

Article

Bearing Fault Diagnosis Based on Randomized Fisher Discriminant Analysis

Hejun Ye ¹, Ping Wu ^{1,2,*}, Yifei Huo ¹, Xuemei Wang ¹, Yuchen He ², Xujie Zhang ³ and Jinfeng Gao ¹ ¹ School of Information Science and Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, China² Key Laboratory of Intelligent Manufacturing Quality Big Data Tracing and Analysis of Zhejiang Province, China Jiliang University, Hangzhou 310018, China³ College of Control Science and Engineering, Zhejiang University, Hangzhou 310058, China

* Correspondence: pingwu@zstu.edu.cn

Abstract: In this paper, a novel randomized Fisher discriminant analysis (RFDA) based bearing fault diagnosis method is proposed. First, several representative time-domain features are extracted from the raw vibration signals. Second, linear Fisher discriminant analysis (FDA) is extended to nonlinear FDA named RFDA by introducing the random feature map to deal with the non-linearity issue. Specifically, the extracted time-domain features data are mapped onto a high-dimensional space using the random feature map function rather than kernel functions. Third, the time-domain features are fed into the built RFDA model to extract the discriminant features for diagnosis. Moreover, a Bayesian inference is employed to identify the class of the collected vibration signals to diagnose the bearing status. The proposed method uses random Fourier features to approximate the kernel matrix in the kernel Fisher discriminant analysis. Through employing randomized Fisher discriminant analysis, the nonlinearity issue is dealt with, and the computational burden is remarkably reduced compared to the kernel Fisher discriminant analysis (KFDA). To illustrate the superior performance of the proposed RFDA-based bearing fault diagnosis method, comparative experiments are conducted on two widely used datasets, the Case Western Reserve University (CWRU) bearing dataset and the Paderborn University (PU) bearing dataset. For the CWRU dataset, the computation time of RFDA is much shorter than KFDA, while the accuracy rate reaches the same level of KFDA. For the PU dataset, the accuracy rate of RFDA is slightly higher than KFDA, and the computation time is only 44.14% of KFDA.

Keywords: bearing; fault diagnosis; random Fourier feature; Fisher discriminant analysis

Citation: Ye, H.; Wu, P.; Huo, Y.; Wang, X.; He, Y.; Zhang, X.; Gao, J. Bearing Fault Diagnosis Based on Randomized Fisher Discriminant Analysis. *Sensors* **2022**, *22*, 8093. <https://doi.org/10.3390/s22218093>

Academic Editors: Kim Phuc Tran, Athanasios Rakitzis and Khanh T. P. Nguyen

Received: 8 September 2022

Accepted: 19 October 2022

Published: 22 October 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

Bearings are an essential and key part that is widely used in modern rotating machinery. As a vital component, the occurrence of bearing faults will result in significant breakdown time, increasing maintenance costs, and even jeopardizing casualties. Therefore, it is critical to precisely and quickly diagnose the bearing status [1–3]. To explore the bearing status, a variety of signals are collected and used, such as acoustic signals [4], vibration signals [5], and current signals [6]. Among them, the vibration signals contain abundant fault energy information, and the data acquisition of bearing vibration signals does not require complex equipment and professionals.

Therefore, the vibration signal is popularly used to monitor the bearing status. Bearing fault diagnosis techniques via vibration signals can be generally categorized into two classes, including signal-analysis-based and data-driven methods. Regarding the signal-analysis-based method, the raw vibration signals are firstly analyzed using signal processing methods such as time-domain analysis [7,8], frequency-domain analysis [9] and time–frequency-domain analysis [10]. Afterward, the bearing status is determined by features extracted from different domains using expert knowledge.

Data-driven methods depend only on the vibration signals, as opposed to signal analysis-based methods. In data-driven methods, the bearing fault diagnosis becomes a pattern recognition problem. To deal with the pattern recognition of high-dimensional data in bearing fault diagnosis, dimensionality reduction techniques have been widely employed, such as principal component analysis (PCA) [11], locality preserving projection (LPP) [12], and recurrence analysis (RA) [13,14]. Moreover, t-distributed stochastic neighbor embedding (t-SNE) is an efficient dimensionality reduction tool. t-SNE identifies close similarities between samples through the relative location of points in the mapped feature space. For t-SNE, the number of features in the reduced space is not restricted by the number of output dimensions [15]. In [15], t-SNE is combined with the multiscale distribution entropy method to extract the low-dimensional nonlinear complexity features from the vibration signals of rolling bearing. Similarly, uniform manifold approximation and projection (UMAP) uses graph layout algorithms to arrange data in a low-dimensional space [16]. In [16], UMAP is combined with feature selection techniques to improve the performance of latent space visualization for chemical process data. As a popular unsupervised learning method, autoencoder (AE) can learn effective features with unlabeled data by minimizing the error between original input and reconstructed input. Deep autoencoder (DAE) has been widely used to extract hierarchical features for bearing fault diagnosis from vibration signal [17]. Variational autoencoder (VAE) generates a latent representation of the data through imposing a distribution over the latent variables [18]. In [18], a semi-supervised learning scheme using variational autoencoder (VAE)-based deep generative models was proposed to deal with the small labeled data problem in the bearing fault diagnosis.

Compared to unsupervised dimensionality reduction techniques, the supervised method, Fisher discriminant analysis (FDA), aims to find the low-dimensional representation from the high-dimensional data to simultaneously maximize the distance between different classes and minimize the distance within the same class. FDA was also named linear discriminant analysis (LDA). More importantly, FDA can fully utilize the labeled information to directly offer the classification results. Therefore, FDA has gained considerable attention to achieving the task of bearing fault diagnosis in recent years. Due to its simplicity and efficiency, FDA has proved its superiority in fault diagnosis.

Jin et al. [19] developed trace ratio linear discriminant analysis (TR-LDA) to address the non-Gaussian data by solving the trace ratio problem. Zhou et al. [20] employed LDA to reduce the dimensionality of 10 statistical features of the raw vibration signals and its transient component using transient-extracting transform (TET) for improving the fault diagnosis performance. To address the nonlinearity issue contained in vibration signals, variants of nonlinear fisher discriminant analysis have been proposed. One of the most frequent extensions is kernel Fisher discriminant analysis (KFDA) [21]. The idea behind KFDA is to map the original data onto a high-dimensional feature space in which linearly separable feature space is expected for further classification. Van et al. [22] presented a wavelet kernel-based local Fisher discriminant analysis (WKLFDA) to extract the nonlinear features from original vibration signals for bearing fault diagnosis. Jiang et al. [23] proposed a semi-supervised kernel marginal Fisher analysis (SSKMFA) method to investigate the inherent manifold structure embedded in data, and simultaneously took the intra-class compactness and the inter-class separability into account. Tao et al. [24] developed a semi-supervised kernel local Fisher discriminant analysis (SSKLFDA) through introducing the regularization term with pseudo labels by utilizing unlabeled data for the supervised dimensionality reduction.

However, a large number of features extracted from kernel methods will increase the computation burden and may lead to poor fault diagnosis performance, due to all training data being involved in the calculation of the kernel matrix. To lessen the computing complexity of KFDA, Li et al. [25] presented a feature vector selection (FVS) approach by using a subset of the samples to present all of the data under the geometrical consideration. Liu et al. [26] combined the kernel feature selection method and KFDA technique to reduce the computation burden and alleviate the impact of irrelevant features in fault diagnosis.

Recently, random feature map was widely studied in large-scale kernel machines [27,28]. Rahimi and Recht suggested a random Fourier feature mapping approach for approximating non-linear kernels by mapping the input data to randomized low-dimensional feature space [27]. Fisher discriminant analysis using random Fourier feature mapping was developed to accelerate kernel Fisher discriminant analysis in [29,30]. In [30], a random feature map was introduced to map the input data to a finite dimension to accelerate FDA and kernel FDA. Moreover, a theoretical guarantee was offered to prove that the FDA algorithms using random projection can derive good generalization ability. In [29], the randomized solution to linear discriminant analysis was developed for processing hyperspectral images to overcome the dimensionality problem. To the best of our knowledge, there is little investigation of Fisher discriminant analysis using random feature map in bearing fault diagnostics.

