12 (b), When the dimension D = 12, the mixed features of the three fault states of the bearing have a certain degree of separability, but the edges of the mixed features of the two fault states on the right are blurred.

TABLE 7. The highest accuracy and corresponding feature dimension x obtained by different classification methods.

Reduced Dimension Model	SVM	BPNN	ELM	BTLSTSVM
Front X-dimensional features	12	25	3	19
Highest precision	0.87	0.94	0.9533	0.9767

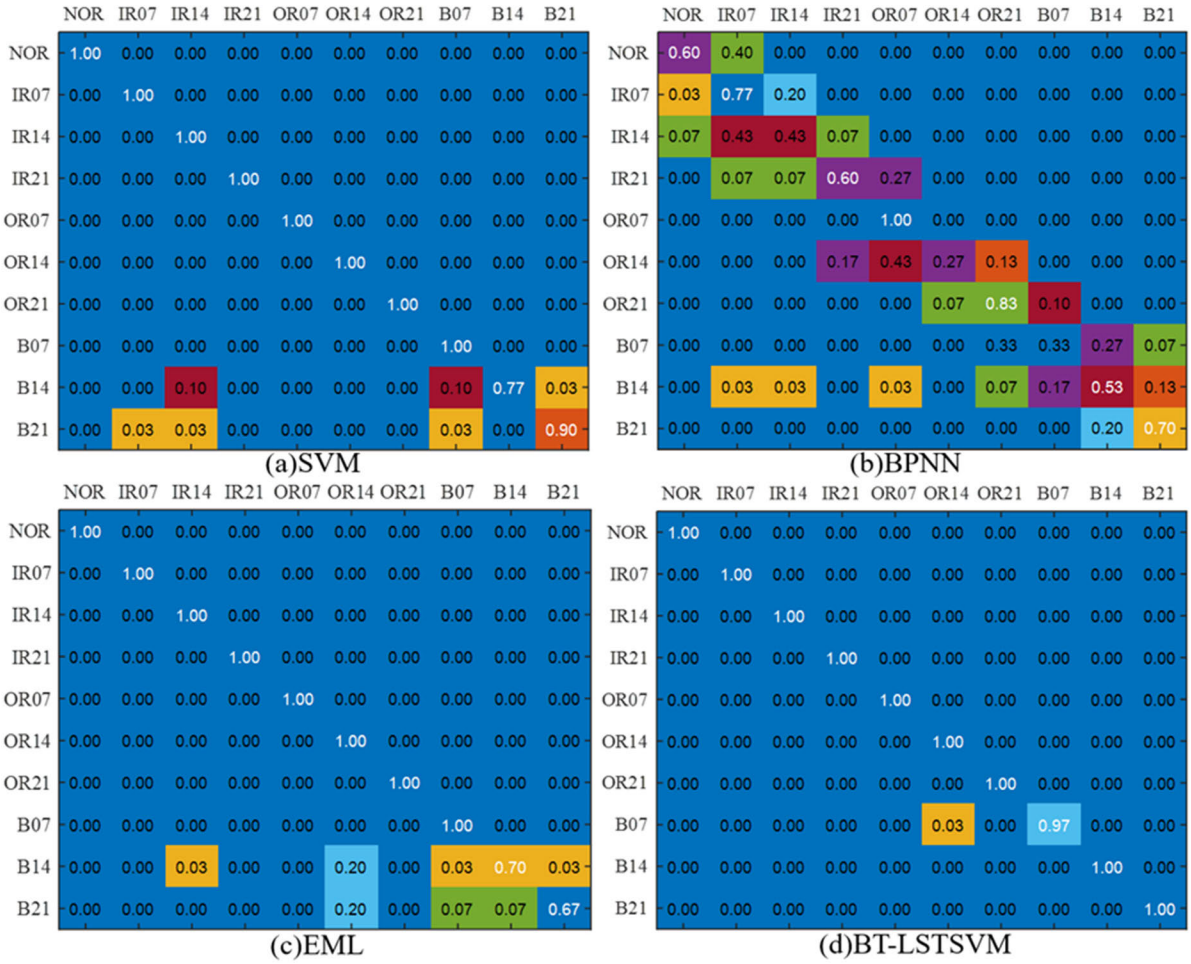


FIGURE 10. The confusion matrix obtained by selecting the highest precision dimension with different classification methods.

TABLE 8. Structural parameters of bearings [53].

Bearing Type	Pitch Diameter	Spherical Diameter	Inner Ring Ball Frequency	Outer Ring Ball Frequency
ER16K	38.52 mm	7.94 mm	9	5.43

When the dimension D=23, samples of the same fault type are clustered together, and the sample boundaries of different fault types are clear. However, the result of the division is 6 categories, while the true number of fault types is only 3. When dimension D = 31, the mixed features of various fault states overlap and are clustered below. When dimension D = 40, the mixed characteristics of various fault states are chaotic and scattered. Therefore, the mixed features obtained by the original extraction are difficult to find the dividing line for dividing sample categories under most dimensions,

TABLE 9. Data type description [53].

	Bearing Health	Inner Ring Fault	Outer Ring Fault
Operating Speed is Increasing	H-A-1	I-A-1	O-A-1
	H-A-2	I-A-2	O-A-2
	H-A-3	I-A-3	O-A-3

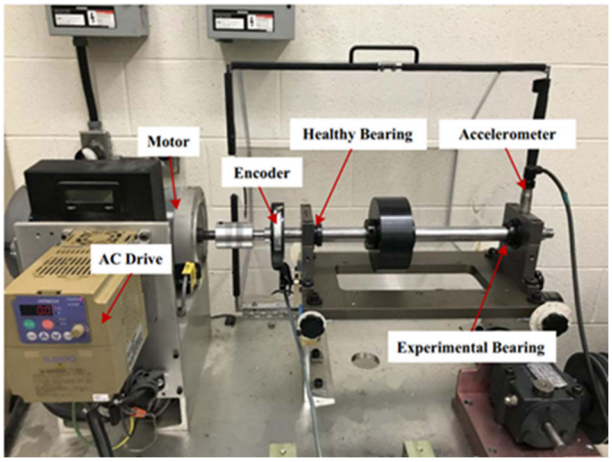


FIGURE 11. Spectra Quest mechanical fault simulator [53].

resulting in low sample separability. Therefore, this paper proposes RUSLP to sort and select the mixed features, filter out the features with small correlations, and improve the reliability of fault diagnosis.

3) FAULT FEATURE SELECTION

To verify the impact of the RUSLP dimensionality reduction method on bearing fault diagnosis performance, we use the NEEEMD decomposition method and the BT-LSTSVM multiple classifiers to compare the accuracy of the RUSLP dimensionality reduction method with LS, MCFS and UDFS methods under the condition of the same mixed high-dimensional fault feature vector set and the same parameters. Among them, the parameter Settings based on the BT-LSTSVM multi-classifier are the same as above. The results are shown in Figure 13.

It can be seen that the fault diagnosis accuracy without the dimension reduction method can reach 98% when $6 \leq x \leq 19$ interval. However, with the increase of feature dimension, the fault diagnosis accuracy decreases to 35%. In addition to LS and UDFS methods, other dimensionality reduction methods can also maintain high fault diagnosis accuracy when the dimensionality continues to increase. Therefore, it is necessary to select an effective dimensionality reduction method to select more compact low-dimensional fault features from the high-dimensional fault feature vector set. For the model built in this paper, except LS, the fault diagnosis accuracy of other dimension reduction methods can reach $\geq 95\%$ in the interval $2 \leq x \leq 19$. Table 10 shows the highest accuracy obtained by each dimension reduction method in Figure 13, and the corresponding feature dimension x. When the LS dimension reduction method is used, the fault diagnosis accuracy is maintained near 33.34% by selecting any feature dimension. This indicates that the LS dimension reduction method is not suitable for the model constructed in this paper. However, when $x = 10$, the fault diagnosis accuracy of 100% can be obtained by using the MCFS dimension reduction method. When $x = 2$, the UDFS dimension reduction method can achieve 99.56% of the fault diagnosis accuracy. When $x = 7$,100% fault diagnosis accuracy can be obtained using the RUSLP dimensionality reduction method. Therefore, for the model constructed in this paper, the other dimensionality reduction methods, except the LS dimensionality reduction method, obtain high fault diagnosis accuracy. However, MCFS can achieve 100% fault diagnosis accuracy only by selecting the first 10 dimensional features, and RUSLP can achieve 100% fault diagnosis accuracy only by selecting the first 7 dimensional features. It shows that RUSLP can well characterize the characteristics of bearing faults with fewer dimensions. The data from the University of Ottawa can again demonstrate the superiority of the dimensionality reduction method cited in this paper by RUSLP.

