

Article

An Improved Fault Diagnosis Approach Using LSSVM for Complex Industrial Systems

Shuyue Guan ¹, Darong Huang ^{1,*}, Shenghui Guo ², Ling Zhao ¹ and Hongtian Chen ³

¹ College of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China; 622200070034@mails.cqjtu.edu.cn (S.G.); zhaoling@cqjtu.edu.cn (L.Z.)

² College of Electronics and Information Engineering, Suzhou University of Science and Technology, Suzhou 215009, China; shguo@usts.edu.cn

³ Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB T6G 1H9, Canada; hc11@ualberta.ca

* Correspondence: drhuang@cqjtu.edu.cn

Abstract: Fault diagnosis is a challenging topic for complex industrial systems due to the varying environments such systems find themselves in. In order to improve the performance of fault diagnosis, this study designs a novel approach by using particle swarm optimization (PSO) with wavelet mutation and least square support (LSSVM). The implementation entails the following three steps. Firstly, the original signals are decomposed through an orthogonal wavelet packet decomposition algorithm. Secondly, the decomposed signals are reconstructed to obtain the fault features. Finally, the extracted features are used as the inputs of the fault diagnosis model established in this research to improve classification accuracy. This joint optimization method not only solves the problem of PSO falling easily into the local extremum, but also improves the classification performance of fault diagnosis effectively. Through experimental verification, the wavelet mutation particle swarm optimization and least square support vector machine (WMP-PSO-LSSVM) fault diagnosis model has a maximum fault recognition efficiency that is 12% higher than LSSVM and 9% higher than extreme learning machine (ELM). The error of the corresponding regression model under the WMP-PSO-LSSVM algorithm is 0.365 less than that of the traditional linear regression model. Therefore, the proposed fault scheme can effectively identify faults that occur in complex industrial systems.

Keywords: fault diagnosis; PSO; wavelet mutation; LSSVM



Citation: Guan, S.; Huang, D.; Guo, S.; Zhao, L.; Chen, H. An Improved Fault Diagnosis Approach Using LSSVM for Complex Industrial Systems. *Machines* **2022**, *10*, 443. <https://doi.org/10.3390/machines10060443>

Academic Editor: Ahmed Abu-Siada

Received: 7 May 2022

Accepted: 31 May 2022

Published: 4 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Fault diagnosis and detection for complex industrial systems has been widely investigated and rapidly developed in recent years [1–5]. In essence, fault diagnosis in industrial environments is pattern recognition based on fault features. In engineering systems, fault diagnosis is usually carried out in two aspects: model-based and data-based [6]. With the progress of science and technology, intelligent pattern recognition algorithms for fault signals have been developed vigorously, such as neural networks [7–9], K-nearest neighbor [10–12], and LSSVM [13–15]. Neural networks have the advantage of being able to approximate arbitrary complex nonlinearities and have good robustness [6,16]. For example, Xu et al. [17] proposed a fault diagnosis method based on neural networks and fuzzy theory for rotating machinery. In [4], a performance degradation and fault detection model for industrial systems was proposed based on transfer learning and federated neural networks, and the analysis illustrated its effectiveness and feasibility for industrial systems. For the purpose of fault detection, Chen et al. [9] established a data-driven fault detection scheme based on two neural networks, which can construct the optimal model adaptively. These methods demonstrate the effectiveness of neural network algorithms in fault diagnosis for dynamic industrial systems [18]. In another respect, vibration signals can be converted into two-dimensional digital images representing the patterns of the permutation entropy of those signals, as in [19], where a deep neural network was

established for pattern recognition. Usually, a neural network algorithm needs a large amount of data training to establish a model with high diagnostic accuracy. However, it is difficult to obtain a large amount of fault data from complex systems in practice. K-nearest neighbor is one of the simplest algorithms based on data-driven classification technology, and it is easy to implement and requires no parameter estimations. It is widely applied in pattern recognition, fault diagnosis, and the multiple classification problem [20–22]. Ma et al. [23] proposed a multilabel learning algorithm based on the K-nearest neighbor algorithm for managing the prognostics and health of rolling bearings, and Gan et al. [24] used the K-nearest classifier to identify different rolling bearing conditions for industrial systems. Nevertheless, K-nearest neighbor is highly dependent on samples, the effect of this defect on classification accuracy cannot be neglected.

Support vector machine (SVM), as a classical pattern identification method, is widely used in various fields. For example, a temporal-based support vector machine for the detection and identification of several toxic gases in a gas mixture was proposed in [25], which also indicates the great potential of SVM. LSSVM, which is a modification of the SVM, was proposed by Suykens and Vandewalle in [26]. Inequality constraints in SVM are replaced by equality constraints in LSSVM, reducing the difficulty of calculation. Zhang et al. [27] combined a generalized frequency response function and LSSVM to achieve fault classification for a nonlinear analog circuit. The results showed that the fault diagnosis method can obtain high recognition accuracy. Product function correntropy and LSSVM were presented in [28] to improve the fault diagnosis performance for rolling bearings in varying industrial conditions. In order to further improve the effectiveness of LSSVM, Zhang et al. [29] used PSO to optimize LSSVM, and their proposed PSO-LSSVM fault diagnosis method had a high recognition rate. Similarly, a fault identification method for rolling bearings in industrial systems was proposed in [30]. In addition, Ren et al. achieved fault detection and diagnosis in complex industrial systems based on PSO-LSSVM, and their experimental results showed that this method can be applied well in the field of industry. As mentioned above, as a classical intelligent optimization algorithm, PSO is widely used due to its convenience of implementation: it does not require that extra attention be paid to parameter tuning. However, the PSO algorithm also has many disadvantages, such as a poor ability to search locally, and its tendency to fall easily into the local extremum [31–33]. To solve this problem, many scholars have made great efforts. For example, Zhang et al. [34] introduced dynamic inertia weights and gradient information to improve PSO. At the same time, a bearing fault diagnosis method via an LSSVM identification model was presented. Liu et al. [35] established a fault detection model based on a chaotic PSO algorithm and a kernel-independent component analysis, and the simulation results showed that the optimization method can avoid the phenomenon of the PSO algorithm's susceptibility to falling into a local extremum. Furthermore, an improved PSO- and SVM-based fault diagnosis methodology was presented in [36] to predict faults in nuclear power plants.

Motivated by the above observations, the first contribution of this study is to design a novel fault diagnosis method based on WMP-PSO-LSSVM that can achieve a high classification accuracy. The second contribution is to solve the problems of the PSO algorithm's susceptibility to falling into a local extremum and its low search precision. In addition, this study adopts the data-driven method to realize the fault diagnosis and prognostics of the actual complex parts in an industrial system, and a contrast experiment shows that the established joint optimization scheme has superior performance and strong robustness, which can promote the development of mechanical fault diagnosis.

The remaining parts of this study include Section 2, which introduces the signal pre-processing and feature extraction methods, which are based on an orthogonal wavelet packet algorithm (WPT); Section 3, in which the WMP-PSO-LSSVM-based fault diagnosis scheme is presented; Section 4, where the effectiveness of this study is verified by actual fault data and comparison experiments; and finally, Section 5, in which the conclusion is given.

2. Signal Decomposition and Feature Extraction-Based Orthogonal Wavelet Packet Transform

Wavelet transforms have been widely used for vibration signal pre-processing for industrial systems. Generally, wavelet transforms only decompose the low-frequency part of the signal, and do not treat the high-frequency portion of the signal at all. However, the detailed information that can characterize the vibration signal usually exists in the high-frequency section. Therefore, the orthogonal wavelet packet transform is introduced to solve this problem. Furthermore, the vibration signal of industrial systems can be decomposed in this way without information loss, which lays a foundation for obtaining high fault diagnosis accuracy. The theoretical basis is described as follows.

In multiresolution analysis, $L^2(R)$ is a square-integrable space and $L^2(R) = \bigoplus_{j \in Z} W_j$, indicating that the multiresolution analysis decomposes $L^2(R)$ into the orthogonal sum of all subspaces $W_j(j \in Z)$, according to the different scale factors j . $W_j(j \in Z)$ is the wavelet subspace of the wavelet function $\psi(t)$. Then, we hope to further subdivide $W_j(j \in Z)$ through a binary fraction. Therefore, the scale subspace V_j and the wavelet subspace W_j can be represented through a new subspace U_j^n , if there are the following conditions:

$$\begin{cases} U_j^0 = V_j, j \in Z \\ U_j^1 = W_j, j \in Z \end{cases} \tag{1}$$

Then, the orthogonal decomposition of the Hilbert space can be expressed as follows:

$$U_{j+1}^0 = U_j^0 \oplus U_j^1 \tag{2}$$

Suppose U_j^n is the wavelet subspace of $u_n(t)$, U_j^{2n} is the wavelet subspace of $u_{2n}(t)$, and $u_n(t)$ is:

$$\begin{cases} u_{2n}(t) = \sqrt{2} \sum_{k \in Z} h(k)u_n(2t - k) \\ u_{2n+1}(t) = \sqrt{2} \sum_{k \in Z} g(k)u_n(2t - k) \end{cases} \tag{3}$$

where $h(k)$ represents the low-pass filter coefficients and $g(k)$ represents the high-pass filter coefficients, and $g(k) = (-1)^k h(1 - k)$. Then, Formula (3) can be rewritten as follows:

$$\begin{cases} u_{2n}(t) = \sqrt{2} \sum_{k \in Z} h_k u_n(2t - k) \\ u_{2n+1}(t) = \sqrt{2} \sum_{k \in Z} g_k u_n(2t - k) \end{cases} \tag{4}$$

where $u_0(t) = \phi(t)$ ($\phi(t)$ is the scale function), $u_1(t) = \psi(t)$ ($\psi(t)$ is the wavelet basis function), and the sequence $\{u_n(t)\}_{n \in Z_+}$ is the orthogonal wavelet packet basis.

Suppose $f(n)$ is the signal to be decomposed. In fact, a wavelet packet transform of $f(n)$ is a projection coefficient on the wavelet packet basis $\{u_n(t)\}_{n \in Z_+}$:

$$p_f(n, j, k) = \langle f(t), u_n(t) \rangle = \int_{-\infty}^{+\infty} f(t) \left[2^{-j/2} \bar{u}_n(2^{-j}t - k) \right] dt \tag{5}$$

where $\{p_s(n, j, k)\}_{k \in Z}$ is the sequence of transformation coefficients of $f(n)$ on U_j^n .

