

Article

Bearing Fault Feature Extraction and Fault Diagnosis Method Based on Feature Fusion

Huibin Zhu ¹, Zhangming He ^{1,2}, Juhui Wei ¹, Jiongqi Wang ¹ and Haiyin Zhou ^{1,*}

¹ College of Liberal Arts and Sciences, National University of Defense Technology, Changsha 410073, China; zhu_huibin@163.com (H.Z.); hzmnudt@163.com (Z.H.); wei_juhui@163.com (J.W.); wjq_gfkd@163.com (J.W.)

² Beijing Institute of Spacecraft System Engineering, China Academy of Space Technology, Beijing 100094, China

* Correspondence: zhouhy@nudt.edu.cn; Tel.: +86-139-7316-5101

Abstract: Bearing is one of the most important parts of rotating machinery with high failure rate, and its working state directly affects the performance of the entire equipment. Hence, it is of great significance to diagnose bearing faults, which can contribute to guaranteeing running stability and maintenance, thus promoting production efficiency and economic benefits. Usually, the bearing fault features are difficult to extract effectively, which results in low diagnosis performance. To solve the problem, this paper proposes a bearing fault feature extraction method and it establishes a bearing fault diagnosis method that is based on feature fusion. The basic idea of the method is as follows: firstly, the time-frequency feature of the bearing signal is extracted through Wavelet Packet Transform (WPT) to form the time-frequency characteristic matrix of the signal; secondly, the Multi-Weight Singular Value Decomposition (MWSVD) is constructed by singular value contribution rate and entropy weight. The features of the time-frequency feature matrix obtained by WPT are further extracted, and the features that are sensitive to fault in the time-frequency feature matrix are retained while the insensitive features are removed; finally, the extracted feature matrix is used as the input of the Support Vector Machine (SVM) classifier for bearing fault diagnosis. The proposed method is validated by data sets from the time-varying bearing data from the University of Ottawa and Case Western Reserve University Bearing Data Center. The results show that the algorithm can effectively diagnose the bearing under the steady-state and unsteady state. This paper proposes that the algorithm has better fault diagnosis capabilities and feature extraction capabilities when compared with methods that are based on traditional feature technology.

Keywords: bearing fault diagnosis; feature extraction; wavelet packet transform; singular value decomposition; entropy weight method; support vector machine



Citation: Zhu, H.; He, Z.; Wei, J.; Wang, J.; Zhou, H. Bearing Fault Feature Extraction and Fault Diagnosis Method Based on Feature Fusion. *Sensors* **2021**, *21*, 2524. <https://doi.org/10.3390/s21072524>

Academic Editor: Yolanda Vidal

Received: 10 February 2021

Accepted: 25 March 2021

Published: 4 April 2021

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



Copyright: © 2021 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

Rotating machinery is one of the most common classes of mechanical equipment and it plays a significant role in industrial applications [1]. As one of the key components in rotating machinery, bearings health directly affects the performance of mechanical equipment [2,3]. According to incomplete statistics, approximately 30% of failures are caused by the bearing fault [4]. Therefore, the fault diagnosis of bearing is of great significance for maintaining the safe operation of equipment [5].

Normally, it cannot be directly diagnosed due to the working environment of the bearing. Sensors can be used to collect digital signals that can reflect the state of the bearing [6–10], such as spectral signals [11], acoustic signals [12], and vibration signals. Spectral signals and acoustic signals can be used for non-destructive flaw detection, and have the advantages of obvious characteristic frequency and good early fault prediction. However, these methods require high professional quality of equipment and operators. The vibration signal of the bearing contains a wealth of fault energy information [13,14], and the collection of the bearing vibration signal does not require complex equipment and

professionals. Therefore, fault diagnosis that is based on vibration signals is a common method for bearing diagnosis [15]. Vibration signals are affected by working conditions and equipment environment, the frequency spectrum is relatively complicated, and there are many interference factors. Therefore, the effective extraction of signal characteristics is the key to bearing fault diagnosis. The commonly used methods for extracting bearing signal features include empirical mode decomposition (EMD) and wavelet transform. EMD is an adaptive time-frequency analysis method without any prior knowledge, which has the ability of adaptive signal decomposition and noise reduction. However, EMD is only an empirical method and it lacks a complete theoretical basis [16]. Besides, in the decomposition process of EMD, modal aliasing is prone to occur due to problems, such as over-envelope, under-envelope, and unreasonable convergence conditions [17,18], which restricts the application of EMD. Wavelet packet transform (WPT) is a kind of wavelet transform. It can divide the frequency band of the signal into multiple scales to obtain information regarding signal in the low-frequency and high-frequency regions. Besides, WPT can adaptively select the corresponding frequency band to match the frequency spectrum of original signal according to the feature of signal, which has a more uniform frequency feature extraction effect [19,20]. Zhong et al. [21] used WPT to decompose the bearing signal, and the decomposed frequency band entropy is used as the input of Support Vector Machine (SVM) to establish a rolling bearing classification model.

The wavelet packet can extract the time-frequency information of the bearing vibration signal without omission, and more comprehensively describe the fault state of the bearing. However, on the one hand, it will increase the dimension of the bearing signal feature matrix and increase the computational complexity of the subsequent diagnosis model; on the other hand, there may be some insensitive features or even invalid features, increasing the probability that sensitive information will be submerged [22]. Therefore, after acquiring the bearing signal feature matrix, it is necessary to be further extracted to remove the irrelevant and redundant features. At present, the methods to remove redundant and irrelevant features of bearing include auto-encoder [23,24], neural networks [25,26], Principal Component Analysis (PCA) [27], kernel PCA [28], and Singular Value Decomposition (SVD). However, although intelligent algorithms, such as self-encoding and neural networks, have been applied to diagnose bearing faults, they have disadvantages, such as low generalization, slow calculation speed, and higher requirements for hardware equipment. For PCA and kernel PCA, on the one hand, PCA needs to be spatially transformed. Furthermore, the features of the original signal will lose their physical meaning through combination transformation; on the other hand, when using PCA, it is necessary to standardize the data. The noise in the data will affect the standardization process of data. SVD solves the dimensionality reduction order through the singular value of the matrix. When compared with PCA, the singular value has good stability and it is not sensitive to changes that are caused by interference, such as noise. It can still collect data information more accurately, even with small interference [29,30]. Kedadouche et al. [31] applied SVD to extract the matrix after WPT and use it as the input of SVM to identify the fault mode of rolling bearings. Cheng et al. [32] invented empirical mode decomposition to decompose the vibration signal of a rotating machine into multiple natural mode functions, and used SVD for the initial features matrix formed by these natural functions to obtain the singular values of matrix and used it for SVM fault diagnosis. Although SVD has good stability, as compared with PCA, the features extracted by SVD have relatively higher computational cost for subsequent diagnosis models. In view of this, Yuan et al. [33] proposed the Weighted Singular Value Decomposition (WSVD) with the ratio of singular values as the weight, and it is applied to radar emitter signals. The results showed that this method can extract the features of radar emitter signals very well. Although this method can effectively reduce the calculation cost of SVD, this method only tried to square the singular value after dimensionality reduction, which cannot fully reflect the information of the data itself and the importance of sensitive features.

This paper presents a study of the fault diagnosis method based on feature fusion when the bearing fault features are difficult to extract effectively which results in low diagnosis performance. Figure 1 shows the flowchart of the bearing fault diagnosis method based on feature fusion. The bearing vibration signal that is collected by the sensor obtains the time-frequency domain characteristics of the bearing through wavelet packet transform (WPT). This time-frequency domain feature is reduced dimension by the Multi-Weight Singular Value Decomposition (MWSVD). The reduced dimensionality features are used in SVM for fault diagnosis. The experimental results show the superiority of this method when compared to some of the traditional feature techniques. The major contributions of this paper include the following:

(1) a feature extraction method that is based on MWSVD is proposed and its effectiveness in two data set is evaluated. In the proposed method, the time-frequency domain information of the vibration signal that is extracted by WPT is best preserved in the low-dimensional space;

(2) the algorithm proposed in this paper is compared with some traditional feature extraction algorithms, combined with support vector machines for fault diagnosis, and the diagnosis effect is compared; and,

(3) a bearing fault diagnosis algorithm based on feature fusion is proposed, which can timely and effectively diagnose bearings in both steady state and non-steady state.

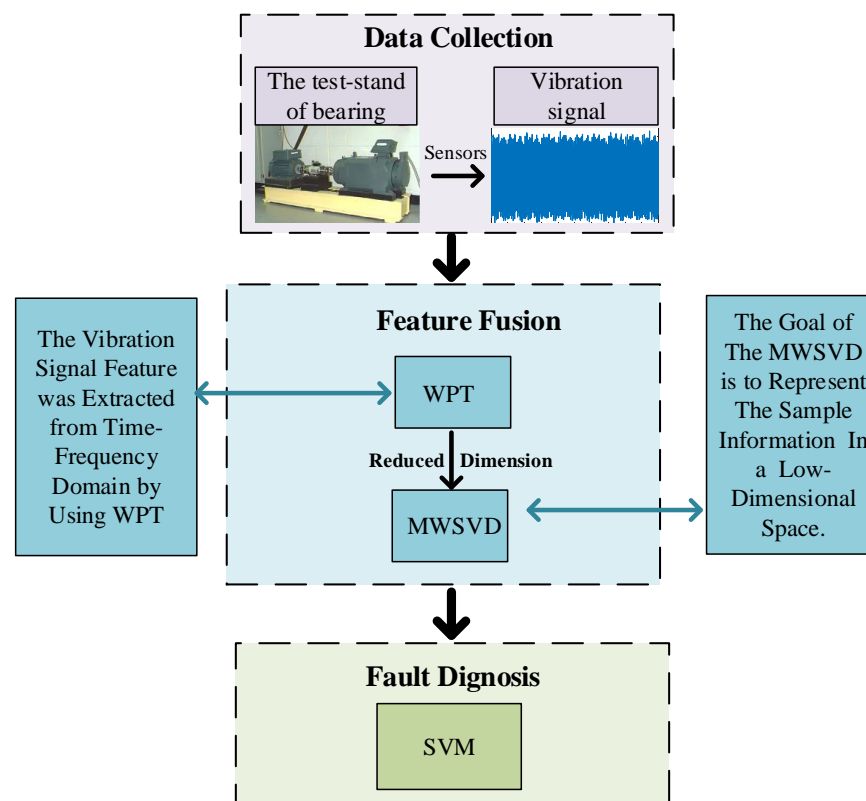


Figure 1. The flowchart of the bearing fault diagnosis method that is based on feature fusion.

The rest of this article is as follows: Section 2 introduces the Weighted Singular Value Decomposition (WSVD) algorithm. Section 3 describes the process of wavelet packet decomposition and weighted singular value decomposition, and it proposes a feature extraction method based on fusion multi-weight singular value decomposition. Section 4 proposes a fault diagnosis method that is based on feature fusion. Section 5 shows the fault diagnosis results of the two data sets and the comparison results with other methods. Section 6 draws the conclusion.