If the LS dimension reduction method is used, 65% of the normal state samples will be incorrectly diagnosed as inner loop fault state, and the fault mode is difficult to be accurately identified. If the MCFS dimensionality reduction method is used, 96% of the normal state samples will be incorrectly diagnosed as the inner ring fault state, and 1% of the inner ring fault samples will be incorrectly diagnosed as the outer ring fault state. If the UDFS dimensionality reduction method is used, 94% of the inner ring fault samples will be incorrectly diagnosed as the outer ring fault state. When using the RUSLP dimensionality reduction method, only 3% of inner ring fault samples will be incorrectly diagnosed as outer ring faults. The experimental results show that the RUSLP can effectively solve the problem of feature confusion and improve the accuracy of fault diagnosis.

4) FAULT IDENTIFICATION

To verify the impact of the BT-LSTSVM multi-classifier method on bearing fault diagnosis performance, the NEEEMD decomposition method and the BT-LSTSVM multi- classifier are used. Under the condition of the same mixed high-dimensional fault feature vector set and the same parameters, we compare the accuracy of the BT-LSTSSVM multi- classifier with SVM, BPNN and ELM methods.

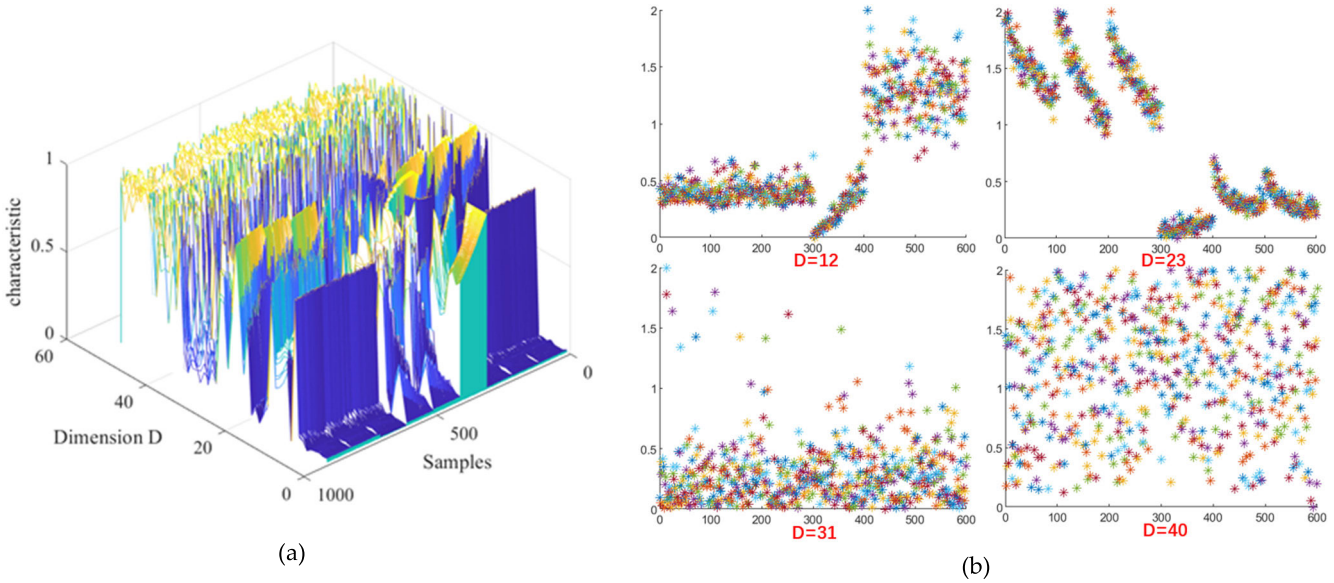


FIGURE 12. Distribution of characteristic samples of Ottawa bearings with different dimension D.

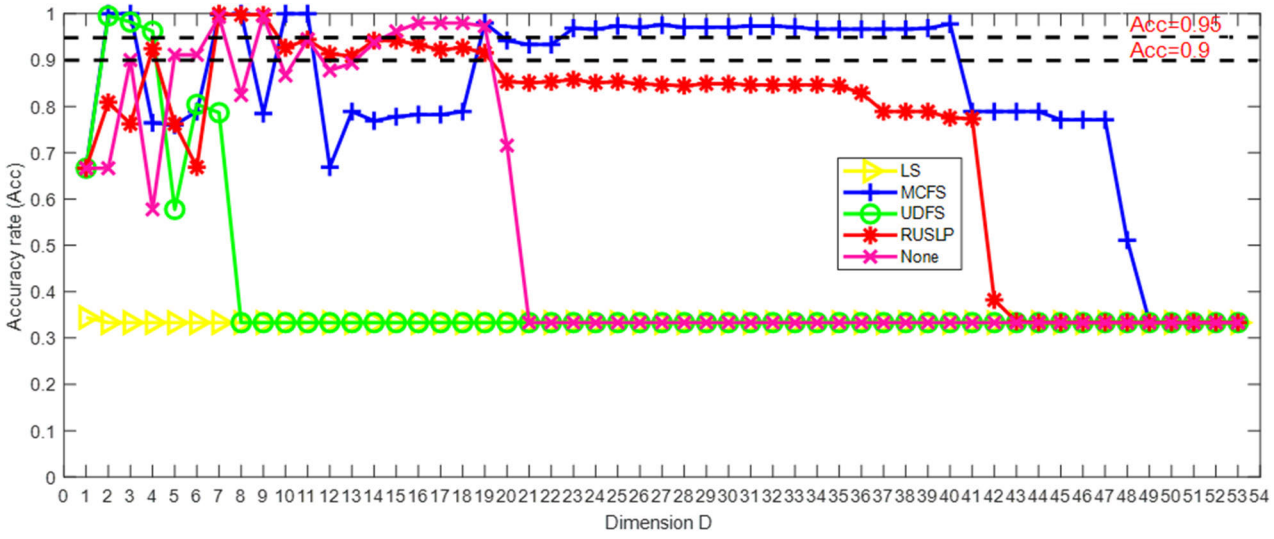


FIGURE 13. Fault diagnosis accuracy under different dimensionality reduction methods.

The parameter settings based on the RUSLP method are the same as above. The results are shown in Figure 15.

It can be seen that except for SVM, the other classification methods can achieve the accuracy of $\geq 95\%$ when the interval $7 \leq x \leq 21$ is used. Table 11 shows the highest accuracy obtained by each classification method in Figure 15 and the corresponding feature dimension x. SVM can extract the first 14 dimensional features with an accuracy of 89.56%. BPNN can extract the first 20 dimensional features with an accuracy of 98.67%. ELM can extract the first 15 dimensional features with an accuracy of 97.56%. BT-LSTSVM can extract the first nine dimensional features with an accuracy of 99.56%. Therefore, under the BT-LSTSVM dimension reduction method, the fault diagnosis accuracy is improved by 10%.

When the SVM, BPNN, ELM and BT-LSTSVM respectively extract the first 14, 20, 15 and 9 dimensions of features, the confusion matrix is shown in Figure 16.

When LS, MCFS, UDFS, and RUSLP methods extract the first 1, 10, 2, and 7 dimensions of features respectively, the Confusion matrix obtained is shown in Figure 14.

If the SVM recognition model is used, 18% of the normal state samples will be incorrectly diagnosed as the outer ring fault state. If the BPNN identification model is used, 11% of the normal state samples will be incorrectly diagnosed as the inner circle fault state, and 18% of the normal state samples

TABLE 10. The highest accuracy and corresponding feature dimension x obtained by different dimensionality reduction methods.

Reduced Dimension Model	LS	MCFS	UDFS	RUSLP
Front X-dimensional features	1	10	2	7
Highest precision	0.3334	1	0.9956	1