Motivated by the above discussions, a new bearing fault diagnosis method is proposed by using randomized Fisher discriminant analysis (RFDA). Specifically, 12 time-domain features are first extracted from original vibration signals. Then, an RFDA model is built by using the extracted 12 time-domain features for fault diagnosis. In RFDA, the high-dimensional data are mapped to a low-dimensional features space using random feature mapping, then the projection matrix with Fisher discriminant analysis is calculated. To identify the state of bearings, a Bayesian inference is employed. The main contributions of this paper lie on the following aspects:

- RFDA, a nonlinear variant of FDA, is utilized for bearing fault diagnosis. The RFDA-based method can achieve similar performance to the KFDA-based method, while the computational burden is remarkably reduced.
- Two widely used bearing datasets are employed to validate the effectiveness of the proposed RFDA-based bearing fault diagnosis method. Results show the superior performance of the proposed method over other related methods.

The remainder of this paper is structured as follows. In Section 2, a brief review of Fisher discriminant analysis and random Fourier feature map is given. In Section 3, details of the proposed RFDA based fault diagnosis method are described. In Section 4, two experimental datasets of bearings are used to assess the performance of the proposed method comparing with other related methods. Conclusions are drawn in the final Section 5.

2. Related Works

2.1. Fisher Discriminant Analysis

The aim of the FDA is to find a linear transformation matrix to separate the projections in the low-dimensional space as much as possible. For this purpose, FDA measures the compactness of each class with the within-class scatter matrix and the distance between classes with the between-class scatter matrix. Through maximizing the ratio of between-class to within-class scatter matrices, the optimal transformation matrix is calculated.

Denote the training dataset as D ,

$$D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\}$$

where the sample $\mathbf{x}_i \in \mathbb{R}^m$ is the m -dimensional vector. y_i represents the category of \mathbf{x}_i . N samples are contained in the training dataset D .

Assumed that there are k types in D . Define $n_j (j = 1, 2, \dots, c)$ as the number of samples of type y_j , and $\mathbf{X}_j (j = 1, 2, \dots, c)$ as the set of all samples which belong to type y_j . Then, the between-class scatter matrix \mathbf{S}_b is defined as

$$\mathbf{S}_b = \sum_{j=1}^c n_j (\boldsymbol{\mu}_j - \boldsymbol{\mu}) (\boldsymbol{\mu}_j - \boldsymbol{\mu})^T \quad (1)$$

where $\boldsymbol{\mu}_j (j = 1, 2, \dots, c)$ is the mean vector of the type y_j , and $\boldsymbol{\mu}$ is the mean vector of all samples.

The within-class scatter matrix \mathbf{S}_w is defined as,

$$\mathbf{S}_w = \sum_{j=1}^c \sum_{\mathbf{x} \in X_j} (\mathbf{x} - \boldsymbol{\mu}_j)(\mathbf{x} - \boldsymbol{\mu}_j)^T \quad (2)$$

To find the optimal projection matrix \mathbf{W} , the optimization problem of FDA is defined as,

$$\arg \max_{\mathbf{W}} J_{FDA} = \frac{\text{Tr}(\mathbf{W}^T \mathbf{S}_b \mathbf{W})}{\text{Tr}(\mathbf{W}^T \mathbf{S}_w \mathbf{W})} \quad (3)$$

where $\text{Tr}(\cdot)$ is the trace operator.

Assumed that \mathbf{S}_w is non-singular, the solution of the optimization problem (3) can be feasibly obtained by using generalized eigenvalue decomposition [31],

$$\mathbf{S}_b \mathbf{w}_j = \rho_j \mathbf{S}_w \mathbf{w}_j \quad (4)$$

where $\mathbf{w}_j \in \mathbb{R}^m$ is the eigenvector and the corresponding eigenvalue is ρ_j . Generally, the eigenvectors which are corresponding to the largest d eigenvalues are retained for the purpose of dimensionality reduction. Thus, the projection matrix $\bar{\mathbf{W}}$ is constructed as $\bar{\mathbf{W}} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_d] \in \mathbb{R}^{m \times d}$.

Using the derived projection matrix $\bar{\mathbf{W}}$, the low-dimensional projection vector $\mathbf{z}_i \in \mathbb{R}^d$ of the original data \mathbf{x}_i is,

$$\mathbf{z}_i = \mathbf{x}_i^T \bar{\mathbf{W}} \quad (5)$$

2.2. Random Fourier Feature Map

To address the issue of nonlinearity, kernel methods are usually employed to extend the linear dimensionality reduction methods to their nonlinear variants. According to Cover's theorem, the original input data $\mathbf{x} \in \mathbb{R}^m$ is mapped to a high-dimensional, even an infinite, reproducing kernel Hilbert space (RKHS) $\in \mathbb{R}^{\mathcal{F}}$ by a given nonlinear mapping function ϕ .

Define the mapping function as,

$$\phi(\mathbf{x}) : \mathbb{R}^m \rightarrow \mathbb{R}^{\mathcal{F}} \quad (6)$$

Since the nonlinear features $\phi(\mathbf{x})$ are created from the implicit mapping function, the core of the kernel methods relies on the implicit lifting through the kernel trick. Using the kernel trick, the nonlinear features are generated by calculating the inner product between pairs of input points.

The inner product between lifted data points $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{x}_j)$ is defined as,

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle \quad (7)$$

where $\langle \cdot \rangle$ represents the inner product operator.

In the kernel method, the kernel matrix can be constructed for N training samples $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ as,

$$\mathbf{K} = [k(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1,2,\dots,N}$$

To generate the features for a testing data point, the kernel matrix \mathbf{K} should be evaluated. As displayed in the kernel matrix \mathbf{K} , all training data are involved. As a result, the disadvantage of the kernel method is that there are large computational and storage costs while facing large training sets.

For addressing this issue, the shift-invariant kernels $k(\mathbf{x}_i, \mathbf{x}_j)$ are related to random non-linear features with the help of Bochner's theorem [32]. Defining $\beta_\omega(\mathbf{x}) = \exp(-j\omega^T(\mathbf{x}))$, the inner product $k(\mathbf{x}_i, \mathbf{x}_k)$ can be expressed as,

$$\begin{aligned} k(\mathbf{x}_i, \mathbf{x}_k) &= \int_{\mathbb{R}^d} p(\omega) \exp(-j\omega^T(\mathbf{x}_i - \mathbf{x}_k)) d\omega \\ &= E_\omega [\beta_\omega(\mathbf{x}_i) \beta_\omega(\mathbf{x}_k)^*] \end{aligned} \quad (8)$$

where $p(\omega)$ is the inverse Fourier transform of k . ω is sampled from the distribution $p(\omega)$. $E_\omega[\cdot]$ is the expectation operator. $\beta_\omega(\mathbf{x}_k)^*$ is the complex conjugate of the inverse Fourier transform of $\beta_\omega(\mathbf{x}_k)$. $\beta_\omega(\mathbf{x}_i) \beta_\omega(\mathbf{x}_k)^*$ can be considered an unbiased estimate of $k(\mathbf{x}_i, \mathbf{x}_k)$.

Furthermore, the kernel $k(\mathbf{x}_i, \mathbf{x}_k)$ is approximated as below,

$$\begin{aligned} k(\mathbf{x}_i, \mathbf{x}_k) &= \int_{\mathbb{R}^d} p(\omega) \exp(-j\omega^T(\mathbf{x}_i - \mathbf{x}_k)) d\omega \\ &= E_\omega [\beta_\omega(\mathbf{x}_i) \beta_\omega(\mathbf{x}_k)^*] \\ &\approx \sum_{j=1}^{\mathcal{D}} \frac{1}{\mathcal{D}} \exp(-j\omega_j^T \mathbf{x}_i) \exp(-j\omega_j^T \mathbf{x}_k) \end{aligned} \quad (9)$$

where $\mathcal{D} \ll N$.

