coatings

Article

Fault Diagnosis of Check Valve Based on KPLS Optimal Feature Selection and Kernel Extreme Learning Machine

Xuyi Yuan [1], Yugang Fan [1,*], Chengjiang Zhou [2,*], Xiaodong Wang [1] and Guanghui Zhang [1]

[1] Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China
[2] School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China
* Correspondence: ygfan@kust.edu.cn (Y.F.); chengjiangzhou@foxmail.com (C.Z.)

Abstract: The check valve is the core part of high-pressure diaphragm pumps. It has complex operation conditions and has difficulty characterizing fault states completely with its single feature. Therefore, a fault signal diagnosis model based on the kernel extreme learning machine (KELM) was constructed to diagnose the check valve. The model adopts a multi-feature extraction method and reduces dimensionality through kernel partial least squares (KPLS). Firstly, we divided the check valve vibration signal into several non-overlapping samples. Then, we extracted 16 time-domain features, 13 frequency-domain features, 16 wavelet packet energy features, and energy entropy features from each sample to construct a multi-feature set characterizing the operation state of the check valve. Next, we used the KPLS method to optimize the 45 dimension multi-feature data and employed the processed feature set to establish a KELM fault diagnosis model. Experiments showed that the method based on KPLS optimal feature selection could fully characterize the operating state of the equipment with an accuracy rate of 96.88%. This result indicates the high accuracy and effectiveness of the multi-feature set constructed with the KELM fault diagnosis model.

Keywords: KPLS; KELM; fault diagnosis; check valve

Citation: Yuan, X.; Fan, Y.; Zhou, C.; Wang, X.; Zhang, G. Fault Diagnosis of Check Valve Based on KPLS Optimal Feature Selection and Kernel Extreme Learning Machine. *Coatings* **2022**, *12*, 1320. https://doi.org/10.3390/coatings12091320

Academic Editors: Ke Feng and Luis Norberto López De Lacalle

Received: 23 July 2022
Accepted: 8 September 2022
Published: 10 September 2022

1. Introduction

As the core piece of equipment in ore transportation pipelines, the check valve directly affects the operation of pipeline systems through its operation status. Research on the fault diagnosis method for check valves is of great significance for the development of the pipeline transportation industry. The operating conditions of check valves are complex, with the vibration signal being a periodic pulse signal affected by environmental noise and other factors. When failure occurs, signal characteristics experience interference and are challenging to extract. Thus, a single feature of a check valve cannot fully characterize the operating state of the equipment.

Characterizing operating states extracting signal features is the basis for fault diagnosis of mechanical equipment. Chen et al. [1] used a continuous wavelet transform (CWT) to preprocess an original vibration signal and constructed a fused convolutional neural network (CNN) with a square pool structure to extract signal features and to realize fault diagnosis of mechanical equipment. Peng et al. [2] proposed a fault classification method based on multi-feature extraction and an improved Mahalanobis–Taguchi System (MTS). The method involves extracting multi-dimensional features using complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) for time, frequency, and adaptive white noise. The authors constructed a DAG-MTS multi-classification model based on the characteristics of the MTS system and a directed acyclic graph (DAG) and applied it to a bearing's fault diagnosis. To solve the problem of noise in a diesel engine's vibration signal and address the difficulty of feature extraction, Jiang et al. [3] proposed a diesel engine fault diagnosis method focusing on the extraction of the wavelet packet energy spectrum and the selection of fuzzy entropy features. The fuzzy entropy selects components out of the feature set extracted from the wavelet packet energy spectrum and inputs the selected features into the least squares twin support vector machine (LSSVM) for fault diagnosis.

Constructing a feature set that characterizes the operating state of equipment and using it to establish a fault diagnosis model is the key to the fault diagnosis of mechanical equipment. Ding et al. [4] proposed a method for scintillation detector fault diagnosis based on the extreme learning machine (ELM), and this method could not only classify the faults of the failed detector but also intelligently determine the severity of various faults. Lee et al. [5] proposed a novel remaining useful life (RUL) estimation method based on systematic feature engineering and the extreme learning machine (ELM) for seven out of eleven bearings; the proposed method reduced the mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) in the RUL estimation by over 50%. Pang et al. [6] proposed an ensembled kernel extreme learning machine that fuses multi-domain features. It selects the most suitable stacked noise reduction autoencoder method through time-domain and frequency-domain feature extraction. It uses a kernel extreme learning machine with deep features for rotating machinery fault diagnosis. Wang et al. [7] proposed an ensemble extreme learning machine (EELM) that consists of two heterogeneous ELM networks. First, it displays the target data using a clustering algorithm. Then, it applies Gaussian-style activation between each target as an input to the back-end classifier to propose a non-empirically specified threshold based on the EELM multi-label classifier. Later, the multiple binary classifiers are combined for composite fault diagnosis. Zhang et al. [8] introduced an online fault diagnosis method that changes the fixed structure of the extreme learning machine into an elastic structure using incremental support vector data description (ISVDD) and an extreme learning machine with an incremental output structure (IOELM). The ISVDD is used to detect a new failure mode, while the IOELM is used to recognize the specific failure mode. Harishvijey et al. [9] proposed an automatic signal classification method for detecting seizures from an EEG signal using an empirical wavelet transform (EWT) feature extraction method, K-principal component analysis (K-PCA)-based feature reduction, and a fuzzy logic-embedded RBF kernel-based ELM. Shen et al. [10] proposed a feature selection and fusion method based on the poll mode and optimized WKPCA. Considering the variation in fault information collected by different sensors, the diagnosis rate in the extreme learning machine (ELM) is taken as the index for the evaluation of each single sensor, and then the sensitivity weight matrix of the features extracted by multiple sensors is constructed after linear normalization. Based on the screened temperature-sensitive points and measured thermal displacement data, an optimized extreme learning machine based on the marine predator algorithm (MPA-ELM) was developed to predict the thermal displacement of an electric spindles model [11].

Huang et al. [12] introduced a kernel function into the ELM to replace the simplicity of the hidden layer. They proposed the KELM, which significantly improves the generalization performance of the ELM. Moreover, Huang et al.'s [13] research on the improvement of the ELM further enhanced the stability, sparsity, and accuracy of the algorithm under both general and specific conditions and accelerated the training speed of the ELM. Chen et al. [14] proposed a new fault diagnosis method based on hierarchical machine learning based on the KELM. A grid search strategy with cross-validation was used to optimize the parameters of hierarchical machine learning (HML). It was applied to detect and identify rotating machinery faults, obtaining excellent results. Based on the KELM, Su et al. [15] added self-adaptive particle swarm optimization (SAPSO) to optimize its parameters and proposed a fault diagnosis method for rotary bearings under mixed working conditions.

The fault diagnosis methods above provide a helpful reference for detecting the operating state of a check valve. However, as check valve operating conditions are complex due to the operating conditions of industrial production and environmental noise, the vibration signal is affected, thus resulting in nonlinear and non-stationary signals. Due to these characteristics, it is difficult to extract the fault characteristics of components comprehensively and accurately with only a single time-domain, frequency-domain/or time-frequency-domain method. Therefore, this paper proposes a new, multi-feature fault feature extraction method for check valves to characterize their fault state information comprehensively.

Based on multi-feature sets that characterize the operating state of one-way valves, a fault diagnosis model based on KELM was established. The model first constructs the multi-feature set to determine the fault state of the one-way valve comprehensively and accurately and then uses the multi-feature set to train the KELM model to diagnose the check valve fault conditions.