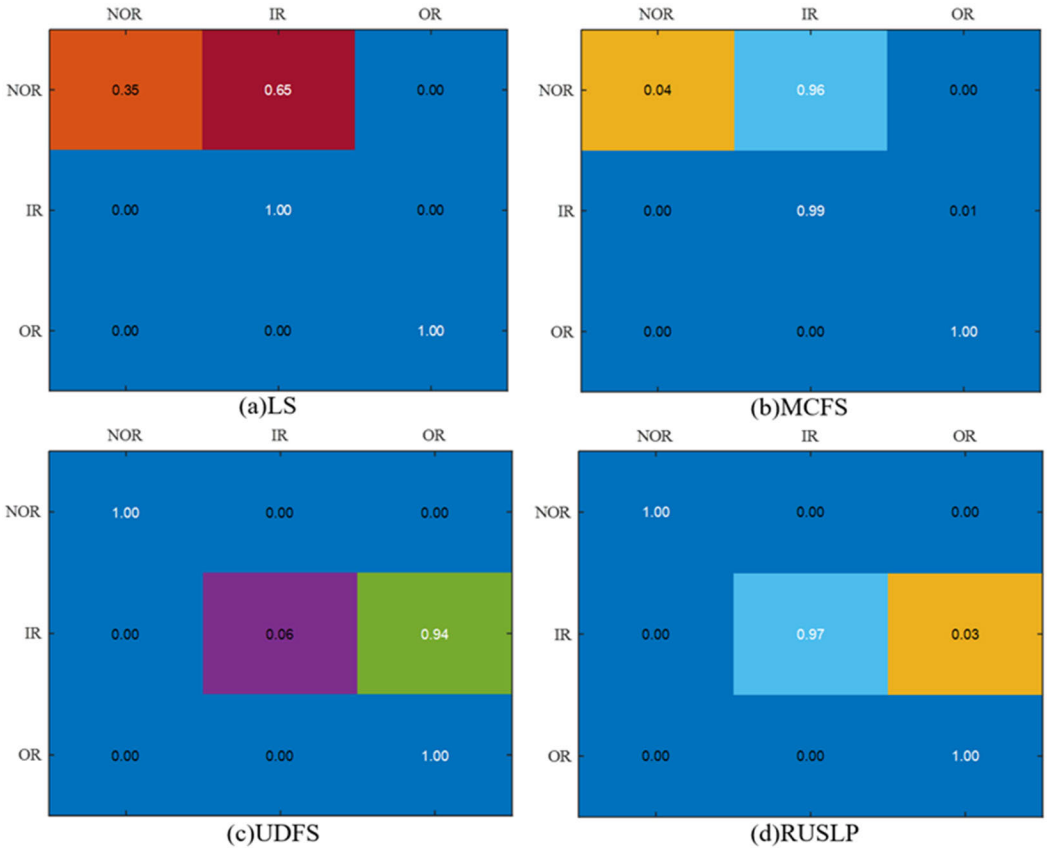


FIGURE 14. Confusion matrix Obtained by different dimension reduction methods taking the dimension with the highest accuracy.

TABLE 11. The highest accuracy and corresponding feature dimension x obtained by different classification methods.

Classification Model	SVM	BPNN	ELM	BTLSTSVM
Front X-dimensional features	14	20	15	9
Highest precision	0.8956	0.9867	0.9756	0.9956

will be incorrectly diagnosed as the outer circle fault state. If the ELM identification model is used, 21% of inner loop fault samples will be incorrectly diagnosed as normal. If the BTLSTSVM multi-classifier is used, only 8% of the normal state samples are incorrectly diagnosed as the outer circle fault state, and only 3% of the inner circle fault samples are incorrectly diagnosed as the normal state. The experimental results show that the BT-LSTSVM multi-classifier can effectively improve the accuracy of bearing fault diagnosis.

D. CHECK VALVE FAULT DIAGNOSIS EXPERIMENT

1) INTRODUCTION TO THE EXPERIMENTAL PLATFORM

Check valves operate in complex slurry conveying environments, and the spool is continuously affected by slurry flushing and its own spring system, so the vibration signal is generally composed of fault signal, multi-part vibration coupling signal and background noise. In this experiment, the vibration signal acquisition system of each state of the check valve is shown in Figure 17.

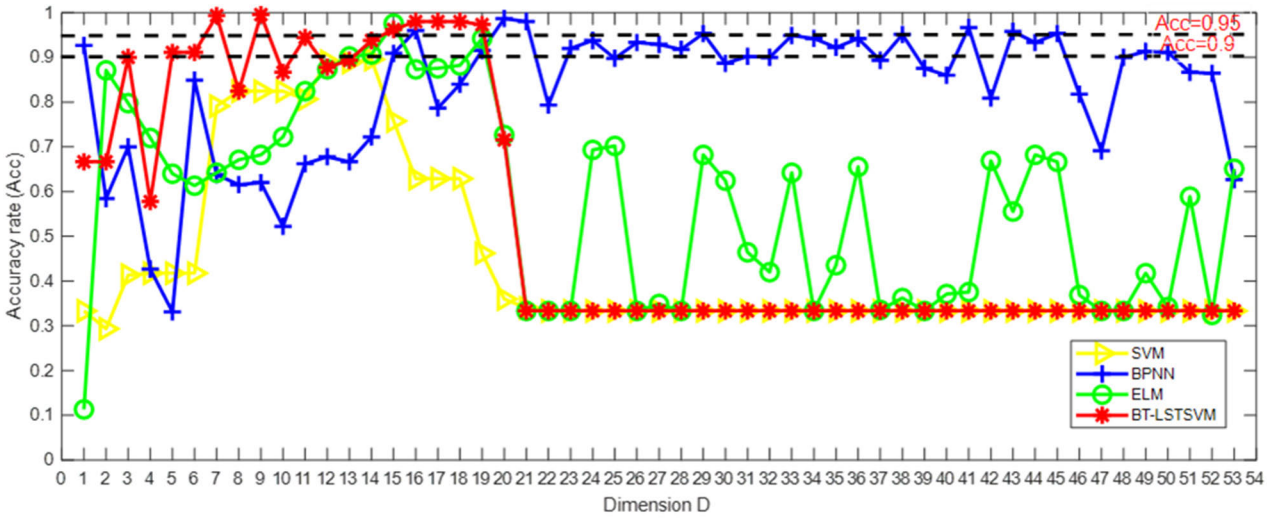


FIGURE 15. Fault diagnosis accuracy under different classification methods.

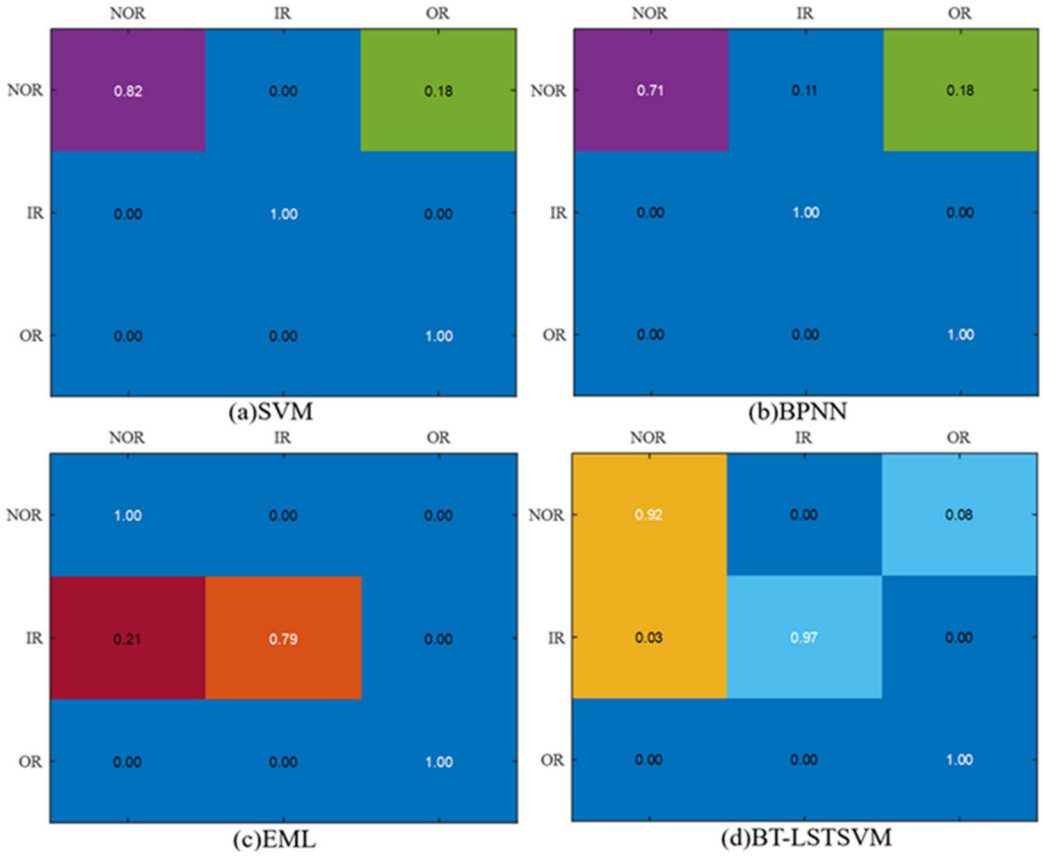


FIGURE 16. Confusion matrix obtained by selecting the highest precision dimension with different classification methods.