Usually, the transformation coefficients $\{p_s(n, j, k)\}_{k \in Z}$ can be calculated through the Mallat algorithm:

$$\begin{cases} p_f(2n, j, k) = \sum_{l \in Z} h_{l-2k} p_f(n, j - 1, l) \\ p_f(2n + 1, j, k) = \sum_{l \in Z} g_{l-2k} p_f(n, j - 1, l) \end{cases} \tag{6}$$

According to the above discussion, the decomposition processing of the original signal is depicted and illustrated in Figure 1.

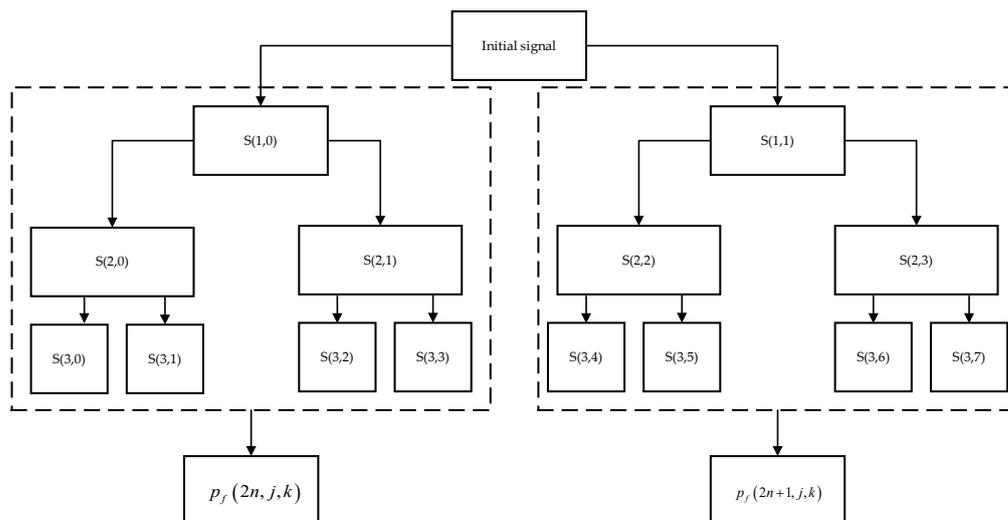


Figure 1. Wavelet decomposition for original signals.

Because of the integrity and orthogonality of the wavelet packet space, the original signal $f(n)$ is almost completely intact after wavelet decomposition, which provides conditions for analyzing signal characteristics.

According to the above definition of the orthogonal wavelet packet transform, the signal $f(n)$ has been projected adaptively into the orthogonal wavelet packet space; then, the obtained component can be regarded as the energy distributed in the corresponding space. If the energy distribution of signals in the space of each orthogonal wavelet packet can be calculated at a certain decomposition level, then the characteristics can be extracted by sorting these energies according to the frequency index of U_j^n . The energy distribution in the time-frequency localization space can be interpreted as follows:

$$E(j, n) = \sum_{k \in Z} [p_f(n, j, k)]^2 \tag{7}$$

Therefore, if the original signal $f(t)$ is decomposed by P levels, the energy feature vector extracted from the original signal can be expressed as follows:

$$E^*(P, f) = [E(P, 0), E(P, 1), \dots, E(P, 2^P - 1)] \tag{8}$$

3. Improved Fault Diagnosis Approach Using WMP SO-LSSVM

3.1. Least Squares Support Vector Machine

The literature of various fields shows that the LSSVM model performs well on various datasets, so it can process the data generated under unknown working conditions in complex industrial systems well. In addition, the complete theoretical basis of LSSVM can also ensure its stability. The principle of LSSVM is as follows:

$$\min_{w, b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \zeta_i^2 \tag{9}$$

$$s.t. y_i (w^T x_i + b) = 1 - \zeta_i, i = 1, 2, \dots, m \tag{10}$$

where $\{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$ are the samples to be observed, w is the perpendicular vector of the line, b is the offset of the hyperplane, C is the regularization parameter, and ζ_i represents the fluctuations in the error of each sample.

To obtain an accurate solution to the above optimal problem, the Lagrange function with slack variables can be established as follows:

$$L(w, b, \zeta, \alpha, \lambda) = \frac{1}{2} \|w\|^2 + C \sum_i \zeta_i^2 + \sum_i \alpha_i (1 - \zeta_i y_i (w^T \varphi_{lssvm}(x_i) + b)) - \sum_i \lambda_i \zeta_i \quad (11)$$

where α_i is the Lagrange multiplier of the original problem, and λ_i is the Lagrange multiplier of the additional slack variables.

Take the derivative of each variable in Formulas (9) and (10) and let them be 0. The following equalities hold:

$$\begin{cases} w = \sum_i^n \alpha_i x_i y_i \\ \sum_{i=1}^n \alpha_i y_i = 0 \\ C - \alpha_i - \lambda_i = 0 \end{cases} \quad (12)$$

Thus, Formula (11) can be rewritten as follows:

$$L(\zeta, \alpha, \lambda) = \sum_i \alpha_i^2 + \sum_{i,j} \alpha_i \alpha_j y_i y_j X_i^T X_j \quad (13)$$

Therefore, the optimal problem of Formulas (9) and (10) can be expressed as follows:

$$\begin{cases} \max_{\alpha} W(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle \\ \text{s.t.} \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad (14)$$

Given the varying conditions of industrial systems, the vibration signal of equipment follows a nonlinear relationship. In order to solve the problem of linear indivisibility in primordial space, it is necessary to transform the failure samples into multi-dimensional distinguishable space by introducing kernel functions. Therefore, Formula (14) can be written as follows:

$$\begin{cases} \max_{\alpha} W(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j}^n y^{(i)} y^{(j)} \alpha_i \alpha_j k(x_i, x_j) \\ \text{s.t.} \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad (15)$$

where $k(x_i, x_j)$ is the kernel function, and the selection of the kernel function has great flexibility. The common kernel functions are described as follows:

1. Linear kernel function:

$$K(x_i, x_j) = x_i \cdot x_j \quad (16)$$

2. Polynomial kernel function:

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^l, l = 1, 2, \dots \quad (17)$$

3. Gaussian kernel function:

$$K(x_i, x_j) = \exp \left[-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right] \quad (18)$$

The Gaussian kernel function selected in this paper can effectively transform the data from the low-dimensional non-separable space to the high-dimensional separable space, and it can further improve the classification accuracy of the model. Another advantage of Gaussian kernels, compared to other kernels, is that the more complex the model, the stronger the performance. In addition, no matter how many dimensions are the

characteristics of each sample point, each sample can be transformed into the total sample quantity dimension after processing by the Gaussian kernel function, which expands the dimension and the diversity of data.

It is natural to notice that LSSVM's classification accuracy is closely related to the penalty factor and parameter σ of the kernel functions. If the kernel function is too small, there will be an over-fitting phenomenon in the classification; otherwise, there will be an under-fitting phenomenon. Similarly, the larger the penalty factor, the more likely it is to overfit; and the smaller the penalty factor is, the more likely it is to underfit. Thus, in order to improve the accuracy of fault diagnosis for industrial systems, an optimized approach, named WMPSO-LSSVM, is proposed in the next section.

3.2. WMPSO-Based Parameters Optimization of LSSVM

As mentioned above, the regularization parameter and kernel functions play an important role in LSSVM. Thus, in this paper, we adopt the proposed WMPSO algorithm to optimize the parameters and establish a desirable model with high classification accuracy. Firstly, the basic model of PSO is as follows:

$$C_i = m \times C_i + c_1 \times rand \times (gbest - \sigma_i) + c_2 \times rand \times (qbest - \sigma_i) \quad (19)$$

$$\sigma_i = \sigma_i + C_i \quad (20)$$

where C_i is the regulation parameter of the LSSVM as well as the current velocity of PSO, and σ_i is the kernel function of the LSSVM as well as the location of particles in PSO. m indicates the weight coefficient, c_1 and c_2 are learning factors, and $rand$ is a random number between 0 and 1. Meanwhile, $gbest$ and $qbest$ store the optimal values corresponding to the penalty coefficient C and the kernel parameter σ , respectively.

Suppose there is a group of particle swarms $S = (S_1, S_2, \dots, S_n)$ in an n -dimensional space; C and σ can be presented as follows:

$$C = (C_1, C_2, \dots, C_i) \quad (21)$$

$$\sigma = (\sigma_1, \sigma_2, \dots, \sigma_i) \quad (22)$$

In this paper, the wavelet function μ^* is used to conduct a random perturbation of all the dimensions of the contemporary optimal value $Q_g^m(t)$ particles, and the perturbation result is taken as the position of the particles. The calculation model is given as follows:

$$\bar{\sigma}^m(t) = \mu^* Q_g^m(t) \quad (23)$$

For the sake of the accuracy of the WMPSO algorithm, the Morlet function was selected as the wavelet base in this study, as shown in Figure 2.

The Morlet wavelet has more accurate and high-resolution spectral estimation, and has thus been widely used. Compared with the Gaussian and Cauchy variations often used in particle swarm optimizations, the Morlet wavelet searches more effectively in the solution space because there is an equal probability of producing positive and negative numbers.

In addition, the Morlet wavelet function changes the local solution more frequently in the solution space, and it is easier to obtain the optimal solution in the local optimization. The Morlet wavelet function can fine-tune the particle, so it is a remarkable choice to select the Morlet wavelet for mutation.

Thus, the wavelet function value applied is expressed as follows:

$$\mu^* = \frac{1}{\sqrt{a}} e^{-\left(\frac{\varphi^*}{a^*}\right)^2 / 2} \cos\left(5\left(\frac{\varphi^*}{a^*}\right)\right) \quad (24)$$

Meanwhile, the scale parameter a^* is calculated by Formula (25):

$$a^* = e^{-\ln(g) \times \left(1 - \frac{t}{T}\right)^{\gamma_{\omega m}}} + \ln(g) \quad (25)$$

where $\gamma_{\omega m}$ is the shape parameter, t is the current iteration number, T is the maximum number of iterations, and g is the limit of a^* .

