

## Research Article

# Trace Ratio Criterion-Based Kernel Discriminant Analysis for Fault Diagnosis of Rolling Element Bearings Using Binary Immune Genetic Algorithm

Wen-An Yang,<sup>1</sup> Maohua Xiao,<sup>2</sup> Wei Zhou,<sup>3</sup> Yu Guo,<sup>1</sup> Wenhe Liao,<sup>1</sup> and Gang Shen<sup>4</sup>

<sup>1</sup>College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

<sup>2</sup>College of Engineering, Nanjing Agricultural University, Nanjing 210031, China

<sup>3</sup>Nanjing Surveying and Mapping Instrument Factory, Nanjing 210003, China

<sup>4</sup>School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou 221116, China

Correspondence should be addressed to Wen-An Yang; [dreamflow@nuaa.edu.cn](mailto:dreamflow@nuaa.edu.cn)

Received 3 July 2015; Revised 19 October 2015; Accepted 27 October 2015

Academic Editor: Dong Wang

Copyright © 2016 Wen-An Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The rolling element bearing is a core component of many systems such as aircraft, train, steamboat, and machine tool, and their failure can lead to reduced capability, downtime, and even catastrophic breakdowns. Due to misoperation, manufacturing deficiencies, or the lack of monitoring and maintenance, it is often found to be the most unreliable component within these systems. Therefore, effective and efficient fault diagnosis of rolling element bearings has an important role in ensuring the continued safe and reliable operation of their host systems. This study presents a trace ratio criterion-based kernel discriminant analysis (TR-KDA) for fault diagnosis of rolling element bearings. The binary immune genetic algorithm (BIGA) is employed to solve the trace ratio problem in TR-KDA. The numerical results obtained using extensive simulation indicate that the proposed TR-KDA using BIGA (called TR-KDA-BIGA) can effectively and efficiently classify different classes of rolling element bearing data, while also providing the capability of real-time visualization that is very useful for the practitioners to monitor the health status of rolling element bearings. Empirical comparisons show that the proposed TR-KDA-BIGA performs better than existing methods in classifying different classes of rolling element bearing data. The proposed TR-KDA-BIGA may be a promising tool for fault diagnosis of rolling element bearings.

## 1. Introduction

The rolling element bearing is a core component of many systems such as aircraft, train, steamboat, and machine tool, and their failure can lead to reduced capability, downtime, and even catastrophic breakdowns [1–6]. Due to misoperation, manufacturing deficiencies, or the lack of monitoring and maintenance, it is often found to be the most unreliable component within these systems. Therefore, effective and efficient fault diagnosis of rolling element bearings has an important role in ensuring the continued safe and reliable operation of their host systems.

Over the past few years, much research effort has been devoted to developing approaches to fault diagnosis of rolling element bearings. When faults occur in rolling element

bearings, vibration signals in the relevant time/frequency-domain have been demonstrated to deviate from their normal ones because of the increased friction and impulsive forces [7–10]. Usually, several dozens or even hundreds of time/frequency-domain features are calculated from the bearing vibration signals to represent the different health status. In the current study, 9 time-domain features and 6 time-frequency-domain features are extracted from the bearing vibration signals to jointly construct a 15-dimension feature vector. In that way, fault diagnosis of rolling element bearings is usually solved as a high-dimensional pattern recognition problem. However, for high-dimensional data, the intrinsic dimension may be small. For example, the number of features responsible for a certain type of fault pattern may be small. Moreover, projection of high-dimensional data onto 2- or

3-dimension subspace can provide real-time visualization, which is convenient for the user to monitor the health status of rolling element bearings. In addition, projection of high-dimensional data onto low dimension subspace also plays a part of data compression, which is helpful for efficient storage and retrieval. Thus, dimensionality reduction techniques are often used to project the high-dimensional feature space to a lower-dimensional space while preserving most of “intrinsic information” contained in the data properties [11–15]. Upon performing dimensionality reduction on the data, its compact representation can be utilized for succeeding tasks (e.g., visualization and classification). Among various dimensionality reduction methods [16–24], principal component analysis (PCA) and linear discriminant analysis (LDA) are the two most common methods [21]. The former is an unsupervised method, which pursues the direction of maximum variance for optimal reconstruction. The latter is a supervised method, which aims to maximize the between-class scatter while minimizing the within-class scatter. Owing to the utilization of labeled information, the latter generally outperforms the former if sufficient labeled samples are provided [21]. In the past few years, a series of studies have been conducted to formulate the LDAs for pattern recognition by Fukunaga [21], Wang et al. [22], Sun and Chen [25], Guo et al. [26], Zhao et al. [27], Jin et al. [28], Jia et al. [29], and so on. Generally, the formulation of LDAs is based on the ratio trace criterion but not trace ratio criterion, because the ratio trace problem is more tractable than the trace ratio problem. Nevertheless, as pointed out by Wang et al. [22], solutions obtained based on ratio trace criterion may deviate from the original intent of the trace ratio problems. To improve the behaviour of LDA implementation, Wang et al. [22], Guo et al. [26], Zhao et al. [27], Jin et al. [28], and Jia et al. [29] presented various trace ratio criterion-based LDAs (TR-LDAs), in which the numerator and denominator of the criterion directly reflect Euclidean distances between of inter- and intraclass samples. Another advantage of trace ratio criterion is that the calculated projection matrix is orthogonal, which can eliminate the redundancy between different projection directions. In addition, the orthogonal projection can thus preserve such similarities without any change when using Euclidean distance to evaluate the similarity between data points [22]. Although the above TR-LDA formulation methods have the aforementioned advantages, they are criticized due to their incapability of dealing with the redundancy among eigenvectors. For example, if the most discriminative eigenvector is duplicated several times, the above TR-LDA formulation methods are prone to selecting all of them. This is problematic for selection of an optimal subset of eigenvectors because other discriminative and complementary eigenvectors will be missed. A classifier with the eigenvectors selected in this way can give rise to poor classification performance. Therefore, the issue of TR-LDA formulation has remained unresolved.