2. Weighted Singular Value Decomposition Method

The principle and steps of weighted singular value decomposition are as follows [33]: Firstly, the data can be normalized by

$$\hat{A}_i = \frac{2 \times (A_i - \bar{A}_i)}{|\max(A_i) - \min(A_i)|} \quad (1)$$

where A_i is the i th row data of matrix A , \hat{A}_i is the i th row data after data normalization, and \bar{A}_i is the mean value of A_i . Secondly, perform SVD decomposition according to the following equation

$$\hat{A} = U_{m \times m} \Sigma_{m \times s} V_{s \times s}^T \quad (2)$$

where \hat{A} is the normalized matrix of A , $\Sigma_{m \times s} = \begin{bmatrix} \Lambda_{s \times s} & O_{(m-s) \times s}^T \end{bmatrix}^T$, $O_{(m-s) \times s}$ is zero matrix, $\Lambda_{s \times s} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_s)$, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_s$ is singular value, $U_{m \times m}$ and $V_{s \times s}$ are the unitary matrix. The order $r < s$ after dimensionality reduction is determined by the cumulative contribution rate of singular value that is greater than 90%. Subsequently, $\Sigma_{m \times s}$ becomes $\Sigma_{r \times r} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$ after dimensionality reduction. The weight is calculated according to the elements in $\Sigma_{r \times r}$

$$w_i = \sigma_i / \sum_{i=1}^r \sigma_i, \quad i = 1, 2, \dots, r \quad (3)$$

Let the weight vector be $[w_i]_{1 \times r} = [w_1, \dots, w_r]$, according to the weight

$$D_{m \times r} = |U_{m \times r}| \otimes w_i = \begin{pmatrix} |u_{11}|w_1 & \dots & |u_{1r}|w_r \\ \vdots & \ddots & \vdots \\ |u_{m1}|w_1 & \dots & |u_{mr}|w_r \end{pmatrix} \quad (4)$$

where u_{ij} is the element of row i and column j of the matrix $U_{m \times r}$. The matrix $D_{m \times r}$ can be normalized according to the following equation to obtain the weighted singular value decomposition matrix B

$$B = \begin{pmatrix} d_{11}/\zeta_1 & \dots & d_{1r}/\zeta_1 \\ \vdots & \ddots & \vdots \\ d_{1m}/\zeta_m & \dots & d_{mr}/\zeta_m \end{pmatrix} \quad (5)$$

where $\zeta_i = \sum_j^r d_{ij}$, $i = 1, \dots, m$, d_{ij} is the element in row i and column j of the matrix $D_{m \times r}$.

3. Fault Feature Extraction Method Based on WPT-MWSVD

3.1. WPT

Suppose that \mathbf{Z} is the set of integers, $L^2(\mathbf{R})$ is a square-integrable real function space, and a series of closed subspace sequence $\{V_l\}_{l \in \mathbf{Z}}$ on $L^2(\mathbf{R})$ is called the multi-resolution analysis of space $L^2(\mathbf{R})$ if the following conditions are met:

- (1) Monotonicity: $V_{l+1} \subset V_l, l \in \mathbf{Z}$;
- (2) Translation invariance: $f(x) \in V_l \Leftrightarrow f(x - \alpha) \in V_l, \alpha \in \mathbf{Z}$;
- (3) Scalability: $f(x) \in V_l \Leftrightarrow f(2x) \in V_{l-1}$;
- (4) Approximation: $\bigcup_{l \in \mathbf{Z}} V_l = L^2(\mathbf{R}), \bigcap_{l \in \mathbf{Z}} V_l = \{0\}$;
- (5) The existence of Riesz base: $\{\varphi(x - k) | k \in \mathbf{Z}\}$ form the Riesz base of V_0 .

If $\{\varphi(x - k) | k \in \mathbf{Z}\}$ is a canonical orthonormal basis of V_0 in $\{V_l\}_{l \in \mathbf{Z}}$ for multi-resolution analysis of $L^2(\mathbf{R})$, then $\{\varphi_{l,k}(x) = 2^{-\frac{l}{2}} \varphi(2^{-l}x - k) | k \in \mathbf{Z}\}_{l \in \mathbf{Z}}$ is the canonical orthonormal basis of $\{V_l\}_{l \in \mathbf{Z}}$ namely $V_l = \text{span}\{2^{-\frac{l}{2}} \varphi(2^{-l}x - k)\}_{k \in \mathbf{Z}}$ [34].

Notice that $V_{l+1} \subset V_l$, if the orthogonal complement of V_{l+1} in V_l is W_{l+1} , that is $V_l = V_{l+1} \oplus W_{l+1}$, then W_{l+1} is called the wavelet subspace of $L^2(R)$

If W_0 is a orthogonal complement of V_0 on V_{-1} , then $\{\psi_{l,k}(x) = 2^{-\frac{l}{2}}\psi(2^{-l}x - k) | k \in \mathbf{Z}\}_{l \in \mathbf{Z}}$ is the orthogonal basis of W_l , that is $W_l = \text{span}\{2^{-\frac{l}{2}}\psi(2^{-l}x - k)\}_{k \in \mathbf{Z}}$ [35].

Suppose that the bearing signal $f(x)$ belongs to V_l , WPT can decompose $f(x)$ in the form of a binary tree. The principle of WPT can be described, as follows [34].

Suppose that $\{V_l\}_{l \in \mathbf{Z}}$ is a multi-resolution analysis of $L^2(R)$, $\varphi(x)$ and $\psi(x)$ are the corresponding orthogonal scaling function and orthogonal wavelet function, and the two-scale equations are satisfied

$$\begin{cases} \varphi(x) = \sqrt{2} \sum_{k \in \mathbf{Z}} h_k \varphi(2x - k) \\ \psi(x) = \sqrt{2} \sum_{k \in \mathbf{Z}} g_k \psi(2x - k) \end{cases} \quad (6)$$

where $h_k(\cdot)$ and $g_k(\cdot)$ are low-pass and high-pass filters, respectively. Let $\mu_0 = \varphi(x)$, $\mu_1 = \psi(x)$, then

$$\begin{cases} \mu_0(x) = \sqrt{2} \sum_{k \in \mathbf{Z}} h_k \mu_0(2x - k) \\ \mu_1(x) = \sqrt{2} \sum_{k \in \mathbf{Z}} g_k \mu_0(2x - k) \end{cases} \quad (7)$$

The above formula is extended to the general situation

$$\begin{cases} \mu_{2n}(x) = \sqrt{2} \sum_{k \in \mathbf{Z}} h_k \mu_n(2x - k) \\ \mu_{2n+1}(x) = \sqrt{2} \sum_{k \in \mathbf{Z}} g_k \mu_n(2x - k) \end{cases} \quad (8)$$

From Equation (8), the function set $\{\mu_n(t) : n = 0, 1, 2, \dots\}$ can be obtained that is called wavelet packet determined by the orthogonal scaling function $\varphi(x)$. The corresponding space of wavelet packet $\{\mu_n(x) : n = 0, 1, 2, \dots\}$ is

$$U_l^n = \overline{\text{span}\{2^{-\frac{l}{2}}\mu_n(2^{-l}x - k), k \in \mathbf{Z}\}_{l \in \mathbf{Z}}} \quad (9)$$

Thus, the following formula is established [34]

$$\begin{aligned} \sqrt{2}\mu_n(2^{-l}x - k) &= \sum_{b \in \mathbf{Z}} \bar{h}_{k-2b} \mu_{2n}(2^{-l-1}x - b) \\ &+ \sum_{b \in \mathbf{Z}} \bar{g}_{k-2b} \mu_{2n+1}(2^{-l-1}x - b), k \in \mathbf{Z} \end{aligned} \quad (10)$$

where $\bar{h}(\cdot)$ and $\bar{g}(\cdot)$ are the complex conjugates of $h_k(\cdot)$ and $g_k(\cdot)$. According to $V_\gamma = V_{\gamma+1} \oplus W_{\gamma+1}$, it can be obtained that

$$U_l^n = U_{l+1}^{2n} \oplus U_{l+1}^{2n+1}, l \in \mathbf{Z} \quad (11)$$

In the case of $l = 3$, the corresponding structural decomposition is shown in Figure 2. Suppose that $f_l^n(x) \in U_l^n$, $f_{l+1}^{2n}(x) \in U_{l+1}^{2n}$, $f_{l+1}^{2n+1}(x) \in U_{l+1}^{2n+1}$, then

$$\begin{aligned} f_l^n(x) &= \sum_{k \in \mathbf{Z}} c_{l,k}^n 2^{-\frac{l}{2}} \mu_n(2^{-l}x - k) = f_{l+1}^{2n}(x) + f_{l+1}^{2n+1}(x) \\ &= \sum_{b \in \mathbf{Z}} c_{l+1,b}^{2n} 2^{-\frac{l+1}{2}} \mu_{2n}(2^{-l-1}x - b) + \sum_{b \in \mathbf{Z}} c_{l+1,b}^{2n+1} 2^{-\frac{l+1}{2}} \mu_{2n+1}(2^{-l-1}x - b) \end{aligned} \quad (12)$$

where $c_{l,k}^n$, $c_{l+1,b}^{2n}$ and $c_{l+1,b}^{2n+1}$ are the coefficients of function $f_l^n(x)$, $f_{l+1}^{2n}(x)$, $f_{l+1}^{2n+1}(x)$ under the corresponding subspace bases. By substituting Equation (10) into Equation (12), it is concluded that

$$\begin{cases} c_{l+1,b}^{2n} = \sum_{k \in \mathbf{Z}} c_{l,k}^n \bar{h}_{k-2b}, b \in \mathbf{Z} \\ c_{l+1,b}^{2n+1} = \sum_{k \in \mathbf{Z}} c_{l,k}^n \bar{g}_{k-2b}, b \in \mathbf{Z} \end{cases} \quad (13)$$

Equation (13) is called the Mallat decomposition algorithm formula of wavelet packet [36]. In application, for the bearing continuous signal $f(x)$, the sample sequence $f(t)$, $t = 1, 2, \dots, m\lambda$ that is obtained by sampling can be directly approximated, as follows

$$c_{0,\lambda}^0 = f(t), t = 1, 2, \dots, m\lambda \quad (14)$$

where $m\lambda$ is the sampling length of the signal. Therefore, as long as the type of wavelet packet function and scales l are selected, all of the wavelet packet coefficients $c_{l,\nu}^n$ of bearing signal sequence $f(t)$, $t = 1, 2, \dots, m\lambda$ under the scales l are obtained by Mallat decomposition algorithm formula, where $\nu = 1, 2, \dots, m\lambda/2^l$, $n = 0, 1, \dots, 2^l - 1$ is the number of nodes corresponding to the scales l [37]. Taking the scale $l = 3$ as an example, Figure 3 shows the corresponding Mallat decomposition process. The characteristic matrix $A_{m \times s} = (a_{ij})$ of bearing signal is constructed by wavelet packet coefficient $c_{l,\nu}^n$, where m is the number of samples of bearing signal, λ is the length of a single sample, and $s = 2^l$ is the number of all wavelet coefficients of a single sample at the scale l . The element a_{ij} in the matrix $A_{m \times s}$ is the j -th wavelet packet coefficient energy at the scale l that is obtained by the i -th sample through WPT. Algorithm 1 and Figure 4 show the specific process.

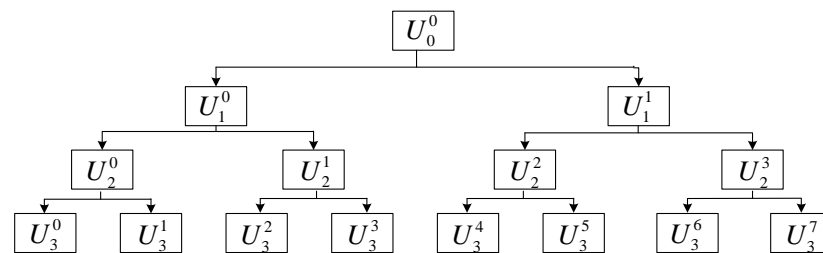


Figure 2. Schematic diagram of the wavelet packet structure decomposition at scale $l = 3$.

Algorithm 1 Wavelet packet decomposition of bearing signal

Input: bearing signal sequence $f(t)$, $t = 1, 2, \dots, m\lambda$, window width λ , wavelet packet scale l , wavelet packet function type

Output: Time-frequency feature matrix $A_{m \times s}$

- 1: Perform sliding window processing on the bearing signal sequence $f(t)$, $t = 1, 2, \dots, m\lambda$;
 - 2: The $f(t)$, $t = 1, 2, \dots, m\lambda$ was divided into m sequence, and each sequence fragment was λ ;
 - 3: **for** $j = 1 : m$ **do**
 - 4: According to the type of wavelet packet function, the wavelet packet coefficient $c_{l,\nu}^n$ of the j -th sample is obtained by Mallat decomposition algorithm formula of wavelet packet;
 - 5: $c_{l,\nu}^n$ is arranged according to the corresponding order under the l -th scale to form the j -th row of $A_{m \times s}$;
 - 6: **end for**
-

Suppose that the sequence $f(t)$, $t = 1, 2, \dots, m\lambda$ is divided into m fragments, each fragment is λ , as shown in Figures 3 and 4.