The red circle marks six PCB 352C33 accelerometers with a sensitivity of 100 mV/g, fixed to the housing of six check valves. The PCB 352C33 sensor is connected to the PS PXI-3342 Type 8-channel data acquisition card via the type 002C10 BNC high shielded wire, as shown in the blue transmission line in the figure, as shown in Figure 17(d). The resolution during data transmission is 24 bit, and the transmission rate is up to 204.8 ks/s. We collect vibration data through PXI-3342 data acquisition card, of which channels 0, 2 and 4 collect vibration signals in the Z direction of

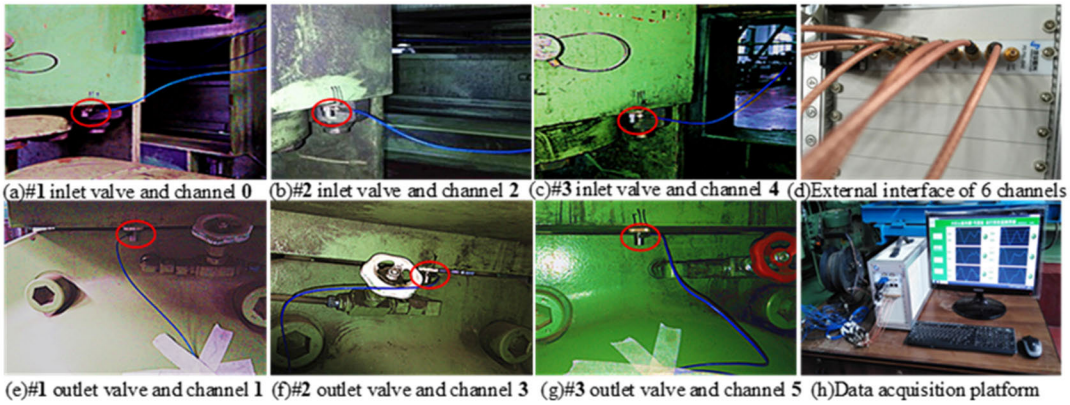


FIGURE 17. Sensor measuring points and data acquisition platform.

the #1, #2 and #3 feed valves respectively, and channels 1, 3 and 5 collect vibration signals in the Z direction of #1, #2 and #3 discharge valves respectively, with a sampling frequency of 2560 Hz and a single channel data length of 76800 points. Among the acquired signals, the time-domain waveforms of a random set of vibration signals with normal state, stuck valve failure, and wear fault are shown in Figure 18.

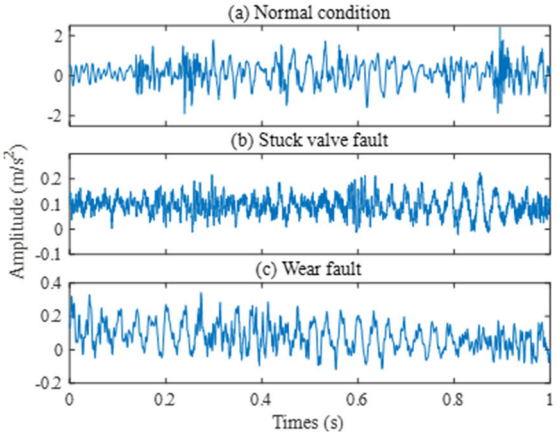


FIGURE 18. Faulty check valve and replaced check valve.

2) FAULT FEATURE EXTRACTION

The vibration signal of the check valve in the actual working condition is used as the experimental data, and the data description refers to the experimental introduction in this section. First, the vibration signal of each state of the check valve is divided into 60 non-overlapping samples of length 1280. Then, the time-domain, frequency-domain characteristics of each original bearing vibration signal are extracted. Secondly, in order to construct the time-frequency domain characteristics of the vibration signal, the wavelet packet energy entropy, packet energy coefficient and Gini coefficient are extracted from the IMF obtained by NEEEMD decomposition. Finally, the extracted time-domain, frequency-domain, and time-frequency domain features are constructed into a fault feature vector set, and the obtained high-dimensional feature matrix size is 180·53, and the maximum dimension D=53. The distribution of characteristic samples under each dimension D is shown in Figure 19 (a) and (b).

In Figure 19 (a), the mixed characteristics of different fault types differ in dimension D, but some features of the same fault type deviate from most of the characteristics of this type, that is, there are outliers. As shown in Figure 19 (b), when the dimension D = 18, the mixed characteristics of the three fault states of the bearing have a certain degree of separability, but the edges of the mixed characteristics of the two fault states below are blurred. When the dimension D=28, samples of the same fault type are clustered together, and the sample boundaries of different fault types are clear. However, the result of the division is 2 categories, and the true number of fault types is 3. When dimension D = 35, the mixed features of various fault states overlap and are clustered below. When dimension D=43, the mixed characteristics of various fault states are chaotic and scattered. Therefore, the mixed features obtained by the original extraction are difficult to find the dividing line for dividing sample categories under most dimensions, resulting in low sample separability. Therefore, this paper proposes RUSLP to sort and select hybrid features, filter out features with small correlation to reduce feature redundancy, and improve fault diagnosis reliability.

3) FAULT FEATURE SELECTION

In order to verify the influence of RUSLP dimensionality reduction method on the diagnostic performance of bearing faults, we compare the accuracy of RUSLP dimensionality reduction method with LS, MCFS and UDFS methods under the same mixed set of high-dimensional fault feature vectors and the same parameters. The parameter settings of the BT-LSTSVM-based multi-classifier are the same as above. The results are shown in Figure 20.

For the model constructed in this paper, except for LS, the fault diagnosis accuracy of other dimensionality reduction

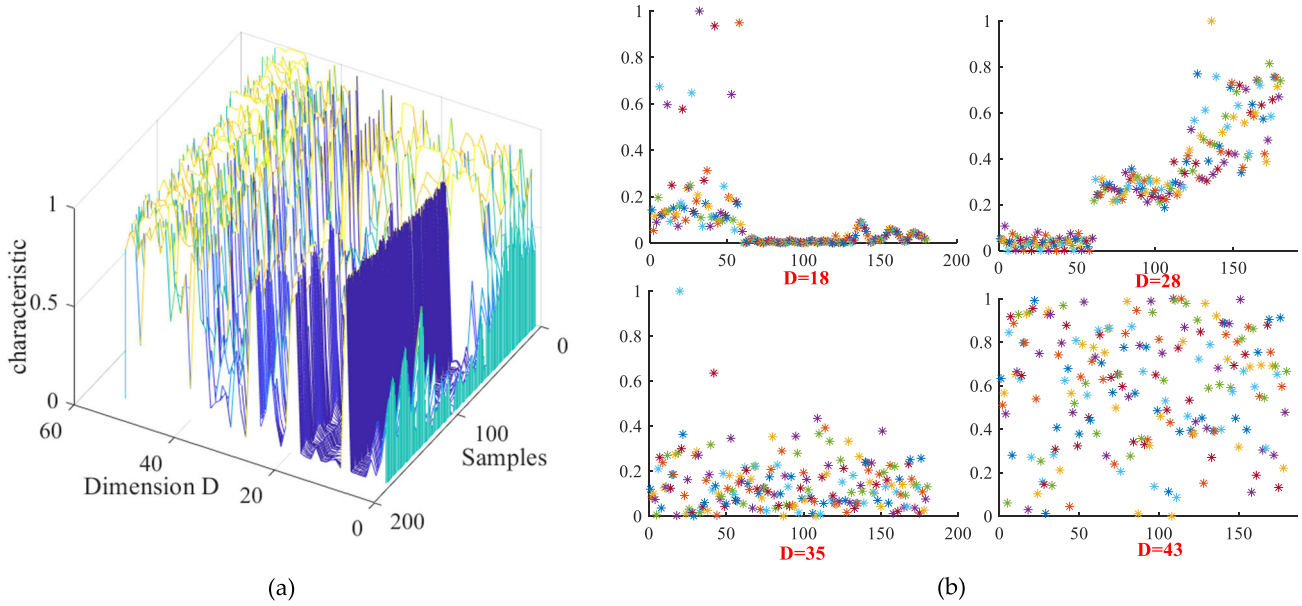


FIGURE 19. Characteristic sample distribution of check valves under different dimensions D.

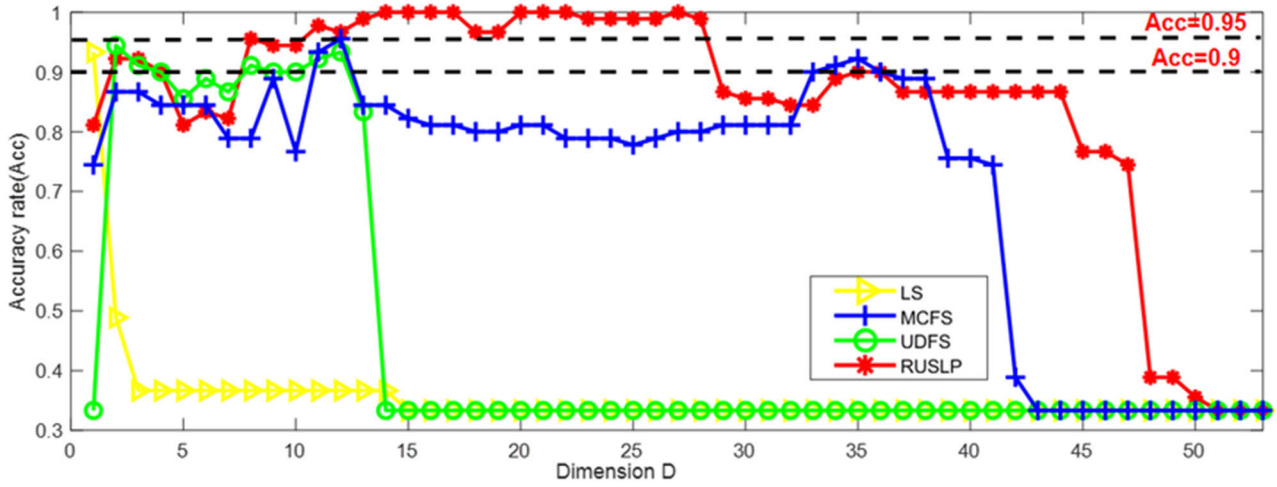


FIGURE 20. Fault diagnosis accuracy under different dimensionality reduction methods.