A review of the related literature also indicates that most of the previous work in the area of applying LDA or TR-LDA to fault diagnosis assumed that samples in each class follow a linear distribution. However, in many

fault diagnosis practices, samples in each class that may follow a nonlinear distribution cannot satisfy the assumption. Without this assumption, the separation of different classes may not be well characterized by the scatter matrices, causing the classification results to be degraded [21]. To solve this problem, kernel trick [30–32], which is to extend many linear methods to its nonlinear kernel version, can be used to extend TR-LDA to handle nonlinear problem. Thus, this study develops a nonlinear kernel version of TR-LDA, that is, trace ratio criterion-based kernel discriminant analysis (TR-KDA), for fault diagnosis of rolling element bearings. However, like many other TR-LDA models, the TR-KDA model presented in this study shares the trace ratio problem in the formulation of projection matrix. Although the above TR-LDA formulation methods have the aforementioned advantages, they are criticized due to the inability to handle redundancy in eigenvector selection. For example, if the most discriminative eigenvector is duplicated several times, the above TR-LDA formulation methods are prone to selecting all of them. This is problematic for selecting the best set of eigenvectors because other discriminative and complementary eigenvectors will be missed. A classifier with the eigenvectors selected in such a way can lead to a poor classification performance. Fortunately, immune genetic algorithm (IGA), a novel evolutionary computation technique developed by Jiao and Wang [33], has the potential to determine a set of discriminative and mutually irredundant eigenvectors. In this study, we propose a method called TR-KDA-BIGA that uses binary IGA (BIGA) to formulate TR-KDA for dimensionality reduction of statistical and wavelet features extracted from the vibration signals and gives rise to effective and efficient fault diagnosis of rolling element bearings. In particular the contributions are to

- (i) use immune evolutionary computation technique such as BIGA to obtain a reduced set of discriminative and mutually irredundant eigenvectors for TR-KDA-BIGA formulation,
- (ii) provide the capability of two-dimensional representation of bearing data that is very useful for the practitioners to monitor the health status of bearings,
- (iii) build a TR-KDA-BIGA model architecture for the vibration measurements for effective and efficient fault diagnosis of rolling element bearings.

The rest of this study is structured as follows. Section 2 briefly reviews the basic concepts of TR-LDA and kernel extension. Section 3 presents a TR-KDA-BIGA method. Section 4 discusses its convergence and initialization. Section 5 conducts performance evaluations of the proposed TR-KDA-BIGA on benchmark problems. Section 6 describes an overall flowchart of the proposed TR-KDA-BIGA for fault diagnosis of rolling element bearings. Section 7 summarizes the conclusions drawn from this study.

## 2. Review of TR-LDA and Kernel Extension

*2.1. Review of TR-LDA.* Suppose we are given a set of  $n$   $d$ -dimensional samples  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ , belonging to  $l$  different classes. The goal of LDA tries to obtain a linear projection matrix  $\mathbf{W} \in \mathbf{R}^{d \times k}$  that can map the original  $d$ -dimensional data  $\mathbf{x}_i$  onto the  $k$ -dimensional data  $\mathbf{y}_i$  (usually  $k \ll d$ ) by maximizing the between-class scatter and meanwhile minimizing the within-class scatter. The between-class scatter matrix  $\mathbf{S}_B$  and the within-class scatter matrix  $\mathbf{S}_W$  are expressed as follows:

$$\begin{aligned} \mathbf{S}_B &= \sum_{i=1}^l n_i (\mathbf{m}^{(i)} - \mathbf{m})(\mathbf{m}^{(i)} - \mathbf{m})^T, \\ \mathbf{S}_W &= \sum_{i=1}^l \left( \sum_{j=1}^{n_i} (\mathbf{x}_j^{(i)} - \mathbf{m}^{(i)})(\mathbf{x}_j^{(i)} - \mathbf{m}^{(i)})^T \right), \end{aligned} \quad (1)$$

where  $\mathbf{m}$  represents the total sample mean vector,  $n_i$  represents the number of samples in the  $i$ th class,  $\mathbf{m}^{(i)}$  represents the average vector of the  $i$ th class, and  $\mathbf{x}_j^{(i)}$  represents the  $j$ th sample in the  $i$ th class. The new mapped feature vectors  $\mathbf{y}_i \in \mathbf{R}^k$  can then be expressed as  $\mathbf{y}_i = \mathbf{W}^T \mathbf{x}_i$ . The original LDA formulation, known as the Fisher LDA [21], only handles binary classification problems. However, many practical applications involve multiclass classification. In order to overcome this issue, a number of researchers have proposed optimization criteria for extending the Fisher LDA to handle multiclass classification problems. The first optimization criterion is in a ratio trace form (referred to as RT-LDA):

$$\mathbf{W}^* = \arg \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \text{Tr} \left( \frac{\mathbf{W}^T \mathbf{S}_B \mathbf{W}}{\mathbf{W}^T \mathbf{S}_W \mathbf{W}} \right), \quad (2)$$

where  $\text{Tr}[\cdot]$  denotes the matrix trace;  $\mathbf{I}$  is an identity matrix. In order to achieve a set of orthogonal normalized vectors, it usually adds the constraint  $\mathbf{W}^T \mathbf{W} = \mathbf{I}$  to (2). The second optimization criterion is in a trace ratio form (referred to as TR-LDA):

$$\mathbf{W}^* = \arg \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \frac{\text{Tr}(\mathbf{W}^T \mathbf{S}_B \mathbf{W})}{\text{Tr}(\mathbf{W}^T \mathbf{S}_W \mathbf{W})}. \quad (3)$$