3.2. Multiple Weighted Singular Value Decomposition Method

The time-frequency matrix $A_{m \times s}$ that is obtained by WPT contains some insensitive features. This paper proposes a multi-weight singular value decomposition algorithm based on WSD in order to effectively extract the sensitive information in the time-frequency feature matrix and eliminate the correlation between variables.

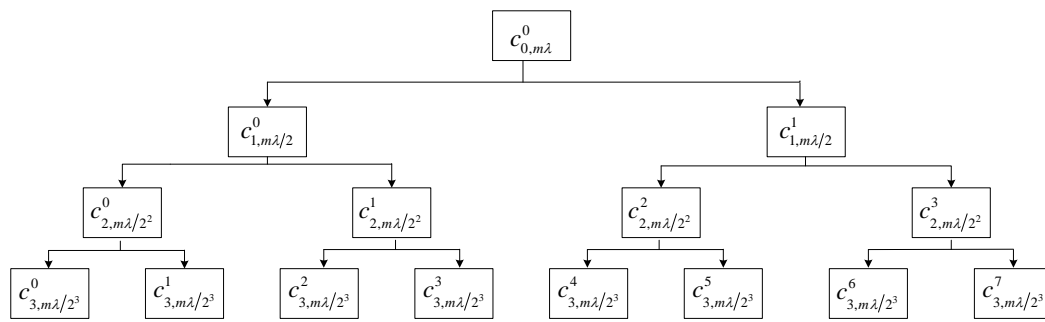


Figure 3. Schematic diagram of Mallat decomposition of wavelet packet at scale $l = 3$.

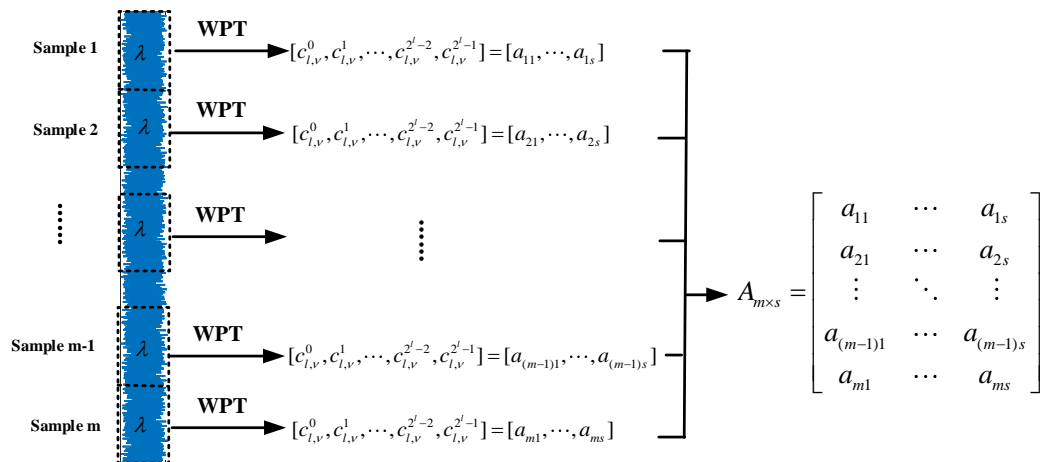


Figure 4. The flowchart of wavelet packet decomposition of bearing signal.

Firstly, the feature matrix $A_{m \times s}$ can be normalized by

$$A_{m \times s}^* = \frac{A_{m \times s} - \bar{A}_{m \times s}}{\text{Var}(A_{m \times s})^{1/2}} \tag{15}$$

where $A_{m \times s}^*$ is the normalized matrix of $A_{m \times s}$, $\bar{A}_{m \times s}$ is the mean of $A_{m \times s}$, $\text{Var}(A_{m \times s})^{1/2}$ is the standard deviation of $A_{m \times s}$. Similar to the WSVD algorithm, the singular value decomposition of the matrix $A_{m \times s}^*$ is performed according to Equation (2) to obtain $U_{m \times m}$, $\Sigma_{m \times s}$ and $V_{s \times s}$. Because the characteristic matrix $A_{m \times s}$ is the projection coefficient of the sample sequence $f(t), t = 1, 2, \dots, m\lambda$ of the bearing signal on the wavelet packet subspace, the matrix $A_{m \times s}$ is a real matrix, $U_{m \times m}$ and $V_{s \times s}$ is orthogonal matrices. Similar to the WSVD algorithm, the order $r < s$ after dimension reduction is determined by the cumulative contribution rate of singular value. Subsequently, $\Sigma_{m \times s}$ becomes $\Sigma_{r \times r}$ after dimensionality reduction. The feature matrix $D_{m \times s}$ after the first weighting is calculated, as follows

$$D_{m \times r} = U_{m \times r} \times \Sigma_{r \times r} \otimes w_i = \begin{pmatrix} u_{11}\sigma_1w_1 & \dots & u_{1r}\sigma_rw_r \\ \vdots & \ddots & \vdots \\ u_{m1}\sigma_1w_1 & \dots & u_{mr}\sigma_rw_r \end{pmatrix} \tag{16}$$

The second weighted weight is obtained by the idea of information entropy. The feature matrix $D_{m \times s}$ needs to be processed according to the following formula before calculating the entropy value

$$d_{ij}^* = \frac{d_{ij} - \min\{d_{1j}, \dots, d_{mj}\}}{\max\{d_{1j}, \dots, d_{mj}\} - \min\{d_{1j}, \dots, d_{mj}\}}, i = 1, \dots, m; j = 1, \dots, r \tag{17}$$

where d_{ij} is the element of row i and column j of the matrix $U_{m \times r}$. d_{ij}^* is the element of row i and column j of the matrix $D_{m \times r}^*$. The information entropy of the matrix $H_j, j = 1, 2, \dots, r$ is calculated, as follows

$$\begin{cases} H_j = - \sum_{i=1}^m p_{ij} \ln p_{ij} / \ln m \\ p_{ij} = d_{ij}^* / \sum_{i=1}^m d_{ij}^* \end{cases} \quad i = 1, 2, \dots, m; j = 1, 2, \dots, r \quad (18)$$

if $p_{ij} = 0, \lim_{p_{ij} \rightarrow 0} p_{ij} \cdot \ln p_{ij} = 0$. According to the information entropy of the matrix $H_j, j = 1, 2, \dots, r$, the entropy weight is calculated, as follows:

$$\omega_j = \frac{1 - H_j}{r - \sum H_j}, j = 1, 2, \dots, r \quad (19)$$

The weighted characteristic matrix is defined using Equation (15)

$$T_{m \times r} = D_{m \times r}^* \otimes \omega_j = \begin{pmatrix} d_{11}^* \omega_1 & \dots & d_{1k}^* \omega_r \\ \vdots & \ddots & \vdots \\ d_{m1}^* \omega_1 & \dots & d_{mk}^* \omega_r \end{pmatrix} \quad (20)$$

Figure 5 shows the specific process of MWSVD.

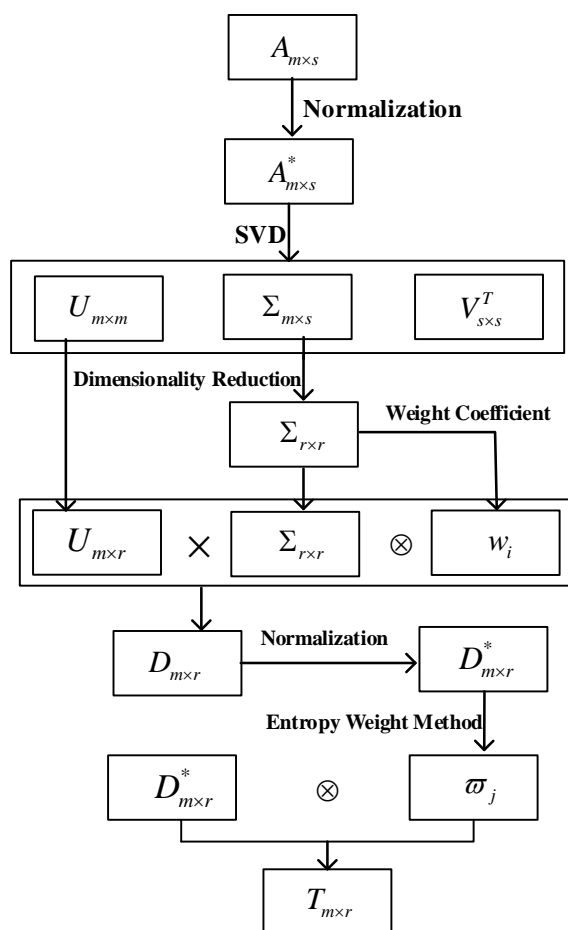


Figure 5. The flowchart of the Multi-Weight Singular Value Decomposition (MWSVD) feature extraction algorithm.

In conclusion, the feature extraction method of bearing fault is given, as follows Algorithm 2.

Algorithm 2 Feature extraction method of bearing fault

Input: bearing signal sequence $f(t), t = 1, 2, \dots, m\lambda$, window width λ , wavelet packet scale l , wavelet packet function type

Output: The feature matrix $T_{m \times r}$

- 1: The time-frequency feature matrix $A_{m \times s}$ is obtained by Algorithm 1;
 - 2: The time-frequency matrix $A_{m \times s}$ is normalized. SVD is decomposed according to Equation (2), and weight is calculated according to Equation (3);
 - 3: The matrix $D_{m \times r}$ is obtained according to Equation (16), the matrix $D_{m \times r}^*$ is obtained according to Equation (17);
 - 4: The entropy weight of the matrix $D_{m \times r}^*$ is obtained according to Equation (18);
 - 5: The characteristic matrix $T_{m \times r}$ of the bearing fault is obtained according to Equation (20).
-

4. Fault Diagnosis Method Based on Feature Fusion

The core idea of SVM is to transform indivisible samples in low-dimensional space into high-dimensional space through a kernel function, and realize the classification between samples by seeking the optimal classification hyperplane [38].

Suppose that sample set $\{(x_i, y_i) | x_i \in R^n, y_i \in \{-1, +1\}, i = 1, 2, \dots, q\}$, where q is the number of training samples, and x_i and y_i are the i -th data points that belong to a binary class y_i .

SVM maps the input of the low-dimensional space to the high-dimensional space by the nonlinear mapping $\theta(\cdot)$ to obtain the linear classification function $f(x) = \omega^T \theta(x) + b$, where ω is the weight and b is the offset. For a binary classification issue with labels -1 and 1 , all of the samples should meet a specific condition, as defined in Equation (21), thus the two types of samples can be completely separated:

$$f(x) = \omega^T \theta(x) + b \begin{cases} > 1 & \text{for } y_i = 1 \\ < -1 & \text{for } y_i = -1 \end{cases} \quad (21)$$

To linearly solve non-separable problems, slack variable ξ_i and penalty factor C are introduced, thus the best classification function is obtained by solving the minimum value of Equation (22)

$$\frac{1}{2} \|\omega\|_2^2 + C \sum_{i=1}^N \xi_i \quad (22)$$

The Lagrange coefficient is introduced, Equation (22) is transformed into a quadratic programming problem to solve

$$L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (23)$$

where $K(x_i, x_j)$ is the kernel function. By solving the smallest $L(\alpha)$, the final classification function is as follows

$$f(x) = \text{sgn}\left(\sum_{i,j=1}^N \alpha_i y_i K(x_i, x_j) + b\right) \quad (24)$$

This paper chooses the Gaussian kernel function as the kernel function of SVM. Its expression is as follows

$$K(x_i, x_j) = \exp\left(-\frac{|x_i - x_j|}{2\varepsilon^2}\right) \quad (25)$$

where ε is the kernel parameter. The penalty parameter C and the kernel parameter ε have an important influence on the classification accuracy and generalization ability. There

is currently no unified theoretical method to find the best combination of the above two parameters. This paper uses the genetic algorithm to find the optimal value of the parameter

This paper uses the following equation to calculate the classification accuracy η of SVM

$$\eta = \frac{\sum_{j=1}^S \beta_j}{S}, \beta_j = \begin{cases} 1 & \text{if predicting label} = \text{actual label} \\ 0 & \text{if predicting label} \neq \text{actual label} \end{cases} \quad (26)$$

where S is the number of samples in the test set. The appeal fault diagnostic model was run τ times and the variance δ of the classification accuracy η is calculated, as follows

$$\begin{cases} \delta = \frac{\sum_{i=1}^{\tau} (\eta_i - \bar{\eta})^2}{\tau - 1} \\ \bar{\eta} = \frac{\sum_{i=1}^{\tau} \eta_i}{\tau} \end{cases} \quad (27)$$

In conclusion, the bearing fault diagnosis method based on feature fusion is proposed

5. Experiments and Analysis Results

5.1. Case A: The Time-Varying Bearing Data from the University of Ottawa

The time-varying bearing data from the University of Ottawa [39]. The experiments are performed on a SpectraQuest machinery fault simulator (MFS-PK5M). The experimental set-up is shown in Figure 6. The shaft is driven by a motor and the rotational speed is controlled by an AC drive. Two ER16K ball bearings are installed to support the shaft, the left one is a healthy bearing and the right one is the experimental bearing, which are replaced by bearings of different health conditions. An accelerometer (ICP accelerometer, Model 623C01) is placed on the housing of the experimental bearing to collect the vibration data. In addition, an incremental encoder (EPC model 775) is installed to measure the shaft rotational speed. To ensure the authenticity of the data, three trials are collected for each experimental setting. In this article, the operating speed condition selected is deceleration. Table 1 shows the operating speed and health status of the selected bearing.