2. Basic Principles

2.1. Feature Extraction

2.1.1. Time-Domain Features

In fault diagnosis, feature fault parameters are generally sensitive to the different information from various states. Usually, fault feature extraction has no specific restrictions on the number and types of feature parameters, so features with high sensitivity to fault information differences and strong reliability are often selected as fault features [16]. The time-domain characteristics of a signal can directly reflect the dynamic changes in the signal's time domain, and they can characterize the fault types of bearings and check valves. Since there may be one or more characteristic parameters corresponding to different states, the selection of the fault feature parameters follows the principles of high sensitivity, high reliability, and feasibility. In this paper, 16 time-domain feature statistics were used and shown in Table 1 below.

2.1.2. Frequency-Domain Features

To characterize the relationship between the frequency and amplitude of a vibration signal, the signal can be transform into the frequency domain with a Fourier transform and the frequency domain characteristics of the signal analyzed in the frequency domain. Frequency domain analysis methods include amplitude spectrum and power spectrum analyses. The frequency-domain characteristic statistics used in this study are shown in Table 1 [17].

Table 1. Composition of multi-domain feature sets.

Multi-Domain Category	Number	Remark
Time-domain features	16	Peak value, mean value, root mean square value, variance, standard deviation, fourth-order center moment, peak factor, kurtosis, pulse factor, margin, waveform factor, the center of gravity frequency, mean square frequency, frequency variance, root mean square frequency, frequency standard deviation
Frequency-domain features	13	Mean, center of gravity frequency, average frequency, maximum value, average phase angle, energy, power, root variance frequency, root mean square frequency, root variance amplitude, maximum phase angle, phase angle range
Time-frequency-domain features	16	Wavelet packet energy features, wavelet packet energy entropy features
Multi-domain features	45	

2.1.3. Energy Characteristics of Wavelet Packets

As a linear transformation method, wavelet packet transform [18] satisfies the law of conservation of energy [19], which is:

$$\int_{-\infty}^{+\infty} |f(t)|^2 dt = \sum_j \sum_k |c_j, k|^2 \tag{1}$$

Since the wavelet packet coefficient contains the dimension of energy, it is used in energy analysis to determine each frequency band's energy size according to the signal's wavelet packet coefficient. Time-frequency-domain analysis methods include the short-time Fourier transform, S transform, and wavelet transform methods.

2.1.4. Energy Entropy Characteristics of Wavelet Packets

Entropy is an index used to measure the degree of disorder of information. The higher the value, the higher the disorder degree of disorder of information is and the smaller the contribution to precision is. When the entropy is smaller, the information contribution is more prominent, and the degree of disorder of information is lower.

If we perform layer j wavelet packet decomposition on the signal, assuming that the decomposition sequence is $X_{i,j}$ and $E_{i,j}$ is the signal sequence energy, then the probability density of the frequency band energy is:

$$P(X_{i,j}) = \frac{E_{i,j}}{E_j} \quad E_j = \sum_{i=1}^{2^j} E_{i,j} \quad \sum_{i=1}^{2^j} P(X_{i,j}) = 1 \tag{2}$$

Then, the band energy entropy of the wavelet packet decomposition [20] is:

$$W_{EE} = -\sum_{i=1}^{n} p(X_{i,j}) \log_2(X_{i,j}) \tag{3}$$

Through analysis, it can be found that the larger the band energy entropy of the wavelet packet decomposition is, the more random the distribution of energy in each band is. The lower the number of bands containing energy and the smaller the W_{EE}, the more regular the energy distribution is.

The multi-domain feature adopted in this paper contained 45 dimensions for the feature components.

2.2. Core Extreme Learning Machine—KELM

The standard extreme learning machine [21] (ELM) is composed of three layers: the input layer, hidden layer, and output layer, respectively. It functions based on single hidden layer feedforward neural networks (SLFNs), but the hidden layer of the SLFN only has a one-layer backpropagation (BP) neural network. The topological structure of the ELM is shown in Figure 1.

As an efficient single hidden layer feedforward neural network, we assume that, for a given n training samples $X = (x_1, x_2, \ldots, x_n) \in R^{d1\times n}$, the labels are $Y = (y_1, y_2, \ldots, y_n) \in R^{n\times d2}$, where d_1 and d_2 represent the dimensions of the input data and output data, respectively. The weight $W = \omega_{ij} \in R^{d1\times L}$ of the hidden layer of the ELM is randomly selected, where L represents the number of neurons in the hidden layer. The calculation of the hidden layer is the same as the calculation of traditional forward propagation networks with $H = g(X, W)$, where $H \in R^{n\times L}$ and $g(\cdot)$ are the activation functions.

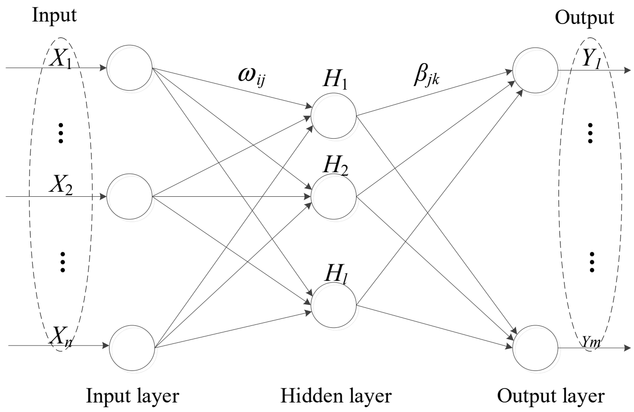

Figure 1. Extreme learning machine model.

The learning objective of an extreme learning machine is to solve the output weight β by minimizing the sum of prediction error loss functions. The objective function is:

$$\min L_{ELM} = \frac{1}{2}\|\beta\|^2 + \frac{C}{2}\|Y - H\beta\|^2 \quad (4)$$

where the C value directly affects the generalization performance of ELM and is a regularization coefficient.

$$H = \begin{bmatrix} g(w_1, b_1, x_1) & \cdots & g(w_L, b_L, x_1) \\ \vdots & & \vdots \\ g(w_1, b_1, x_N) & \cdots & g(w_L, b_L, x_N) \end{bmatrix}_{N\times L} \quad (5)$$

$$\beta = \left[\beta_1^T, \beta_2^T \ldots, \beta_L^T\right]^T_{L\times m} \quad (6)$$

$$Y = \left[y_1^T, y_2^T \ldots, y_L^T\right]^T_{N\times m} \quad (7)$$

Take the derivative β of Equation (1) and set it to 0, and the calculation formula of the output weight β can be obtained as follows:

$$\beta = \begin{cases} (\frac{I}{C}+H^T H)^{-1} H^T Y,\ N \geq L \\ H^T(\frac{I}{C}+HH^T)^{-1} Y,\ N < L \end{cases} \quad (8)$$

where I is the identity matrix.

Compared with the ELM, the KELM introduces the kernel function [12], which replaces the feature mapping of the hidden layer in ELM. The idea is to map the input sample data to the high-dimensional space and replace the inner product operation in the transformed high-dimensional space with the kernel operation in the original input space [22].

The composition of the kernel matrix Ω_{ELM} is as follows:

$$\begin{cases} \Omega_{ELM} = HH^T \\ \Omega_{ELM(i,j)} = h(x_i)\cdot h(x_j) = K(x_i, x_j) \end{cases} \quad (9)$$

where x_i and x_j are sample input vectors, and $K(x_i, x_j)$ indicates the kernel functions. When Gaussian kernel functions are used:

$$K(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{\gamma^2}) \quad (10)$$

where γ is the nuclear parameter. According to the KELM formula above, the weight β of the connection between the output function and the hidden and output layers is:

$$\begin{cases} y(x) = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix}^T (1/C + \Omega_{ELM})^{-1} T \\ \beta = (1/C + \Omega_{ELM})^{-1} T \end{cases} \tag{11}$$

2.3. KPLS Kernel Partial Least Squares Regression

The essential function of PLS is dealing with linear problems. When there is a nonlinear relationship between data, the PLS method is generally not adopted. The KPLS method can solve nonlinear problems by selecting the kernel function.