The optimization problem in (3) can be solved directly through the generalized eigenvalue decomposition (GED) method [22]:

$$\mathbf{S}_B \mathbf{w}_i = \tau_i \mathbf{S}_W \mathbf{w}_i, \quad (4)$$

where  $\tau_i$  is the  $i$ th largest eigenvalue,  $\mathbf{w}_i$  is the eigenvector corresponding to  $\tau_i$ , and  $\mathbf{w}_i$  constitutes the  $i$ th column vector of the matrix  $\mathbf{W}$ . Although a closed-form solution for (3) can be approximately obtained with the GED, it does not necessarily guarantee best trace ratio optimization. Thus, this approximation of ratio trace optimization to trace ratio optimization may lead to classification capability loss of the derived optimal low-dimensional feature space. Moreover, the physical meaning of the trace ratio form is clearer than that of

the ratio trace form. However, the optimization problem in (3) is generally nonconvex and a closed-form solution for it does not exist. Fortunately, a recent study conducted by Guo et al. [26] showed that, using the trace difference function  $z(\lambda) = \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \text{Tr}[\mathbf{W}^T (\mathbf{S}_B - \lambda \mathbf{S}_W) \mathbf{W}]$ , the trace ratio problem can be solved equivalently by finding zero points of the equation  $z(\lambda) = 0$ . Following up Guo et al.'s work, Wang et al. presented an iterative method named ITR algorithm to solve the trace ratio problem [22]. The ITR algorithm optimizes the objective function  $\lambda = \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \text{Tr}[\mathbf{W}^T \mathbf{S}_B \mathbf{W}] / \text{Tr}[\mathbf{W}^T \mathbf{S}_W \mathbf{W}]$  in an iterative and incremental manner. The  $\mathbf{W}$  in the  $t$ th iteration step (referred to as  $\mathbf{W}_t$ ) is obtained through solving the trace difference problem  $\arg \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \text{Tr}[\mathbf{W}^T (\mathbf{S}_B - \lambda_t \mathbf{S}_W) \mathbf{W}]$ , where  $\lambda_t$  represents the trace ratio value derived from the  $\mathbf{W}$  in the previous iteration step (referred to as  $\mathbf{W}_{t-1}$ ). However, the initialization for the  $\mathbf{W}$  influences substantially the convergence performance of the ITR algorithm. A good initialization can generally make the ITR algorithm yield a quick convergence. A bad initialization usually increases the number of iterations. Moreover, in ITR algorithm, although it seems that the  $\mathbf{W}$  formed with these eigenvectors corresponding to the  $k$  largest eigenvalues of  $\mathbf{S}_B - \lambda_t \mathbf{S}_W$  can maximize the trace difference  $\text{Tr}[\mathbf{W}^T (\mathbf{S}_B - \lambda_t \mathbf{S}_W) \mathbf{W}]$ , it cannot necessarily maximize the trace ratio  $\text{Tr}[\mathbf{W}^T \mathbf{S}_B \mathbf{W}] / \text{Tr}[\mathbf{W}^T \mathbf{S}_W \mathbf{W}]$ . On the other hand, from the perspective of fault diagnosis, the aim is mainly to find a set of projection vectors that can pose the highest levels of discrimination in the different fault patterns. Thus, these eigenvectors with the largest eigenvalues are not necessarily representative for discriminating one class from others as previously mentioned in Section 1. To overcome the above shortcomings, this study presents a BIGA-based solution method for trace ratio criterion (to be in detail discussed in Section 3).

*2.2. Kernel Extension.* In some applications, it is insufficient to model the data using the TR-LDA, which is a linear discriminating method. To address the issue of nonlinearities in the data, this section presents a nonlinear discriminating method using kernel trick [30–32], that is, TR-KDA. The so-called kernel trick is to map the original data to a high-dimensional Hilbert space through a nonlinear mapping function  $\phi$ . Let  $\phi(\mathbf{X})$  denote the data matrix in the Hilbert space:  $\phi(\mathbf{X}) = [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_n)]$ . The function form of the mapping does not need to be known since it is implicitly defined by the choice of kernel function  $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ , that is, the inner product in the kernel-induced feature space. The kernel function  $K$  may be any positive kernel satisfying Mercer's condition. Radial basis function (RBF) kernel function, one of the most popular kernel functions employed in various kernelled learning algorithms, is adopted in this study. Then, (3) in Hilbert space can be written as follows:

$$\mathbf{W}^{\phi*} = \arg \max_{(\mathbf{W}^{\phi})^T \mathbf{W}^{\phi} = \mathbf{I}} \frac{\text{Tr}((\mathbf{W}^{\phi})^T \mathbf{S}_B^{\phi} \mathbf{W}^{\phi})}{\text{Tr}((\mathbf{W}^{\phi})^T \mathbf{S}_W^{\phi} \mathbf{W}^{\phi})}, \quad (5)$$

where  $\mathbf{W}^{\phi}$ ,  $\mathbf{S}_B^{\phi}$ , and  $\mathbf{S}_W^{\phi}$  are the matrices in Hilbert space corresponding to  $\mathbf{W}$ ,  $\mathbf{S}_B$ , and  $\mathbf{S}_W$  in (3), respectively. Notably,

we can show that matrices  $\mathbf{S}_B$  and  $\mathbf{S}_W$  in (3) can be essentially expressed as  $\mathbf{S}_B = \mathbf{X}\mathbf{L}_B\mathbf{X}^T$  and  $\mathbf{S}_W = \mathbf{X}\mathbf{L}_W\mathbf{X}^T$  through simple manipulation, respectively.  $\mathbf{X}$  is the vector, where  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ . Matrices  $\mathbf{L}_B$  and  $\mathbf{L}_W$  are the graph Laplacian matrices [34] of the weighted undirected graphs reflecting the between-class and within-class relationship of the samples. Consider