According to Euler's formula, it obtains

$$\exp(-j\omega^T \mathbf{x}) = \cos(\omega \mathbf{x}) + j \sin(\omega \mathbf{x}) \quad (10)$$

To obtain a real-value random feature for k , the distribution $p(\omega)$ and kernel k should be real. Thus, only the real part of the exponential would remain in Equation (11). By replacing $\exp(-j\omega^T(\mathbf{x}_i - \mathbf{x}_k))$ with $\cos(\omega^T(\mathbf{x}_i - \mathbf{x}_k))$ [27], then

$$\begin{aligned} k(\mathbf{x}_i, \mathbf{x}_k) &\approx \sum_{j=1}^{\mathcal{D}} \frac{1}{\mathcal{D}} \exp(-j\omega_j^T \mathbf{x}_i) \exp(-j\omega_j^T \mathbf{x}_k) \\ &= \left\langle \frac{1}{\sqrt{\mathcal{D}}} \mathbf{z}(\mathbf{x}_i), \frac{1}{\sqrt{\mathcal{D}}} \mathbf{z}(\mathbf{x}_k) \right\rangle \end{aligned} \quad (11)$$

Based on Equation (11), the original data are explicitly projected onto a low-dimensional Euclidean inner product space through randomized feature map in [27], instead of implicitly mapping function ϕ . Following this idea, the inner product $k(\mathbf{x}_i, \mathbf{x}_j)$ is approximated by defining an explicit random feature map function $z(\mathbf{x}) : \mathbb{R}^m \rightarrow \mathbb{R}^{\mathcal{D}}$, where

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle \approx \mathbf{z}(\mathbf{x}_i)^T \mathbf{z}(\mathbf{x}_j) \quad (12)$$

where $z_\omega(\mathbf{x}) = \sqrt{2} \cos(\omega^T \mathbf{x} + b)$ with b is drawn from uniform distribution $\mathcal{U}[0, 2\pi]$. It is noted that the parameters b are used to make the expectation of the inner product of $z_\omega(\mathbf{x})$ close to the shift-invariant kernel. For further explanation of the parameters b , refer to [27,33].

Thus, the inner product $\mathbf{z}_\omega(\mathbf{x}_i)^T \mathbf{z}_\omega(\mathbf{x}_k)$ is expressed as $\mathbf{z}_\omega(\mathbf{x}_i)^T \mathbf{z}_\omega(\mathbf{x}_k) = \frac{1}{\mathcal{D}} \sum_{j=1}^{\mathcal{D}} z_{\omega_j}(\mathbf{x}_i) z_{\omega_j}(\mathbf{x}_k)$ where \mathcal{D} samples of z_{ω_j} are randomly chosen. Define \mathcal{D} -dimensional vector $\mathbf{z}_\omega(\mathbf{x})$

$$\mathbf{z}_\omega(\mathbf{x}) = \sqrt{\frac{2}{\mathcal{D}}} \begin{bmatrix} \cos(\omega_1^T \mathbf{x} + b_1) \\ \cos(\omega_2^T \mathbf{x} + b_2) \\ \vdots \\ \cos(\omega_{\mathcal{D}}^T \mathbf{x} + b_{\mathcal{D}}) \end{bmatrix} \quad (13)$$

Based on the random feature map \mathbf{z} , a \mathcal{D} -dimensional feature is obtained from the original data \mathbf{x} . Then, linear FDA is conducted in the \mathcal{D} -dimensional feature space. The details of Fisher discriminant analysis with random feature map are elaborated in the next section.

Remark 1. The bound error between the kernel matrix $\mathbf{K}_{i,k} = k(\mathbf{x}_i, \mathbf{x}_k)$ and $\bar{\mathbf{K}}_{i,k} = \mathbf{z}(\mathbf{x}_i)^T \mathbf{z}(\mathbf{x}_k)$ was discovered in spectral norm as in [28],

$$\mathbf{E}\|\mathbf{K} - \bar{\mathbf{K}}\| \leq \sqrt{\frac{3N^2 \log N}{\mathcal{D}}} + \frac{2N \log N}{\mathcal{D}} \quad (14)$$

Obviously, the error becomes small as the dimensionality \mathcal{D} is selected as a large number. Yet, the computational cost will also increase. To balance the computational burden and the error, the dimensionality \mathcal{D} of random features can be determined by cross-validation in the offline training phase.

3. RFDA-Based Fault Diagnosis

The RFDA-based fault diagnosis is developed in this section. First, the time-domain features are extracted from the collected raw vibration signals. Then, an RFDA model is trained using these time-domain features. To identify the class of vibration signals, a Bayesian inference is employed.

3.1. Time-Domain Feature Extraction

Time-domain statistical features, such as the impulsive factor, kurtosis, skewness, peak factor, and root mean square, are usually employed to detect and identify the bearing damage. In this study, 12 time-domain features are computed from each signal to exhibit different distribution characteristics of the raw vibration signals. Table 1 lists the detailed descriptions of the adopted time-domain features.

Table 1. The 12 time-domain features.

Features	Equations
Peak	$x_p = \max(\mathbf{x})$
Peak-to-peak	$x_{pp} = \max(\mathbf{x}) - \min(\mathbf{x})$
Mean	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$
Absolute mean amplitude	$ \bar{x} = \frac{\sum_{i=1}^n x_i }{n}$
Square root amplitude	$x_{sra} = \left(\frac{\sqrt{\sum_{i=1}^n x_i }}{n} \right)^2$
Variance	$x_v = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$
Standard deviation	$x_{std} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$
Root mean square	$x_{rms} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}$
Kurtosis	$x_k = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{(n-1)x_{std}^4}$
Skewness	$x_{ske} = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)x_{std}^3}$
Peak factor	$x_{pf} = \frac{x_p}{x_{rms}}$
Impulsive factor	$x_{if} = \frac{x_p}{ \bar{x} }$

where $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ denotes the row signal series of each sample, n is the corresponding length of the vibration signal. In this study, the sample of vibration signals contains 1024 data points.

From Table 1, the time-domain features of the raw vibration signals $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ are extracted and denoted as $\mathbf{h} = f(\mathbf{x}) = \{x_p, x_{pp}, \bar{x}, |\bar{x}|, x_{sra}, x_v, x_{std}, x_{rms}, x_k, x_{ske}, x_{pf}, x_{if}\}^T \in \mathbb{R}^{12}$ for the RFDA modeling, where $f(\cdot)$ is the function of the time-domain feature extraction.

3.2. RFDA Model Training

Assumed that N_{all} samples are collected from fault-free and $c - 1$ faulty classes, where each sample has n data points, then $N_{all} \mathbf{h}$ are computed according to Table 1.

For the purpose of training the RFDA model, the extracted time-domain features \mathbf{h} are normalized,

$$\hat{\mathbf{h}} = \frac{\mathbf{h} - \mathbf{h}_{mean}}{\mathbf{h}_{std}} \quad (15)$$

where \mathbf{h}_{mean} and \mathbf{h}_{std} are the mean and standard variance of the time-domain features \mathbf{h} extracted from normal data, respectively.

To address the issue of nonlinearity in data, the random Fourier feature of \mathbf{h} is extracted using random feature map,

$$\mathbf{z}(\hat{\mathbf{h}}) = \sqrt{\frac{2}{m}} \begin{bmatrix} \cos(\omega_1^T \hat{\mathbf{h}} + b_1) \\ \vdots \\ \cos(\omega_m^T \hat{\mathbf{h}} + b_m) \end{bmatrix} \in \mathbb{R}^m \quad (16)$$

where m is the dimension of random features.

In kernel-based fault diagnosis, the Gaussian radial basis function (RBF) kernel is usually employed due to its generalization ability. The RBF kernel is defined as,

$$K(\delta) = \exp\left(-\frac{\|\delta\|_2^2}{s}\right) \quad (17)$$

where s is the kernel width of RBF kernel. Corresponding to the selected RBF kernel, the parameter ω of random Fourier features $\mathbf{z}(\hat{\mathbf{h}})$ is sampled from the following Gaussian distribution,

$$\omega \sim \mathcal{N}(\mathbf{0}, 2\mathbf{I}/s) \quad (18)$$

where \mathbf{I} is the identity matrix with appropriate dimension.

Then, using Equations (15)–(17) random Fourier features, the between-class scatter matrix (\mathbf{S}_{hb}) and within-class scatter matrix (\mathbf{S}_{hw}) are formed,

$$\mathbf{S}_{hb} = \sum_{j=1}^c N_j \left(\widehat{\mathbf{z}}_{\mathbf{h}_{m,j}} - \widehat{\mathbf{z}}_{\mathbf{h}_m} \right) \left(\widehat{\mathbf{z}}_{\mathbf{h}_{m,j}} - \widehat{\mathbf{z}}_{\mathbf{h}_m} \right)^T \quad (19)$$

where $\widehat{\mathbf{z}}_{\mathbf{h}_{m,j}}$ is the mean of random Fourier features, which belong to the $j = 1, \dots, c$ th class. $\widehat{\mathbf{z}}_{\mathbf{h}_m}$ is the mean of all extracted features.

$$\mathbf{S}_{hw} = \sum_{j=1}^c \mathbf{S}_{hwj} = \sum_{j=1}^c \sum_{\hat{\mathbf{z}}_{\mathbf{h}} \in \hat{\mathbf{H}}_j} \left(\hat{\mathbf{z}}_{\mathbf{h}} - \widehat{\mathbf{z}}_{\mathbf{h}_{m,j}} \right) \left(\hat{\mathbf{z}}_{\mathbf{h}} - \widehat{\mathbf{z}}_{\mathbf{h}_{m,j}} \right)^T \quad (20)$$

where $\hat{\mathbf{H}}_j$ represents the domain of the j th class.