TABLE 12. The highest accuracy and corresponding feature dimension x obtained by different dimensionality reduction methods.

Reduced Dimension Model	LS	MCFS	UDFS	RUSLP
Front X-dimensional features	1	12	2	14
Highest precision	0.9333	0.9556	0.9501	1

TABLE 13. The highest accuracy and corresponding feature dimension x obtained by different classification methods.

Classification Model	SVM	BPNN	ELM	BTLSTSVM
Front X-dimensional features	3	4	4	31
Highest precision	0.7778	0.9778	0.9778	1

methods in the interval of $2 \leq x \leq 15$ can reach 95%. Table 12 shows the highest accuracy obtained by each dimensionality reduction method in Figure 20, and the corresponding feature dimension x. When $x = 1$, the fault

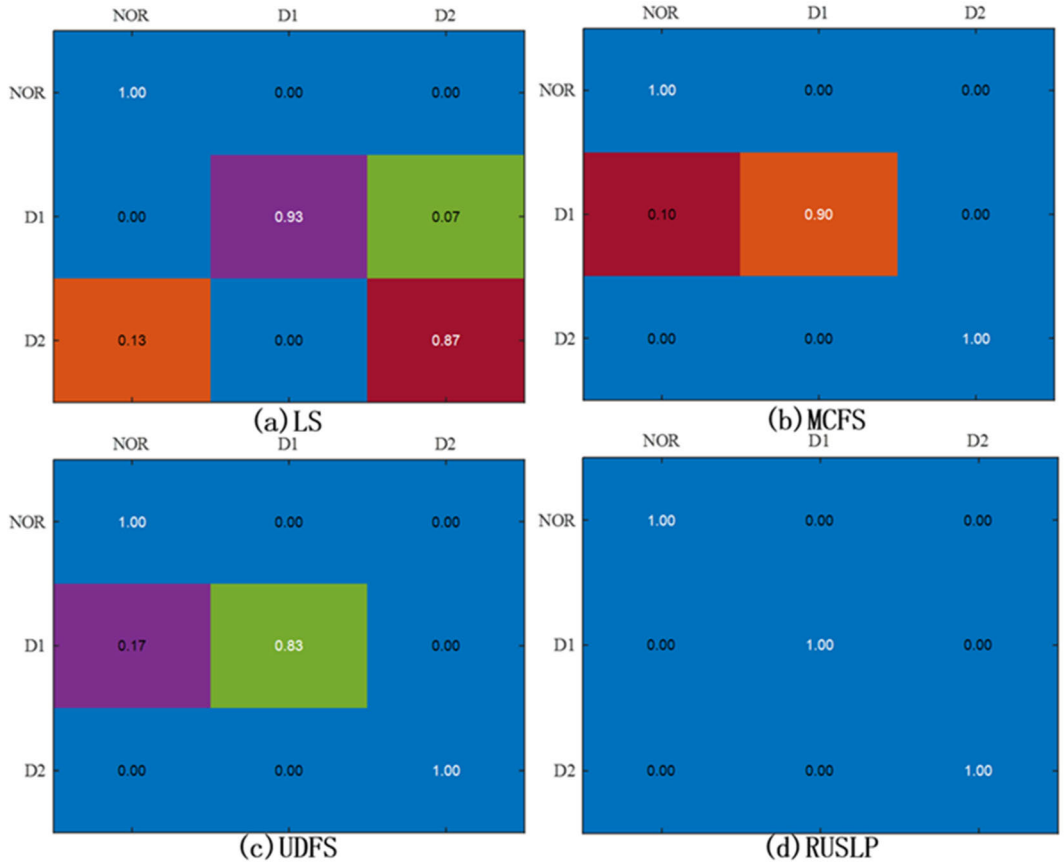


FIGURE 21. Confusion matrix obtained by selecting the highest precision dimension with different classification methods.

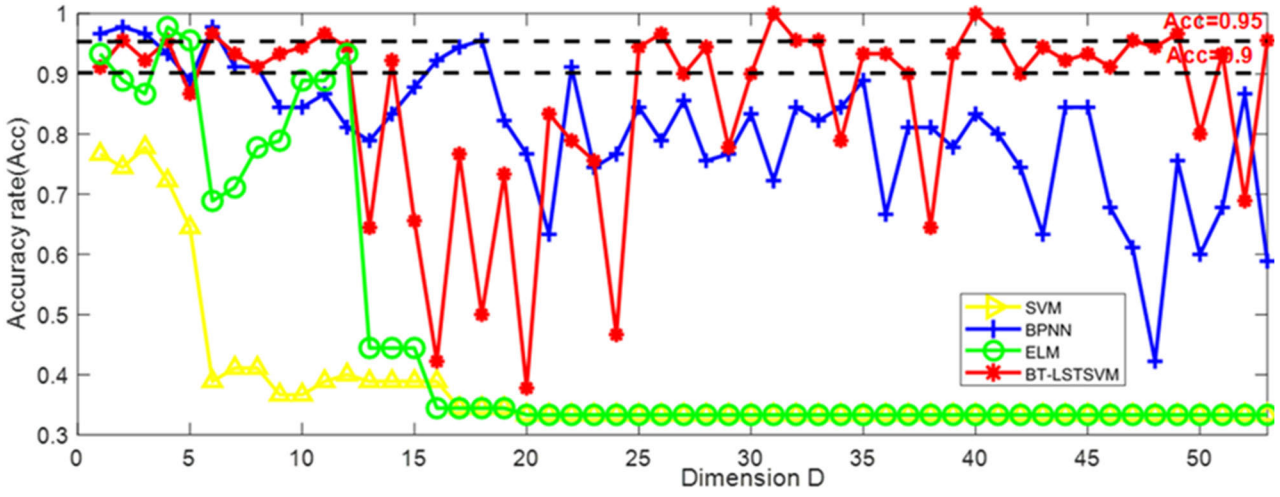


FIGURE 22. Fault diagnosis accuracy under different classification methods.

diagnosis accuracy of LS dimension reduction method is maintained near 93.33%. With the increase of feature dimension, the fault diagnosis accuracy of LS dimension reduction method drops sharply to 36.67%, indicating that LS dimension reduction method is not suitable for the model constructed in this paper. However, when $x = 12$, the fault diagnosis accuracy of 95.56% can be obtained by using MCFS dimensionality reduction method. When $x = 2$, the fault diagnosis accuracy of 95.01% can be obtained by using UDFS dimensionality reduction method. When $x = 14$, 100% accuracy of fault diagnosis can be obtained by using RUSLP dimensionality reduction method. Therefore, for the