$$\begin{aligned}
\mathbf{S}_B &= \sum_{i=1}^l n_i (\mathbf{m}^{(i)} - \mathbf{m}) (\mathbf{m}^{(i)} - \mathbf{m})^T \\
&= \left( \sum_{i=1}^l n_i \mathbf{m}^{(i)} (\mathbf{m}^{(i)})^T \right) - \mathbf{m} \left( \sum_{i=1}^l n_i (\mathbf{m}^{(i)})^T \right) \\
&\quad - \left( \sum_{i=1}^l n_i \mathbf{m}^{(i)} \right) \mathbf{m}^T + \left( \sum_{i=1}^l n_i \right) \mathbf{m} \mathbf{m}^T \\
&= \left( \sum_{i=1}^l \frac{1}{n_i} (\mathbf{x}_1^{(i)} + \mathbf{x}_2^{(i)} + \dots + \mathbf{x}_{n_i}^{(i)}) \right. \\
&\quad \cdot \left. (\mathbf{x}_1^{(i)} + \mathbf{x}_2^{(i)} + \dots + \mathbf{x}_{n_i}^{(i)})^T \right) - 2n\mathbf{m}\mathbf{m}^T + n\mathbf{m}\mathbf{m}^T \\
&= \left( \sum_{i=1}^l \sum_{j,q=1}^{n_i} \frac{1}{n_i} \mathbf{x}_j^{(i)} (\mathbf{x}_q^{(i)})^T \right) - n\mathbf{m}\mathbf{m}^T = \mathbf{X}\mathbf{G}\mathbf{X}^T \\
&\quad - n\mathbf{m}\mathbf{m}^T = \mathbf{X}\mathbf{G}\mathbf{X}^T - \mathbf{X} \left( \frac{1}{n} \mathbf{e}\mathbf{e}^T \right) \mathbf{X}^T = \mathbf{X} \left( \mathbf{G} - \frac{1}{n} \right. \\
&\quad \cdot \left. \mathbf{e}\mathbf{e}^T \right) \mathbf{X}^T,
\end{aligned} \tag{6}$$

where  $\mathbf{e} = [1, 1, \dots, 1]^T$  is an  $n$ -dimensional vector. We can simplify the above equation even further by defining that

$$(\mathbf{G})_{ij} = \begin{cases} \frac{1}{n_q} & \text{if samples } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ belong to the } q\text{th class} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

$$\mathbf{L}_B = \mathbf{G} - \frac{1}{n} \mathbf{e}\mathbf{e}^T. \tag{8}$$

Thus, we get

$$\mathbf{S}_B = \mathbf{X}\mathbf{L}_B\mathbf{X}^T. \tag{9}$$

Then, the matrix  $\mathbf{S}_W$  can similarly be computed as follows:

$$\begin{aligned}
\mathbf{S}_W &= \sum_{i=1}^l \left( \sum_{j=1}^{n_i} (\mathbf{x}_j^{(i)} - \mathbf{m}^{(i)}) (\mathbf{x}_j^{(i)} - \mathbf{m}^{(i)})^T \right) \\
&= \sum_{i=1}^l \left( \sum_{j=1}^{n_i} \mathbf{x}_j^{(i)} (\mathbf{x}_j^{(i)})^T - \mathbf{m}^{(i)} (\mathbf{x}_j^{(i)})^T - \mathbf{x}_j^{(i)} (\mathbf{m}^{(i)})^T \right. \\
&\quad \left. + \mathbf{m}^{(i)} (\mathbf{m}^{(i)})^T \right) = \sum_{i=1}^l \left( \sum_{j=1}^{n_i} \mathbf{x}_j^{(i)} (\mathbf{x}_j^{(i)})^T \right.
\end{aligned}$$

$$\begin{aligned}
&\quad \left. - n_i \mathbf{m}^{(i)} (\mathbf{m}^{(i)})^T \right) = \sum_{i=1}^l \left( \mathbf{X}_i \mathbf{X}_i^T - \frac{1}{n_i} (\mathbf{x}_1^{(i)} + \mathbf{x}_2^{(i)} \right. \\
&\quad \left. + \dots + \mathbf{x}_{n_i}^{(i)}) (\mathbf{x}_1^{(i)} + \mathbf{x}_2^{(i)} + \dots + \mathbf{x}_{n_i}^{(i)})^T \right) = \sum_{i=1}^l \left( \mathbf{X}_i \mathbf{X}_i^T \right. \\
&\quad \left. - \frac{1}{n_i} \mathbf{X}_i (\mathbf{e}_i \mathbf{e}_i^T) \mathbf{X}_i^T \right) = \sum_{i=1}^l \left( \mathbf{X}_i \left( \mathbf{I} - \frac{1}{n_i} \mathbf{e}_i \mathbf{e}_i^T \right) \mathbf{X}_i^T \right) \\
&= \sum_{i=1}^l \mathbf{X}_i \mathbf{L}_i \mathbf{X}_i^T,
\end{aligned} \tag{10}$$

where  $\mathbf{X}_i = [\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_{n_i}^{(i)}]$  is a  $d \times n_i$  matrix,  $\mathbf{e}_i = [1, 1, \dots, 1]^T$  is an  $n_i$ -dimensional vector,  $\mathbf{I}$  is the identity matrix,  $\mathbf{L}_i = \mathbf{I} - 1/n_i \mathbf{e}_i \mathbf{e}_i^T$  is an  $n_i \times n_i$  matrix, and  $\mathbf{X}_i \mathbf{L}_i \mathbf{X}_i^T$  is the data covariance matrix of the  $i$ th class. Based on (7), the above equation can be simplified similarly by defining

$$\mathbf{L}_W = \mathbf{I} - \mathbf{G}. \tag{11}$$

Thus, we get

$$\mathbf{S}_W = \mathbf{X}\mathbf{L}_W\mathbf{X}^T. \tag{12}$$

Using the definitions in (9) and (12), (5) can be rewritten as follows:

$$\begin{aligned}
\mathbf{W}^{\phi^*} &= \arg \max_{(\mathbf{W}^{\phi})^T \mathbf{W}^{\phi} = \mathbf{I}} \frac{\text{Tr} \left( (\mathbf{W}^{\phi})^T (\phi(\mathbf{X}) \mathbf{L}_B (\phi(\mathbf{X}))^T) \mathbf{W}^{\phi} \right)}{\text{Tr} \left( (\mathbf{W}^{\phi})^T (\phi(\mathbf{X}) \mathbf{L}_W (\phi(\mathbf{X}))^T) \mathbf{W}^{\phi} \right)}. \tag{13}
\end{aligned}$$