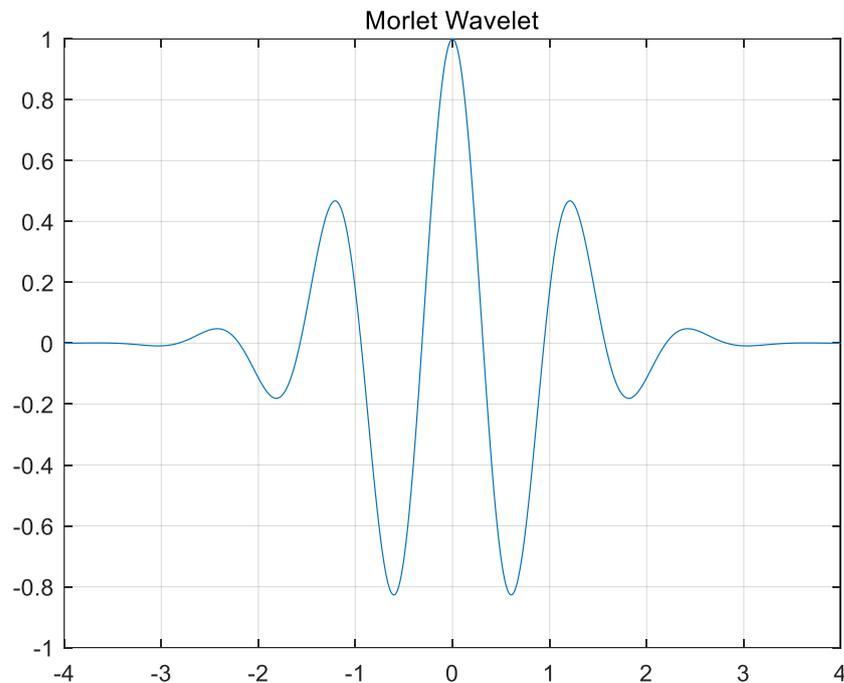


Figure 2. Morlet-wavelet function.

Therefore, after the perturbation by using the wavelet mutation function, the new positions of the particles are $\bar{\sigma}^m = (\bar{\sigma}_1^m, \bar{\sigma}_2^m, \dots, \bar{\sigma}_n^m)$. Once the position and kernel parameter σ are determined, the regularization factor C can be confirmed according to Formula (19). The optimization process for the parameters in this study is given in Algorithm 1:

Algorithm 1 The process of the WMP SO parameters' optimization

Initialize $\sigma_i \setminus \setminus \sigma_i$ is the position of the i_{th} particle

Calculate fitness function $\setminus \setminus$ Individual extreme values of particles can be calculated by fitness function

while $i \leq T$ **do** $\setminus \setminus T$ is the maximum number of iterations performed by the algorithm
 $i = i + 1$

for $j = 1$ to n **do**

 Update velocity C_i based on Equation (19)

 Update position σ_i based on Equation (20)

if $p_m > rand$ **then**

 Calculate a^* based on Equation (25)

$\varphi^* = 2.5 * a^* * rand(1, 30)$

 Calculate μ^* based on Equation (24)

 Update position σ_i based on Equation (23)

end if

 Calculate fitness function

 Update Q_i and Q_g

end for

end while

3.3. Design of WMP SO-LSSVM-Based Fault Diagnosis Scheme for Industrial Systems

Based on the above analysis, the WMP SO-LSSVM-based data-driven fault diagnosis approach is designed as follows:

1. Decompose the composite fault data of industrial systems based on the orthogonal wavelet packet algorithm and extract the fault characteristics;
2. Take the extracted characteristics as the input to the WMP SO-LSSVM identification model, training to obtain the regularization coefficient C and kernel parameter σ . The training process is summarized as follows:
 - Initialize the following parameters: the evolution algebra of the particles, the learning factors c_1 and c_2 , the regularization factor C , the kernel parameter σ , and the historical optimal kernel parameter Q_σ ;
 - Calculate the new information of the C and σ , and update a new generation of the particles;
 - Calculate the fitness value of the particles according to the fitness function, and update the individual and global optimal values of C and σ on this basis;
 - Evaluate whether the maximum number of iterations or searching boundaries has been reached. If so, store the C and σ , and construct the WMP SO-LSSVM-based identification model;
3. Take the extracted characteristics as the input to the WMP SO-LSSVM identification model, testing to obtain the classification result.

The corresponding flowchart is presented in Figure 3.

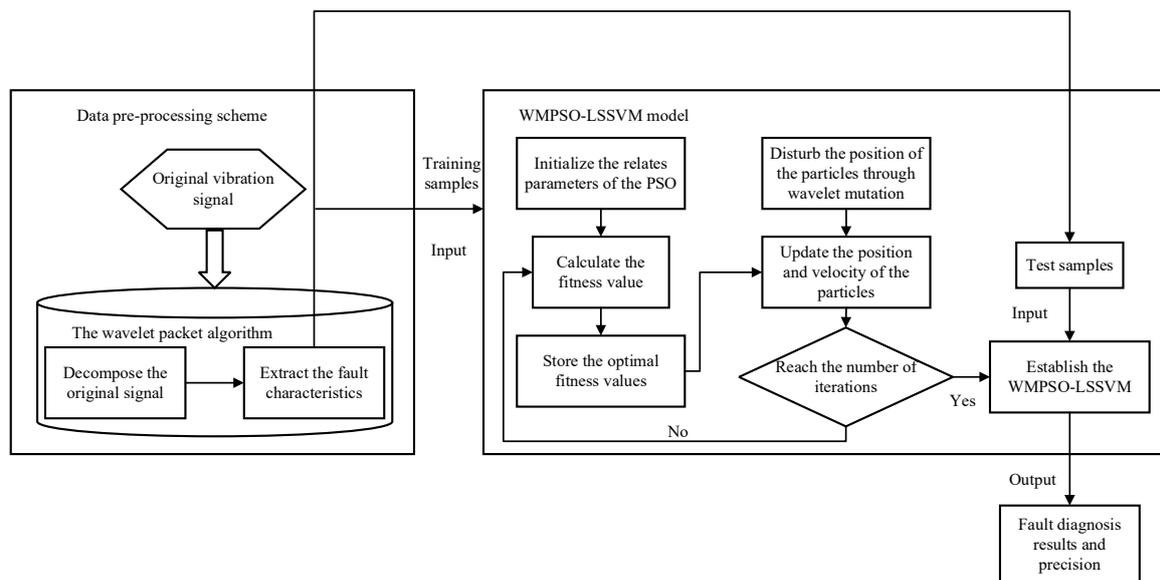


Figure 3. The flowchart of the proposed WMP SO-LSSVM algorithm.

4. Experimental Applications for Industrial Systems Based on WMP SO-LSSVM

The effectiveness and superiority of this study for industrial systems are evaluated on a database taken from the Guangdong Provincial Key Laboratory of Petrochemical Equipment Fault Diagnosis of China. Meanwhile, some comparative experiments are used to further prove the fault diagnosis performance of the proposed method.

As shown in Figure 4, the industrial system studied in this section is the main fan motor of a steam turbine, and the specific research object of this system is the gearbox containing the rolling bearings. The actual data of the gearbox and bearings are obtained from the intelligent fault diagnosis system, which consists of an acceleration sensor, a preamplifier (PMP), an explosion-proof BOX (BOX), a data collector (butylated hydroxytoluene), and a server (PC-1).

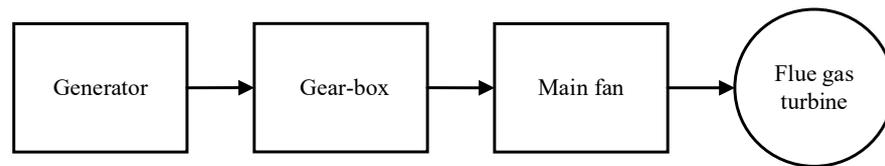


Figure 4. Schematic diagram of the main fan system.

In addition, the acceleration sensor is installed on the generator to obtain the vibration signals; the role of the BOX is to protect the preamplifier; the preamplifier is installed in the BOX for signal amplification; the data collector is installed in the steam turbine of the main fan for signal acquisition and processing; and the server is used for data storage and management.

The accelerometer used to measure the vibration acceleration mainly contains the following information. The highest amplitude is 50 g, the channel number is 6, the maximum transmission distance is 300 m, the working power supply is 18–30 VDC, and the working current is constant (2–10 mA). The actual industrial system operation environment and data collection situation are shown in Figures 5–8.



Figure 5. The on-site industrial environment.



Figure 6. The local-data acquisition system.

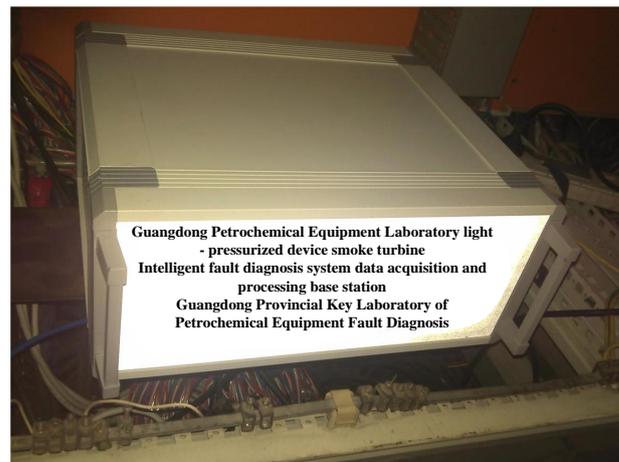


Figure 7. The data acquisition base station.



Figure 8. The data acquisition platform.

The data collected by the intelligent fault diagnosis system mainly include seven states, which are different fault combinations of gears and bearings. Their fault modes and corresponding indicators are shown in Table 1, and the waveforms of the part of the original vibration signals are shown in Figures 9–12.