Similar to Equation (4), to seek the projection matrix \mathbf{W}_h that maximizes the distance between samples that belong to different classes and minimize the distance between samples that belong to the same classes, the following generalized eigenvalue problem is to be solved:

$$\mathbf{S}_{hb} \mathbf{w}_{hi} = \lambda_i \mathbf{S}_{hw} \mathbf{w}_{hi} \quad (21)$$

where λ_i is the eigenvalue corresponding to the eigenvector \mathbf{w}_{hi} .

Using eigenvalue decomposition, the solution of Equation (20) can be derived. To reduce the dimensionality, only eigenvectors corresponding to large eigenvalues are retained. Thus, the projection matrix $\bar{\mathbf{W}}_h$ is formed as $\bar{\mathbf{W}}_h = [\mathbf{w}_{h1} \ \mathbf{w}_{h2} \ \dots \ \mathbf{w}_{hd}] \in \mathbb{R}^{m \times d}$, where d is the number of retained latent variables.

Remark 2. The selection of the number of retained latent variables d can be determined by minimizing the information criterion similar to Akaike's information criterion (AIC) [34]. In discriminant analysis, the number of retained latent variables d is usually set as $c - 1$, where c is the number of sample classes.

3.3. RFDA-Based Bearing Fault Diagnosis Scheme

Through the projection matrix $\bar{\mathbf{W}}_h$ derived from the RFDA model, the projection \mathbf{d}_{ht} of a new vibration sample \mathbf{x}_t can be calculated as,

$$\mathbf{d}_{ht} = \bar{\mathbf{W}}_h^T \mathbf{z}(f(\mathbf{x}_t)) \quad (22)$$

Based on the established RFDA model, k-nearest neighbors (kNN) and Bayesian inference are often adopted to classify the class of the new samples. Due to its simplicity and low computational burden, Bayesian inference is used in this study. Bayesian inference is based on the posterior probability to determine the class of new data samples. Assuming that the projections \mathbf{d}_{ht} are Gaussian distributed, then the class membership can be determined by the Fisher discriminant function, which is defined as [35],

$$\begin{aligned} \mathbf{g}_j(\mathbf{d}_{ht}) = & -\frac{1}{2} [(\mathbf{d}_{ht} - \mathbf{d}_{m,j})^T (\frac{1}{n_j - 1} \mathbf{W}_n^T \mathbf{S}_{hw,j} \mathbf{W}_n)^{-1} (\mathbf{d}_{ht} - \mathbf{d}_{m,j}) \\ & - \ln[\det(\frac{1}{n_j - 1} \mathbf{W}_h^T \mathbf{S}_{hw,j} \mathbf{W}_h)]] \end{aligned} \quad (23)$$

where $\mathbf{d}_{m,j} = \mathbf{W}_b^T \hat{\mathbf{h}}_{m,j}$.

In accordance with the Fisher discriminant function, the class of the test data sample \mathbf{x}_t is classified to the class, where the value of the Fisher discriminant function is the largest:

$$C(\mathbf{x}_t) = \arg \max_{1 \leq j \leq c} g_j(\bar{\mathbf{W}}_h^T \mathbf{z}(f(\mathbf{x}_t))) \quad (24)$$

The flowchart of the proposed RFDA-based bearing diagnosis is depicted in Figure 1.

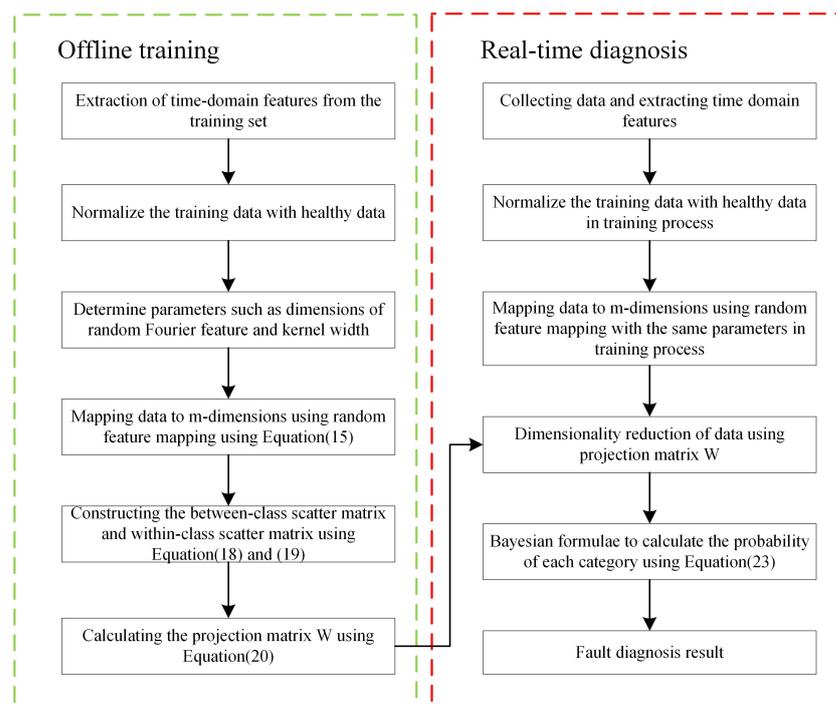


Figure 1. Flowchart of the proposed RFDA-based bearing diagnosis.

4. Experiments and Results

In this section, to demonstrate the applicability and effectiveness of the proposed RFDA-based fault diagnosis method, experiments on two popular bearing benchmark datasets are conducted. For comparison study, FDA and KFDA [36] are used. The simulation environment is Intel Core i7-8750h CPU@2.20 GHz CPU and 24 GB RAM running under Windows 10 with MATLAB R2020b.

4.1. Case 1: CWRU Dataset

As a commonly and widely used bearing dataset for the evaluation of fault diagnosis performance, the CWRU datasets were collected and provided by the Case Western Reserve University bearing data center on the experimental facility shown in Figure 2.

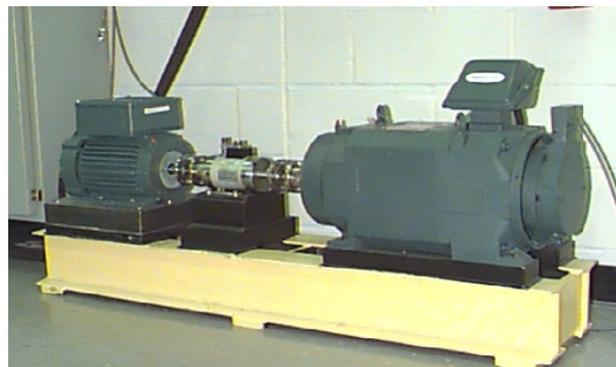


Figure 2. The experimental test rig of CWRU dataset.

In the CWRU dataset, the data with the acquisition frequency of 12 K are selected, and several single-point faults are introduced on the outer ring, the inner ring, and rolling element, with fault diameters of 0.007 for each position. For each scenario, 50 samples are collected. A sample is composed of 1024-point vibration sequences without overlap. The number of training and testing datasets and the descriptions of fault scenarios are shown in Table 2.

Table 2. The number of training and testing datasets and the descriptions of fault scenarios: CWRU dataset.

Bearing State	Fault Location	Train Number	Test Number	Characteristic Frequency (Hz)
Health	/	50	188	29.95
Fault 1	inner ring	50	68	162.1852
Fault 2	rolling element	50	69	141.0907667
Fault 3	outer ring	50	69	107.305

All the vibration signals were collected under the same motor loads at 1797 rpm and 0 HP.

To assess the fault diagnosis performance quantitatively, the widely used index (i.e., accuracy rate) is adopted in this work. The definition of accuracy rate is

$$Acc = \frac{TN + TP}{TN + FN + FP + TP} \quad (25)$$

where TN , FN , FP and TP denote the number of true negative, false negative, false positive, and true positive outcomes, respectively.