In order to pursue the matrix  $\mathbf{W}^{\phi^*}$ , solving the above equation involves decomposition of  $\phi(\mathbf{X})$  into an orthogonal matrix  $\mathbf{Q}$  (satisfying  $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ ) and a right triangular matrix  $\mathbf{R}$  such that  $\mathbf{Q} * \mathbf{R} = \phi(\mathbf{X})$ . We have

$$(\phi(\mathbf{X}))^T \phi(\mathbf{X}) = \mathbf{R}^T \mathbf{R}. \tag{14}$$

Let us map  $\mathbf{W}^{\phi}$  into the span of  $\mathbf{Q}$ .  $\mathbf{Q}$  is currently an orthogonal basis of  $\phi(\mathbf{X})$ , so we have

$$\mathbf{W}^{\phi} = \mathbf{Q}\mathbf{V}^{\phi}, \tag{15}$$

where  $\mathbf{V}^{\phi} = \mathbf{R}^{k \times n}$  is an orthogonal matrix satisfying  $(\mathbf{V}^{\phi})^T \mathbf{V}^{\phi} = \mathbf{I}$ . Using the definitions in (14) and (15), (13) can be further rewritten as follows:

$$\mathbf{V}^{\phi^*} = \arg \max_{(\mathbf{V}^{\phi})^T \mathbf{V}^{\phi} = \mathbf{I}} \frac{\text{Tr} \left( (\mathbf{V}^{\phi})^T \mathbf{R} \mathbf{L}_B \mathbf{R}^T \mathbf{V}^{\phi} \right)}{\text{Tr} \left( (\mathbf{V}^{\phi})^T \mathbf{R} \mathbf{L}_W \mathbf{R}^T \mathbf{V}^{\phi} \right)}. \tag{16}$$

Let  $\tilde{\mathbf{S}}_B = \mathbf{R} \mathbf{L}_B \mathbf{R}^T$  and  $\tilde{\mathbf{S}}_W = \mathbf{R} \mathbf{L}_W \mathbf{R}^T$ ; then (16) can be further rewritten as follows:

$$\mathbf{V}^{\phi^*} = \arg \max_{(\mathbf{V}^{\phi})^T \mathbf{V}^{\phi} = \mathbf{I}} \frac{\text{Tr} \left( (\mathbf{V}^{\phi})^T \tilde{\mathbf{S}}_B \mathbf{V}^{\phi} \right)}{\text{Tr} \left( (\mathbf{V}^{\phi})^T \tilde{\mathbf{S}}_W \mathbf{V}^{\phi} \right)}. \tag{17}$$

After the matrix  $\mathbf{V}^{\phi*}$  is obtained with the BIGA-based solution method (to be in detail discussed in Section 3), the output points in the reduced data space can thus be expressed as

$$\mathbf{Y}^{\phi*} = (\mathbf{W}^{\phi*})^T \phi(\mathbf{X}) = (\mathbf{V}^{\phi*})^T \mathbf{Q}^T \mathbf{Q} \mathbf{R} = (\mathbf{V}^{\phi*})^T \mathbf{R}. \quad (18)$$

### 3. The Proposed TR-KDA-BIGA

As previously mentioned, construction of TR-KDA needs to select  $k$  out of  $d$  eigenvectors to form the matrix  $\mathbf{V}^{\phi}$  for dimensionality reduction. However, finding a subset of eigenvectors based on the trace ratio criterion is not an easy task since the space of possible subsets is very large especially when  $d$  is a large number. Thus, it is not impractical to use exhaustive search to find an optimal subset of  $k$  eigenvectors. Instead, in this study, the BIGA is utilized to select  $k$  out of  $d$  eigenvectors of  $\mathbf{V}_t^{\phi}$  as the bases for projection matrix formulation based on the trace ratio criterion such that the trace ratio value  $\lambda = \text{Tr}[(\mathbf{V}^{\phi})^T \tilde{\mathbf{S}}_B \mathbf{V}^{\phi}] / \text{Tr}[(\mathbf{V}^{\phi})^T \tilde{\mathbf{S}}_W \mathbf{V}^{\phi}]$  can be maximized. Immune genetic algorithm, originally developed by Jiao and Wang [33], is a novel genetic algorithm based on the biological immune theory, which combined the immune mechanism with the evolutionary mechanism. In what follows, further discussion of the proposed TR-KDA-BIGA is carried out.

**3.1. Chromosome Encoding.** Encoding a solution of a problem into a chromosome is an important issue when using BIGAs. In this study, every chromosome in a BIGA corresponds to a discrete binary selector  $\mathbf{u} = [u_1, u_2, \dots, u_d]$ , where each gene in the *chromosome* is “1,” indicating an eigenvector  $\mathbf{v}_i^{\phi}$  ( $i = 1, 2, \dots, d$ ) of  $\tilde{\mathbf{S}}_B - \lambda_t \tilde{\mathbf{S}}_W$  appearing in forming the projection matrix  $\mathbf{V}^{\phi}$  of the  $t$ th step, while “0” denotes its absence. Thus, the length of the chromosome is  $d$ .

**3.2. Genetic Operators.** Genetic operators give every chromosome the chance to become the fittest chromosome of its generation. If it is difficult to reach the target of trace ratio optimization, crossover and mutation may introduce degeneracy into generations of chromosomes.

**3.2.1. Crossover Operator.** Crossover operator in a BIGA is employed to generate two new children chromosomes based on two existing parent chromosomes selected from the current population in terms of a prespecified crossover rate. In this study, “one-point” crossover operator was adopted to randomly select a cut point to exchange the parts between the cut point and the end of the string of the parent chromosomes. Specifically, suppose that two parent chromosomes  $P_1$  and  $P_2$  selected randomly from the population are undergoing the crossover operation at a randomly selected crossover point  $g$  ( $1 \leq g \leq d$ ), where

$$\begin{aligned} P_1 &= (u_{i,1}, u_{i,2}, \dots, u_{i,d}), \\ P_2 &= (u_{j,1}, u_{j,2}, \dots, u_{j,d}). \end{aligned} \quad (19)$$

Consequently, the offspring is generated by one-point crossover on the genes of two parents selected randomly from the population. We can thus get the two offspring chromosomes  $C_1$  and  $C_2$ :

$$\begin{aligned} C_1 &= (u_{i,1}, u_{i,2}, \dots, u_{i,g-1}, u_{i,g}, u_{j,g+1}, \dots, u_{j,d}), \\ C_2 &= (u_{j,1}, u_{j,2}, \dots, u_{j,g-1}, u_{j,g}, u_{i,g+1}, \dots, u_{i,d}). \end{aligned} \quad (20)$$

However, the exchange procedure is not simply exchanging their genetic information between gene segments after the crossover points. We must keep the number of eigenvectors to be included in the subset equal to  $k$ . In this study, therefore, a simple but effective crossover operator strategy in this study is performed in order to ensure that the crossover operator does not change the total number of “1” genes in chromosomes.