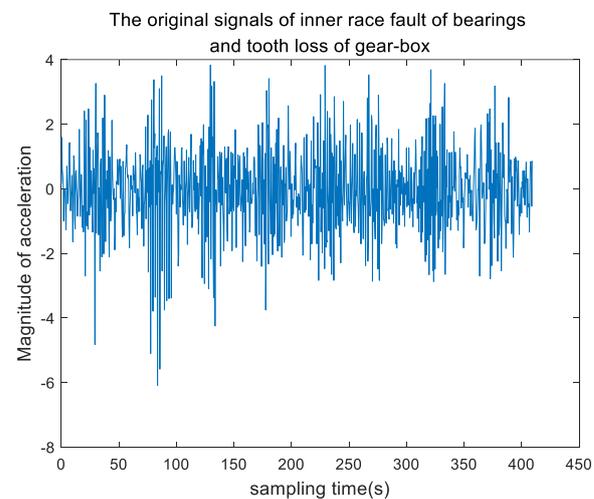


Figure 9. The original signals of the inner race fault of the bearings and the tooth loss of gear-box.

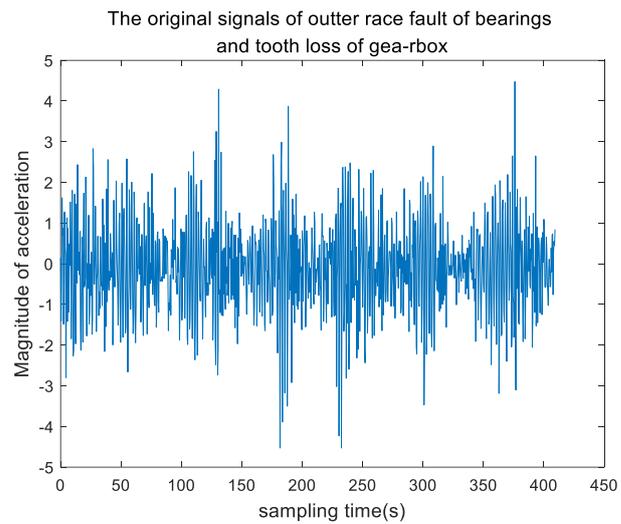


Figure 10. The original signals of the outer race fault of the bearings and the tooth loss of the gear-box.

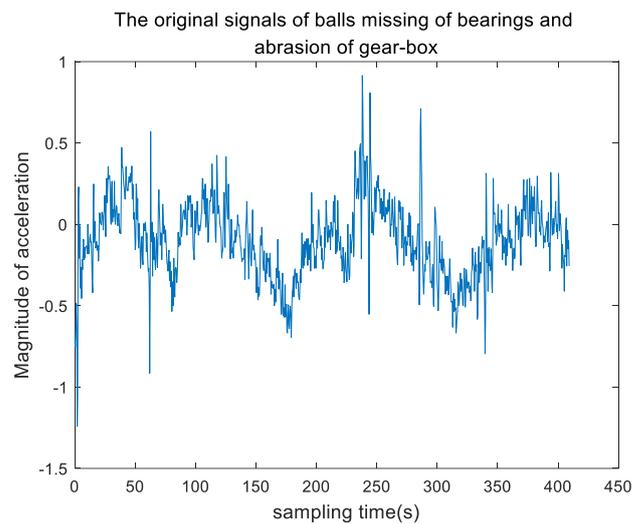


Figure 11. The original signals of balls that are missing bearings and the abrasion of the gear-box.

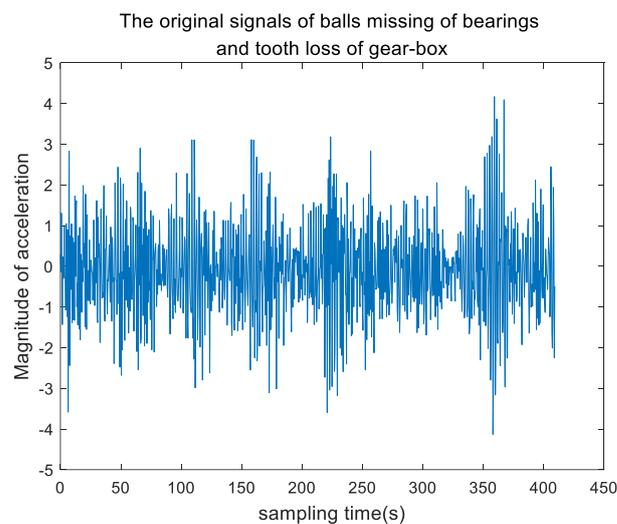


Figure 12. The original signals of balls that are missing bearings and the tooth loss of the gear-box.

Table 1. Seven fault states of the key components for entire systems.

Description of Seven States	Vibration Index	Impulsion Index	Tolerance Index	Peak Index	Kurtosis Index
State 1: missing gear teeth and outer ring wear of right bearing	1.1975 1.3132	2.5531 6.8919	2.9015 8.2115	2.1319 5.3947	3.0860 4.1036
State 2: missing gear teeth and lack of balls on left bearing	1.2293 1.2920	3.1451 4.9894	3.6689 5.9483	2.5414 3.9279	2.7140 3.5757
State 3: missing gear teeth and outer ring wear on left bearing	1.2657 1.3558	4.3240 7.5935	5.1791 9.1797	3.3671 5.7598	3.4370 5.4632
State 4: missing gear teeth and inner ring wear on right bearing	1.2438 1.3082	3.2264 5.6916	3.7968 6.8665	2.5912 4.3945	2.8526 4.3278
State 5: wear of gear and inner ring wear on left bearing	1.2252 1.3433	2.2448 4.2110	2.6442 4.9972	1.8041 3.3652	2.3961 4.6594
State 6: wear of gear and lack of balls on left bearing	1.2257 1.3227	2.6885 5.3905	3.3278 6.7998	2.4035 4.1221	2.7392 8.0007
State 7: wear of gear and outer ring wear on left bearing	1.3007 1.3742	4.3120 7.4453	5.1996 9.0964	3.3152 5.5460	3.6755 5.4385

The numbers in bold in the table represent the time domain index of the faulty component. Look at the numbers in the table. If the data in the table appears to be significantly asynchronous, this can be used to distinguish component failures. Taking the waveform indicator as an example, 1.2920 is obviously out of sync with all the numbers in the second row of the waveform column, and 1.3007 is also out of sync with the numbers in the first row of the waveform column, so it can be used as the basis for division.

Therefore, according to the indicators in bold in Table 1, the following analysis can be obtained.

- States 2 and 7 can be distinguished via the vibration index;
- States 3 and 5 can be distinguished via the impulsion and tolerance indices;
- States 5 and 7 can be distinguished via the impulsion and tolerance indices;
- States 3 and 5 can be distinguished via the peak index;
- States 2 and 7 can also be distinguished via the kurtosis index, as can states 2 and 3.

Then, the original signals are decomposed into three layers using the wavelet packet decomposition algorithm, and the node coefficients are calculated according to Formula (5). The corresponding results are given in Figure 13. In addition, the wavelet packet coefficients of the third layer, consisting of nodes 7 to 14 and calculated according to Formula (6), are shown in Figure 14.

The spectral distributions of the non-stationary vibration signals of the gearbox and bearings are closely related to their characteristic structures. Therefore, the energy distributions in the wavelet packet space of the original vibration signals decomposed by the wavelet packet are the fault features of the gears and bearings to be extracted. The parts of the characteristic extraction results are shown in Figure 15.

Finally, by using 75% of the extracted fault features as the input to establish the optimal WMPPO-LSSVM and by inputting the test samples into the model, the classification results can be obtained. The experimental results of LSSVM, PSO-LSSVM, and WMPPO-LSSVM are given in Figures 16–18, respectively.

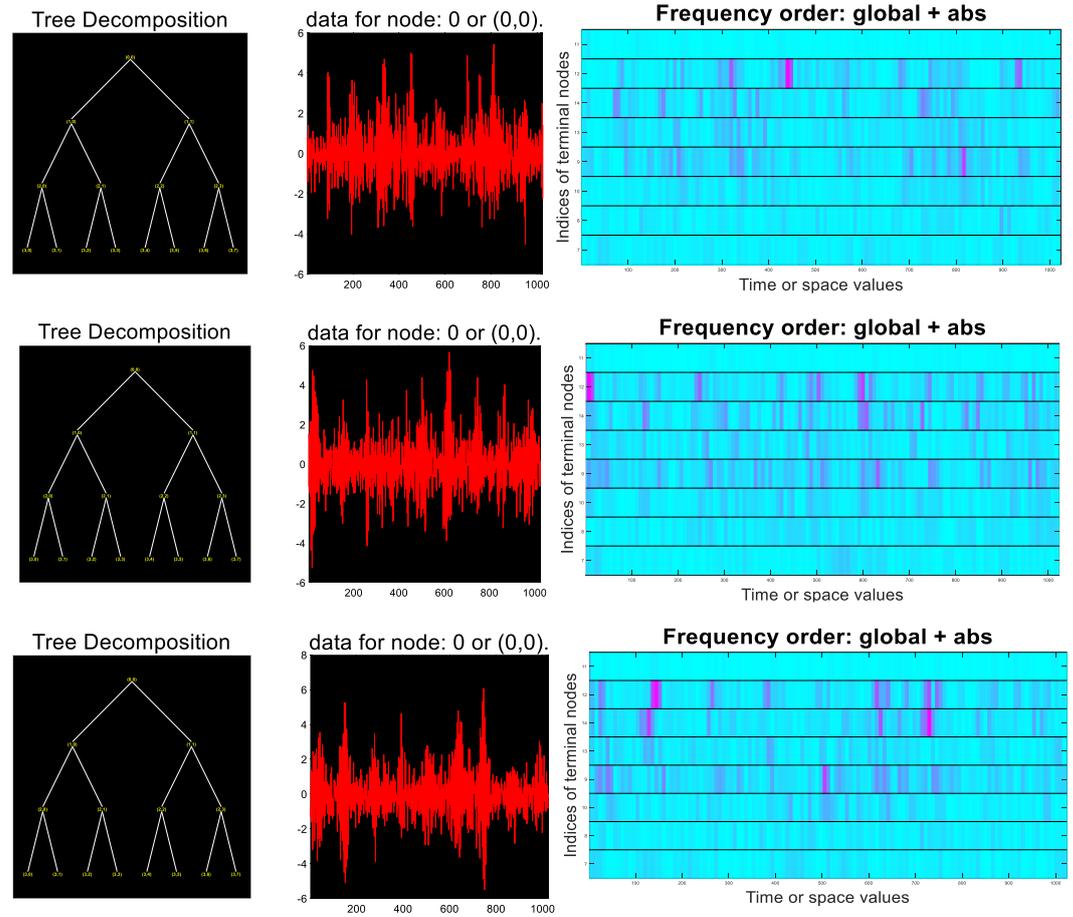


Figure 13. The decomposition results of the vibration signals.