Assuming a nonlinear mapping $\phi : xi \in R^m \to \phi(xi) \in H$ from original spatial variables $\{xi\}_{i=1}^n$ to a feature space H, Rosipal and Trejo [23] used the relationship between the reproducing kernel Hilbert space (RKHS) and the feature space and developed the linear PLS method into KPLS. The KPLS algorithm can be expressed as follows.

The realization process of KPLS algorithm according to Rosipal and Trejo.
1. Randomly set the initial value u (u can be set to be equal to any column in the Y matrix)
2. According to $w_i = \phi_i^T / \lvert\lvert \phi_i^T u_i \rvert\rvert$, calculate the weight vector w_i
3. Calculate the score vector t_i according to the formulas $t_i = \phi_i w_i = \phi_i \phi_i^T u_i / \sqrt{u_i^T \phi_i \phi_i^T}$ and $\phi_i \phi_i^T u_i / \sqrt{u_i^T \phi_i \phi_i^T} = K_i K_i / \sqrt{u_i^T K_i u_i}$. Unify the vector t_i with the formula $t_i / \lvert\lvert t_i \rvert\rvert \to t_i$
4. According to $q_i = Y_i t_i / \lvert\lvert t_i^T t_i \rvert\rvert$ and $u_i = Y_i q_i / (q_i^T q_i)$, calculate the feature vectors
5. Repeat steps 2–5 until convergence
6. Calculate the matrices K and Y as $$\begin{cases} K_{i+1} = (I - \frac{tit_i^T}{t_i^T ti}) Ki (I - \frac{tit_i^T}{t_i^T ti}) \\ Y_{i+1} = (I - \frac{tit_i^T}{t_i^T ti}) Yi \end{cases} \tag{12}$$
7. Repeat steps 2–7 until all feature vectors have been calculated

3. KPLS Optimal Feature Selection and KELM Fault Diagnosis Method

The specific steps of the fault diagnosis method based on multi-feature extraction and the improved KELM proposed in this paper are as follows. The flow chart is shown in Figure 2.

The realization process of multi-feature and improved KELM fault diagnosis
1. Collect vibration signals of various states of parts
2. Divide the collected vibration signal data, divide the non-overlapping samples into 60 segments, and extract fault features from each segment
3. Construct a high-dimensional feature space and extract 16 time-domain features, 13 frequency-domain features, 8 wavelet energy features, and 8 wavelet packet energy entropy features
4. Input the obtained high-dimensional features into the KELM model for training and testing. Fifty percent are selected as training samples and fifty percent as test samples
5. Adopt a multi-domain KELM fault diagnosis model to identify the fault information

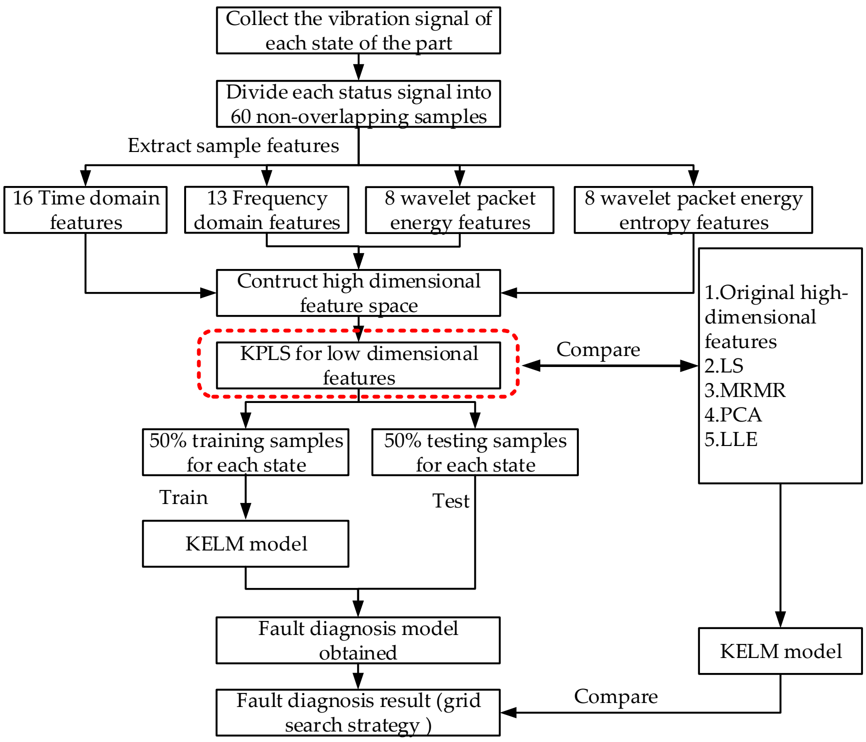

Figure 2. The realization process of KPLS-KELM flow chart.

The advantages and disadvantages of different classification methods are shown in Table 2.

Table 2. The advantages and disadvantages of the KELM and ELM.

	Advantages	Disadvantages
ELM	1. Fast 2. Small sample	1. Low generalization ability
KELM	1. High generalization ability 2. Introduction of kernel function to deal with multi-classification problems 3. Faster calculation speed	

The advantages and disadvantages of the partial least square (LS), max-relevance and min-redundancy (MRMR), principal component analysis (PCA), and locally linear embedding (LLE) dimensionality reduction methods are shown in Table 3.

Table 3. The advantages and disadvantages of dimensionality reduction methods.

	Advantages	Disadvantages
LS	1. Simple 2. Linear	1. Overfitting
MRMR	1. Feature selection based on maximum statistical dependency criteria	1. Underestimates the usefulness of features
PCA	1. Simple and quick 2. Linear methods	1.Difficult to find the right solution
LLE	1. Maintains local linear relationship of samples 2. Low computational complexity	1. Sensitive to the selection of nearest neighbor sample number 2. The manifold learned by the LLE algorithm can only be non-closed
KPLS	1. Introduces kernel function to solve nonlinear problems	

4. Experimental Simulation and Analysis

4.1. Analysis of Test Data

In this part of the study, we used bearing data from Case Western Reserve University in the United States for bearing fault diagnosis [24] to verify the effectiveness of the method proposed in this paper and the effectiveness of the multi-domain feature extraction. All the experiment were implemented using MATLAB 2018A and run on the same Windows 10 machine with an Intel(R) Core (TM) I9-9880h, 2.30GHz CPU and 16GB RAM.

The ten fault states in the rolling bearing fault diagnosis were artificially added, as shown in Table 4, and the three fault diameters of the inner ring, the outer ring, and the rolling element were 0.07 ft, 0.014 ft, and 0.021 ft, respectively. Their time-domain waveform is shown in Figure 3 below.