Let

$$\begin{aligned} n_{P_1} &= \sum_{q=g+1}^d u_{i,q}, \\ n_{P_2} &= \sum_{q=g+1}^d u_{j,q}. \end{aligned} \quad (21)$$

When  $n_{P_1}$  is not equal to  $n_{P_2}$ , the following retention criterion will be conducted:

- (1) If  $n_{P_1}$  is larger than  $n_{P_2}$ , randomly select  $(n_{P_1} - n_{P_2})$  genes with “0-bit” from the current offspring chromosome  $C_1$  and reset these  $(n_{P_1} - n_{P_2})$  selected genes to “1-bit,” and then randomly select  $(n_{P_1} - n_{P_2})$  genes with “1-bit” from the current offspring chromosome  $C_2$  and reset these  $(n_{P_1} - n_{P_2})$  selected genes to “0-bit.”
- (2) If  $n_{P_1}$  is smaller than  $n_{P_2}$ , randomly select  $(n_{P_2} - n_{P_1})$  genes with “1-bit” from the current offspring chromosome  $C_1$  and reset these  $(n_{P_2} - n_{P_1})$  selected genes to “0-bit,” and then randomly select  $(n_{P_2} - n_{P_1})$  genes with “0-bit” from the current offspring chromosome  $C_2$  and reset these  $(n_{P_2} - n_{P_1})$  selected genes to “1-bit.”

**3.2.2. Mutation Operator.** Mutation operator in a BIGA is used primarily as a mechanism for maintaining diversity in the population. For each gene in a chromosome that is undergoing the mutation, a real-valued number is randomly selected within the range of  $[0, 1]$ . If the real-valued number is less than the prespecified mutation rate, then the gene will change from “0-bit” to “1-bit” and vice versa. Upon adding (or removing) one eigenvector in that way, we shall randomly remove (or add) a different one such that the number of eigenvectors to be included in the subset is equal to  $k$ . The mutation operator helps the chromosomes to guide the search in new areas.

**3.3. Immune Operators.** The immune ability of BIGAs is realized through two kinds of immune operators: a vaccination and an immune selection. The vaccination is responsible for improving individuals’ overall fitness levels. The immune selection is responsible for prevention of deterioration.

**3.3.1. Vaccination Operator.** Given a chromosome  $\mathbf{u}$ , vaccination operation in a BIGA is employed to modify the genes on some bits according to a priori knowledge such that individuals with higher fitness have a greater probability of being selected. Let  $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{n_0})$  be a population; the vaccination operation on  $\mathbf{U}$  means that the operation is performed on  $n_\alpha = \alpha n_0$  chromosomes selected from  $\mathbf{U}$  according to the proportion of  $\alpha$ , where  $n_0$  represents the population size of a BIGA. A vaccine is abstracted from the prior knowledge of the pending problem, whose information amount and validity play an important role in the performance of the algorithm.

**3.3.2. Immune Selection Operator.** The immune selection operation consists of the following two steps. The first step is the immunity test: if the fitness of a chromosome  $\mathbf{u}$  is smaller than that of the parent chromosome, which indicates that degeneration occurred during crossover and mutation, then the parent chromosome will be used for the next competition. The second step is the annealing selection [35]: a chromosome  $\mathbf{u}_i$  is selected from the current offspring population  $\mathbf{U}_k$  to join with the new parents with the probability as follows:

$$P(\mathbf{u}_i) = \frac{\exp(f(\mathbf{u}_i)/T_k)}{\sum_{i=1}^{n_0} \exp(f(\mathbf{u}_i)/T_k)}, \quad (22)$$

where  $f(\mathbf{u}_i)$  is the fitness of the individual  $\mathbf{u}_i$  and  $\{T_k\}$  is the temperature-controlled series tending towards 0.

**3.4. Fitness Evaluation.** Fitness evaluation plays a critical role in selecting offspring chromosomes from the current population for the next generation. In this study, the fitness function for eigenvector selection is defined as

$$f(\mathbf{u}) = \frac{u_1 h_1 + u_2 h_2 + \dots + u_d h_d}{u_1 g_1 + u_2 g_2 + \dots + u_d g_d} = \frac{\mathbf{u}\mathbf{h}^T}{\mathbf{u}\mathbf{g}^T}, \quad (23)$$

where  $h_i$  denotes the  $\mathbf{v}_i^T \tilde{\mathbf{S}}_B \mathbf{v}_i$  value for the  $i$ th eigenvector,  $g_i$  denotes the  $\mathbf{v}_i^T \tilde{\mathbf{S}}_W \mathbf{v}_i$  value for the  $i$ th eigenvector,  $\mathbf{h} = [h_1, h_2, \dots, h_d]$ ,  $\mathbf{g} = [g_1, g_2, \dots, g_d]$ ,  $\mathbf{u} = [u_1, u_2, \dots, u_d]$ ,  $u_i \in \{0, 1\}$ ,  $\mathbf{u}\mathbf{1}^T = k$ , and  $i = 1, 2, \dots, d$ . Notably,  $\mathbf{u}$  is called the binary selector and  $k$  is the desired lower feature dimension. Finally, according to the evolved binary selector  $\mathbf{u}$ , we can thus form the projection matrix  $\mathbf{V}^\phi$  of the  $t$ th step by choosing the  $k$  eigenvectors with  $u_i = 1$  ( $i = 1, 2, \dots, d$ ). The procedures of the proposed TR-KDA-BIGA are summarized in the procedures of the proposed TR-KDA-BIGA part. The computational flow of the BIGA obtained using the aforementioned genetic and immune operators is also provided in the computational flow of the BIGA part.