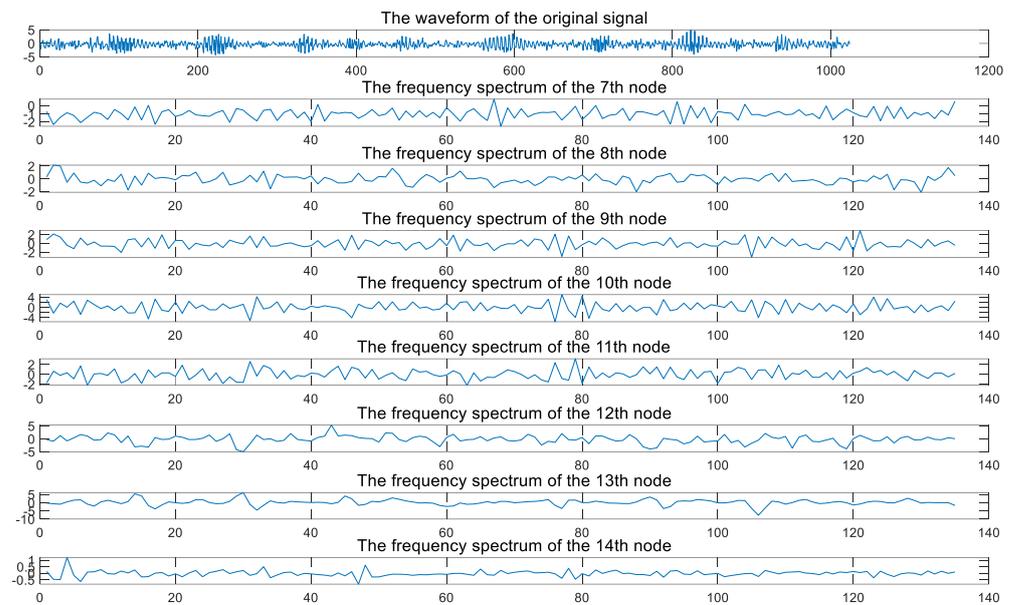


Figure 14. The node coefficients of the wavelet-packet algorithm.

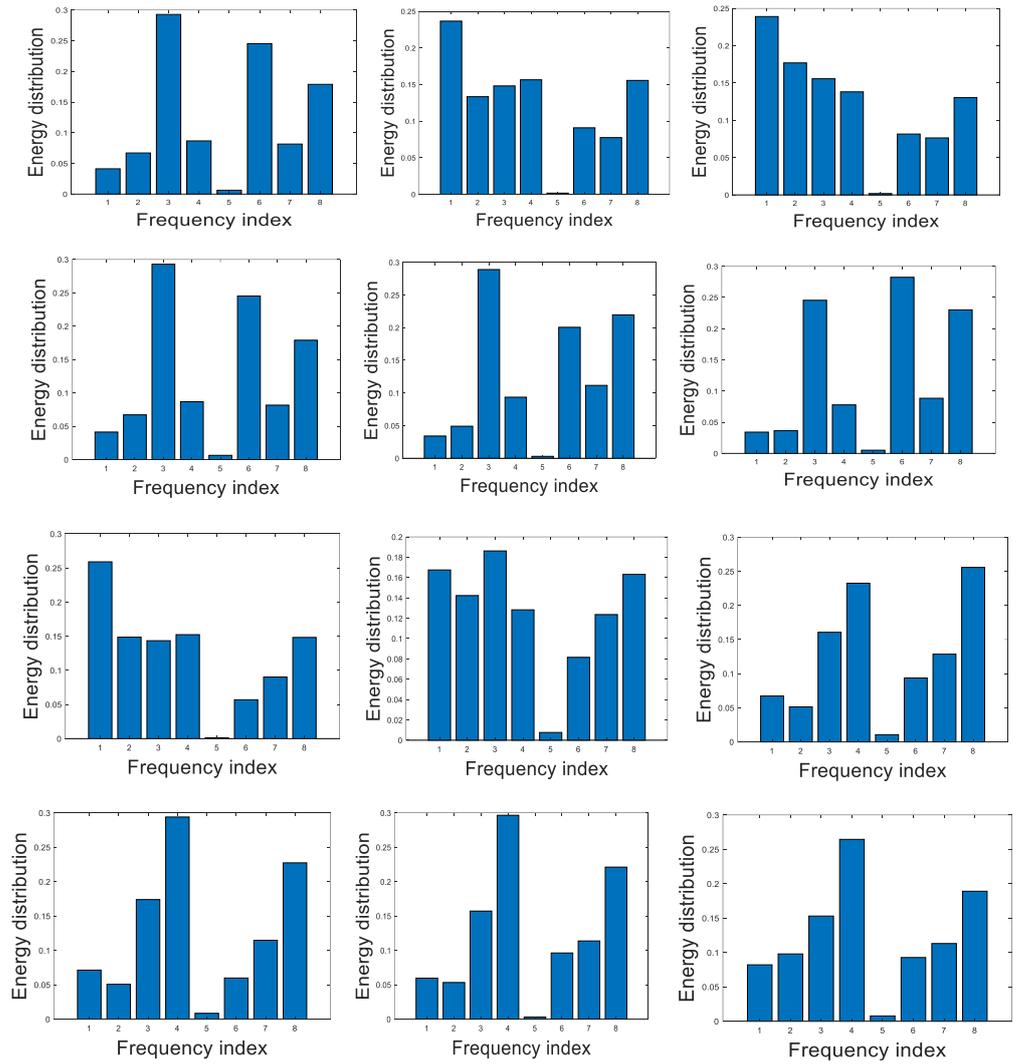


Figure 15. The fault characteristic extraction results of the gear-box and bearings.

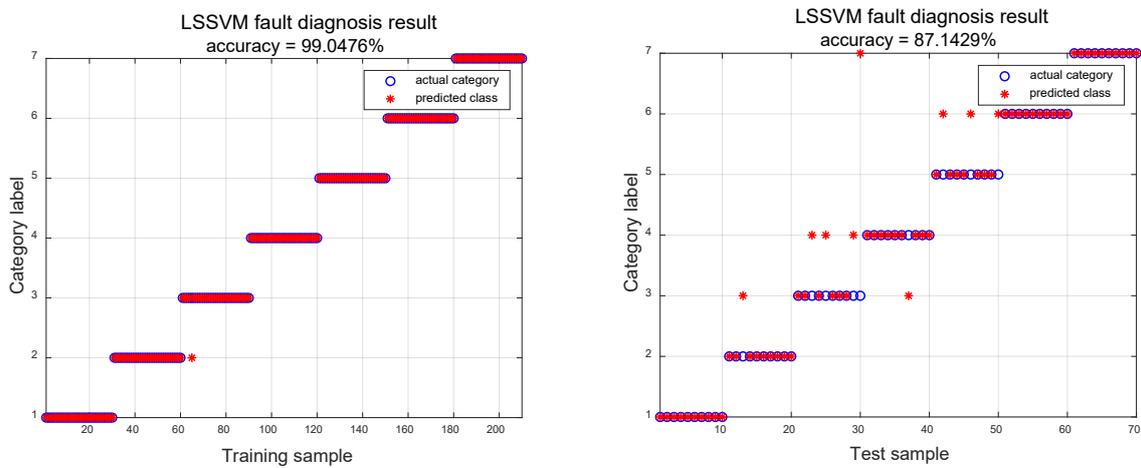


Figure 16. The classification results of LSSVM.

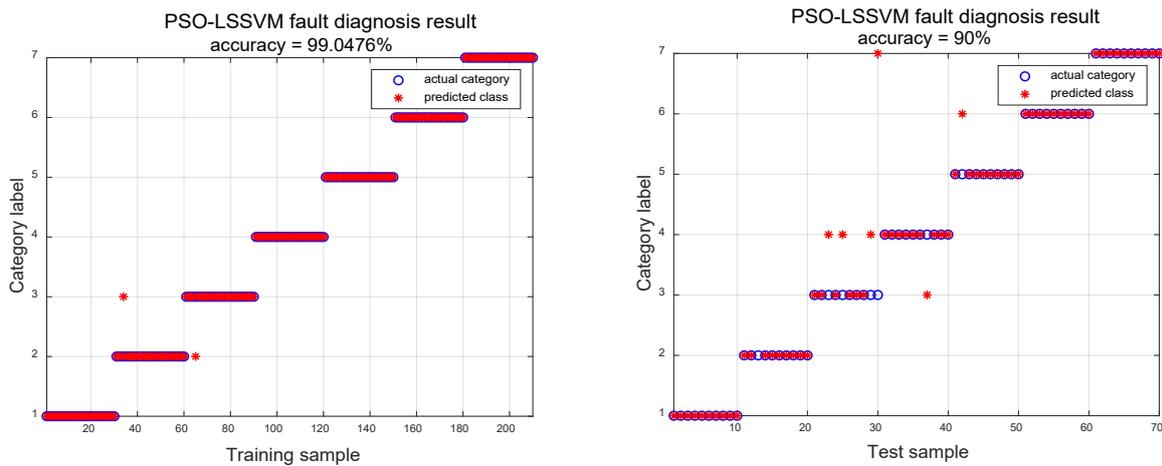


Figure 17. The classification results of PSO-LSSVM.