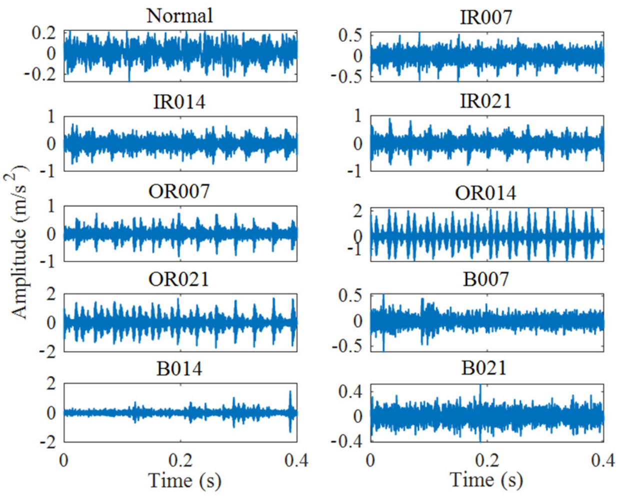

Figure 3. Time-domain diagrams of bearing in ten states (1).

We used the vibration signal of the fan terminal bearing at the motor speed of 1797 r/min as the experimental data, with the data description as shown in Table 4.

Table 4. Experimental bearing sample attributes.

Outer Ring (ft)	Inner Ring (ft)	Rolling Element (ft)
0.014	0.07	0.021

As can be seen from Figure 3, IR007, IR014, IR021, OR007, OR014, OR021, and other signals demonstrated periodic impacts, while the signals of B007, B014, and B021 showed no obvious periodic hints. No amplitude difference between the signals was apparent and neither were the characteristics of the impact nor its period. Hence, the proposed mixed domain could identify the bearing fault type and the degree of the fault.

For the multi-domain feature extraction method proposed in this paper, we divided the vibration signals of each state into 60 non-overlapping samples with a length of 2000 and extracted multi-domain features from each sample separately.

There were thus 16 types of time-domain features and 13 types of frequency-domain features, as shown in Table 1. For the 16 time-frequency domain features, the wavelet packet was decomposed using the wavelet type db5, and the number of decomposed layers was 3. Then, the energy and energy entropy features were extracted from the components obtained by each layer of the wavelet packets. The energy entropy features of the wavelet packets in the time-frequency domain were obtained. The obtained feature is shown in the box diagram in Figure 4.

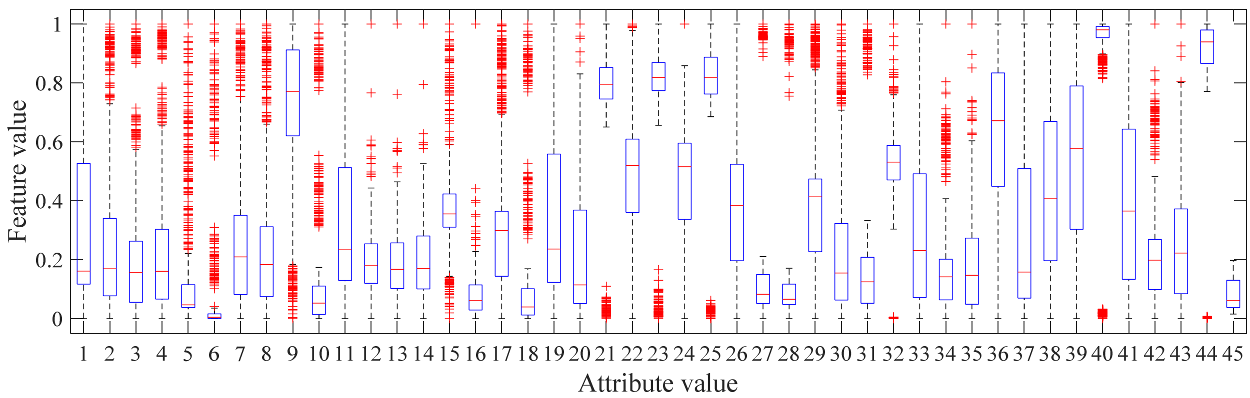

Figure 4. Multi-domain characteristic box plot of bearing in ten states.

After being standardized, most of these features were distributed in the range of 0.2 to 0.8. After the feature extraction, we input the extracted 45 dimension, high-dimensional features into the KELM for fault diagnosis. To verify the effectiveness of the proposed multi-domain features, we extracted 16 types of time-domain features, 13 types of frequency-domain features, 16 types of time-frequency-domain features, and feature sets of the 45 multi-domain features after the dimensionality reduction by KPLS. Due to complex operation conditions, algorithms have difficulty characterizing the fault state completely with its single feature. The combined or multi-domain features could reveal the state information, but there was much redundant information that reduced the accuracy and efficiency of the diagnostic model, so extracting features with moderate dimensions and high sensitivity to each state is the key. LS and MRMR select features that are sensitive to faults or have large contributions, while in feature selection methods such as the Pearson correlation coefficient, distance criterion, and information gain, the emphasis is on the physical meaning of the original features remaining unchanged. PCA employs feature dimension reduction to obtain a compact manifold structure based on feature mapping or feature fusion, and it can extract nonlinear features with the variance in global distribution information unchanged, but it cannot maintain local manifold information, so it is only suitable for linear dimension reduction. The kernel function in KPCA maps data to high dimensional space to obtain nonlinear principal components with higher separability, but the kernel function has a great influence on the results. LLE determines the similarity of neighborhood points using the Euclidean distance, but ignoring the relationship between data leads to unreasonable neighborhood construction.

When comparing the method proposed in this paper with other dimensionality reduction methods, the results shown in Figure 5 were obtained.

The data used for training the KELM model before the dimensionality reduction can be seen in Table 3, including the ten fault states from Figure 3. The total number of samples was 600×45, and each sample contained 45 feature points, comprising 60% of the data used for training. Thus, the number of training sets was 360×45, and the number of testing sets was 240×45. The training accuracy was 100%, and the testing accuracy was 91.67%. The data used for the KELM are shown in Table 5.

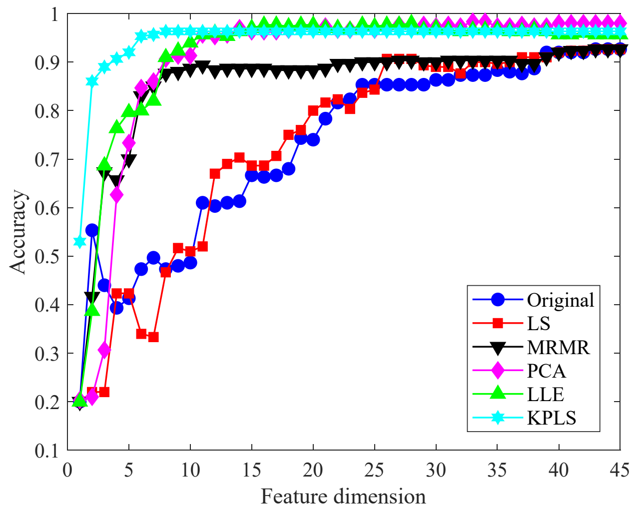

Figure 5. The bearing fault diagnosis accuracy for the KELM with different dimensionality reduction methods.

Table 5. The data used for the KELM.

Total Samples (Number)	Training Sets (Number)	Testing Sets (Number)	Learn Time (s)	Training Accuracy	Testing Accuracy
600 × 45	360 × 45	240 × 45	0.0051	100%	91.67%

We input these feature sets into KELM for fault diagnosis analysis, and the classification results are shown in Figure 6.