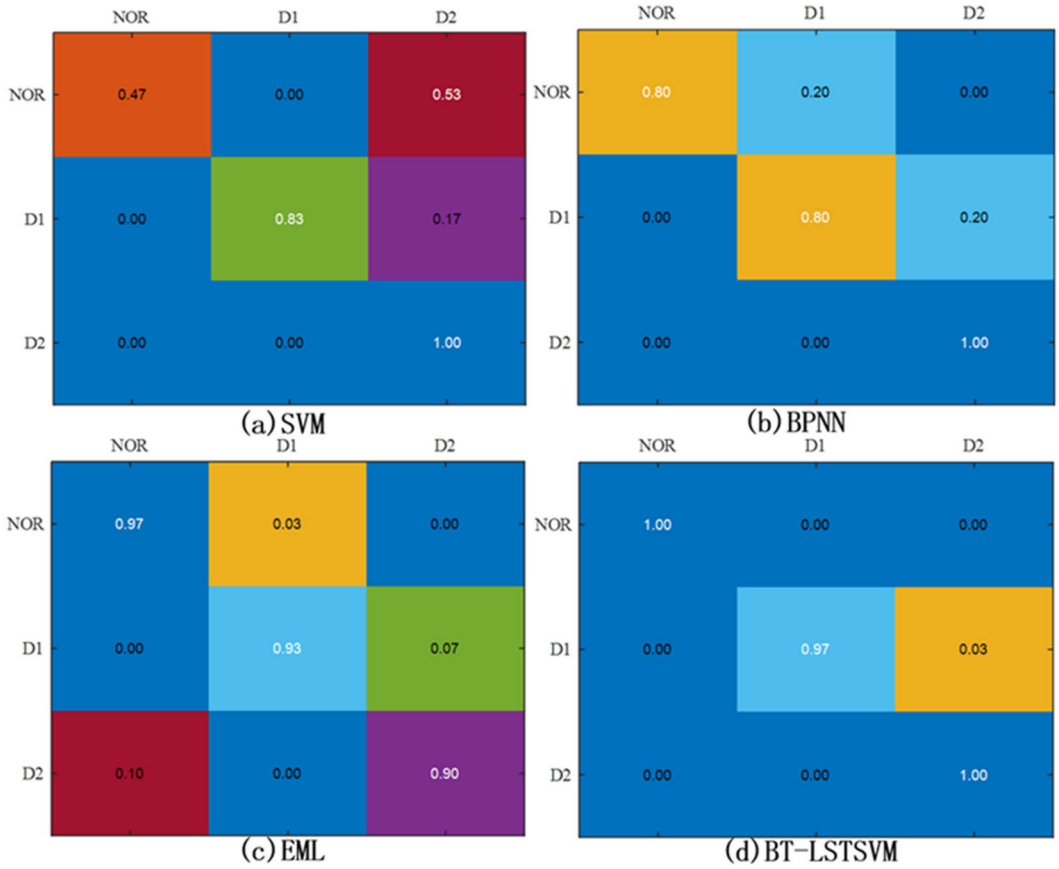


FIGURE 23. Confusion matrix obtained by selecting the highest precision dimension with different classification methods.

model constructed in this paper, except LS dimensionality reduction method, other dimensionality reduction methods have obtained high fault diagnosis accuracy. However, the RUSLP method can achieve 100% accuracy of fault diagnosis. The superiority of the dimensionality reduction method of RUSLP used in this paper can be proved again by the data of check valve in actual working conditions.

When LS, MCFS, UDFS and RUSLP methods extract the features of the first 1 dimension, the first 12 dimension, the first 2 dimension and the first 14 dimension respectively, the confusion matrix obtained is shown in Figure 21. If LS dimensionality reduction is used, 7% of the card valve fault samples will be incorrectly diagnosed as wear faults, and 13% of wear fault samples will be incorrectly diagnosed as normal state. If the MCFS dimensionality reduction method is used, 10% of the sample of the valve failure will be incorrectly diagnosed as normal. If UDFS dimensionality reduction is used, 17% of the sample of valve faults is incorrectly diagnosed as normal. If the RUSLP dimensionality reduction method is used, 100% of the samples can be identified by classification. Experimental results show that the RUSLP dimensionality reduction method can effectively solve the problem of feature confusion and the fault diagnosis accuracy is improved.

4) FAULT IDENTIFICATION

In order to verify the influence of the BT-LSTSVM multi-classifier method on the diagnostic performance of bearing faults, we compare the accuracy of the BT-LSTSVM multi-classifier with the SVM, BPNN and ELM methods under the same mixed high-dimensional fault feature vector set and the same parameters. The parameter settings based on the RUSLP method are the same as above. The results are shown in Figure 22.

It can be seen that except for SVM, other classification methods can achieve 95% accuracy in the interval of $2 \leq x \leq 32$. Table 13 shows the highest accuracy obtained by each classification method in Figure 22, and the corresponding feature dimension x. SVM extracts the first 3 dimensional features, and the accuracy can be obtained 77.78%. BPNN extracts the first 4 dimensional features and obtains an accuracy of 97.78%. ELM extracts the first 4dimensional features and obtains an accuracy of 97.78%. BT-LSTSVM extracts the first 31 dimensional features with 100% accuracy. Therefore, under the BT-LSTSVM method, the fault diagnosis accuracy is improved by 22%.

When the SVM, BPNN, EML and BT-LSTSVM respectively extract the first 3, 4, 4 and 31 dimensions of features, the confusion matrix is shown in Figure 23. If the SVM

identification model is used, 53% of normal state samples will be incorrectly diagnosed as wear faults, and 17% of card valve faults will be incorrectly diagnosed as wear faults. If the BPNN identification model is used, 20% of the normal state samples will be incorrectly diagnosed as a valve fault, and 20% of the sample will be incorrectly diagnosed as a wear fault. If the ELM identification model is used, 3% of normal state samples will be incorrectly diagnosed as stuck valve faults, 7% of card valve faults will be incorrectly diagnosed as wear faults, and 10% of wear fault samples will be incorrectly diagnosed as normal states. If the BT-LSTSVM multi-classifier is used, only 3% of the card valve failure samples were incorrectly diagnosed as wear fault conditions. Experimental results show that the BT-LSTSVM multi-classifier can effectively improve the fault diagnosis accuracy of bearings.

V. CONCLUSION

Because the bearing vibration signal contains different frequency interference data, the traditional method can not get satisfactory results. Therefore, in order to solve the defects of previous methods in fault diagnosis, this paper proposes a new fault diagnosis method based on NEEEMD-RUSLP feature selection and BTLSTSVM. Experimental analysis of feature extraction, feature selection, and classification identification was carried out through Case Western Reserve University data, University of Ottawa data, and check valve data under actual industrial production environment. the effectiveness and superiority of this method are proved.

1) In the simulation signal experiment, the IMF obtained by NEEEMD is least affected by noise and contains the most obvious fault impact. The results indicate that NEEEMD can suppress white noise in vibration signals by subtracting the intermediate frequency of the same white noise from the final intermediate frequency, and NEEEMD has outstanding advantages in signal decomposition accuracy, decomposition efficiency and noise suppression, and is suitable for fault feature extraction in the background of strong noise.
2) In fault feature extraction, most of the dimensionality features obtained by the proposed method have good separability and fewer feature aliasing phenomena. The results show that wavelet packet decomposition can suppress some noise during feature extraction of wavelet packet energy entropy and packet energy coefficient based on NEEEMD, and the Gini coefficient can extract the periodic impact characteristics contained in IMF, and the constructed high-dimensional features have superior performance. In feature selection, the diagnostic accuracy obtained by the first x-dimensional low-dimensional features of RUSLP is higher than that obtained by other feature selection methods. The results show that the RUSLP feature selection strategy can select more relevant fault features by enhancing the feature importance information in the projection matrix.
3) In the fault identification experiment, the proposed BTLSTSVM can obtain the best diagnostic accuracy in a suitable dimension, that is, close to 100%, and the feature dimension required by BTLSTSVM is relatively small. The results indicate that the proposed multi classification LSTSVM based on binary tree (BT) not only inherits the advantages of fewer classifiers and higher efficiency of BT, but also has the advantage of high recognition accuracy of LSTSVM.

After improving the method, the bearing fault diagnosis accuracy in Case Western Storage data increased by 10.67%, and the bearing fault diagnosis accuracy in Ottawa data increased by 10%. The accuracy of fault diagnosis in check valve data is improved by 22%. The experimental results show that the proposed method not only has outstanding advantages in the fault diagnosis of rotating machinery equipment in the laboratory environment, but also has good performance in the fault diagnosis of reciprocating machinery in the industrial production environment. Therefore, the method proposed in this paper is expected to be used for mechanical bearing fault diagnosis under actual production conditions. However, the proposed fault diagnosis method inevitably has some shortcomings. Although the proposed feature extraction strategy has noise reduction performance and impact feature extraction performance, the configuration of high-dimensional features is complex, which relies on manual experience and the obtained fault features do not have the ability to transfer and diagnose under different working conditions. Therefore, a deep feature extraction method with transfer capability will be the focus of the next step of research. Although the proposed BTLSTSVM can effectively identify multiple fault modes and the recognition results are not sensitive to hyper parameters, the recognition errors of the upper nodes in BTLSTSVM will be transmitted to the lower nodes and have a significant impact on the recognition accuracy of the lower nodes, which means there is an error accumulation problem. The more categories, the more serious the problem of error accumulation. In addition, the hyper parameters in BTLSTSVM can also affect recognition accuracy and performance to a certain extent. Therefore, solving the problem of error accumulation in BTLSTSVM and implementing an adaptive BTLSTSVM model is the next research focus.

AUTHOR CONTRIBUTIONS

Conceptualization: Chengjiang Zhou; Methodology: Rongrong Lu and Miao Xu; Software: Rongrong Lu and Miao Xu; Validation: Zhaodong Zhang, Qihua Yang, and Shanyou He; Formal Analysis: Chengjiang Zhou; Investigation: Chengjiang Zhou; Resources: Min Mao and Jingzong Yang; Data Curation: Rongrong Lu, Shanyou He, and Zhaodong Zhang; Writing—Original Draft Preparation: Qihua Yang and Miao Xu; Writing—Review and Editing:

Chengjiang Zhou; Visualization: Chengjiang Zhou; Supervision: Min Mao and Jingzong Yang; Project Administration: Min Mao; Funding Acquisition: Jingzong Yang. All authors have read and agreed to the published version of the manuscript.

INSTITUTIONAL REVIEW BOARD STATEMENT

Not applicable.