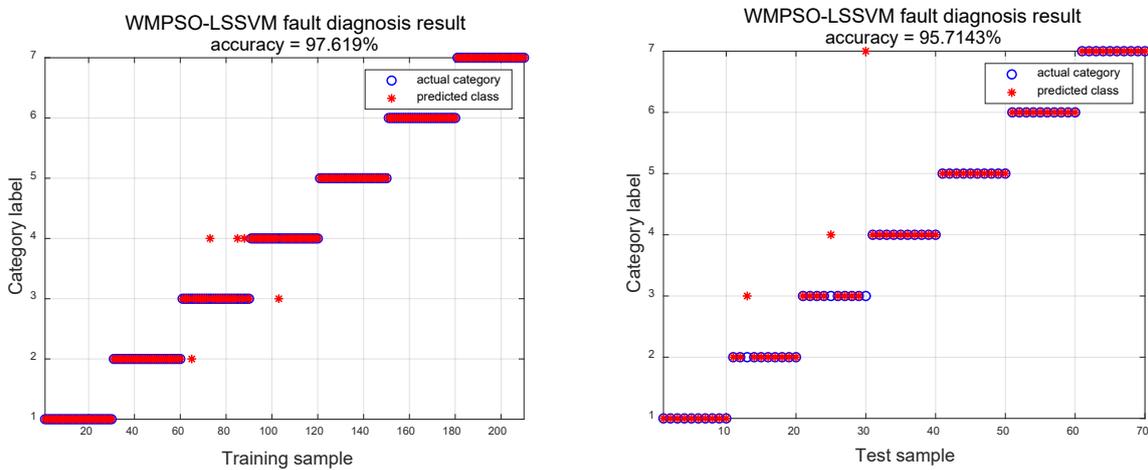


Figure 18. The classification results of WMPPO-LSSVM.

In order to further verify the superiority of the WVPPO-LSSVM classification model for key components of industrial systems, ELM and the traditional BP network are used for comparison purposes; the experimental results are shown in Table 2 and Figures 19–22.

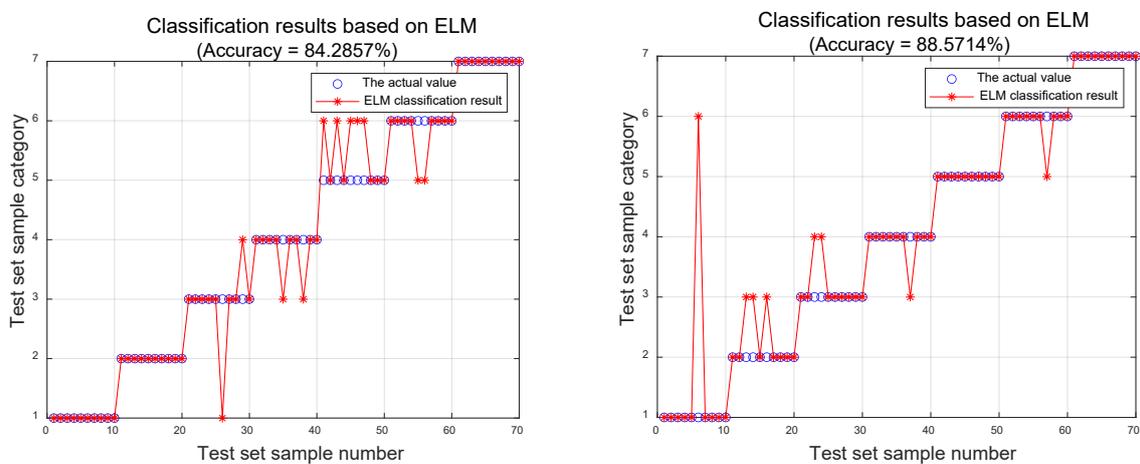


Figure 19. Fault diagnosis results based on ELM (1).

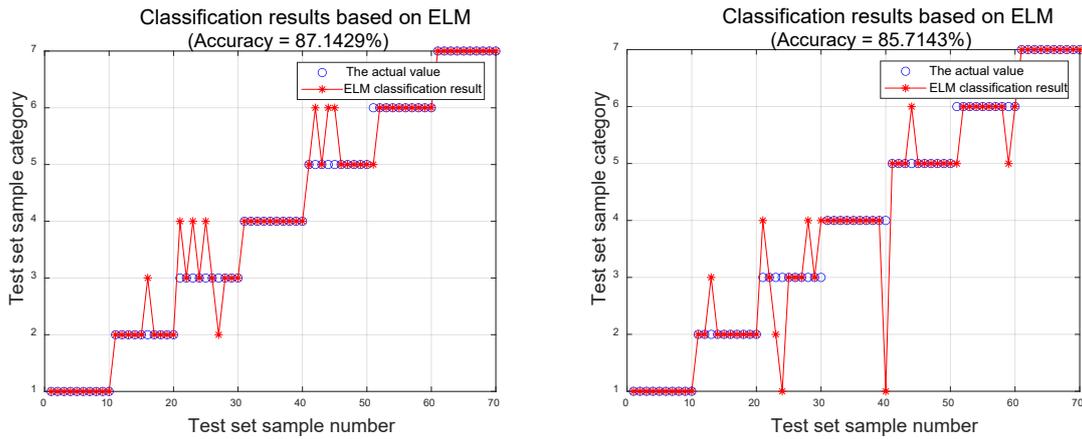


Figure 20. Fault diagnosis results based on ELM (2).

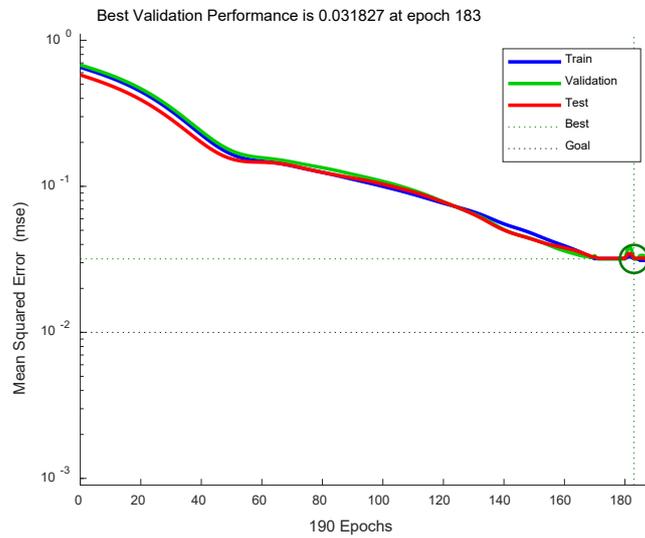


Figure 21. Training performance of the neural-networks.

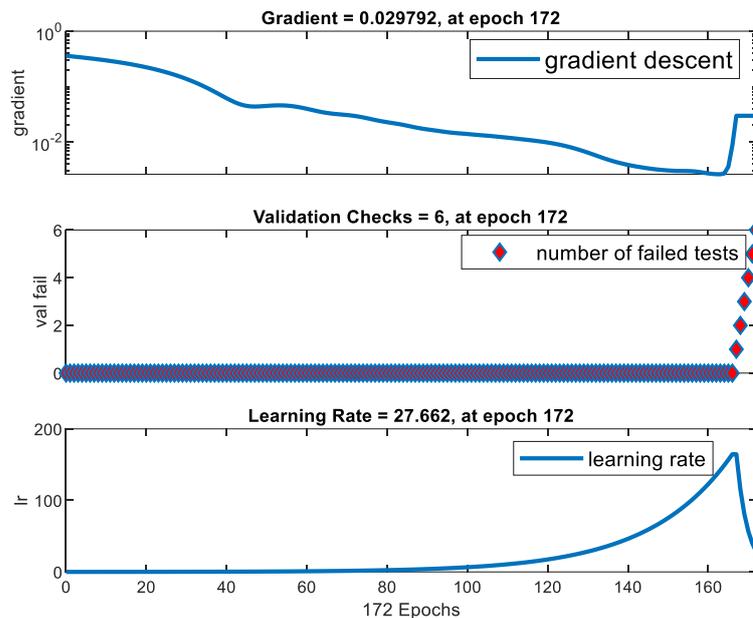


Figure 22. Training state of the neural-networks.

Table 2. This table contrasts the results of the three mechanisms.

Classification Method	BP	ELM	LSSVM	PSO-LSSVM	WMP SO-LSSVM
Classification accuracy (%)	64.29	86.50	84.17	90.00	95.71

To evaluate the performance of the WMP SO-LSSVM classification model, the confusion matrices of the WMP SO-LSSVM and ELM are presented, respectively, in Figures 23 and 24.

In the Figures 23 and 24, the blue square represents the number of correctly classified samples, while the pink square represents the number of incorrectly classified samples. For example, in Figure 23, there is only one incorrectly classified sample for the second type, and the remaining nine are correctly classified. The more diagonally distributed samples in the matrix, the better the performance of the model. And according to the results, the WMP SO-LSSVM has a higher precision than ELM.

In order to further verify the effectiveness of the proposed algorithm, the corresponding WMP SO-LSSVM regression model for the bearings and gearbox is established, and the composite fault characteristic trend is predicted. The comparative results are shown in Figures 25–28 and Table 3.

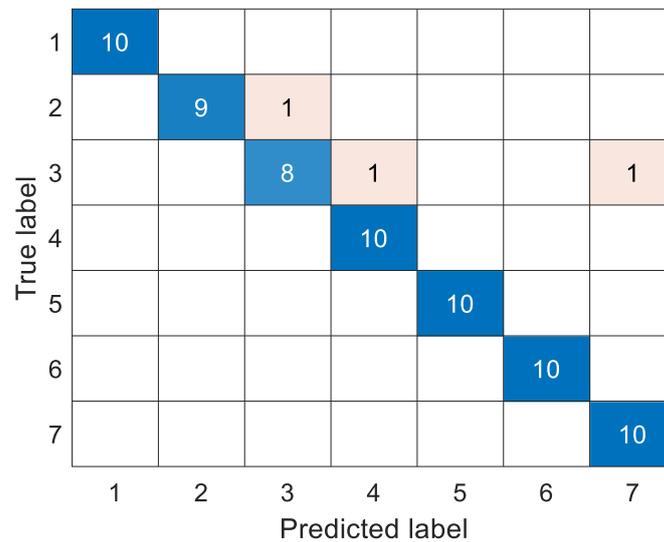


Figure 23. The confusion matrix of the WMP SO-LSSVM model.