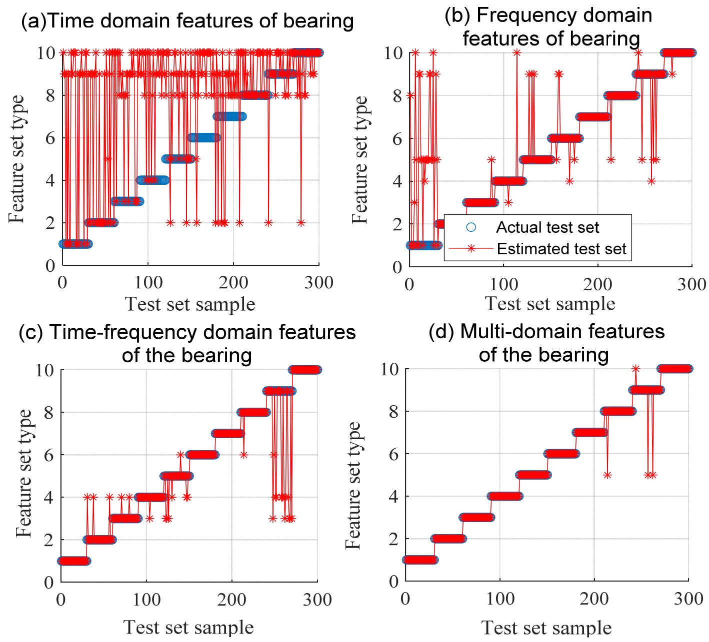

Figure 6. Comparison of diagnoses of the bearing's time-domain, frequency-domain, time-frequency-domain, and multi-domain features.

Comparing the four graphs in Figure 6, it can be seen that the diagnostic results for the multi-domain features of the bearing in Figure 6d are better than the time-frequency-domain features in Figure 6c. Figure 6c shows a better diagnostic result than the frequency domain feature from Figure 6b and the time-frequency domain feature in Figure 6a, while Figure 6b displays a better diagnostic result than Figure 6a. The results shown in Table 6 indicate that the proposed multi-domain feature extraction was better than the time-domain, frequency-domain, and time-frequency-domain features and could achieve satisfactory diagnosis results in the experiment on the bearings with ten fault states. By analyzing in detail the diagnostic results of Figure 6d, we can see that one sample in the third type of rolling element with 007 states was wrongly classified as the fifth type of the outer ring with 007 states. Two samples for the ninth type of the rolling element B0014 were improperly classified as the fifth outer ring with OR007 states. One sample was also wrongly classified into the tenth state of the rolling element with 021 states. There were four wrongly classified samples in total. The accuracy rates for the time domain, frequency domain, time and frequency domain, and multi-feature domain using KPLS were 30.00%, 86.67%, 91.00%, and 97.33%, respectively. As a combination of time-domain, frequency-domain, and time-frequency-domain features, multi-domain features can characterize a fault state fully. The results show that the extraction of 45 multi-domain fault features proposed in this paper had the best effect in bearing fault diagnosis.

Table 6. Diagnosis results under different feature sets.

Feature Set	Diagnostic Time (Seconds)	Accuracy (%)
Time domain	0.0035	30.00
Frequency domain	0.0055	86.67
Time-frequency domain	0.0037	91.00
Multi-feature via KPLS	0.0039	97.33

To eliminate contingency, and as the KELM diagnosis results are often affected by node parameters, we conducted another bearing fault diagnosis experiment with different hidden layer node numbers using the feature extraction method above: time domain, frequency domain, and time-frequency multi-feature domain. The results are shown in Figure 7.

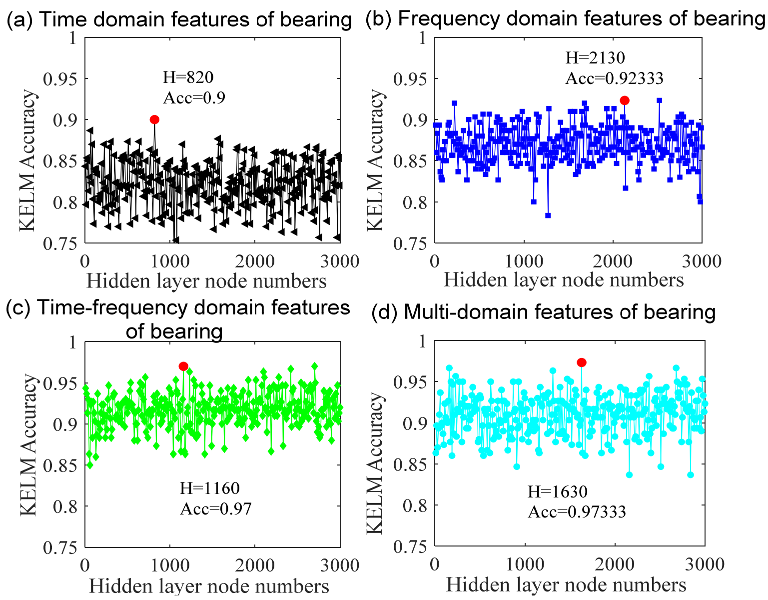

Figure 7. Diagnosis results for the time-domain, frequency-domain, time-frequency domain, and multi-domain features of the bearing with different hidden layers in KELM.

As shown in Figure 7, with the extraction of multi-domain features, while the KELM hidden layer node changed from 0 to 3000 (Table 7), the average fault diagnosis accuracy was maintained at more than 90%. When the hidden layer node value H = 1630, the fault diagnosis accuracy was 97.333%. The results show that the proposed multi-domain feature extraction method had the best fault diagnosis accuracy. It was superior to the time-domain, frequency-domain, and time-frequency-domain feature extraction methods, no matter how the hidden layer nodes changed. At the same time, as shown in Figure 7a, when extracting time-domain features, the average diagnostic accuracy was close to 80%.

Table 7. Corresponding results for different nodes of hidden layer H.

	Number of Nodes in Hidden Layer H						
H	500	820	1000	1500	1630	2130	3000
Time domain	47.65%	58.50%	70.45%	92.22%	91.66%	89.55%	91.66%
Frequency domain	45.57%	67.89%	74.76%	86.56%	83.67%	88.90%	86.43%
Time-frequency domain	65.44%	69.90%	74.34%	84.38%	80.23%	87.77%	85.57%
Multi-domain via KPLS	96.88%	93.45%	92.75%	96.88%	97.33%	93.56%	95.53%

When extracting frequency-domain features, the average diagnostic accuracy was close to 85%. When extracting time-frequency-domain features, the average diagnostic accuracy was close to 90%. When adopting the multi-domain feature proposed, the diagnostic accuracy was higher than or equal to 90%, with the maximum diagnostic accuracy being 97.333%. In this scenario, the number of hidden layer nodes was 1630. The experiment result indicates that the proposed multi-domain feature extraction method had the best diagnostic accuracy.

4.2. Data Analysis of Diaphragm Pump Check Valve

Figure 8a,b show the sensors that were fixed on the shells of the inlet and outlet valves. For each valve, there was one acceleration sensor of the type PCB352C33 (sensitivity: 100 mV/g) and one sound pressure sensor of the type MP021 (50 mV/Pa), respectively. The acceleration sensor collected the shell vibration signal along the Z-axis using three channels, while the sound pressure sensor collected the sound signal along the Y-axis direction.

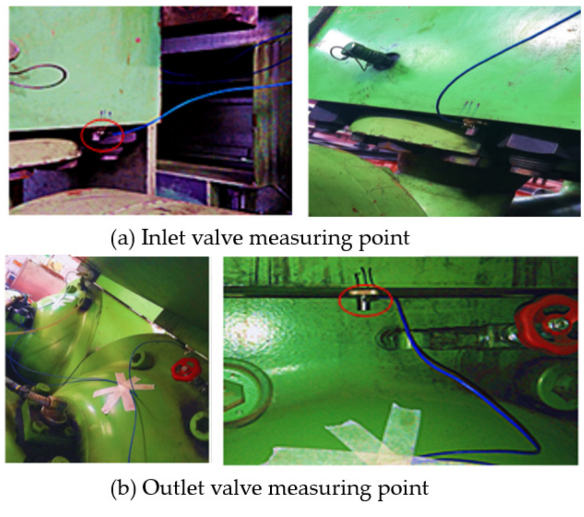

Figure 8. Acceleration sensor positions.