To determine the kernel width of KFDA, the cross-validation method is used. Figure 3 shows the fault classification results of the KFDA model with different kernel widths. From Figure 3, the optimal kernel width is selected as 700. The computation time is kept stable (i.e., 0.24 s). To make the comparison fair, the same kernel width s is used for RFDA. Since the selection of dimension m not only affects the performance of approximation

of kernel matrix, but also requires different computational burdens in RFDA-based fault diagnosis method. To obtain optimal parameters of RFDA, the accuracy rates with different parameters are plotted in Figure 4. We can set $s = 700$, $m = 50$ from Figure 4. The number of maintained latent variables is selected as 3 for all methods.

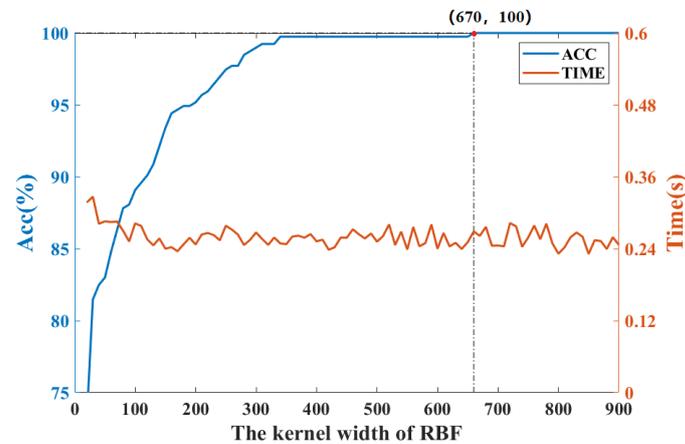


Figure 3. Accuracy rate and computation time vs. kernel width of RBF: CWRU dataset.

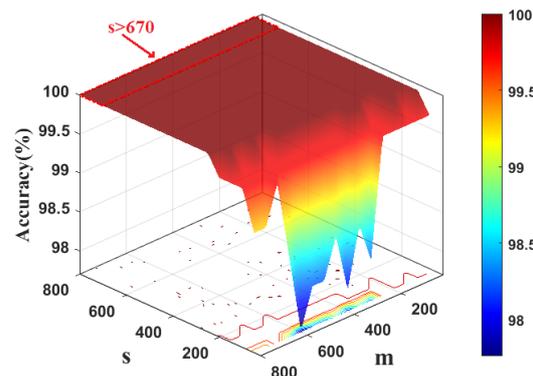


Figure 4. Computation time and accuracy rate vs. kernel width of RBF and dimensionality of random Fourier feature: CWRU dataset.

To visualize the features vividly, the extracted features are displayed in Figure 5. From Figure 5, it can be observed that the distances between different clusters corresponding to specific scenarios are large. Compared to FDA, RFDA and KFDA can provide more efficient discriminant performance, since the distance between the data of the same category is smaller.

The accuracy rates and computation times of FDA, KFDA, and RFDA using the testing dataset are given in Table 3. In Table 3, the accuracy rate and computation time are the average results of 100 experiments. From the data in Table 3, the accuracy rates of FDA, RFDA, and KFDA can achieve 100%. The reason is that the magnitudes of the faults are large in the CWRU data, making them easily classified. On the other hand, the average computation time of RFDA is 0.1232 s, which is at the same level as FDA. However, the computation time of KFDA is almost two times that of RFDA. Thus, the computation burden of the calculation of kernel matrix is hugely reduced in RFDA through the adoption of the random Fourier feature.

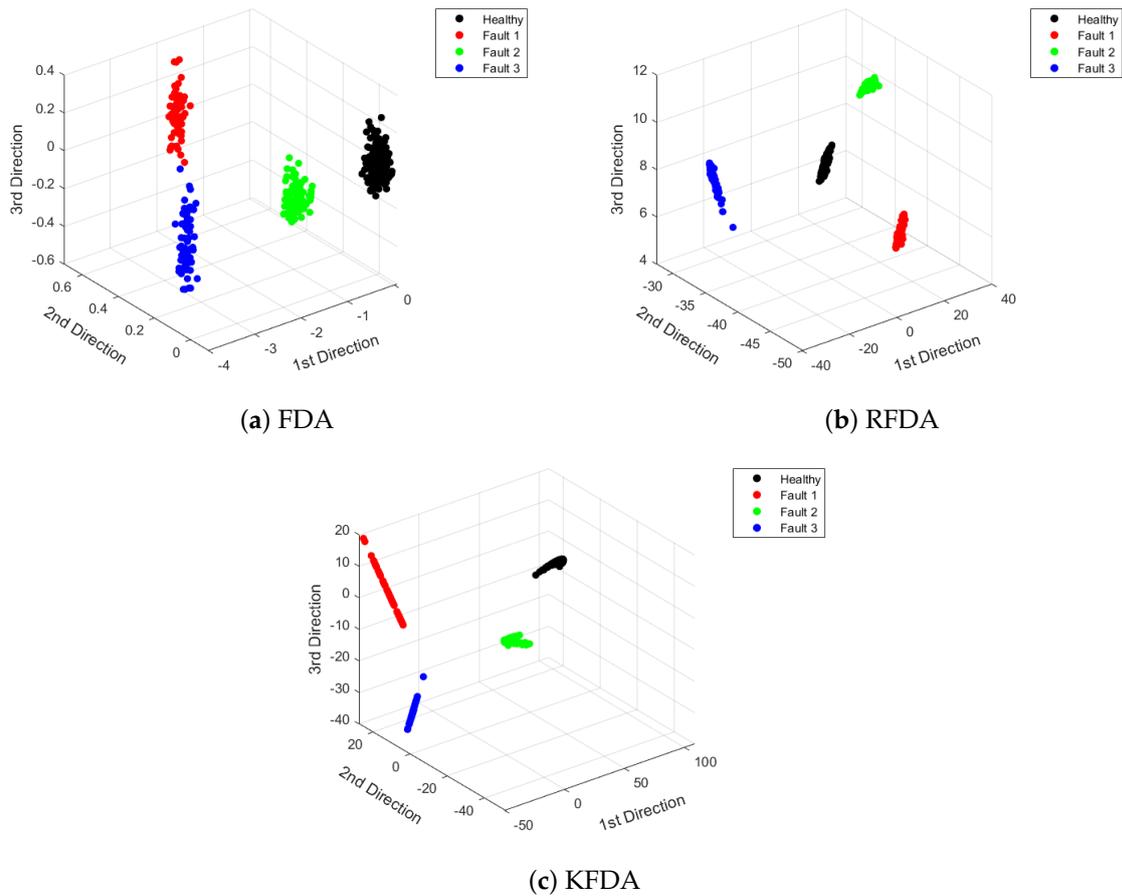


Figure 5. Visualization of dimensionality reduction: CWRU dataset.

Table 3. Computation times and accuracy rates using FDA, KFDA and RFDA: CWRU.

Method	Mean Acc	Mean Time (s)
FDA	100%	0.1146
KFDA	100%	0.2647
RFDA	100%	0.1232

4.2. Case 2: PU Dataset

PU dataset is another popular bearing dataset provided by the Paderborn University Bearing Data Center for comparing the fault diagnosis performance [37]. The test rig of PU datasets is displayed in Figure 6.

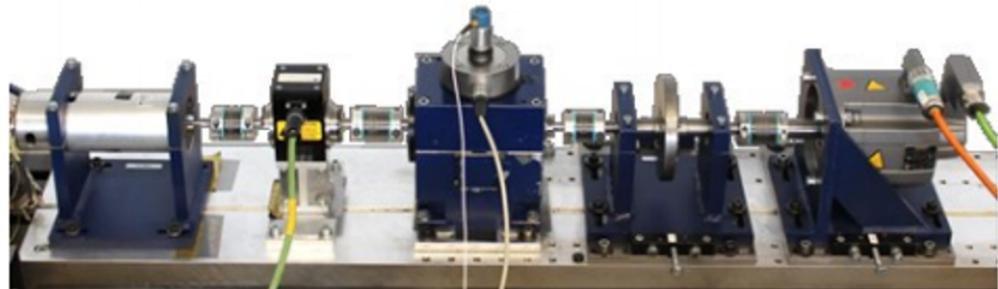


Figure 6. The experimental test rig of PU dataset test rig.

There are 14 types of real failure data in PU dataset. For details of the working conditions, refer to [37]. In this study, the vibration data were collected under the working conditions of rotating speed 1500 rpm, load torque 0.7 nm, and radial force 1000 N.

The sampling frequency of the vibration data is 64 KHz, and the sampling time is 4 s. In the training dataset, 100 samples are included in each class, whereas 150 samples are included in the test dataset for each class. A sample is composed of 1024-point vibration sequences without overlap.