Table 1. Each Fault State and Corresponding Label in the Data Set.

Status	Operating Rotational Speed	Label
Normal Status	from 28.6 Hz to 13.9 Hz	1
Inner Race Fault	from 25.8 Hz to 12.0 Hz	2
Out Race Fault	from 25.4 Hz to 10.3 Hz	3

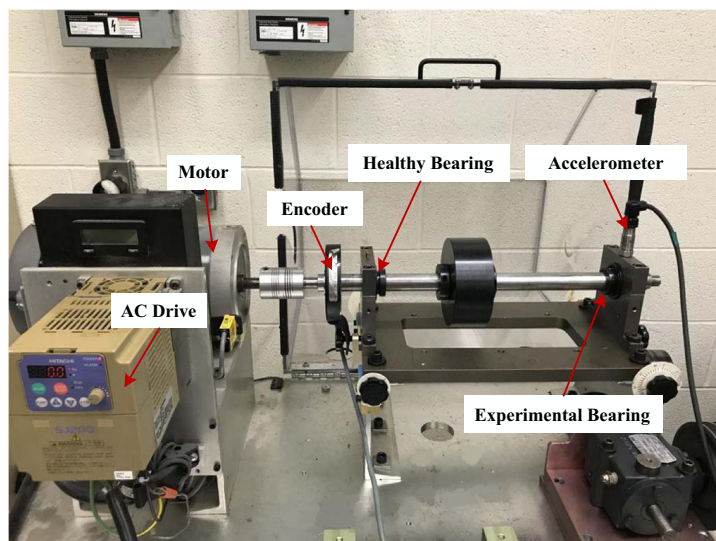


Figure 6. Time-varying bearing experimental device at the University of Ottawa.

In this paper, Table 1 shows each fault state and its corresponding label. Signals are sampled at 200 KHz. For each state, 76,800 points are collected and labeled in turn. There are also some other research data in the data set, which are not described because they are not used in this article. Table 2 shows the experimental environment and experimental parameters of this article.

Table 2. Experimental Environment and Parameters.

Experimental Environment		Experimental Modal Parameters	
Operating System	Windows 10 Home Chinese Version	The Size of The Sliding Window	$\lambda = 1024$
CPU	Intel Core i7-8550U @2.00 GHz	Wavelet Function	db10 [40]
RAM	8.00 GB	wavelet packet scale	$l = 4$
System Type	64 bit	The Number of Runs	$\tau = 100$
SVM Classifier	LIBSVM by Lin Zhiren of National Taiwan University [41]	Cumulative Contribution Rate of Singular Values	90%

According to Algorithm 3, 60 groups are randomly selected from the state category of each bearing as the training set, and 60 groups are used as the test set, which are labeled according to the state category that they belong to. To illustrate the effectiveness of the proposed method, PCA, SVD, and WSVD [30] are selected as the comparison. Furthermore, the SVM classifier is obtained from the training set data, and the classification accuracy and diagnosis time of the test set data by the SVM classifier are used as the criteria for assessing the optimal diagnosis method. To further illustrate the effectiveness of MWSVD method that is proposed in this paper, after feature extraction three feature extraction methods are visualized and analyzed to observe the effects of feature extraction.

Algorithm 3 Bearing fault diagnosis method based on Feature Fusion

Input: bearing signal sequence $f(t), t = 1, 2, \dots, m\lambda$, window width λ , wavelet packet scale l , wavelet packet function type, Number of runs τ .

Output: classification accuracy η , calculation time, Variance of classification accuracy δ .

- 1: The feature matrix of bearing fault is obtained $T_{m \times r}$ by Algorithm 2;
- 2: The feature matrix $T_{m \times r}$ is randomly divided into the training set and test set, and different state types are labeled;
- 3: The SVM classifier is trained by the training set to obtain the SVM-based classification model;
- 4: The test set is input to the SVM-based classification model to obtain the predicted label of the test set. The actual label and predicted label of the test set are calculated according to Equation (26) to calculate the classification accuracy η of the diagnostic model. The total running time of bearing signal from MVSVD feature extraction to training SVM classification model to test result is calculated;
- 5: The model is run τ times in sequence to get the classification accuracy η each time, and the variance δ of classification accuracy is obtained according to Equation (27).

In this paper, a genetic algorithm is used to find the optimal parameters of SVM in the training set after five-fold cross-validation. Table 3 shows the optimal parameters and classification results of SVM on the training set under the optimal parameters (CA accuracy).

Table 3. The optimal parameters of SVM

Fault Diagnosis Model	Penalty Parameter C	Kernel Parameter ϵ	CA Accuracy(%)
WPT-PCA+SVM	25.48	990.5	55
WPT-SVD+SVM	5.65	4.13	81.67
WPT-WSVD+SVM	1.19	103.44	66.67
WPT-MWSVD+SVM	25.15	212.47	81.67

Figure 7 is the Receiver Operating Characteristic Curve (ROC) curve diagram of the four algorithms on bearing fault diagnosis in a single experiment. It shows that WPT-MWSVD+SVM can effectively diagnose the fault of the bearing inner race and outer race as compared to the other three methods. Figure 8 shows the Classification confusion matrix of four algorithms on bearing fault diagnosis in a single experiment. It can be seen that: WPT-WSVD+SVM and WPT-MWSVD+SVM can effectively distinguish the normal state of the bearing; WPT-MWSVD+SVM can effectively diagnose the inner race fault, and the diagnosis effect is better than the other three methods; WPT-SVD+SVM and WPT-MWSVD+SVM can both effectively distinguish outer race faults, and the diagnostic effect is much better than the other two methods. The four algorithms are run for 100 times in sequence, and the experimental results are shown in Figures 9 and 10, Table 4. It can be seen that the average classification accuracy of this method is 87.87%, which is higher than the other three methods, and the average time used is 16.32 s, which is significantly lower than the other three methods. This shows that the proposed method has better computational efficiency and diagnostic accuracy. Besides, it can be seen from Figure 9b that the fluctuation of the classification accuracy of this method is small. Table 4 shows the variance of classification accuracy. To examine the computational cost of this method in analyzing experimental data, refer to reference [42], and calculate the computational efficiency of the model based on the classification time; the related processing times are listed in Table 4. The results show that the average diagnosis time for WPT-MWSVD+SVM model diagnosis to collect 1 s sample data only takes 10.63 s. Therefore, the method proposed in this paper is superior to the other three methods for bearing fault diagnosis. Because the difference between the four fault diagnosis methods lies on the extraction of bearing fault features, this shows that the MWSVD feature extraction method proposed in this paper can effectively extract sensitive features of bearing information and it has good feature extraction capabilities.

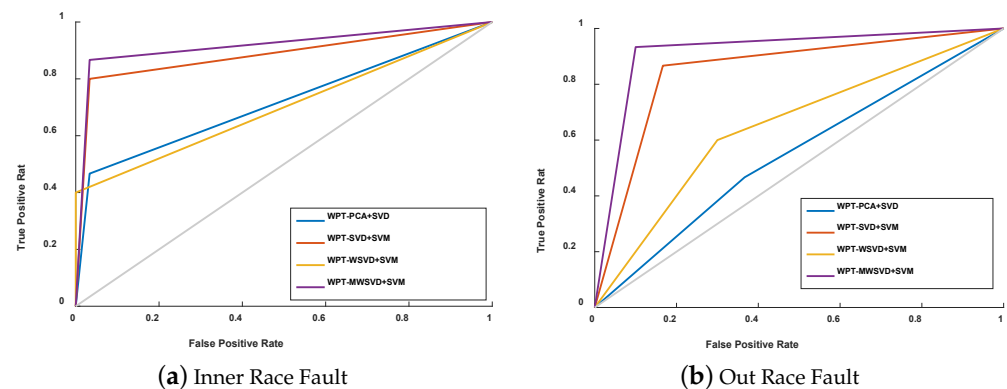


Figure 7. ROC curve diagram of the four algorithms on the fault diagnosis of the bearing inner race (a) and outer race (b).



Figure 8. Classification confusion matrix of four algorithms: (a) WPT-PCA+SVM; (b) WPT-SVD+SVM; (c) WPT-MSVD+SVM; and, (d) WPT-WMSVD+SVM.

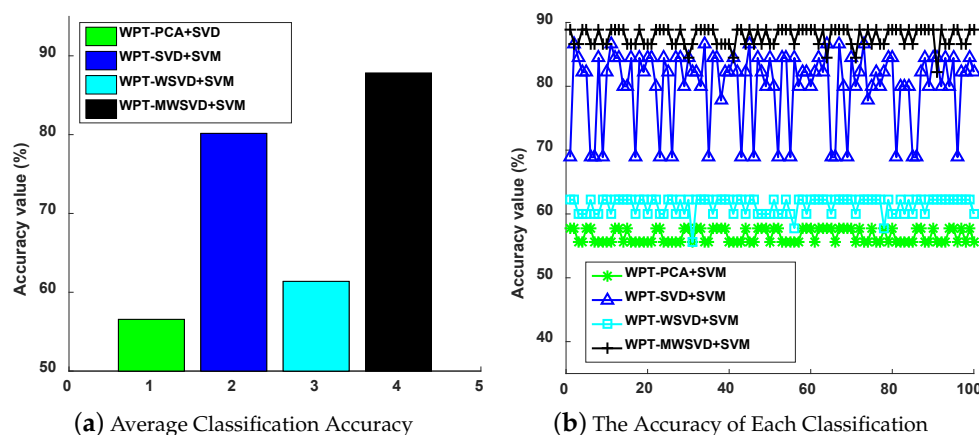


Figure 9. The average classification accuracy of 100 runs (a) and each classification accuracy (b).

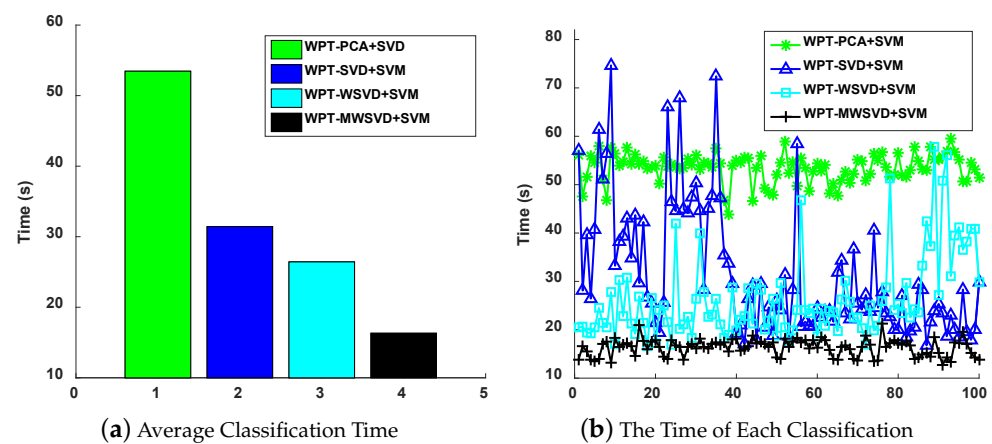


Figure 10. The average classification time of 100 runs (a) and each classification time (b).