*The Procedures of the Proposed TR-KDA-BIGA.* The procedures are as follows:

- (1) Construct the kernel matrix  $\mathbf{K} = (\phi(\mathbf{X}))^T \phi(\mathbf{X})$ .
- (2) Perform Cholesky decomposition to the kernel matrix  $\mathbf{K} = \mathbf{R}^T \mathbf{R}$ .
- (3) Form the kernel scatter matrixes as  $\tilde{\mathbf{S}}_B = \mathbf{R}\mathbf{L}_B \mathbf{R}^T$  and  $\tilde{\mathbf{S}}_W = \mathbf{R}\mathbf{L}_W \mathbf{R}^T$ .

- (4) Set iterations number  $t$  to 1.
- (5) Set the initial trace ratio value  $\lambda_t$  to  $\text{Tr}(\tilde{\mathbf{S}}_B)/\text{Tr}(\tilde{\mathbf{S}}_W)$ .
- (6) Compute the eigendecomposition of  $\tilde{\mathbf{S}}_B - \lambda_t \tilde{\mathbf{S}}_W$  as  $(\tilde{\mathbf{S}}_B - \lambda_t \tilde{\mathbf{S}}_W) \mathbf{v}_i^\phi = \tau_i \mathbf{v}_i^\phi$ , where  $\mathbf{v}_i^\phi$  ( $i = 1, 2, \dots, d$ ) is the eigenvector of  $\tilde{\mathbf{S}}_B - \lambda_t \tilde{\mathbf{S}}_W$ .
- (7) Calculate  $h_i = (\mathbf{v}_i^\phi)^T \tilde{\mathbf{S}}_B \mathbf{v}_i^\phi$  and  $g_i = (\mathbf{v}_i^\phi)^T \tilde{\mathbf{S}}_W \mathbf{v}_i^\phi$  for each eigenvector  $\mathbf{v}_i^\phi$  ( $i = 1, 2, \dots, d$ ).
- (8) Generate a population of BIGA selectors.
- (9) Evolve the population where the fitness of a BIGA selector  $\mathbf{u}$  is measured as  $f(\mathbf{u}) = \mathbf{u}\mathbf{h}^T/\mathbf{u}\mathbf{g}^T$ .
- (10)  $\mathbf{u}^*$  is the evolved best BIGA selector.
- (11) Form the projection matrix  $\mathbf{V}_t^\phi$  by choosing the  $k$  eigenvectors  $\mathbf{v}_i^\phi$  with  $u_i^* = 1$  ( $i = 1, 2, \dots, d$ ).
- (12) Update the trace ratio value  $\lambda_{t+1} = \text{Tr}[(\mathbf{V}_t^\phi)^T \tilde{\mathbf{S}}_B \mathbf{V}_t^\phi] / \text{Tr}[(\mathbf{V}_t^\phi)^T \tilde{\mathbf{S}}_W \mathbf{V}_t^\phi]$ ,  $t = t + 1$ , and go to step (6). Repeat this procedure until a convergence condition was established when the trace ratio value does not increase in consecutive 5 iterations.
- (13) Output  $\mathbf{Y}^{\phi*} = (\mathbf{V}^{\phi*})^T \mathbf{R}$ .

*The Computational Flow of the BIGA.* The computational flow is as follows:

- (1) Set  $l$  (time of generation) to 1.
- (2) Initialize randomly the original population  $A_l$ .
- (3) Evaluate each chromosome in the original population  $A_l$ .
- (4) Abstract vaccines according to the prior knowledge.
- (5) Check for termination criteria. If the fixed number of generations is not reached or the optimal chromosome found thus far is not satisfied, then go to the next step. Otherwise, output the optimal chromosome as the final solutions for further decision-making.
- (6) Perform crossover operation on the  $A_l$  and then generate the population  $B_l$ .
- (7) Perform mutation operation on the  $B_l$  and then generate the population results  $C_l$ .
- (8) Perform vaccination operation on the  $C_l$  and then generate the population  $D_l$ .
- (9) Perform immune selection operation on the  $D_l$  and then generate the next generational population  $A_{l+1}$ . Go to step (3).

## 4. The Convergence of the Proposed TR-KDA-BIGA

In this section, we analyze the convergence of the proposed TR-LDA-BIGA. Before doing this task, it should be worth noting that the BIGA is convergent. It has been demonstrated by Jiao and Wang [33] that as long as enough iteration has

been completed, the immune genetic population converges towards the true optimum with probability one.

Recall the trace difference function

$$z(\lambda) = \max_{(\mathbf{V}^\phi)^\top \mathbf{V}^\phi = \mathbf{I}} \text{Tr} \left[ (\mathbf{V}^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda \tilde{\mathbf{S}}_W) \mathbf{V}^\phi \right]; \quad (24)$$

it follows that

$$\begin{aligned} z(\lambda_{t+1}) &= \max_{(\mathbf{V}^\phi)^\top \mathbf{V}^\phi = \mathbf{I}} \text{Tr} \left[ (\mathbf{V}^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda_{t+1} \tilde{\mathbf{S}}_W) \mathbf{V}^\phi \right] \implies \\ z(\lambda_{t+1}) &= \text{Tr} \left[ (\mathbf{V}_{t+1}^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda_{t+1} \tilde{\mathbf{S}}_W) \mathbf{V}_{t+1}^\phi \right]. \end{aligned} \quad (25)$$

Since  $\lambda_{t+1} = \text{Tr}[(\mathbf{V}_t^\phi)^\top \tilde{\mathbf{S}}_B \mathbf{V}_t^\phi] / \text{Tr}[(\mathbf{V}_t^\phi)^\top \tilde{\mathbf{S}}_W \mathbf{V}_t^\phi]$  as previously mentioned, we get

$$\text{Tr} \left[ (\mathbf{V}_t^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda_{t+1} \tilde{\mathbf{S}}_W) \mathbf{V}_t^\phi \right] = 0. \quad (26)$$