In this study, 8 different types of real failure data, as well as health data, were selected for experiments. The details of used data and descriptions of faults are shown in Tables 4 and 5, respectively.

Table 4. The number of training and testing datasets and the descriptions of fault scenarios: PU dataset.

Bearing State	Bearing Code	Train Number	Test Number
Health	K004	100	150
Fault 1	KA04	100	150
Fault 2	KA16	100	150
Fault 3	KA22	100	150
Fault 4	KA30	100	150
Fault 5	KB23	100	150
Fault 6	KB24	100	150
Fault 7	KB27	100	150
Fault 8	KI16	100	150

Table 5. Detailed description of PU datasets.

Bearing State	Bearing Code	Fault Position	Description
Health	K004	Healthy	Run-in period 5 h
Fault 1	KA04	Outer ring (SP, S, Level 1)	Caused by fatigue and pitting
Fault 2	KA16	Outer ring (SP, R, Level 2)	Caused by fatigue and pitting
Fault 3	KA22	Outer ring (SP, S, Level 1)	Caused by fatigue and pitting
Fault 4	KA30	Outer ring (D, R, Level 1)	Caused by plastic deform and indentation
Fault 5	KB23	Outer ring and inner ring (SP, M, Level 2)	Caused by fatigue and pitting
Fault 6	KB24	Outer ring and inner ring (D, M, Level 3)	Caused by fatigue and pitting
Fault 7	KB27	Outer ring and inner ring (D, M, Level 1)	Caused by plastic deform and indentation
Fault 8	KI16	Inner ring (SP, S, Level 1)	Caused by fatigue and pitting

SP—single point fault; S—single damage; D—distributed fault; R—repetitive damage; M—multiple damage.

Similarly, Figure 7 depicts the fault classification results of the KFDA model with different kernel widths. It can be found that KFDA can achieve the best performance while the kernel width is selected as 610. For RFDA, the optimal parameters s, m are selected as $s = 450, m = 100$ as displayed in Figure 8. The number of retained latent variables for all methods is set as 8.

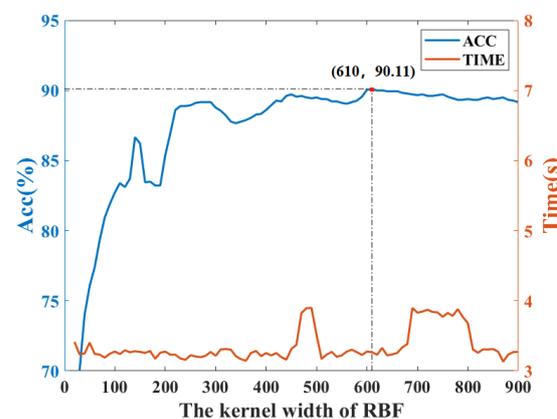


Figure 7. Computation time and accuracy rate vs. kernel width of RBF: PU dataset.

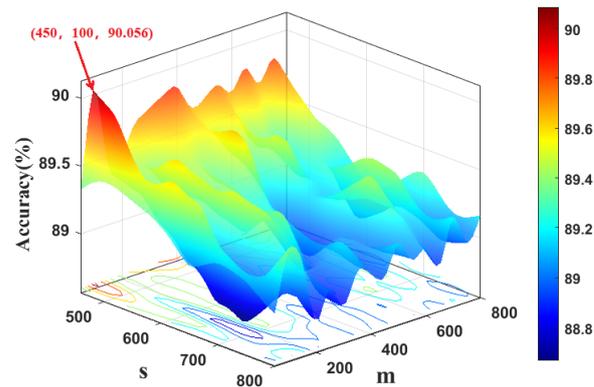


Figure 8. Computation time and accuracy rate vs. kernel width of RBF and dimensionality of random Fourier feature: PU dataset.

Figure 9 is used to display the projections onto the first three FDA loading vectors by standard FDA, KFDA and RFDA. As plotted in Figure 9, it can be observed that the distances between the clusters of different classes are closer. Thus, the fault diagnosis performance will degrade. The reason is that the magnitudes of faults in the PU datasets are smaller. The vibration signals contain more noise. However, the clusters derived by RFDA and KFDA have larger discrepancy than FDA. Thus, higher accuracy rates can be offered. Nevertheless, KFDA and RFDA can provide better discriminant ability than FDA.

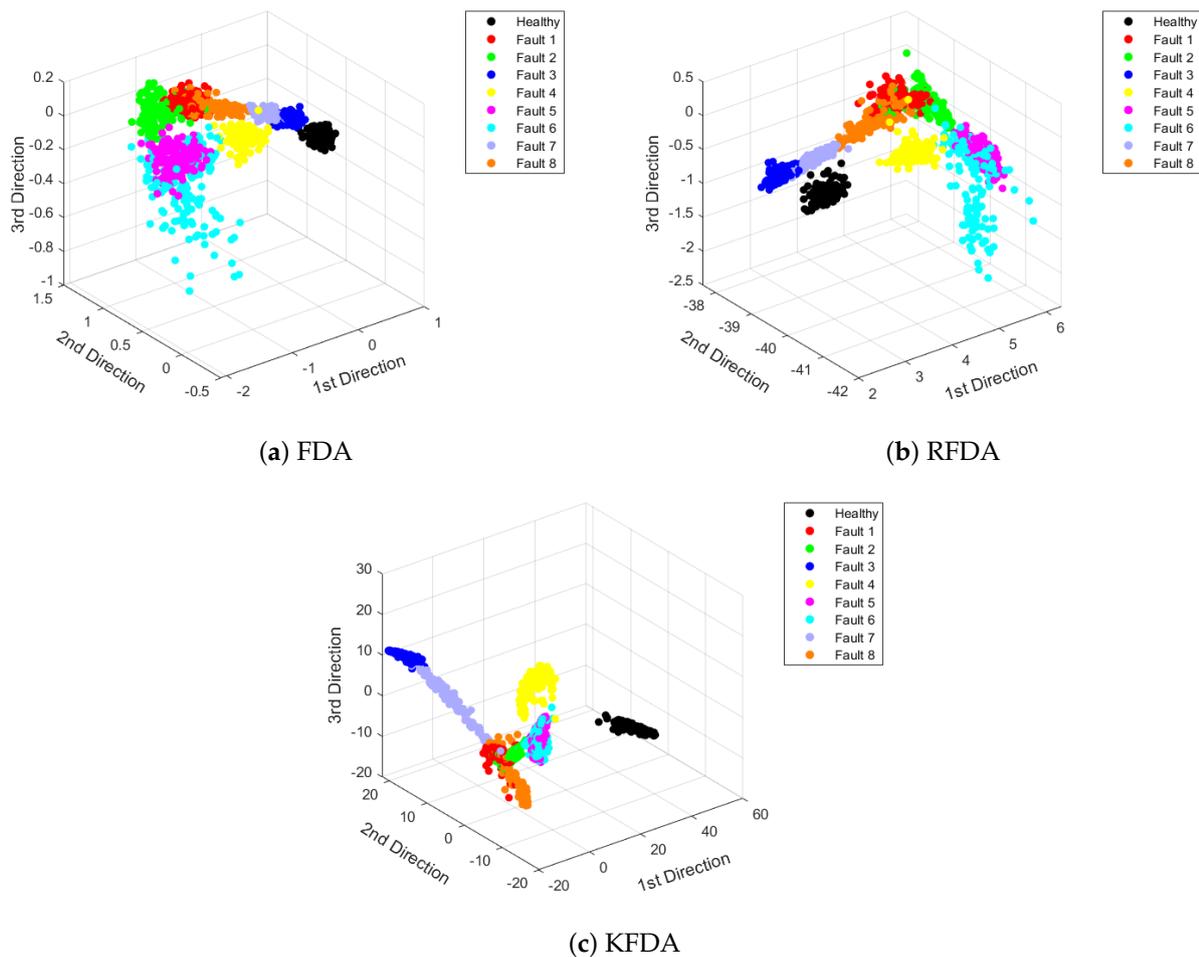


Figure 9. Visualization of dimensionality reduction: PU dataset.

Table 6 offers the accuracy rates and computation times of FDA, KFDA, and RFDA using the testing dataset. From the data in Table 6, the accuracy rates of FDA and KFDA are below 90%. RFDA can provide a slightly higher accuracy rate (i.e., 90.05%) than KFDA. However, KFDA and RFDA can derive better fault diagnosis performance than FDA. For computation time, FDA requires the lowest computation burden compared to KFDA and RFDA. As demonstrated in the case study of PU, the computation time of RFDA is close to that of the FDA. The average computation time of RFDA is 1.43 s. Due to the need for calculating the kernel matrix, the average computation time of KFDA (i.e., 3.24 s) is longer than that of FDA and RFDA.