The four methods are visualized and analyzed to further illustrate the feature extraction capability of MWSVD, which are shown in Figure 11. It can be seen that the number of principal components extracted by PCA is less than the number of singular values extracted by the other three methods under the same cumulative singular value contribution rate. WPT-MWSVD has a more scattered distribution of data samples as compared with the other three methods, which can not only effectively improve the classification accuracy of subsequent fault diagnosis, but also effectively shorten the fault diagnosis time, which corresponds to the results presented in Table 4. Therefore, the MWSVD method that is constructed in this paper can effectively extract bearing signal features and improve the classification ability of SVM classifier.

Table 4. Classification results under different models.

Fault Diagnosis Model	The Average Accuracy η of 100 Runs (%)	Sample Variance of Accuracy δ	The Average Time of 100 Runs (s)	Average Processing Time for 1 s of Data
WPT-PCA+SVM	56.56	1.23	53.46	34.80
WPT-SVD+SVM	80.16	34.04	31.44	20.47
WPT-WSVD+SVM	61.38	1.67	26.43	17.21
WPT-MWSVD+SVM	87.80	1.94	16.32	10.62

5.2. Case B: Case Western Reserve University Bearing Data Set

The data used in this case are taken from the Case Western Reserve University Bearing Data Set [43]. Figure 12 presents a schematic diagram of the experimental platform, the test stand consists of a 2 hp motor (left), a torque transducer/encoder (center), a dynamometer (right), and control electronics (not shown). The test bearings support the motor shaft. Vibration data were collected using accelerometers, which were attached to the housing with magnetic bases. The test bearing are SKF6205-2RS deep groove ball bearing.

In this article, we choose the vibration acceleration signals were collected under the condition that rotor speeds 1730 r/min with sampling frequency 12 kHz. These data include three fault levels and four fault states. For each state under each fault level, 76,800 points are collected and labeled in turn. Table 5 presents the details.

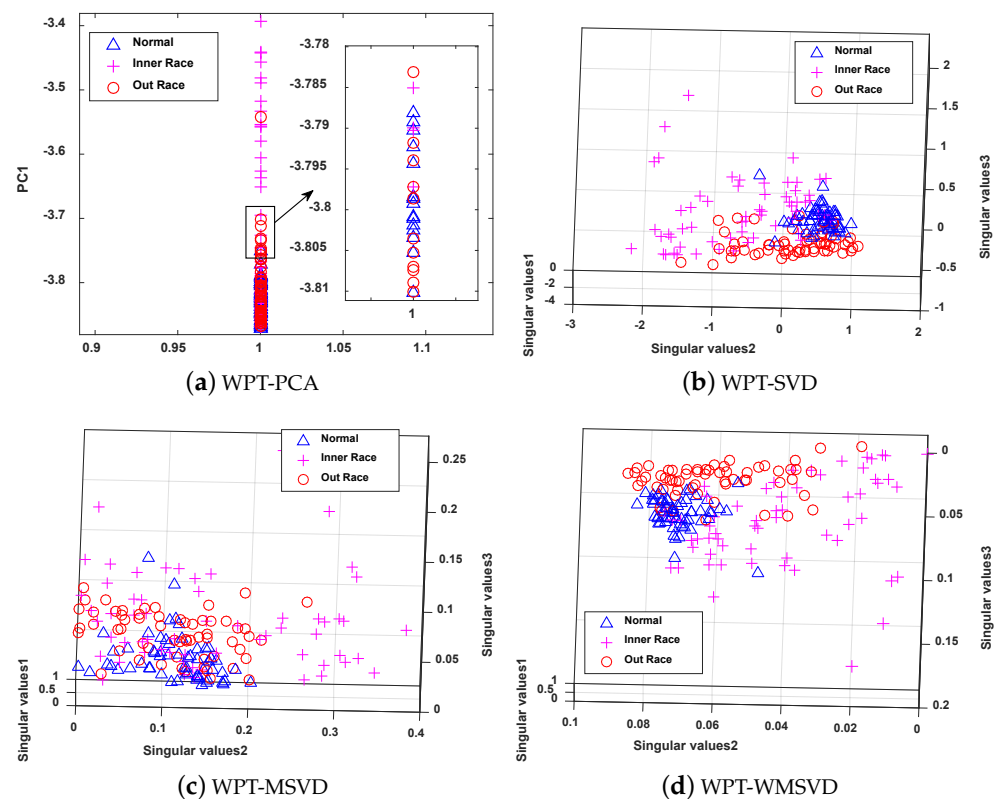


Figure 11. The Visualization of feature extraction effect of four methods: (a) Distribution of the data samples based on WPT-PCA; (b) Distribution of the data samples based on WPT-SVD; (c) Distribution of the data samples based on WPT-WSVD; (d) Distribution of the data samples based on WPT-WMSVD.

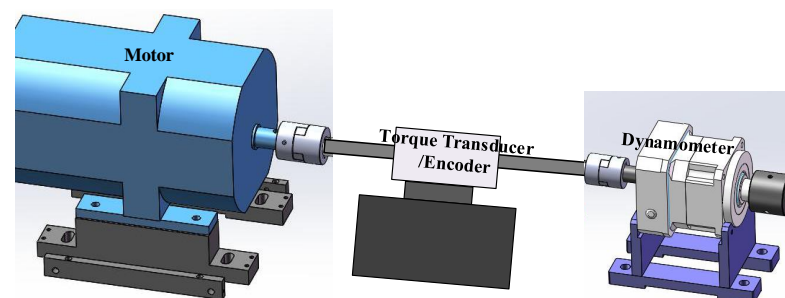


Figure 12. The schematic diagram of bearing experimental platform of Case Western Reserve University.

In this case, the wavelet basis function we choose is db6, and the number of wavelet layers is $l = 3$. The experimental environment, experimental parameters, selection of training set and test set, selection of comparison model, and evaluation criteria of Section 5.2 are the same as Section 5.1, except that wavelet function, wavelet packet scale, and the selected data set.

In this paper, a genetic algorithm is used to find the optimal parameters of SVM in the training set after five-fold cross-validation. Table 6 shows the optimal parameters and classification results of SVM on the training set under the optimal parameters (CA accuracy).

Table 5. Failure status label for each degree of fault.

The Severity of the Fault	Fault Size of Samples	Status	Label
Minor Fault	0.1778 (mm)	Normal Status	1
		Inner Race Fault	2
		Ball Fault	3
		Out Race Fault	4
General Fault	0.3556 (mm)	Normal Status	1
		Inner Race Fault	2
		Ball Fault	3
		Out Race Fault	4
Serious Fault	0.5334 (mm)	Normal Status	1
		Inner Race Fault	2
		Ball Fault	3
		Out Race Fault	4

Table 6. The optimal parameters of Support Vector Machine (SVM).

The Severity of the Fault	Fault Diagnosis Model	Penalty Parameter C	Kernel Parameter ϵ	CA Accuracy (%)
Minor Fault	WPT-PCA+SVM	0.54	89.44	98.33
	WPT-SVD+SVM	1.42	31.46	98.33
	WPT-WSVD+SVM	42.97	4.83	97.50
	WPT-MWSVD+SVM	4.71	691.86	100
General Fault	WPT-PCA+SVM	79.78	39.67	75.42
	WPT-SVD+SVM	0.20	2.32	100
	WPT-WSVD+SVM	5.96	19.02	84.17
	WPT-MWSVD+SVM	4.36	167.58	100
Serious Fault	WPT-PCA+SVM	4.89	34.84	87.08
	WPT-SVD+SVM	4.21	9.62	98.33
	WPT-WSVD+SVM	8.88	57.90	98.75
	WPT-MWSVD+SVM	4.06	623.54	100

Figure 13 presents the ROC curve diagram of the four algorithms on bearing fault diagnosis in a single experiment. It can be seen that Minor fault: the ROC curve area of the other three methods is larger than that of WPT-WSVD+SVM under inner race fault and outer race fault, and the ROC curve area of WPT-MWSVD+SVM is larger than the other three methods under ball fault; general failure: the curve areas of WPT-SVD+SVM and WPT-MWSVD+SVM are larger than the other two methods under three types of failures; serious failures: the ROC curve area of WPT-MWSVD+SVM is larger than the other three methods under inner race fault and outer race fault. This shows that WPT-MWSVD+SVM has better bearing fault diagnosis capabilities. Figure 14 is the classification confusion matrix of four kinds of bearing fault diagnosis in a single experiment under different fault degrees. It can be seen that Minor fault: the diagnostic ability of WPT-MWSVD+SVM is better than the other three methods under ball fault; general failure: the diagnostic capabilities of WPT-WSVD+SVM and WPT-MWSVD+SVM are much better than the other two methods under inner race fault and ball fault; serious failures: the diagnostic capabilities of WPT-WSVD+SVM and WPT-MWSVD+SVM are better than the other two methods under inner race fault. WPT-MWSVD+SVM can effectively diagnose outer race fault, and the diagnostic effect is better than the other three methods. The four algorithms are run 100 times, in turn, and the experimental results are shown in Figures 15 and 16 and Table 7. The results show that this method has the advantages of high classification accuracy and short calculation time under the three failure levels. Besides, in Figure 15, the fluctuation of the classification

accuracy of this method is small. It shows that the algorithm in this paper can enhance the sensitive features of bearing signals and reduce the interference of insensitive features on the diagnosis model after twice weighting. Therefore, the diagnosis model that is proposed in this paper has great accuracy. Table 7 lists the related processing times. The results show that the average diagnosis time for WPT-MWSVD+SVM model diagnosis to collect 1 s sample data only takes 10.63 s. Therefore, the method that is proposed in this paper is superior to the other three methods for bearing fault diagnosis.

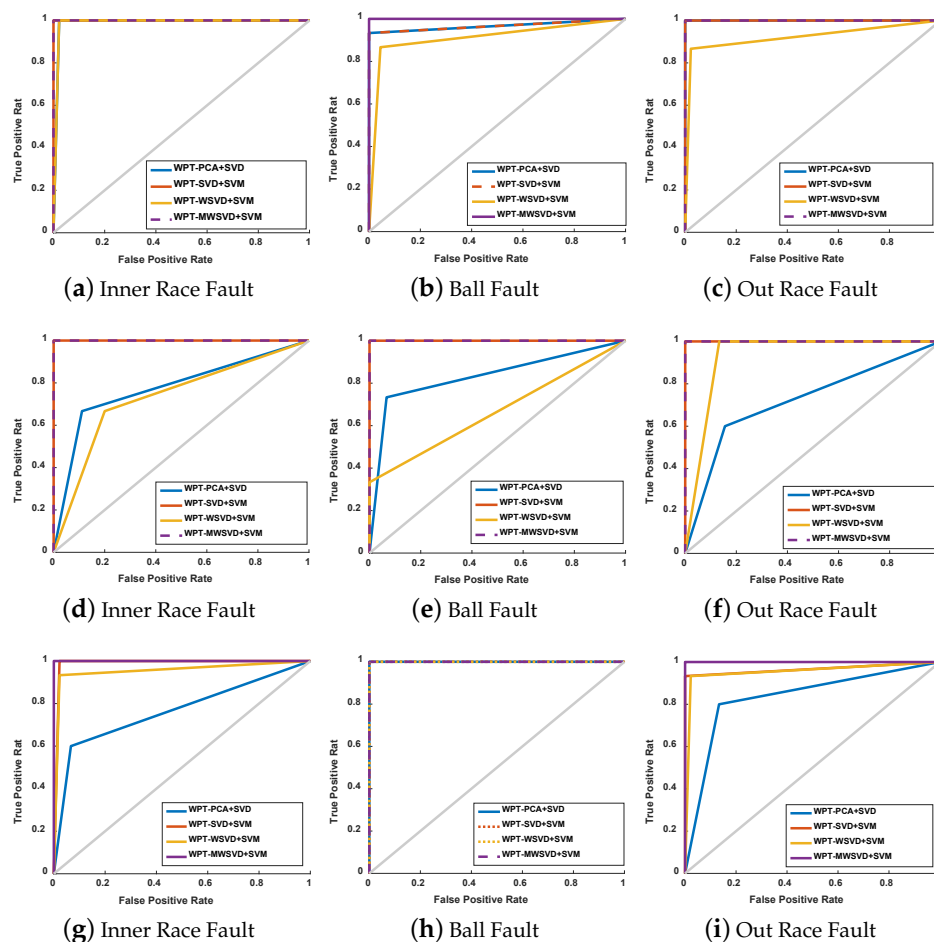


Figure 13. ROC curve diagram of the four algorithms: the minor fault diagnosis of the bearing inner race (a), Ball (b) and outer race (c); the general fault diagnosis of the bearing inner race (d), Ball (e) and outer race (f); the serious fault diagnosis of the bearing inner race (g), Ball (h), and outer race (i).