Figure 9 shows the vibration signal acquisition device for the check valve. The eight-channel analog signal was amplified, filtered, and converted into A/D by the data acquisition card and sent to the PS PXI-3050EXT 2.7 ghz controller; then, the signal was transferred to the PS PXIE-9108Ext eight-slot industrial computer and stored in the hard disk. When the diaphragm pump check valve ran normally (first 500 h), eight-channel data were collected at the sampling rate of 2560 Hz every 1 h. When the check valve was potentially damaged (500 h~1000 h), eight-channel data were collected every 10 min. When the one-way valve potentially underwent serious failure or damage (after 1000 h), we collected eight-channel data every 2 min.

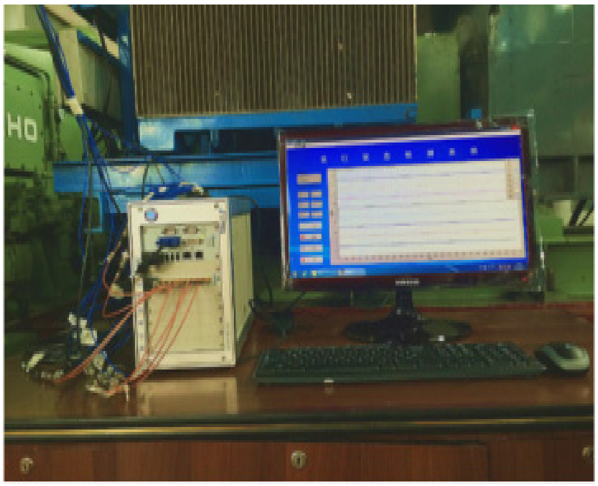

Figure 9. Data acquisition device.

Due to reasons of safety and cost, specific experiments could not be carried out; the damage states of the check valve depended on the actual working conditions. After the check valve was replaced by technicians who work on the site, we checked the damage and recorded the basic fault size and location. The typical damage is shown in Figure 10; Figure 10a shows a stuck valve fault, Figure 10b shows a wear fault, and Figure 10c shows a worn valve fault.

(a) Stuck Valve fault (b) Wear fault (c) Worn valve seat (d)Replaced check valve

Figure 10. Fault check valve.

When the check valve was in a normal state, stuck valve failure state, or wear breakdown failure state, we randomly generated a set of time-domain diagrams of vibration signals (as shown in Figure 11). It was found that there were some fault impulses in the middle of the time-domain graph. However, since the impulse period was not apparent, as the noise in the local waveform was inevitable, it was difficult to analyze the type and the cause of failure based on the time-domain waveform alone.

Therefore, the following experiment was used for the vibration signal sample and the multi-domain feature extraction method employed to perform fault diagnosis on the check valve. First, the vibration signal of the check valve in each state was divided into 60 non-overlapping samples, and the number of data points in each sample was 1280. For each non-overlapping piece, we extracted 45 multi-domain features, of which samples 1 to 16 were time-domain features, samples 17 to 29 were frequency-domain features, and samples 30 to 45 were time-frequency-domain features. Among the time-frequency-domain features, we extracted the energy features for the first eight time-frequency-domain features and the energy entropy features for the second eight time-frequency domain features.

The results for the multi-domain features are shown in the boxplot in Figure 12. After normalization, most of the 45 characteristics of the check valve were distributed in the range from 0 to 0.2. Some samples were distributed in the range from 0.2 to 0.8, and very few were distributed in the range from 0.8 to 1.

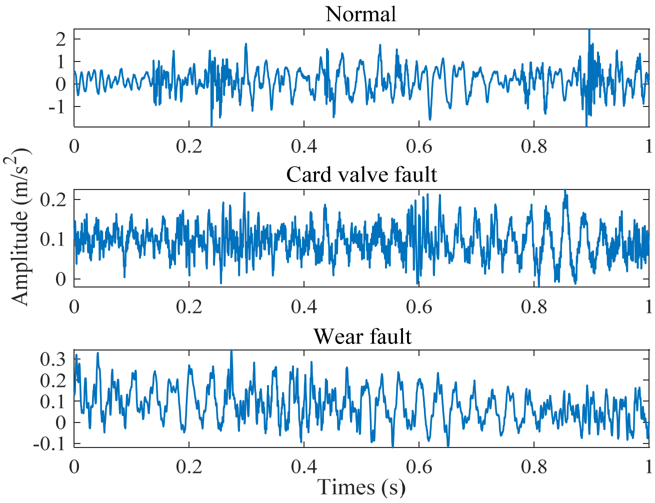

Figure 11. Time-domain diagram of three states of one-way valve.

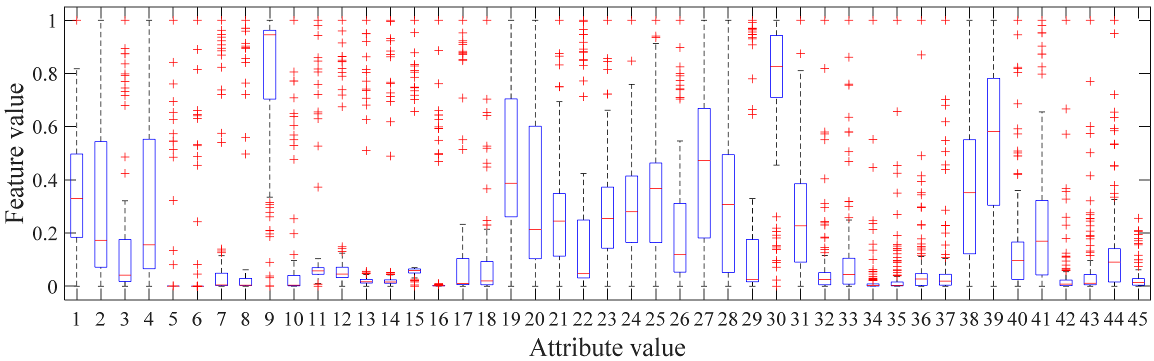

Figure 12. Box diagram of three states for multi-domain features from check valve.

As the number of dimensions of the multi-domain features reached 45, we adopted the KPLS method to reduce the dimensionality of the multi-domain features.

After the KPLS dimensionality reduction, the diagnosis accuracy rate reached more than 95% with only eight dimensions. Figure 13 shows the accuracy results for all the selected dimensionality reduction methods: LS, MRMR, PCA, and LLE.

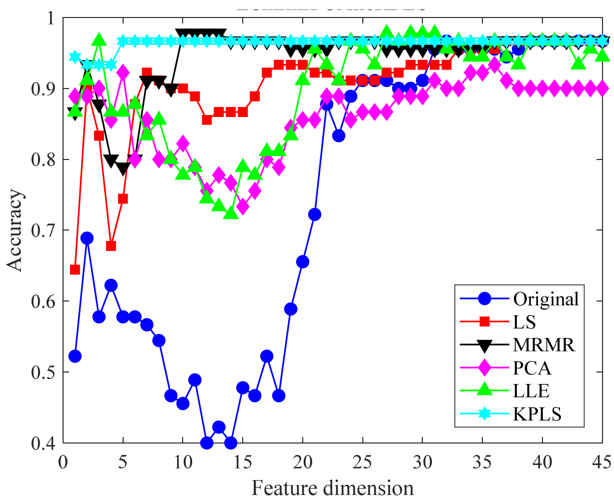

Figure 13. Check valve fault diagnosis accuracy obtained by KELM with different dimensionality reduction methods.