Table 6. Computation times and accuracy rates using FDA, KFDA and RFDA: PU.

Method	Mean Acc	Mean Time (s)
FDA	86.72%	1.35
KFDA	89.72%	3.24
RFDA	90.05%	1.43

To investigate the performance of the proposed RFDA based fault diagnostic system further, the confusion matrices are plotted in Figure 10. From the data in Figure 10, values in the black squares on the diagonal indicate the classification accuracy, while values in the other gray-white squares indicate the error rate of the corresponding sample type. Most of the samples are correctly classified. However, it can be seen that the samples collected under Fault 6 are misclassified as Fault 5 for all methods. Among the comparable methods, fewer samples of Fault 6 are misclassified using RFDA. Overall, the RFDA-based method can offer satisfying results in the field of computation time and accuracy rate.

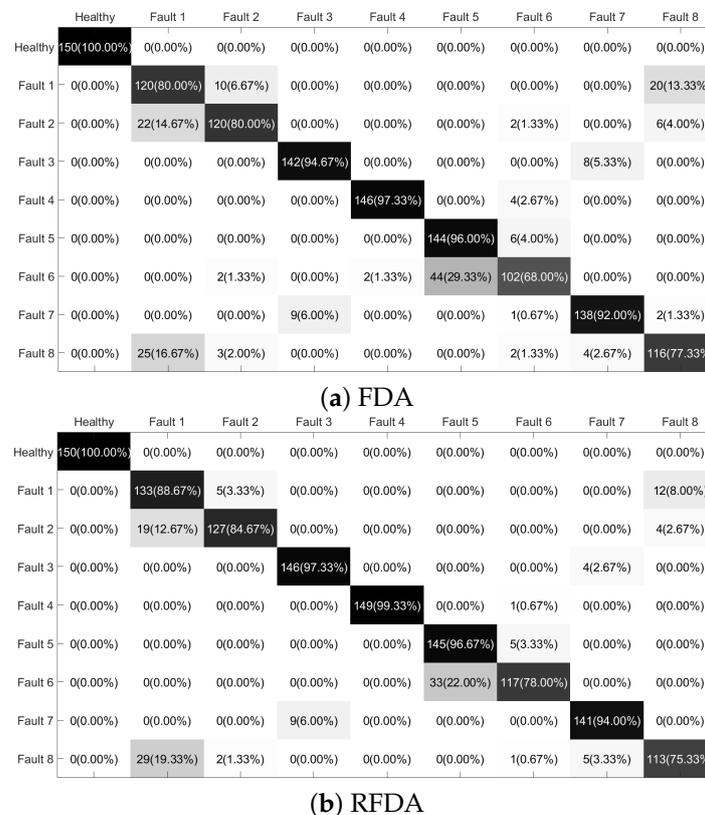


Figure 10. Cont.

	Healthy	Fault 1	Fault 2	Fault 3	Fault 4	Fault 5	Fault 6	Fault 7	Fault 8
Healthy	150(100.00%)	0(0.00%)	0(0.00%)	0(0.00%)	0(0.00%)	0(0.00%)	0(0.00%)	0(0.00%)	0(0.00%)
Fault 1	0(0.00%)	133(88.67%)	5(3.33%)	0(0.00%)	0(0.00%)	0(0.00%)	0(0.00%)	0(0.00%)	12(8.00%)
Fault 2	0(0.00%)	15(10.00%)	130(86.67%)	0(0.00%)	0(0.00%)	0(0.00%)	0(0.00%)	0(0.00%)	5(3.33%)
Fault 3	0(0.00%)	0(0.00%)	0(0.00%)	146(97.33%)	0(0.00%)	0(0.00%)	0(0.00%)	4(2.67%)	0(0.00%)
Fault 4	0(0.00%)	0(0.00%)	0(0.00%)	0(0.00%)	150(100.00%)	0(0.00%)	0(0.00%)	0(0.00%)	0(0.00%)
Fault 5	0(0.00%)	0(0.00%)	0(0.00%)	0(0.00%)	0(0.00%)	133(88.67%)	17(11.33%)	0(0.00%)	0(0.00%)
Fault 6	0(0.00%)	0(0.00%)	2(1.33%)	0(0.00%)	0(0.00%)	23(15.33%)	125(83.33%)	0(0.00%)	0(0.00%)
Fault 7	0(0.00%)	0(0.00%)	0(0.00%)	5(3.33%)	0(0.00%)	0(0.00%)	0(0.00%)	145(96.67%)	0(0.00%)
Fault 8	0(0.00%)	34(22.67%)	3(2.00%)	0(0.00%)	0(0.00%)	0(0.00%)	0(0.00%)	3(2.00%)	110(73.33%)

(c) KFDA

Figure 10. Confusion matrices: PU dataset.

5. Conclusions

This paper proposed a randomized Fisher discriminant analysis (RFDA) algorithm to diagnose the bearing fault. First, 12 time-domain feature sets were constructed by extracting the original vibration signals. Then the RFDA model was established for fault diagnosis by introducing a random Fourier feature. To classify the class of vibration data, Bayesian inference was applied. Experiments on the PU and CWRU datasets were conducted to assess the performance of the RFDA-based fault diagnosis, compared with FDA and KFDA methods. Results verified that the RFDA-based method can provide better fault diagnosis performance using less computation time.

Despite this, there are still some studies to continue in future work. First, the performance can be improved by the reduction of similar cases (i.e., defects for the same part) for the PU dataset. The exploration of manifold structure in the bearing dataset would be another alternative way to improve the RFDA methodology. Additionally, the diagnostic performance is influenced by the magnitudes of different defects. The level of defect should be taken into consideration. The incipient fault detection and diagnosis is another hot topic in the bearing fault diagnosis community. There are two types of choices of random Fourier features [33]. It would be interesting to compare and study the performance of different forms of random Fourier features in the fault diagnosis.

Author Contributions: Conceptualization, H.Y. and P.W.; validation, Y.H. (Yifei Huo), X.W., Y.H. (Yuchen He), X.Z. and J.G.; investigation, H.Y., P.W., Y.H. (Yifei Huo), X.W., Y.H. (Yuchen He), and X.Z.; data curation, H.Y., Y.H. (Yifei Huo), X.W., Y.H. (Yuchen He), X.Z. and J.G.; writing—original draft preparation, H.Y.; writing—review and editing, P.W. and J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Laboratory of Intelligent Manufacturing Quality Big Data Tracing and Analysis of Zhejiang Province, China Jiliang University (Grant No. ZNZZSZ-CJLU2022-03), the Social Development Project of Zhejiang Provincial Public Technology Research grant number LGF19F030004 and LGG21F030015, the Fundamental Research Funds of Zhejiang Sci-Tech University grant number 2021Q024 and the Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University, China grant number ICT2021B45, National Natural Science Foundation of China under Grant 62073296.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This work was supported in part by the Key Laboratory of Intelligent Manufacturing Quality Big Data Tracing and Analysis of Zhejiang Province, China Jiliang University (Grant No. ZNZZSZ-CJLU2022-03), the Social Development Project of Zhejiang Provincial Public Technology Research (Grant No. LGF19F030004, LGG21F030015), the Fundamental Research Funds of Zhejiang Sci-Tech University (No.2021Q024), and the Open Research Project of the State Key Lab-

oratory of Industrial Control Technology, Zhejiang University, China (No.ICT2021B45). National Natural Science Foundation of China under Grant (No.62073296)