INFORMED CONSENT STATEMENT

Informed consent was obtained from all subjects involved in the study.

DATA AVAILABILITY STATEMENT

The data used to support the findings of this study are available from the corresponding author upon request.

ACKNOWLEDGMENT

The authors would like to thank the team for their guidance and also thank the Case West Reserve University and the University of Ottawa for their bearing datasets. They also thank Yunnan Dahongshan Pipeline Company Ltd., for providing check valve vibration signal data and platform equipment. They also thank the reviewers for taking the time to review the paper in a busy schedule.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] H. Tao, J. Qiu, Y. Chen, V. Stojanovic, and L. Cheng, "Unsupervised cross-domain rolling bearing fault diagnosis based on time-frequency information fusion," *J. Franklin Inst.*, vol. 360, no. 2, pp. 1454–1477, Jan. 2023.

[2] Y. Sun, S. Li, and X. Wang, "Bearing fault diagnosis based on EMD and improved Chebyshev distance in SDP image," *Measurement*, vol. 176, May 2021, Art. no. 109100.

[3] M. Motahari-Nezhad and S. M. Jafari, "Bearing remaining useful life prediction under starved lubricating condition using time domain acoustic emission signal processing," *Expert Syst. Appl.*, vol. 168, Apr. 2021, Art. no. 114391.

[4] J. P. Kumar, P. S. Chauhan, and P. P. Pandit, "Time domain vibration analysis techniques for condition monitoring of rolling element bearing: A review," *Mater. Today, Proc.*, vol. 62, pp. 6336–6340, Jan. 2022.

[5] H. Wu, J. Li, Q. Zhang, J. Tao, and Z. Meng, "Intelligent fault diagnosis of rolling bearings under varying operating conditions based on domain-adversarial neural network and attention mechanism," *ISA Trans.*, vol. 130, pp. 477–489, Nov. 2022.

[6] H. Li, X. Wu, T. Liu, S. Li, B. Zhang, G. Zhou, and T. Huang, "Composite fault diagnosis for rolling bearing based on parameter-optimized VMD," *Measurement*, vol. 201, Sep. 2022, Art. no. 111637.

[7] B. Wang, Y. Guo, Z. Zhang, D. Wang, J. Wang, and Y. Zhang, "Developing and applying OEGOA-VMD algorithm for feature extraction for early fault detection in cryogenic rolling bearing," *Measurement*, vol. 216, Jul. 2023, Art. no. 112908.

[8] Z. Tang, M. Wang, T. Ouyang, and F. Che, "A wind turbine bearing fault diagnosis method based on fused depth features in time–frequency domain," *Energy Rep.*, vol. 8, pp. 12727–12739, Nov. 2022.

[9] J. Chen and H. Liu, "A multi-gradient hierarchical domain adaptation network for transfer diagnosis of bearing faults," *Expert Syst. Appl.*, vol. 225, Sep. 2023, Art. no. 120139.

[10] Z. Jiang, K. Zhang, L. Xiang, G. Yu, and Y. Xu, "A time-frequency spectral amplitude modulation method and its applications in rolling bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 185, Feb. 2023, Art. no. 109832.

[11] G. Vashishtha and R. Kumar, "An amended grey wolf optimization with mutation strategy to diagnose bucket defects in Pelton wheel," *Measurement*, vol. 187, Jan. 2022, Art. no. 110272.

[12] Z. Lao, D. He, Z. Wei, H. Shang, Z. Jin, J. Miao, and C. Ren, "Intelligent fault diagnosis for rail transit switch machine based on adaptive feature selection and improved LightGBM," *Eng. Failure Anal.*, vol. 148, Jun. 2023, Art. no. 107219.

[13] D. He, C. Liu, Z. Jin, R. Ma, Y. Chen, and S. Shan, "Fault diagnosis of flywheel bearing based on parameter optimization variational mode decomposition energy entropy and deep learning," *Energy*, vol. 239, Jan. 2022, Art. no. 122108.

[14] M. S. R. M. Saufi and K. A. Hassan, "Remaining useful life prediction using an integrated Laplacian-LSTM network on machinery components," *Appl. Soft Comput.*, vol. 112, Nov. 2021, Art. no. 107817.

[15] J. Zheng, H. Pan, J. Tong, and Q. Liu, "Generalized refined composite multiscale fuzzy entropy and multi-cluster feature selection based intelligent fault diagnosis of rolling bearing," *ISA Trans.*, vol. 123, pp. 136–151, Apr. 2022.

[16] A. Hashemi, M. B. Dowlatshahi, and H. Nezamabadi-Pour, "An efficient Pareto-based feature selection algorithm for multi-label classification," *Inf. Sci.*, vol. 581, pp. 428–447, Dec. 2021.

[17] S. Ma, G. Cheng, Y. Li, and R. Zhao, "Dimension reduction method of high-dimensional fault datasets based on C_M_t-SNE under unsupervised background," *Measurement*, vol. 214, Jun. 2023, Art. no. 112835.

[18] A. Wahid, D. M. Khan, I. Hussain, S. A. Khan, and Z. Khan, "Unsupervised feature selection with robust data reconstruction (UFS-RDR) and outlier detection," *Expert Syst. Appl.*, vol. 201, Sep. 2022, Art. no. 117008.

[19] Y. Guo, H. Sun, and S. Hao, "Adaptive dictionary and structure learning for unsupervised feature selection," *Inf. Process. Manage.*, vol. 59, no. 3, May 2022, Art. no. 102931.

[20] C. Zeng, H. Chen, T. Li, and J. Wan, "Robust unsupervised feature selection via sparse and minimum-redundant subspace learning with dual regularization," *Neurocomputing*, vol. 511, pp. 1–21, Oct. 2022.

[21] P. Zhu, X. Hou, K. Tang, Y. Liu, Y.-P. Zhao, and Z. Wang, "Unsupervised feature selection through combining graph learning and $l_{2,0}$-norm constraint," *Inf. Sci.*, vol. 622, pp. 68–82, Apr. 2023.

[22] H. R. F. Fotso, C. V. A. Kazé, and G. D. Kenmoé, "Real-time rolling bearing power loss in wind turbine gearbox modeling and prediction based on calculations and artificial neural network," *Tribol. Int.*, vol. 163, Nov. 2021, Art. no. 107171.

[23] M. Ghorvei, M. Kavianpour, M. T. H. Beheshti, and A. Ramezani, "Spatial graph convolutional neural network via structured subdomain adaptation and domain adversarial learning for bearing fault diagnosis," *Neurocomputing*, vol. 517, pp. 44–61, Jan. 2023.

[24] J. Wang, T. Li, C. Sun, R. Yan, and X. Chen, "Improved spiking neural network for intershaft bearing fault diagnosis," *J. Manuf. Syst.*, vol. 65, pp. 208–219, Oct. 2022.

[25] S. Xie, H. Tan, Y. Li, Z. Feng, and Z. Cao, "Locally generalized preserving projection and flexible grey wolf optimizer-based ELM for fault diagnosis of rolling bearing," *Measurement*, vol. 202, Oct. 2022, Art. no. 111828.

[26] J. Gong, X. Yang, H. Wang, J. Shen, W. Liu, and F. Zhou, "Coordinated method fusing improved bubble entropy and artificial gorilla troops optimizer optimized KELM for rolling bearing fault diagnosis," *Appl. Acoust.*, vol. 195, Jun. 2022, Art. no. 108844.

[27] O. Yaman, F. Yol, and A. Altinors, "A fault detection method based on embedded feature extraction and SVM classification for UAV motors," *Microprocessors Microsyst.*, vol. 94, Oct. 2022, Art. no. 104683.

[28] S. Gao, S. Zhang, Y. Zhang, and Y. Gao, "Operational reliability evaluation and prediction of rolling bearing based on isometric mapping and NoCuSa-LSSVM," *Rel. Eng. Syst. Saf.*, vol. 201, Sep. 2020, Art. no. 106968.

[29] C. He, T. Wu, C. Liu, and T. Chen, "A novel method of composite multiscale weighted permutation entropy and machine learning for fault complex system fault diagnosis," *Measurement*, vol. 158, Jul. 2020, Art. no. 107748.

[30] J. Wei, H. Huang, L. Yao, Y. Hu, Q. Fan, and D. Huang, "New imbalanced fault diagnosis framework based on cluster-MWMOTE and MFO-optimized LS-SVM using limited and complex bearing data," *Eng. Appl. Artif. Intell.*, vol. 96, Nov. 2020, Art. no. 103966.

[31] H. S. Dhiman, D. Deb, S. M. Muyeen, and I. Kamwa, "Wind turbine gearbox anomaly detection based on adaptive threshold and twin support vector machines," *IEEE Trans. Energy Convers.*, vol. 36, no. 4, pp. 3462–3469, Dec. 2021.