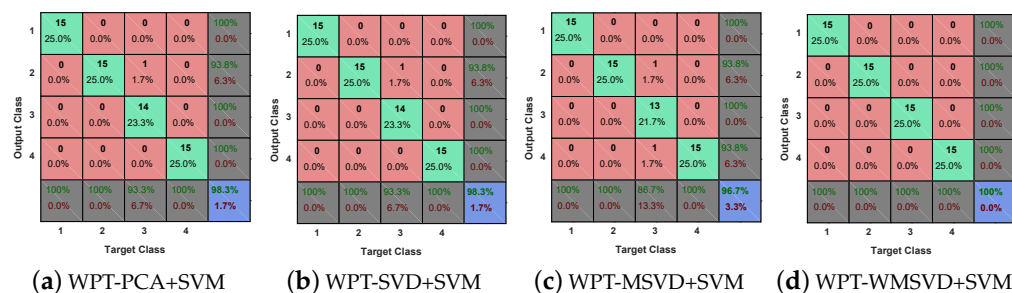


Figure 14. Cont.

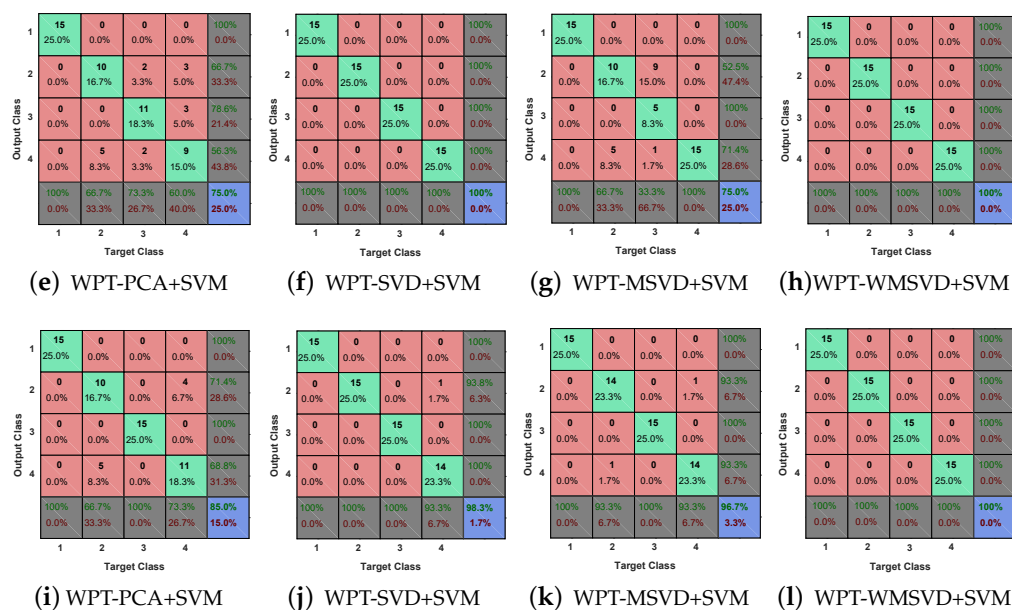


Figure 14. Classification confusion matrix of four algorithms: the minor fault (a) WPT-PCA+SVM; (b) WPT-SVD+SVM; (c) WPT-MSVD+SVM; (d) WPT-WMSVD+SVM; the general fault (e) WPT-PCA+SVM; (f) WPT-SVD+SVM; (g) WPT-MSVD+SVM; (h) WPT-WMSVD+SVM; the serious fault (i) WPT-PCA+SVM; (j) WPT-SVD+SVM; (k) WPT-MSVD+SVM; and, (l) WPT-WMSVD+SVM.

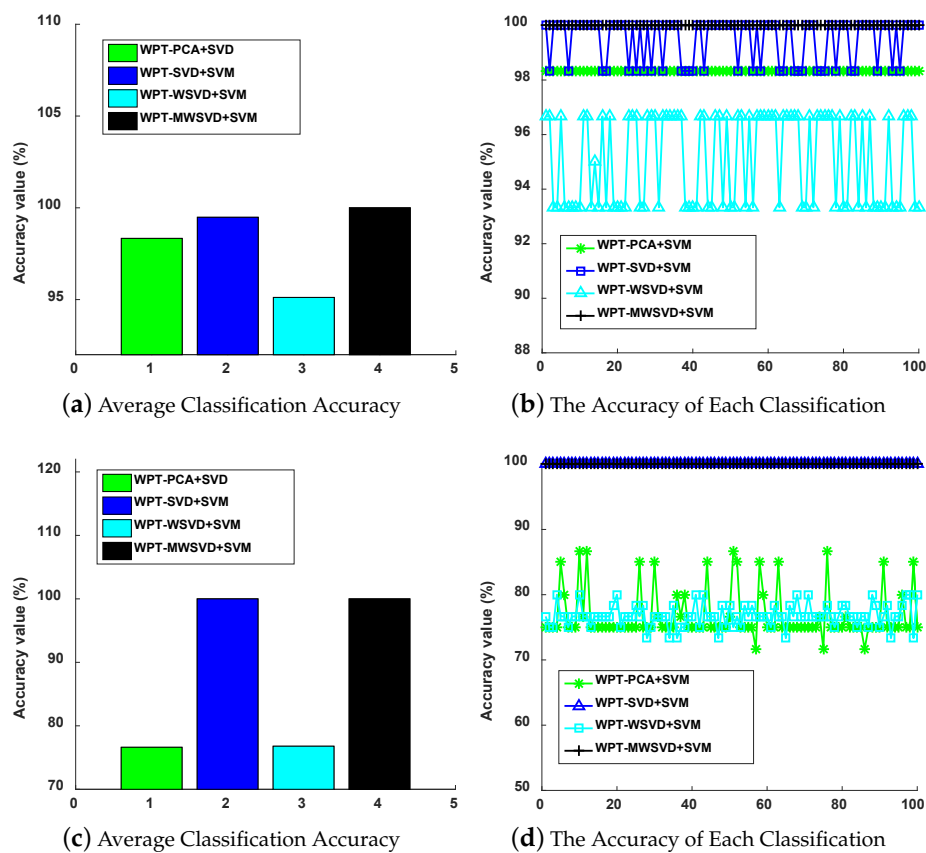


Figure 15. Cont.

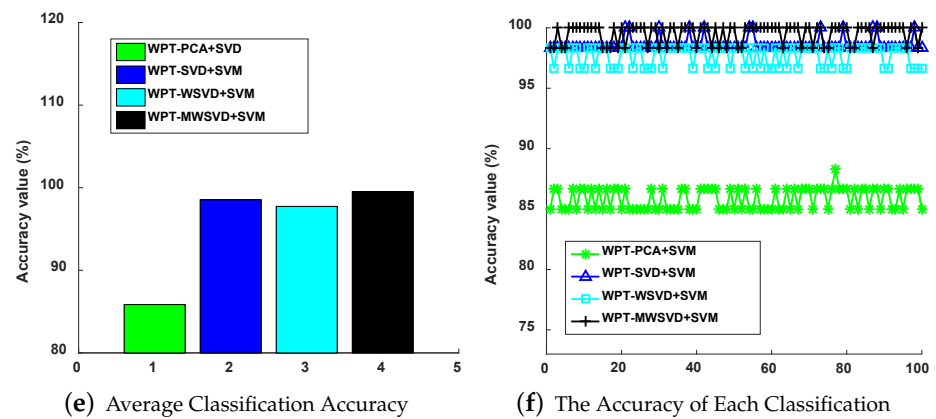


Figure 15. The degree of fault is minor fault (a), general fault (c) and serious fault (e); The degree of fault is minor fault (b), general fault (d), and serious fault (f).

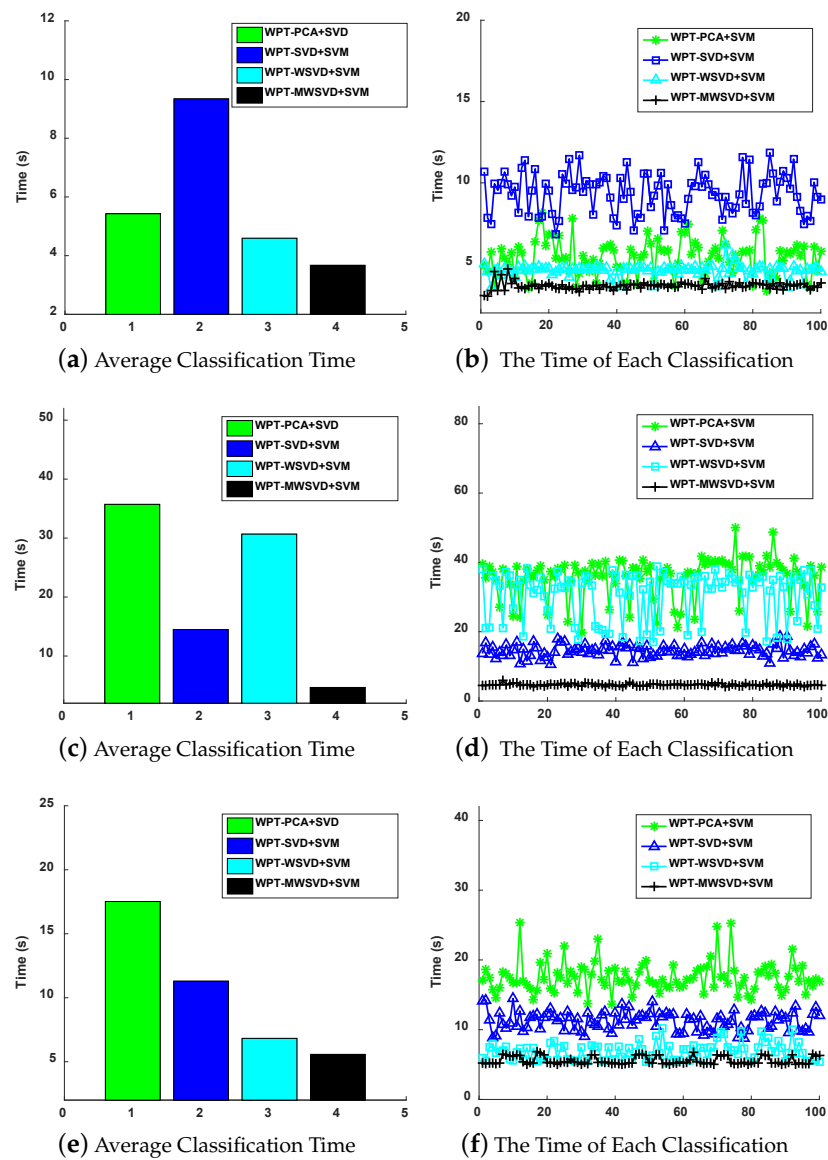


Figure 16. The degree of fault is minor fault (a), general fault (c) and serious fault (e); The degree of fault is minor fault (b), general fault (d), and serious fault (f).