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

References

1. Hoang, D.T.; Kang, H.J. A survey on deep learning based bearing fault diagnosis. *Neurocomputing* **2019**, *335*, 327–335. [[CrossRef](#)]
2. Zhang, S.; Zhang, S.; Wang, B.; Habetler, T.G. Deep learning algorithms for bearing fault diagnostics—A comprehensive review. *IEEE Access* **2020**, *8*, 29857–29881. [[CrossRef](#)]
3. Zhang, Y.; Gao, Q.; Lu, Y.; Sun, D.; Xia, Y.; Peng, X. A novel intelligent method for bearing fault diagnosis based on Hermitian scale-energy spectrum. *IEEE Sens. J.* **2018**, *18*, 6743–6755. [[CrossRef](#)]
4. Liu, L.; Chen, L.; Wang, Z.; Liu, D. Early fault detection of planetary gearbox based on acoustic emission and improved variational mode decomposition. *IEEE Sens. J.* **2020**, *21*, 1735–1745. [[CrossRef](#)]
5. Li, X.; Yang, Y.; Pan, H.; Cheng, J.; Cheng, J. A novel deep stacking least squares support vector machine for rolling bearing fault diagnosis. *Comput. Ind.* **2019**, *110*, 36–47. [[CrossRef](#)]
6. Naha, A.; Samanta, A.K.; Routray, A.; Deb, A.K. Low complexity motor current signature analysis using sub-Nyquist strategy with reduced data length. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 3249–3259. [[CrossRef](#)]
7. Nikula, R.P.; Karioja, K.; Pylvänäinen, M.; Leiviskä, K. Automation of low-speed bearing fault diagnosis based on autocorrelation of time domain features. *Mech. Syst. Sig. Process.* **2020**, *138*, 106572. [[CrossRef](#)]
8. Nayana, B.; Geethanjali, P. Analysis of statistical time-domain features effectiveness in identification of bearing faults from vibration signal. *IEEE Sens. J.* **2017**, *17*, 5618–5625. [[CrossRef](#)]
9. Lu, Y.; Wang, Z.; Zhu, D.; Gao, Q.; Sun, D. Bearing fault diagnosis based on clustering and sparse representation in frequency domain. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–14. [[CrossRef](#)]
10. Zhang, Y.; Xing, K.; Bai, R.; Sun, D.; Meng, Z. An enhanced convolutional neural network for bearing fault diagnosis based on time–frequency image. *Measurement* **2020**, *157*, 107667. [[CrossRef](#)]
11. De Moura, E.; Souto, C.; Silva, A.; Irmao, M. Evaluation of principal component analysis and neural network performance for bearing fault diagnosis from vibration signal processed by RS and DF analyses. *Mech. Syst. Sig. Process.* **2011**, *25*, 1765–1772. [[CrossRef](#)]
12. Yu, J. Local and nonlocal preserving projection for bearing defect classification and performance assessment. *IEEE Trans. Ind. Electron.* **2011**, *59*, 2363–2376. [[CrossRef](#)]
13. Ambroźkiewicz, B.; Syta, A.; Gassner, A.; Georgiadis, A.; Litak, G.; Meier, N. The influence of the radial internal clearance on the dynamic response of self-aligning ball bearings. *Mech. Syst. Sig. Process.* **2022**, *171*, 108954. [[CrossRef](#)]
14. Syta, A.; Czarnigowski, J.; Jakliński, P. Detection of cylinder misfire in an aircraft engine using linear and non-linear signal analysis. *Measurement* **2021**, *174*, 108982. [[CrossRef](#)]
15. Tu, D.; Zheng, J.; Jiang, Z.; Pan, H. Multiscale distribution entropy and t-distributed stochastic neighbor embedding-based fault diagnosis of rolling bearings. *Entropy* **2018**, *20*, 360. [[CrossRef](#)]
16. Joswiak, M.; Peng, Y.; Castillo, I.; Chiang, L.H. Dimensionality reduction for visualizing industrial chemical process data. *Control Eng. Pract.* **2019**, *93*, 104189. [[CrossRef](#)]
17. Mao, W.; Feng, W.; Liu, Y.; Zhang, D.; Liang, X. A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis. *Mech. Syst. Sig. Process.* **2021**, *150*, 107233. [[CrossRef](#)]
18. Zhang, S.; Ye, F.; Wang, B.; Habetler, T.G. Semi-Supervised Bearing Fault Diagnosis and Classification Using Variational Autoencoder-Based Deep Generative Models. *IEEE Sens. J.* **2021**, *21*, 6476–6486. [[CrossRef](#)]
19. Jin, X.; Zhao, M.; Chow, T.W.; Pecht, M. Motor bearing fault diagnosis using trace ratio linear discriminant analysis. *IEEE Trans. Ind. Electron.* **2013**, *61*, 2441–2451. [[CrossRef](#)]
20. Zhou, Y.; Yan, S.; Ren, Y.; Liu, S. Rolling bearing fault diagnosis using transient-extracting transform and linear discriminant analysis. *Measurement* **2021**, *178*, 109298. [[CrossRef](#)]
21. Mika, S.; Ratsch, G.; Weston, J.; Scholkopf, B.; Mullers, K.R. Fisher discriminant analysis with kernels. In Proceedings of the Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No. 98th8468), Madison, WI, USA, 25 August 1999; IEEE: Piscataway, NJ, USA, 1999; pp. 41–48.
22. Van, M.; Kang, H.J. Bearing defect classification based on individual wavelet local fisher discriminant analysis with particle swarm optimization. *IEEE Trans. Ind. Inf.* **2015**, *12*, 124–135. [[CrossRef](#)]
23. Jiang, L.; Xuan, J.; Shi, T. Feature extraction based on semi-supervised kernel Marginal Fisher analysis and its application in bearing fault diagnosis. *Mech. Syst. Sig. Process.* **2013**, *41*, 113–126. [[CrossRef](#)]
24. Tao, X.; Ren, C.; Li, Q.; Guo, W.; Liu, R.; He, Q.; Zou, J. Bearing defect diagnosis based on semi-supervised kernel Local Fisher Discriminant Analysis using pseudo labels. *ISA Trans.* **2021**, *110*, 394–412. [[CrossRef](#)]
25. Li, J.; Cui, P. Improved kernel fisher discriminant analysis for fault diagnosis. *Expert Syst. Appl.* **2009**, *36*, 1423–1432. [[CrossRef](#)]
26. Liu, Z.; Qu, J.; Zuo, M.J.; Xu, H.b. Fault level diagnosis for planetary gearboxes using hybrid kernel feature selection and kernel Fisher discriminant analysis. *Int. J. Adv. Manuf. Technol.* **2013**, *67*, 1217–1230. [[CrossRef](#)]

27. Rahimi, A.; Recht, B. Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems 20 (NIPS 2007)*; Citeseer: Princeton, NJ, USA, 2007; Volume 20, pp. 1177–1184.
28. Lopez-Paz, D.; Sra, S.; Smola, A.; Ghahramani, Z.; Schölkopf, B. Randomized nonlinear component analysis. In *Proceedings of the International Conference on Machine Learning, PMLR, Beijing, China, 22–24 June 2014*; pp. 1359–1367.
29. Jayaprakash, C.; Damodaran, B.B.; Viswanathan, S.; Soman, K.P. Randomized independent component analysis and linear discriminant analysis dimensionality reduction methods for hyperspectral image classification. *J. Appl. Remote Sens.* **2020**, *14*, 036507. [[CrossRef](#)]
30. Ye, H.; Li, Y.; Chen, C.; Zhang, Z. Fast Fisher discriminant analysis with randomized algorithms. *Pattern Recognit.* **2017**, *72*, 82–92. [[CrossRef](#)]
31. Fukunaga, K. *Introduction to Statistical Pattern Recognition*; Elsevier: Amsterdam, The Netherlands, 2013.
32. Funktionen, M. Stieltjessche Integrale und harmonische Analyse. *Math. Ann.* **1933**, *108*, 378–410.
33. Sutherland, D.J.; Schneider, J. On the error of random Fourier features. *arXiv* **2015**, arXiv:1506.02785.
34. Chiang, L.H.; Russell, E.L.; Braatz, R.D. Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemom. Intell. Lab. Syst.* **2000**, *50*, 243–252. [[CrossRef](#)]
35. Ge, Z.; Zhong, S.; Zhang, Y. Semisupervised Kernel Learning for FDA Model and its Application for Fault Classification in Industrial Processes. *IEEE Trans. Ind. Inf.* **2016**, *12*, 1403–1411. [[CrossRef](#)]
36. Zhang, J.F.; Huang, Z.C. Kernel Fisher discriminant analysis for bearing fault diagnosis. In *Proceedings of the 2005 International Conference on Machine Learning and Cybernetics, Guangzhou, China, 18–21 August 2005*; IEEE: Piscataway, NJ, USA, 2005; Volume 5, pp. 3216–3220.
37. Lessmeier, C.; Kimotho, J.K.; Zimmer, D.; Sextro, W. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. In *Proceedings of the European Conference of the Prognostics and Health Management Society, Bilbao, Spain, 5–8 July 2016*; Citeseer: Princeton, NJ, USA, 2016; pp. 5–8.