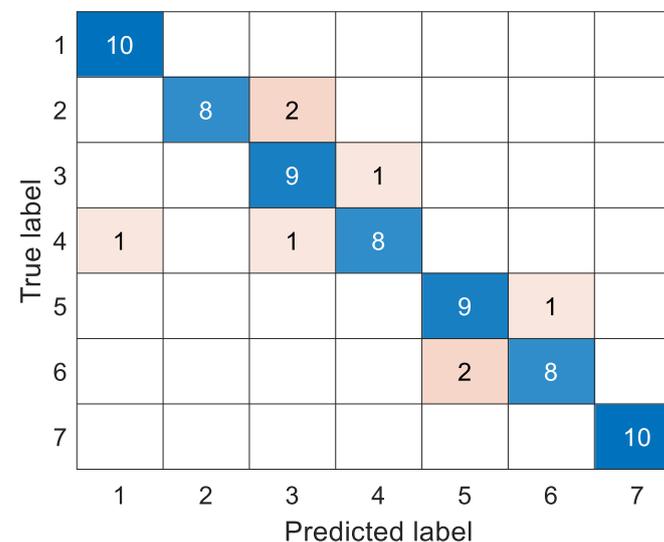


Figure 24. The confusion matrix result of the ELM model.

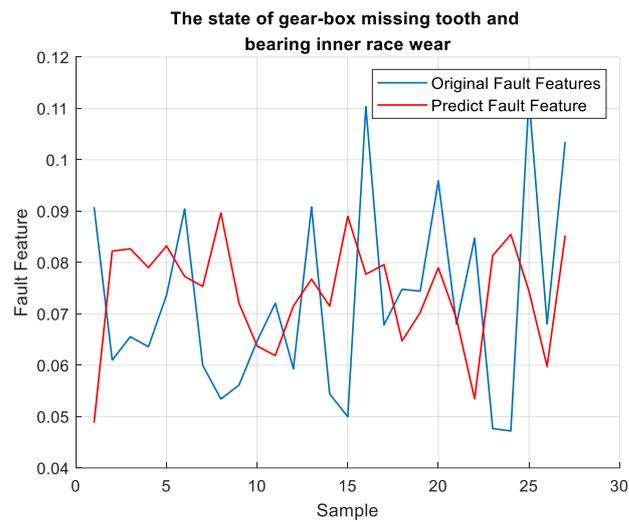


Figure 25. Bearing inner ring wear and gear tooth loss.

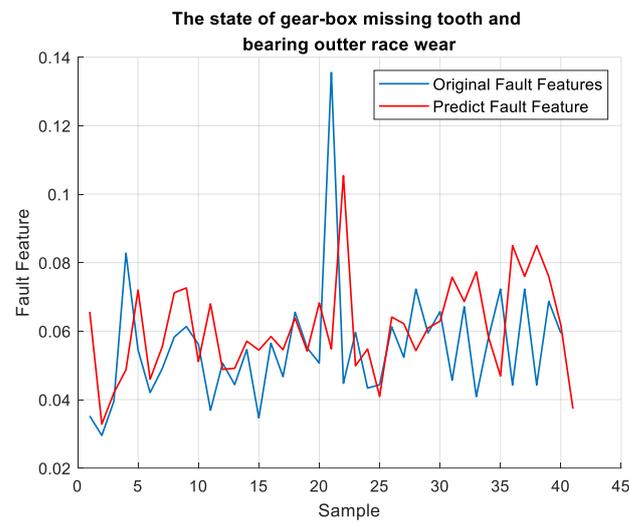


Figure 26. Bearing outer ring wear and gear tooth loss.

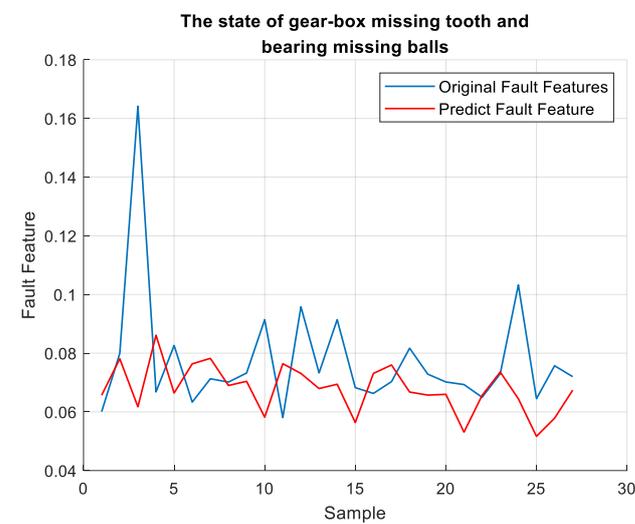


Figure 27. Bearing missing balls and gear tooth loss.

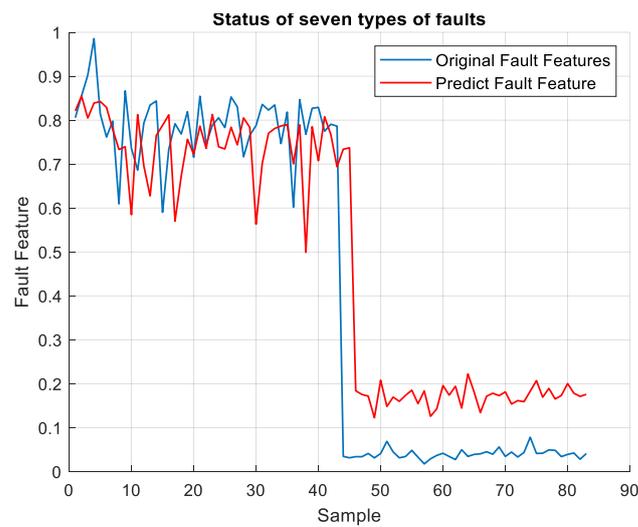


Figure 28. The state of seven types of faults.

Table 3. Comparison between the WMP SO-LSSVM regression model and the linear regression model.

Fault Types	Method	
	WMP SO-LSSVM	Linear Regression
Bearing inner ring wear and gear tooth loss	0.0707154	0.411682
Bearing outer ring wear and gear tooth loss	0.00146932	0.2976
Bearings missing balls and gear tooth loss	0.00260635	0.545191
Seven types of fault features	0.0224879	0.304906

Since the weight and the deviation of ELM are randomly generated, the inconsistent networks generated each time will eventually lead to a large performance difference, although the learning speed of ELM is fast and its generalization performance is good. Furthermore, because the BP neural network is a gradient descent method, its optimized objective function is extremely complex, and there will be a zig-zag phenomenon in the training process, which makes the BP algorithm inefficient. The accuracy of the BP neural network also depends largely on the sample size, and the number of fault samples obtained from industrial systems is small. Thus, it is not suitable for limited fault data of complex industrial systems.

In addition, it can be seen from the comparative experimental results that the WMP SO-LSSVM model has strong performance. The introduction of the Gaussian kernel function in WMP SO-LSSVM can expand the diversity and dimension of limited data and solve the defect of traditional neural networks' unsuitability for small samples. At the same time, the model can not only classify complex fault data effectively, but can also predict the complex fault characteristic trend, which has good applicability to complex fault data in industrial systems.

5. Conclusions

In this research, aiming to address the difficulty of the low precision of fault diagnosis methods for industrial systems, a new fault diagnosis methodology, named WMP SO-LSSVM, is proposed. Based on the decomposition of fault signals for feature extraction, the gearbox and bearings derived from the composable components are taken as the specific objects, and the vibration can be decomposed without information loss based on WPT. By comparing the proposed method with the existing pattern recognition methods, the results show that the WMP SO-LSSVM method can achieve higher classification accuracy for multiple fault modes in industrial systems.

In addition, PSO optimized by the wavelet mutation is combined with the LSSVM algorithm to realize the further optimization of the regularization parameter and kernel

function in the LSSVM, thereby improving the fault diagnosis accuracy. Particles that jump out of the local extreme value through the wavelet mutation algorithm will seek the optimal solution of parameters in the global space, so the optimal hyperplane of the LSSVM model can be established. As demonstrated via the comparative experiments, the accuracy of the WMP SO-LSSVM is almost 12% higher than that of the LSSVM, and is 9% higher than the ELM; moreover, the average error of the regression is 0.365 less than that of the traditional linear regression model, implying the potency of this scheme.

However, how to better select the parameters in the wavelet mutation function adaptively is not yet resolved in this work. Further research on the optimization of parameters in wavelet mutation is warranted.

In summary, the WMP SO-LSSVM proposed in this paper can significantly improve the fault diagnosis accuracy for complex industrial systems, and therefore, it offers better operability and scalability in the actual industrial environment.

Author Contributions: S.G. (Shuyue Guan): writing-original draft, methodology, investigation; D.H.: funding acquisition, supervision; S.G. (Shenghui Guo): writing—review & editing, methodology; L.Z.: methodology; H.C.: writing—review & editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of P.R. China, under Grants 61663008 and 62073051, the Chong-qing Technology Innovation, Application Special Key Project, under Grant cstc2019jscx-mbdxX0015, and the 2018 Reliable control and safety maintenance of dynamic system under Grant JDDSTD2018001.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data is provided by the Guangdong Provincial Key Laboratory of Petrochemical Equipment Fault Diagnosis of China for providing the dataset of the rolling bearings.

Acknowledgments: The authors would like to thank the Guangdong Provincial Key Laboratory of Petrochemical Equipment Fault Diagnosis of China for providing the dataset of the rolling bearings.