Table 7. Failure status label for each degree of fault.

The Severity of the Fault	Fault Diagnosis Model	The Average Accuracy η of 100 Runs (%)	Sample Variance of Accuracy δ	The Average Time of 100 Runs(s)	The Average Processing Time for 1 s of Data
Minor Fault	WPT-PCA+SVM	98.33	0	5.43	0.21
	WPT-SVD+SVM	99.48	0.60	9.34	0.36
	WPT-WSVD+SVM	95.11	2.76	4.59	0.18
	WPT-MWSVD+SVM	100	0	3.67	0.14
General Fault	WPT-PCA+SVM	76.62	13.72	35.72	1.38
	WPT-SVD+SVM	100	0	14.50	0.57
	WPT-WSVD+SVM	76.78	2.93	30.66	1.20
	WPT-MWSVD+SVM	100	0	4.67	0.18
Serious Fault	WPT-PCA+SVM	85.85	0.76	17.52	0.68
	WPT-SVD+SVM	98.53	0.30	11.29	0.44
	WPT-WSVD+SVM	97.73	0.65	6.82	0.27
	WPT-MWSVD+SVM	99.52	0.58	5.57	0.22

The four methods are visualized and analyzed to further illustrate the feature extraction capability of the MWSVD method constructed in this paper, and the results are shown in Figures 17–19. It can be seen that the number of principal components extracted by PCA is less than the number of singular values extracted by the other three methods under the same cumulative singular value contribution rate. WPT-MWSVD has a more scattered distribution of data samples as compared with the other three methods, which can not only effectively improve the classification accuracy of subsequent fault diagnosis, but also effectively shorten the fault diagnosis time, which corresponds to the results shown in Table 7. Therefore, the MWSVD method that is constructed in this paper can effectively extract bearing signal features and improve the classification ability of SVM classifier.

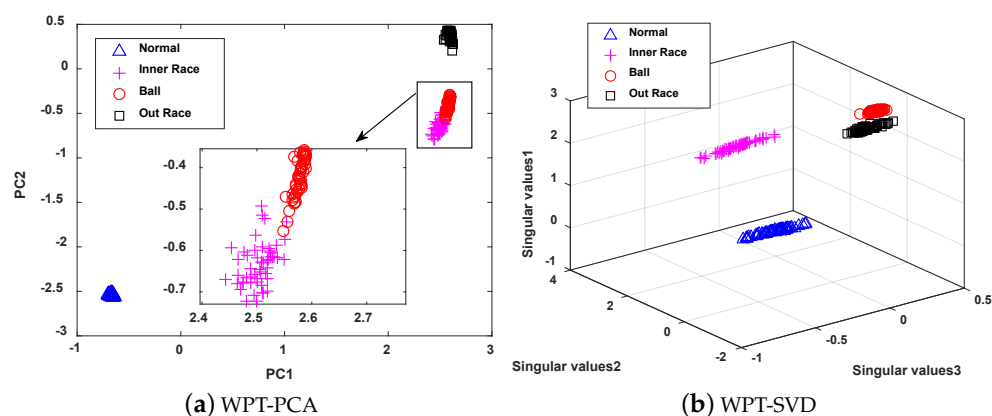


Figure 17. Cont.

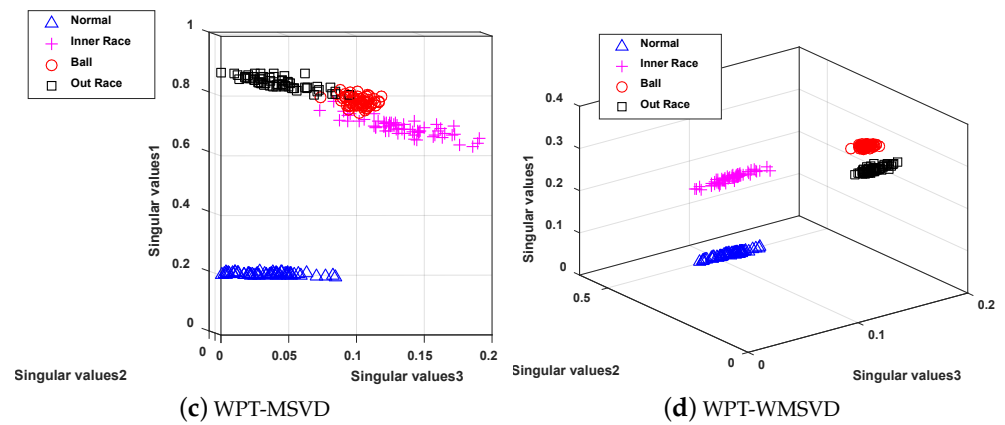


Figure 17. The Visualization of feature extraction effect of four methods under minor faults: (a) Distribution of the data samples based on WPT-PCA; (b) Distribution of the data samples based on WPT-SVD; (c) Distribution of the data samples based on WPT-WSVD; and, (d) Distribution of the data samples based on WPT-MWSVD.

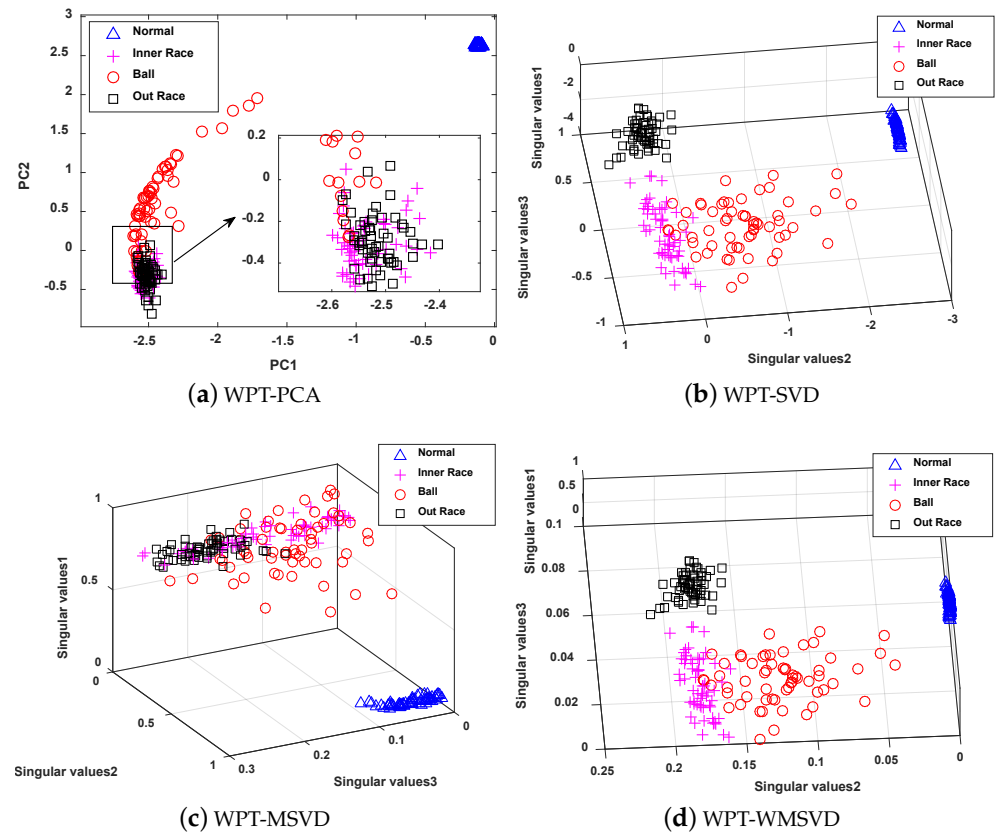


Figure 18. The Visualization of feature extraction effect of four methods under general faults: (a) Distribution of the data samples based on WPT-PCA; (b) Distribution of the data samples based on WPT-SVD; (c) Distribution of the data samples based on WPT-WSVD; and, (d) Distribution of the data samples based on WPT-MWSVD.

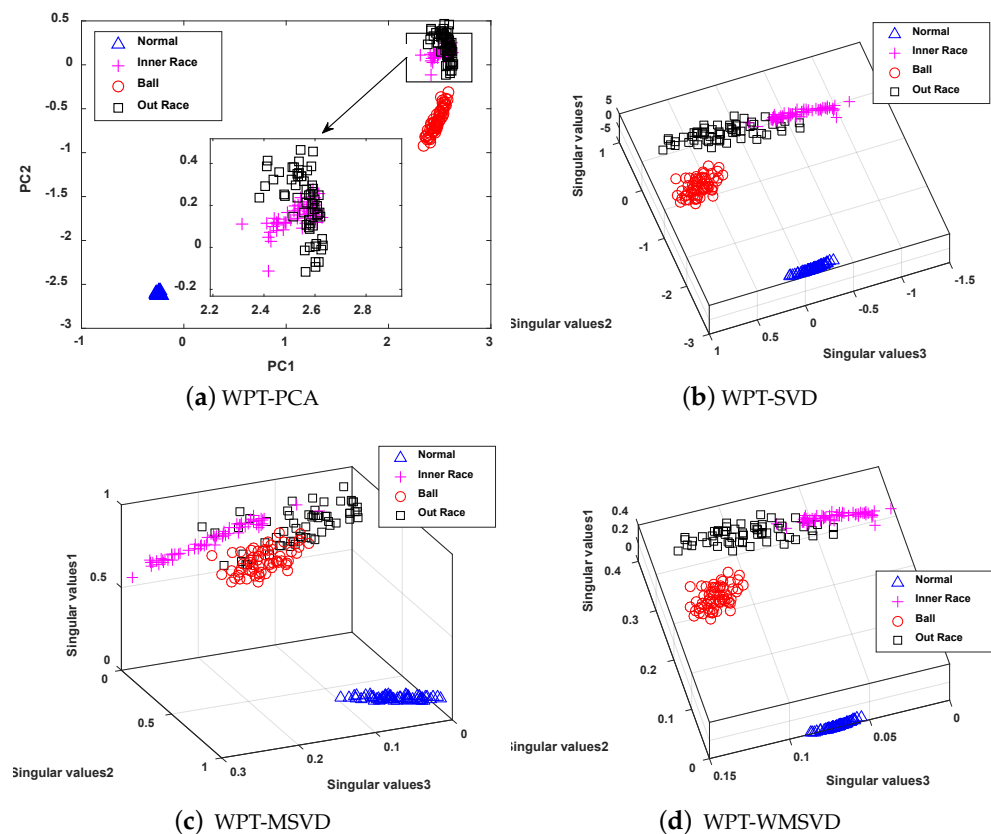


Figure 19. The Visualization of feature extraction effect of four methods under serious faults: (a) Distribution of the data samples based on WPT-PCA; (b) Distribution of the data samples based on WPT-SVD; (c) Distribution of the data samples based on WPT-WSVD; (d) Distribution of the data samples based on WPT-MWSVD.

To sum up, we can see that:

1. in different bearing data sets or different failure degrees, the four algorithms are run 100 times in sequence. The bearing fault diagnosis method that is based on feature fusion proposed in this paper has a high average classification accuracy rate. This model has a shorter average time than the other three fault diagnosis methods, and the average diagnosis time for the model diagnosis to collect 1-s sample data is the lowest. This shows that the method in this paper not only has higher accuracy, but also lower computational cost in bearing fault diagnosis; and,
2. In different bearing data sets or different failure degrees, the four feature extraction algorithms are visualized and analyzed. The results show that, as compared with the traditional feature extraction methods, the MWSVD feature extraction method proposed in this paper can retain more bearing signals information. Besides, the feature distribution of bearing signal extracted in this paper is relatively divergent. This means that the MWSVD feature extraction method proposed in this paper can effectively extract bearing signal features, reduce the computational complexity of subsequent diagnostic models, and improve the diagnostic capabilities of subsequent diagnostic models.