The data used for training the KELM model before the dimensionality reduction can be seen in Table 8. We can see three fault categories in Figure 11; the total number of samples was 180 × 45, each dataset contained 45 feature points, and 60% of the data were used for training. Thus, the number of training sets was 108 × 45, the number of testing sets was 72 × 45, the training accuracy was 100%, and the testing accuracy was 86.11%.

Table 8. The data used for training the KELM.

Total Samples (Number)	Training Sets (Number)	Testing Sets (Number)	Learn Time (s)	Training Accuracy	Testing Accuracy
180 × 45	108 × 45	72 × 45	0.0014	100%	86.11%

To verify the effectiveness of the multi-domain feature extraction results for the check valve, we compared the fault diagnosis results with the time-domain feature, frequency feature, and time-frequency-domain features. The results obtained by the KELM are shown in Figure 14.

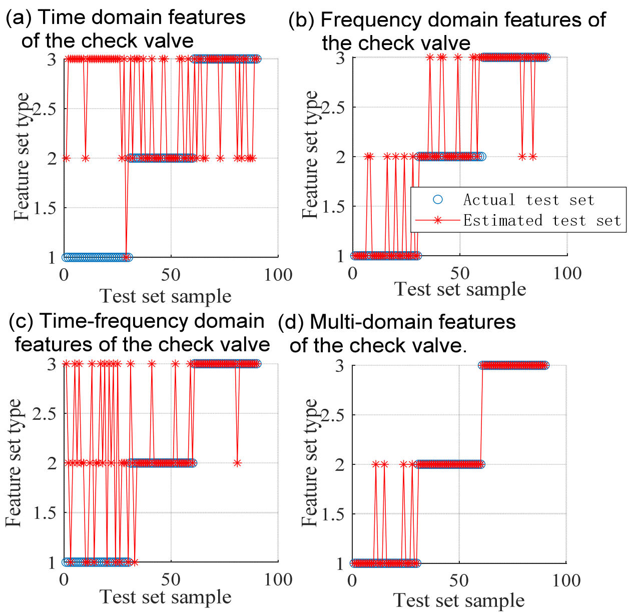

Figure 14. Comparison of diagnostic results for the time-domain, frequency-domain, time-frequency-domain, and multi-domain features of the check valve.

As shown in Figure 14d, four regular sample points were wrongly classified as type 2 (stuck valve failure state). When the number of hidden layer nodes was 200, the final fault diagnosis result was 96.88%.

Compared with the time-domain features, the frequency-domain features, and the time-frequency-domain features for the check valve—which demonstrated accuracies of 45.56%, 82.22%, and 68.89%, respectively—the fault diagnosis results obtained with the multi-domain features, as shown in the Figure 14d, were improved to 96.88%, raising the accuracy rate by 51.32%, 14.66%, and 27.99%. The results shown in Table 9 indicate that the proposed multi-domain feature extraction achieved the optimal diagnosis result in the check valve fault diagnosis experiment, which proves the proposed method's effectiveness.

Table 9. Diagnosis results with different feature sets.

Feature Set	Diagnostic Time (s)	Accuracy (%)
Time domain	0.0039	45.56
Frequency domain	0.0041	82.22
Time frequency domain	0.0039	68.89
KPLS multi-domain feature	0.0047	96.88

As the fault diagnosis of the KELM was affected by the number of hidden layer nodes, we compared the fault diagnosis results of the check valve using KELM when the hidden layer nodes were 0 to 3000, as shown in Figure 15 and Table 10 below.

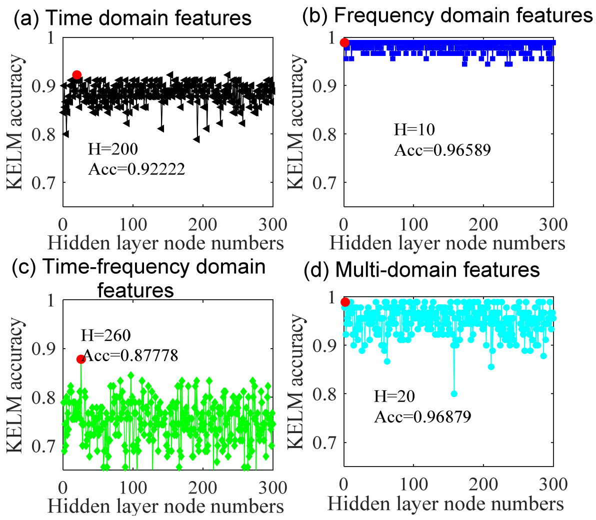

Figure 15. Diagnosis results for check valve with the time-domain, frequency-domain, time-frequency-domain, and multi-domain features using different numbers of hidden layers in KELM.

Table 10. Corresponding results for node numbers different numbers of hidden layers.

	Number of Nodes in Hidden Layer H						
H	10	20	100	200	250	260	300
Time domain	49.66%	57.89%	82.22%	92.22%	85.53%	91.22%	90.32%
Frequency domain	96.59%	94.83%	95.55%	94.78%	93.68%	94.45%	96.20%
Time-frequency domain	65.45%	69.89%	72.32%	74.94%	83.83%	87.77%	87.85%
Multi-domain	92.45%	96.88%	95.56%	96.88%	93.45%	96.50%	96.23%

As shown in Figure 15d, no matter how the hidden layer nodes change, the average diagnosis result was close to 95%, with the maximum diagnosis accuracy being 96.879%. In this scenario, the number of nodes in the hidden layer was ten, which was better than the fault diagnosis results for the time domain, frequency domain, and time-frequency domain. Therefore, the multi-domain feature extraction method proposed in this paper achieved better results in bearing and check valve fault diagnosis, proving the method's effectiveness.

After analyzing different feature sets, we used different classification algorithms to conduct an experimental analysis of multi-domain feature sets. The results are shown in Figure 16 below. It can be seen that the overall accuracy of the proposed KELM exceeded 90%, which was significantly better than the back propagation neural network (BPNN) and ELM.

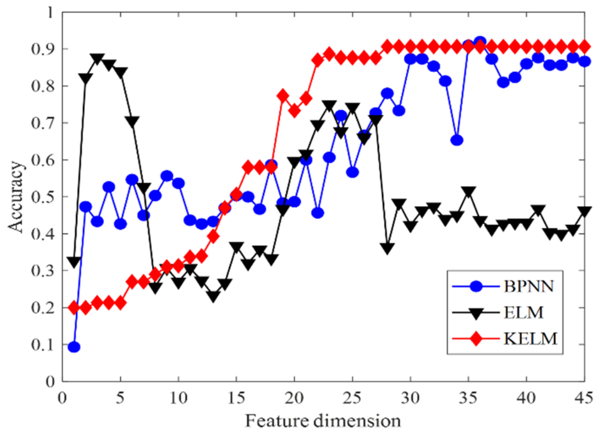

Figure 16. Diagnostic accuracy obtained with KPLS dimensionality reduction features.