[32] X. Bai, T. Tao, L. Gao, C. Tao, and Y. Liu, "Wind turbine blade icing diagnosis using RFECV-TSVM pseudo-sample processing," *Renew. Energy*, vol. 211, pp. 412–419, Jul. 2023.
[33] F. Yuan, J. Guo, Z. Xiao, B. Zeng, W. Zhu, and S. Huang, "A transformer fault diagnosis model based on chemical reaction optimization and twin support vector machine," *Energies*, vol. 12, no. 5, p. 960, Mar. 2019, doi: 10.3390/en12050960.
[34] M. Arun Kumar and M. Gopal, "Least squares twin support vector machines for pattern classification," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7535–7543, May 2009.
[35] M. A. Ganaie and M. Tanveer, "LSTSVM classifier with enhanced features from pre-trained functional link network," *Appl. Soft Comput.*, vol. 93, Aug. 2020, Art. no. 106305.
[36] J. Ali, M. Aldhaifallah, K. S. Nisar, A. A. Aljabr, and M. Tanveer, "Regularized least squares twin SVM for multiclass classification," *Big Data Res.*, vol. 27, Feb. 2022, Art. no. 100295.
[37] C. Zhou, Y. Jia, S. Zhao, Q. Yang, Y. Liu, Z. Zhang, and T. Wang, "A mechanical part fault diagnosis method based on improved multiscale weighted permutation entropy and multiclass LSTSVM," *Measurement*, vol. 214, Jun. 2023, Art. no. 112671.
[38] W.-J. Chen, Y.-H. Shao, N.-Y. Deng, and Z.-L. Feng, "Laplacian least squares twin support vector machine for semi-supervised classification," *Neurocomputing*, vol. 145, pp. 465–476, Dec. 2014.
[39] C. Yuan and L. Yang, "Capped $L_{2,p}$-norm metric based robust least squares twin support vector machine for pattern classification," *Neural Netw.*, vol. 142, pp. 457–478, Oct. 2021.
[40] G. Yu, J. Ma, and C. Xie, "Hessian scatter regularized twin support vector machine for semi-supervised classification," *Eng. Appl. Artif. Intell.*, vol. 119, Mar. 2023, Art. no. 105751.
[41] A. Othman, N. Iqbal, S. M. Hanafy, and U. B. Waheed, "Automated event detection and denoising method for passive seismic data using residual deep convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5900711.
[42] N. Iqbal, "DeepSeg: Deep segmental denoising neural network for seismic data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 7, pp. 3397–3404, Jul. 2023.
[43] Z. Wei, D. He, Z. Jin, B. Liu, S. Shan, Y. Chen, and J. Miao, "Density-based affinity propagation tensor clustering for intelligent fault diagnosis of train bogie bearing," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 6, pp. 6053–6064, Jun. 2023.
[44] Z. Jin, D. He, R. Ma, X. Zou, Y. Chen, and S. Shan, "Fault diagnosis of train rotating parts based on multi-objective VMD optimization and ensemble learning," *Digit. Signal Process.*, vol. 121, Mar. 2022, Art. no. 103312.
[45] X. Yu, Z. Liang, Y. Wang, H. Yin, X. Liu, W. Yu, and Y. Huang, "A wavelet packet transform-based deep feature transfer learning method for bearing fault diagnosis under different working conditions," *Measurement*, vol. 201, Sep. 2022, Art. no. 111597.
[46] H. Wen, W. Guo, and X. Li, "A novel deep clustering network using multi-representation autoencoder and adversarial learning for large cross-domain fault diagnosis of rolling bearings," *Expert Syst. Appl.*, vol. 225, Sep. 2023, Art. no. 120066.
[47] A. Faysal, W. K. Ngui, M. H. Lim, and M. S. Leong, "Noise eliminated ensemble empirical mode decomposition scalogram analysis for rotating machinery fault diagnosis," *Sensors*, vol. 21, no. 23, p. 8114, Dec. 2021.
[48] A. Faysal, W. K. Ngui, and M. H. Lim, "Noise eliminated ensemble empirical mode decomposition for bearing fault diagnosis," *J. Vib. Eng. Technol.*, vol. 9, pp. 2229–2245, Aug. 2021.
[49] A. Borré, L. O. Seman, E. Camponogara, S. F. Stefenon, V. C. Mariani, and L. D. S. Coelho, "Machine fault detection using a hybrid CNN-LSTM attention-based model," *Sensors*, vol. 23, no. 9, p. 4512, May 2023.
[50] X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, Nov. 2018.
[51] C. Luo, J. Zheng, T. Li, H. Chen, Y. Huang, and X. Peng, "Orthogonally constrained matrix factorization for robust unsupervised feature selection with local preserving," *Inf. Sci.*, vol. 586, pp. 662–675, Mar. 2022.
[52] C. Zhou, H. Li, J. Yang, Q. Yang, L. Yang, S. He, and X. Yuan, "Fuzzy regular least squares twin support vector machine and its application in fault diagnosis," *Expert Syst. Appl.*, vol. 231, Nov. 2023, Art. no. 120804.
[53] H. Huang and N. Baddour, "Bearing vibration data collected under time-varying rotational speed conditions," *Data Brief*, vol. 21, pp. 1745–1749, Dec. 2018.

RONGRONG LU is currently pursuing the bachelor's degree in data science and big data technology with Yunnan Normal University, China.

Her current research interests include signal processing and fault diagnosis research, mechanical equipment fault detection, and condition monitoring of electromechanical equipment.

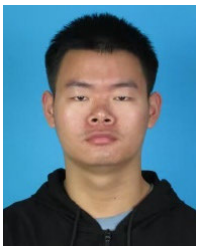
Ms. Lu's awards and honors include the National Encouragement Scholarship and the Professional Quality Award from Yunnan Normal University.

MIAO XU received the bachelor's degree in computer science and technology from Yunnan Normal University, China, in 2023.

His current research interests include signal processing and fault diagnosis research, mechanical equipment fault detection, and condition monitoring of electromechanical equipment.

CHENGJIANG ZHOU received the bachelor's degree in automation from the North China University of Science and Technology, in 2014, and the master's degree in control theory and control engineering and the Ph.D. degree in metallurgical control engineering from the Kunming University of Science and Technology, in 2017 and 2020, respectively.

He is currently a Teacher with the Information School, Yunnan Normal University, China. He has published more than 30 academic articles, including more than 20 SCI articles. His current research interests include signal processing and fault diagnosis research, mechanical equipment fault detection, mechanical equipment fault diagnosis, and mechanical fault diagnosis based on image processing. His research results can be found in the academic homepage at https://www.researchgate.net/profile/Chengjiang-Zhou/research.



ZHAODONG ZHANG is currently pursuing the bachelor's degree in computer science and technology (teacher education) with Yunnan Normal University, China.

His current research interests include signal processing and fault diagnosis research, mechanical equipment fault detection, and condition monitoring of electromechanical equipment.

Mr. Zhang's awards and honors include the University-Level Second Prize of the National English Competition for College Students, in November 2022 and June 2023, and the Practical Ability Award from Yunnan Normal University.

SHANYOU HE is currently pursuing the bachelor's degree in computer science and technology with Yunnan Normal University, China.

Her current research interests include signal processing and fault diagnosis research, mechanical equipment fault detection, and condition monitoring of electromechanical equipment.

Ms. He's honors include the Excellence Award in the Preliminary Round of the Fourth Huajiao Cup National College Student Mathematics Competition, in 2021.

QIHUA YANG is currently pursuing the bachelor's degree in data science and big data technology with Yunnan Normal University, China.

Her current research interests include signal processing and fault diagnosis research, mechanical equipment fault detection, and condition monitoring of electromechanical equipment.

Ms. Yang's honors include the Merit Student of Yunnan Normal University of China, in June 2021, the Merit Student of Yunnan Province of China, in June 2022, and the Provincial First Prize of China Mathematical Modeling Competition, in December 2022.

MIN MAO received the bachelor's degree in electrical engineering and automation from Zhejiang Wanli University, China, in 2013, and the master's degree in instrument science and technology from the Kunming University of Science and Technology, China, in 2017.

He is currently a Teacher with the Faculty of Information Engineering, Quzhou College of Technology, China. He has published more than ten academic articles, including more than two SCI articles. His current research interests include dynamic signal processing, intelligent fault diagnosis and prediction, and mechanical and electrical equipment status monitoring. His research results can be found in the academic homepage at https://orcid.org/0000-0002-0209-4077.

JINGZONG YANG received the bachelor's degree in electronic information science and technology from Yunnan Normal University, China, in 2013, and the Ph.D. degree in metallurgical control engineering from the Kunming University of Science and Technology, China, in 2018.

He is currently a Teacher with the College of Big Data, Baoshan University, China. He has published more than 40 academic articles, including more than 20 SCI and EI retrieval articles. His current research interests include mechanical equipment fault detection and mechanical equipment fault diagnosis. His research results can be found in the academic homepage at https://www.webofscience.com/wos/author/record/GNP-4449-2022.