6. Conclusions

To cope with the problem that it is difficult to extract feature vector effectively in rolling bearing fault diagnosis, our work is as follows: firstly, this paper constructs an SVD feature extraction method that is based on the fusion of multiple weights through the contribution rate of singular values and entropy weights. On the one hand, this method makes up for the problem that the traditional PCA algorithm loses its physical meaning

due to the combination transformation of features in the process of removing feature redundancy, and it reduces the impact of noise on the data; on the other hand, it makes up for the problem that the effects of the features extracted by the traditional SVD algorithm have a high computational cost for subsequent models. Secondly, this paper combines it with the SVM classifier to propose a bearing fault diagnosis method that is based on feature fusion. Finally, the time-varying bearing data of the University of Ottawa and the data set of Case Western Reserve University bearing data center are used in the experiment. It shows that, under the condition of the steady-state and non-steady-state of bearing, under different sampling frequency and sampling time of bearing signal, and under a different degree of damage of bearing, MWSVD can effectively extract the sensitive features in the bearing and reduce the interference of non-sensitive features to the diagnosis model. WPT-MWSVD+SVM diagnosis models can quickly and accurately identify bearing faults, have good model adaptability, high calculation accuracy and calculation efficiency, and they have great application potential. Besides, the SVD-based MWSVD feature extraction algorithm is also suitable for other aspects of dimensionality reduction requirements, which will be the author's next research direction.

Author Contributions: H.Z. (Huibin Zhu) proposed the main idea, finished the draft manuscript; Z.H. participated in the discussion of diagnostic thinking and correction of English grammar, J.W. (Juhui Wei) drew the figures and tables, J.W. (Jiongqi Wang) conceived of the experiments, H.Z. (Haiyin Zhou) analyzed the data. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by civil space pre-research project grant number B 0103 and National Natural Science Foundation of China grant number 61903366 and Provincial Natural Science Foundation grant number 2019JJ50745.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Abbreviations

The following abbreviations are used in this manuscript:

WPT	Wavelet Packet Transform
MWSVD	Multi-Weight Singular Value Decomposition
SVM	Support Vector Machine
PCA	Principal Component Analysis
SVD	Singular Value Decomposition
WSVD	Weighted Singular Value Decomposition

References

1. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Pr.* **2018**, *108*, 33–476. [[CrossRef](#)]
2. Li, J.; Wang, Y.; Zi, Y.; Jiang, S. RA Local Weighted Multi-Instance Multilabel Network for Fault Diagnosis of Rolling Bearings Using Encoder Signal. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 8580–8589.
3. Li, Y.; Yang, Y.; Wang, X.; Liu, B.; Liang, X. Early fault diagnosis of rolling bearings based on hierarchical symbol dynamic entropy and binary tree support vector machine. *J. Sound Vib.* **2018**, *428*, 72–86. [[CrossRef](#)]
4. Zhang, B.; Zhang, L.; Xu, J. Degradation Feature Selection for Remaining Useful Life Prediction of Rolling Element Bearings. *Qual. Reliab. Eng. Int.* **2016**, *32*, 547–554. [[CrossRef](#)]
5. Wei, Y.; Li, Y.; Xu, M.; Huang, W. A Review of Early Fault Diagnosis Approaches and Their Applications in Rotating Machinery. *Entropy* **2019**, *21*, 409–435. [[CrossRef](#)]
6. Wang, X.; Xiao, F. An Improved Multisensor Data Fusion Method and Its Application in Fault Diagnosis. *IEEE Access* **2018**, *7*, 3928–3937. [[CrossRef](#)]
7. Liu, Q.; Wang, H. A case study on multisensor data fusion for imbalance diagnosis of rotating. *AI EDAM* **2001**, *15*, 203–210. [[CrossRef](#)]

8. Jafari, H.; Poshtan, J.; Sadeghi, H. Application of fuzzy data fusion theory in fault diagnosis of rotating machinery. *Proc. Inst. Mech. Eng. Part I J. Syst. Control. Eng.* **2018**, *8*, 1015–1024. [[CrossRef](#)]
9. Akilu, Y.; Jyoti, S. Effective Vibration-based Condition Monitoring (eVCM) of Rotating Machines. *J. Qual. Maint. Eng.* **2017**, *23*, 279–296.
10. Moslem, A.; Jaskaran, S.; Inaki, B.; Jay, L. Multisensor data fusion for gearbox fault diagnosis using 2-D convolutional neural network and motor current signature. *Mech. Syst. Signal Pr.* **2020**, *144*, 1–18.
11. Akilu, Y.; Jyoti, S.; Keri, E. HOS analysis of measured vibration data on rotating machines with different simulated faults. In *Advances in Condition Monitoring of Machinery in Non-Stationary Operations*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 81–89.
12. Adam, G. Recognition of Acoustic Signals of Loaded Synchronous Motor Using FFT, MSAF-5 and LSVM. *Arch. Acoust.* **2015**, *40*, 197–203.
13. Huang, W.; Sun, H.; Luo, J. Periodic feature oriented adapted dictionary free OMP for rolling element bearing incipient fault diagnosis. *Mech. Syst. Signal Pr.* **2019**, *126*, 137–160. [[CrossRef](#)]
14. Shen, C.; Wang, D.; Kong, F.; Tse, P. Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier. *Measurement* **2013**, *46*, 1551–1564. [[CrossRef](#)]
15. Mbo’O, C.P.; Hameyer, K. Fault Diagnosis of Bearing Damage by Means of the Linear Discriminant Analysis of Stator Current Features From the Frequency Selection. *IEEE Trans. Ind. Appl.* **2016**, *52*, 3861–3868. [[CrossRef](#)]
16. Zheng, J.; Pan, H. Mean-optimized mode decomposition: An improved EMD approach for non-stationary signal processing. *ISA Trans.* **2020**, *106*, 392–401. [[CrossRef](#)]
17. Wang, J.; Du, G.; Zhu, Z.; Shen, C.; He, Q. Fault diagnosis of rotating machines based on the EMD manifold. *Mech. Syst. Signal Pr.* **2020**, *135*, 135–156. [[CrossRef](#)]
18. Ye, X.; Hu, Y.; Shen, J.; Chen, C.; Zhai, G. An Adaptive Optimized TVF-EMD Based on a Sparsity-Impact Measure Index for Bearing Incipient Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 99–110.
19. Fei, S. Fault diagnosis of bearing based on wavelet packet transform-phase space reconstruction-singular value decomposition and svm classifier. *Arab. J. Sci. Eng.* **2017**, *42*, 1967–1975. [[CrossRef](#)]
20. Feng, Z.; Liang, M.; Chu, F. Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples. *Mech. Syst. Signal Pr.* **2013**, *38*, 165–205. [[CrossRef](#)]
21. Zhong, Y.; Fan, H.; Lu, J.; Pang, L. Research on Fault Diagnosis of Rolling Bearing Based on Wavelet Packet Transform and IPSO-SVM. In *Proceedings of the 2018 IEEE International Conference on Industrial Engineering and Engineering Management*, Bangkok, Thailand, 16–19 December 2018; pp. 1682–1686.
22. Kang, M.; Kim, J.; Kim, J.; Tan, A.; Kim, E.; Choi, B. Reliable Fault Diagnosis for Low-Speed Bearings Using Individually Trained Support Vector Machines With Kernel Discriminative Feature Analysis. *IEEE Trans. Power Electron.* **2015**, *30*, 2786–2797. [[CrossRef](#)]
23. Shi, P.; Guo, X.; Han, D.; Fu, R. A sparse auto-encoder method based on compressed sensing and wavelet packet energy entropy for rolling bearing intelligent fault diagnosis. *J. Mech. Sci. Technol* **2020**, *34*, 1455–1458. [[CrossRef](#)]
24. Kong, X.; Mao, G.; Wang, Q.; Ma, H.; Yang, W. A multi-ensemble method based on deep auto-encoders for fault diagnosis of rolling bearings. *Measurement* **2020**, *151*, 107–132. [[CrossRef](#)]
25. Ma, M.; Li, T.; Xu, L.; Liu, L.; Silva, C. Fault Diagnosis for Rotating Machinery Using Multiple Sensors and Convolutional Neural Networks. *IEEE-ASME Trans. Mech.* **2018**, *23*, 101–110.
26. Chen, J.; Wang, J.; Zhu, J.; Lee, T.; Silva, C. Unsupervised Cross-domain Fault Diagnosis Using Feature Representation Alignment Networks for Rotating Machinery. *IEEE-ASME Trans. Mech.* **2020**. [[CrossRef](#)]
27. Chaabi, L.; Lemzadmi, A.; Djebala, A.; Bouhalais, M.L.; Ouelaa, N. Fault diagnosis of rolling bearings in non-stationary running conditions using improved CEEMDAN and multivariate denoising based on wavelet and principal component analyses. *Int. J. Adv. Manuf. Tech.* **2020**, *107*, 3859–3873. [[CrossRef](#)]
28. Hu, X.; Xiao, Z.; Liu, D.; Tang, Y.; Malik, P.; Xia, X. KPCA and AE Based Local-Global Feature Extraction Method for Vibration Signals of Rotating Machinery. *Math. Probl. Eng.* **2020**, *2020*, 1–17. [[CrossRef](#)]
29. Shen, F.; Chen, C.; Yan, R.; Gao, R. Bearing fault diagnosis based on SVD feature extraction and transfer learning classification. In *2015 Prognostics and System Health Management Conference*, Beijing, China, 21–23 October 2015; pp. 1–6.
30. Liu, H.; Qi, X.; Xu, K.; Xiao, X. A method for identifying weak faults of rolling bearings around railway based on pca & svd-lmd. In *Proceedings of the 38st Chinese Control Conference*, Guangzhou, China, 27–30 July 2019; pp. 4854–4859.
31. Kedadouche, M.; Liu, Z. Fault feature extraction and classification based on WPT and SVD: Application to element bearings with artificially created faults under variable conditions. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2016**, *231*, 4186–4196. [[CrossRef](#)]
32. Cheng, J.; Yu, D.; Tang, J.; Yang, Y. Application of SVM and SVD technique based on EMD to the fault diagnosis of the rotating machinery. *Shock. Vib.* **2013**, *16*, 89–98. [[CrossRef](#)]
33. Yuan, B.; Yao, P.; Zheng, T. Radar emitter signal identification based on weighted normalized singular-value decomposition. *J. Radars* **2019**, *8*, 44–53.
34. Wang, J. *Wavelet Analysis*, 1st ed.; Northwest A&F University Press Co. Ltd.: Xi’an, China, 2004; pp. 89–98.

35. Fu, W.; Shao, K.; Tian, J.; Wang, K. Fault Diagnosis for Rolling Bearings Based on Composite Multiscale Fine-Sorted Dispersion Entropy and SVM With Hybrid Mutation SCA-HHO Algorithm Optimization. *IEEE Access* **2020**, *8*, 13086–13104. [[CrossRef](#)]
36. Mallat, S. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal.* **1989**, *11*, 647–693. [[CrossRef](#)]
37. He, Q. Vibration signal classification by wavelet packet energy flow manifold learning. *Sound Vib.* **2013**, *332*, 1881–18941. [[CrossRef](#)]
38. Pal, M.; Foody, G. Feature selection for classification of hyperspectral data by svm. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2297–2307. [[CrossRef](#)]
39. Huang, H.; Baddour, N. Bearing vibration data collected under time-varying rotational speed conditions. *Data Brief* **2018**, *21*, 1745–1749. [[CrossRef](#)]
40. Guo, Y. Research on Wavelet Based Selection for Vibration Signal Processing. Master's Thesis, Hefei University of Technology, Hefei, China, 2003.
41. Chang, C.; Lin, C. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2007**, *2*, 1–27. [[CrossRef](#)]
42. Huang, H.; Natalie, B.; Liang, M. Bearing fault diagnosis under unknown time-varying rotational speed conditions via multiple time-frequency curve extraction. *J. Sound Vib.* **2018**, *414*, 43–60. [[CrossRef](#)]
43. The Case Western Reserve University Bearing Data Center. Available online: <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file> (accessed on 25 September 2019).