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

Notations

$L^2(R)$	square-integrable space
W_j	wavelet subspace
$\psi(t)$	wavelet function in wavelet packet algorithm
V_j	scale subspace
U_j^n	Hilbert space
$u_n(t)$	orthogonal wavelet packet basis
$h(k)$	low-pass filter coefficients
$g(k)$	high-pass filter coefficients
$\phi(t)$	scale function in wavelet packet
$f(n)$	original signal
$p_f(n, j, k)$	a sequence of transformation coefficients in wavelet packet
$E(j, n)$	energy distribution
w	the perpendicular vector in LSSVM
b	an offset of the hyperplane in LSSVM
C	regularization parameter in LSSVM
ζ	the fluctuations of the error in LSSVM
α	Lagrange multiplier of the original problem
λ	Lagrange multiplier of the additional slack variables
$K(x_i, x_j)$	kernel function
C_i	the velocity of the i th particle
σ_i	the position of the i th particle

σ	kernel parameter of the Gaussian kernel function
m	weight coefficient in PSO
c_1	learning factor in PSO
c_2	learning factor in PSO
$rand$	random number uniformly distributed in [0, 1]
$gbest$	the best particle that indicates the global best
$qbest$	the best particle that indicates the local best
S	particle swarm
μ^*	wavelet function in the mutation wavelet algorithm
a^*	scale parameter in the mutation wavelet algorithm
γ_{wm}	shape parameter
t	the current iteration number
T	the maximum number of iterations
g	limit of scale parameter
$\bar{\sigma}^m$	the new position of the disturbed particle
p_m	the mutation rate
$Q_g^m(t)$	the global best of the i th particle
φ^*	wavelet function basis in Morlet
Q_g	the best particle that indicates the global best of the disturbed particle
Q_i	the best particle that indicates the individual best of the disturbed particle
Q_σ	the historical optimal kernel parameter

References

1. Yuan, H.D.; Wu, N.L.; Chen, X.Y.; Wang, Y. Fault Diagnosis of Rolling Bearing Based on Shift Invariant Sparse Feature and Optimized Support Vector Machine. *Machines* **2021**, *9*, 98. [\[CrossRef\]](#)
2. Espinoza-Sepulveda, N.; Sinha, J. Mathematical Validation of Experimentally Optimized Parameters Used in a Vibration-Based Machine-Learning Model for Fault Diagnosis in Rotating Machines. *Machines* **2021**, *9*, 155. [\[CrossRef\]](#)
3. Nguyen, V.; Hoang, D.T.; Tran, X.T.; Van M.; Kang, H.J. A Bearing Fault Diagnosis Method Using Multi-Branch Deep Neural Network. *Machines* **2022**, *9*, 345. [\[CrossRef\]](#)
4. Chen, H.; Chai, Z.; Jiang, B.; Huang, B. Data-driven fault detection for dynamic systems with performance degradation: A unified transfer learning framework. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 3504712. [\[CrossRef\]](#)
5. Ran, G.; Liu, J.; Li, C.; Lam, H.-K.; Li, D.; Chen, H. Fuzzy-Model-Based Asynchronous Fault Detection for Markov Jump Systems with Partially Unknown Transition Probabilities: An Adaptive Event-Triggered Approach. *IEEE Trans. Fuzzy Syst.* **2022**, 1–10. [\[CrossRef\]](#)
6. Gao, Z.W.; Cecati, C.; Ding, S.X. A Survey of Fault Diagnosis and Fault-Tolerant Techniques-Part I: Fault Diagnosis with Model-Based and Signal-Based Approaches. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3757–3767. [\[CrossRef\]](#)
7. Chadha, G.S.; Panambilly, A.; Schwung, A.; Ding, S.X. Bidirectional deep recurrent neural networks for process fault classification. *ISA Trans.* **2021**, *106*, 330–342. [\[CrossRef\]](#)
8. Wang, L.; Zhang, Z.J.; Long, H.; Xu, J.; Liu, R. Wind Turbine Gearbox Failure Identification with Deep Neural Networks. *IEEE Trans. Ind. Inform.* **2017**, *13*, 1360–1368. [\[CrossRef\]](#)
9. Chen, H.T.; Li, L.L.; Shang, C.; Huang, B. Fault Detection for Nonlinear Dynamic Systems with Consideration of Modeling Errors: A Data-Driven Approach. *IEEE Trans. Cybern.* **2022**, 1–11. [\[CrossRef\]](#)
10. Javaid, A.; Javaid, N.; Wadud, Z.; Saba, T.; Sheta, O.E.; Saleem, M.Q.; Alzahrani, M.E. Machine Learning Algorithms and Fault Detection for Improved Belief Function Based Decision Fusion in Wireless Sensor Networks. *Sensors* **2019**, *19*, 1334. [\[CrossRef\]](#)
11. Moreno, A.P.; Santiago, O.L.; de Lazaro, J.M.B.; Moreno, E.G. Comparative Evaluation of Classification Methods Used in the Fault Diagnosis of Industrial Processes. *IEEE Lat. Am. Trans.* **2013**, *11*, 682–689. [\[CrossRef\]](#)
12. Ebrahimi, B.M.; Faiz, J. Diagnosis and performance analysis of three-phase permanent magnet synchronous motors with static, dynamic and mixed eccentricity. *IET Electr. Power Appl.* **2010**, *4*, 53–66. [\[CrossRef\]](#)
13. Zuo, M.J.; Xiang, G.; Hu, S. Fault diagnosis of the constant current remote power supply system in CUIs based on the improved water cycle algorithm. *India J. Geo-Mar. Sci.* **2021**, *50*, 914–921.
14. Zhao, Y.P.; Wang, J.J.; Li, X.Y.; Peng, G.J.; Yang, Z. Extended least squares support vector machine with applications to fault diagnosis of aircraft engine. *ISA Trans.* **2020**, *97*, 189–201. [\[CrossRef\]](#)
15. Heidari, M.; Homaei, H.; Golestanian, H.; Heidari, A. Fault diagnosis of gearboxes using wavelet support vector machine, least square support vector machine and wavelet packet transform. *J. Vibroeng.* **2016**, *18*, 860–875.
16. Milosevic, N.; Rackovic, M. Classification based on Missing Features in Deep Convolutional Neural Networks. *Neural Netw. World* **2019**, *29*, 221–234. [\[CrossRef\]](#)
17. Chen, H.; Chen, Z.; Chai, Z.; Jiang, B.; Huang, B. A single-side neural network-aided canonical correlation analysis with applications to fault diagnosis. *IEEE Trans. Cybern.* **2021**, 1–13. [\[CrossRef\]](#)

18. Chen, H.T.; Chai, Z.; Dogru, O.; Jiang, B.; Huang, B. Data-Driven Designs of Fault Detection Systems via Neural Network-Aided Learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *1–12*. [[CrossRef](#)]
19. Landauskas, M.; Cao, M.; Ragulskis, M. Permutation entropy-based 2D feature extraction for bearing fault diagnosis. *Nonlinear Dyn.* **2020**, *102*, 1717–1731. [[CrossRef](#)]
20. Kherif, O.; Benmhamed, Y.; Teguair, M.; Boubakeur, A.; Ghoneim, S.S. Accuracy Improvement of Power Transformer Faults Diagnostic Using KNN Classifier with Decision Tree Principle. *IEEE Access* **2021**, *9*, 81693–81701. [[CrossRef](#)]
21. Liu, Y.W.; Cheng, Y.Q.; Zhang, Z.Z.; Wu, J. Multi-information Fusion Fault Diagnosis Based on KNN and Improved Evidence Theory. *J. Vib. Eng. Technol.* **2022**, *10*, 841–852. [[CrossRef](#)]
22. Liu, T.Y.; Luo, H.; Kaynak, O.; Yin, S. A Novel Control-Performance-Oriented Data-Driven Fault Classification Approach. *IEEE Syst. J.* **2020**, *14*, 1830–1839. [[CrossRef](#)]
23. Ma, X.; Hu, Y.; Wang, M.H.; Li, F.; Wang, Y. Degradation State Partition and Compound Fault Diagnosis of Rolling Bearing Based on Personalized Multilabel Learning. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–11. [[CrossRef](#)]
24. Gan, X.; Lu, H.; Yang, G.Y.; Liu, J. Rolling Bearing Diagnosis Based on Composite Multiscale Weighted Permutation Entropy. *Entropy* **2018**, *20*, 821. [[CrossRef](#)] [[PubMed](#)]
25. Djeziri, M.A.; Djedidi, O.; Morati, N.; Seguin, J.L.; Bendahan, M.; Contaret, T. A temporal-based SVM approach for the detection and identification of pollutant gases in a gas mixture. *Appl. Intell.* **2021**, *52*, 6065–6078. [[CrossRef](#)]
26. Suykens, J.A.K.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process Lett.* **1999**, *9*, 293–300. [[CrossRef](#)]
27. Zhang, J.L. Fault Diagnosis of Nonlinear Analog Circuit Based on Generalized Frequency Response Function and LSSVM Classifier Fusion. *Math. Probl. Eng.* **2020**, *2020*, 8274570. [[CrossRef](#)]
28. Fu, Y.X.; Jia, L.M.; Qin, Y.; Yang, J. Product function correntropy and its application in rolling bearing fault identification. *Measurement* **2017**, *97*, 88–99. [[CrossRef](#)]
29. Zhang, K.; Su, J.P.; Sun, S.; Liu, Z.; Wang, J.; Du, M.; Liu, Z.; Zhang, Q. Compressor fault diagnosis system based on PCA-PSO-LSSVM algorithm. *Sci. Prog.* **2021**, *104*, 1–16. [[CrossRef](#)] [[PubMed](#)]
30. Xu, H.B.; Chen, G.H. An intelligent fault identification method of rolling bearings based on LSSVM optimized by improved PSO. *Mech. Syst. Signal Process.* **2013**, *35*, 167–175. [[CrossRef](#)]
31. Deng, W.; Yao, R.; Zhao, H.M.; Yang, X.; Li, G. A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm. *Soft Comput.* **2019**, *23*, 2445–2462. [[CrossRef](#)]
32. Yuan, H.D.; Chen, J.; Dong, G.M. Bearing Fault Diagnosis Based on Improved Locality-Constrained Linear Coding and Adaptive PSO-Optimized SVM. *Math. Probl. Eng.* **2017**, *2017*, 7257603. [[CrossRef](#)]
33. Zhang, D.D.; Xiang, W.G.; Cao, Q.W.; Chen, S. Application of incremental support vector regression based on optimal training subset and improved particle swarm optimization algorithm in real-time sensor fault diagnosis. *Appl. Intell.* **2020**, *51*, 3323–3338. [[CrossRef](#)]
34. Zhang, Q.F.; Chen, S.; Fan, Z.P. Bearing fault diagnosis based on improved particle swarm optimized VMD and SVM models. *Adv. Mech. Eng.* **2021**, *13*, 1–12. [[CrossRef](#)]
35. Liu, M.G.; Li, X.S.; Lou, C.Y.; Jiang, J. A Fault Detection Method Based on CPSO-Improved KICA. *Entropy* **2019**, *21*, 668. [[CrossRef](#)] [[PubMed](#)]
36. Wang, H.; Peng, M.J.; Hines, J.W.; Zheng, G.Y.; Liu, Y.K.; Upadhyaya, B.R. A hybrid fault diagnosis methodology with support vector machine and improved particle swarm optimization for nuclear power plants. *ISA Trans.* **2020**, *95*, 358–371. [[CrossRef](#)] [[PubMed](#)]