5. Conclusions

Investigating the problem of a check valve's fault state being difficult to classify, this paper proposed a diagnosis method based on multi-domain features and KELM. It adopted the method to analyze bearing test fault data and check valve fault data. The conclusions are as follows:

1. When the time-domain, frequency-domain, and time-frequency-domain features were used alone for bearing fault diagnosis, the diagnostic accuracies were 30.00%, 86.67%, and 91.00%, respectively. With the multi-domain feature extraction method after KPLS dimensionality reduction, the accuracy was improved to 97.33%;
2. When the bearing fault diagnosis test was carried out with different numbers of hidden layer nodes, the accuracy was increased from 45.56%, 82.22%, and 68.89% to 97.33% with multi-domain features and KPLS;
3. The proposed KPLS-KELM algorithm could accurately and effectively extract the fault information for the check valve, and the accuracy reached 95%. Compared with the ELM method, KELM is superior for the traditional time-domain, frequency-domain, and time-frequency-domain analysis methods and has higher accuracy.

The accuracy of KELM is affected by the kernel parameters and penalty coefficients. Achieving fast and accurate parameters for different objects is the focus of future research. Application of the theory in practice would be the ultimate end of this research, and online fault diagnosis and big data analysis also need to be considered next.

Though the proposed method achieves better diagnosis results and provides superior accuracy for feature extraction and better robustness in fault diagnosis, the stability of the multi-feature method with coarse-grained data could be further improved. Therefore, it needs more time for training and classification. Next, we will develop a fast training algorithm to facilitate model training.

Author Contributions: Conceptualization, X.Y. and Y.F.; methodology, X.Y.; software, X.Y. and G.Z.; validation, Y.F. and X.W.; formal analysis, C.Z.; investigation, X.Y.; resources, X.Y.; data curation, X.Y.; writing—original draft preparation, Y.F.; writing—review and editing, X.Y.; visualization, X.Y.; supervision, X.W. and C.Z.; project administration, X.Y.; funding acquisition, X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University, China (no. ICT2022B06), the PhD research startup foundation of Yunnan Normal University (no. 01000205020503131), and the Fundamental Research Program of Yunnan Province (no. 202101AU070061), Project of Educational Commission of Yunnan Province of China.

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.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, Z.; Gryllias, K.; Li, W. Mechanical fault diagnosis using Convolutional Neural Networks and Extreme Learning Machine. *Mech. Syst. Signal Process.* **2019**, *133*, 106272. [CrossRef]
2. Peng, Z.; Cheng, L.; Zhan, J.; Yao, Q. Fault classification method for rolling bearings based on the mlllti-feature extraction and modified Mahalanobis-Taguchi system. *J. Vib. Shock* **2020**, *39*, 249–256.
3. Jiang, J.; Hu, Y.; Ke, Y. Fault diagnosis of diesel engines based on wavelet packet energy spectrum feature extraction and fuzzy entropy feature selection. *J. Vib. Shock* **2020**, *39*, 273–277.
4. Ding, T.; Yan, Y.; Li, X.; Liu, L. Study of scintillation detector fault diagnosis based on ELM method. *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrometers Detect. Assoc. Equip.* **2022**, *1032*, 166637. [CrossRef]
5. Lee, J.; Sun, Z.; Tan, T.B.; Mendez, J.; Flores-Cerrillo, J.; Wang, J.; He, Q.P. Remaining Useful Life Estimation for Ball Bearings Using Feature Engineering and Extreme Learning Machine. *IFAC-PapersOnLine* **2022**, *55*, 198–203. [CrossRef]
6. Pang, S.; Yang, X.; Zhang, X.; Lin, X. Fault diagnosis of rotating machinery with ensemble kernel extreme learning machine based on fused multi-domain features. *ISA Trans.* **2020**, *98*, 320–337. [CrossRef]
7. Wang, X.B.; Zhang, X.; Li, Z. Ensemble extreme learning machines for compound-fault diagnosis of rotating machinery. *Knowl.-Based Syst.* **2019**, *188*, 105012. [CrossRef]
8. Yin, G.; Zhang, Y.T.; Li, Z.N.; Ren, G.Q.; Fan, H.B. Online fault diagnosis method based on incremental support vector data description and extreme learning machine with incremental output structure. *Neurocomputing* **2014**, *128*, 224–231. [CrossRef]
9. Harishvijey, A.; Raja, J.B. Automated technique for EEG signal processing to detect seizure with optimized Variable Gaussian Filter and Fuzzy RBFELM classifier. *Biomed. Signal Process. Control* **2022**, *74*, 103450. [CrossRef]
10. Shen, J.; Xu, F. Method of fault feature selection and fusion based on poll mode and optimized weighted KPCA for bearings. *Measurement* **2022**, *194*, 110950. [CrossRef]
11. Li, Z.; Wang, B.; Zhu, B.; Wang, Q.; Zhu, W. Thermal error modeling of electrical spindle based on optimized ELM with marine predator algorithm. *Case Stud. Therm. Eng.* **2022**, *38*, 102326. [CrossRef]
12. Huang, G.B. An insight into extreme learning machines: Random neurons, random features and kernels. *Cogn. Comput.* **2014**, *6*, 376–390. [CrossRef]
13. Huang, G.; Huang, G.B.; Song, S.; You, K. Trends in extreme learning machines: A review. *Neural Netw.* **2015**, *61*, 32–48. [CrossRef] [PubMed]
14. Chen, Q.; Wei, H.; Rashid, M.; Cai, Z. Kernel extreme learning machine based hierarchical machine learning for multi-type and concurrent fault diagnosis. *Measurement* **2021**, *184*, 109923. [CrossRef]
15. Su, H.; Xiang, L.; Hu, A.; Gao, B.; Yang, X. A novel hybrid method based on KELM with SAPSO for fault diagnosis of rolling bearing under variable operating conditions. *Measurement* **2021**, *177*, 109276. [CrossRef]
16. Kang, S.; Ma, D.; Wang, Y.; Lan, C.; Chen, Q.; Mikulovich, V. Method of assessing the state of a rolling bearing based on the relative compensation distance of multiple-domain features and locally linear embedding. *Mech. Syst. Signal Process.* **2017**, *86*, 40–57. [CrossRef]
17. Xue, X.; Zhou, J. A hybrid fault diagnosis approach based on mixed-domain state features for rotating machinery. *Isa Trans.* **2016**, *66*, 284–295. [CrossRef]
18. Attoui, I.; Fergani, N. A New Time-frequency Method for the Identification and classification of Boutasseta faults. *J. Jilin Univ. (Nat. Sci. Ed.) Sound Vib.* **2017**, *397*, 241–265. [CrossRef]
19. Yao, B.; Su, J.; Wu, L.; Guan, Y. Modified Local Linear Embedding Algorithm for Rolling Element Bearing Fault Diagnosis. *Appl. Sci.* **2017**, *7*, 1178. [CrossRef]
20. Chen, R.; Chen, S.; Yang, L.; Wang, J.; Xu, X.; Luo, T. Looseness diagnosis method for connecting bolt of fan foundation based on sensitive mixed-domain features A Mathematical Model for Solving mathematical Congestion. *Comput. Modeling Comput. Modeling* **2005**, *54*, 554.
21. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [CrossRef]
22. Peng, Y.; Li, Q.; Kong, W.; Qin, F.; Zhang, J.; Cichocki, A. A joint optimization framework to semi-supervised RVFL and ELM networks for efficient data classification. *Appl. Soft Comput.* **2020**, *97*, 106756. [CrossRef]
23. Rosipal, R.; Trejo, L.J.; Wheeler, K. Locally-Based Kernal Pls Smoothing to Non-parametric Regression Curve Fitting. 2002; pp. 1–36. Available online: https://ntrs.nasa.gov/citations/20030015244 (accessed on 23 June 2022).
24. Loparo, K. Bearings Vibration Data Set. Case Western Reserve University. Available online: https://engineering.case.edu/bearingdatacenter/download-data-file (accessed on 23 June 2022).