Consider the inequality

$$\begin{aligned} \text{Tr} \left[ (\mathbf{V}_{t+1}^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda_{t+1} \tilde{\mathbf{S}}_W) \mathbf{V}_{t+1}^\phi \right] \\ \geq \text{Tr} \left[ (\mathbf{V}_t^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda_{t+1} \tilde{\mathbf{S}}_W) \mathbf{V}_t^\phi \right], \end{aligned} \quad (27)$$

and the equation

$$\text{Tr} \left[ (\mathbf{V}_t^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda_{t+1} \tilde{\mathbf{S}}_W) \mathbf{V}_t^\phi \right] = 0, \quad (28)$$

and we have

$$\begin{aligned} z(\lambda_{t+1}) &\geq 0, \\ \text{Tr} \left[ (\mathbf{V}_{t+1}^\phi)^\top \tilde{\mathbf{S}}_B \mathbf{V}_{t+1}^\phi \right] - \lambda_{t+1} \text{Tr} \left[ (\mathbf{V}_{t+1}^\phi)^\top \tilde{\mathbf{S}}_W \mathbf{V}_{t+1}^\phi \right] \\ &\geq 0. \end{aligned} \quad (29)$$

Consequently,

$$\begin{aligned} \frac{\text{Tr} \left[ (\mathbf{V}_{t+1}^\phi)^\top \tilde{\mathbf{S}}_B \mathbf{V}_{t+1}^\phi \right]}{\text{Tr} \left[ (\mathbf{V}_{t+1}^\phi)^\top \tilde{\mathbf{S}}_W \mathbf{V}_{t+1}^\phi \right]} &\geq \lambda_{t+1} \implies \\ \lambda_{t+2} &\geq \lambda_{t+1}. \end{aligned} \quad (30)$$

Substituting the subscript  $t + 1$  by  $t$  yields

$$\lambda_{t+1} \geq \lambda_t. \quad (31)$$

So we obtain the following inequality which gives the first expression of convergence of the proposed TR-KDA-BIGA.

Further, suppose that  $\lambda^*$  is the optimal trace ratio value; it follows that

$$z(\lambda^*) = \text{Tr} \left[ (\mathbf{V}^{\phi*})^\top (\tilde{\mathbf{S}}_B - \lambda^* \tilde{\mathbf{S}}_W) \mathbf{V}^{\phi*} \right] = 0, \quad (32)$$

TABLE 1: Specification of benchmark problems.

Data set	#samples	#dim.	#class
Heart-statlog	270	13	2
Ionosphere	351	34	2
Iris	150	4	3
Wine	178	13	3
Waveform	5000	40	3
Balance	625	4	3
SCCTS	600	60	6

where  $\mathbf{V}^{\phi*}$  is the optimal projection matrix. We therefore have

$$\begin{aligned} z(\lambda^*) &= \text{Tr} \left[ (\mathbf{V}^{\phi*})^\top (\tilde{\mathbf{S}}_B - \lambda^* \tilde{\mathbf{S}}_W) \mathbf{V}^{\phi*} \right] \\ &\geq \text{Tr} \left[ (\mathbf{V}_t^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda^* \tilde{\mathbf{S}}_W) \mathbf{V}_t^\phi \right] \\ &= \text{Tr} \left[ (\mathbf{V}_t^\phi)^\top (\tilde{\mathbf{S}}_B - \lambda_{t+1} \tilde{\mathbf{S}}_W) \mathbf{V}_t^\phi \right] \\ &\quad + (\lambda_{t+1} - \lambda^*) \text{Tr} \left[ (\mathbf{V}_t^\phi)^\top \tilde{\mathbf{S}}_W \mathbf{V}_t^\phi \right] \\ &= z(\lambda_{t+1}) + (\lambda_{t+1} - \lambda^*) \text{Tr} \left[ (\mathbf{V}_t^\phi)^\top \tilde{\mathbf{S}}_W \mathbf{V}_t^\phi \right]. \end{aligned} \quad (33)$$

Consider  $z(\lambda^*) = 0$ ,  $z(\lambda_{t+1}) \geq 0$  and  $\tilde{\mathbf{S}}_W$  is semipositive definite; we have

$$\lambda_{t+1} - \lambda^* \leq 0. \quad (34)$$

So we obtain the following inequality which gives the second expression of convergence of the proposed TR-KDA-BIGA:

$$\lambda_{t+1} \leq \lambda^*. \quad (35)$$

We conclude therefore that, for a particular initial trace ratio value  $\lambda_t$ , the updated value  $\lambda_{t+1}$  can always satisfy (1)  $\lambda_{t+1} \geq \lambda_t$  and (2)  $\lambda_{t+1} \leq \lambda^*$ .

## 5. Performance Evaluation on Benchmark Problems

In order to extensively verify the performance of the proposed TR-KDA-BIGA, it is first tested on wide types of commonly used benchmark problems taken from the UCI machine learning repository and evaluated with the classification rate (i.e., the number of correctly identified training examples/total number of training examples) by comparison with other existing methods such as PCA, LDA, KPCA [30, 31], KDA [32], and TR-LDA. These data sets include Heart-statlog, Ionosphere, Iris, Wine, Waveform, Balance, and Synthetic Control Chart Time Series (SCCTS) data sets (Table 1), which are of small sizes, low dimensions, large sizes, and/or high dimensions. For comparative study, we randomly select 50% data points from each data set as training set and the rest of the data points as test set. All methods use training set

TABLE 2: Results of the classification rate for the benchmark problems (mean  $\pm$  derivation).

Data set	PCA	KPCA	LDA	KDA	TR-LDA	TR-KDA-BIGA
Heart-statlog	59.7 $\pm$ 4.6	60.0 $\pm$ 4.2	75.5 $\pm$ 3.3	74.2 $\pm$ 2.7	76.1 $\pm$ 3.1	75.8 $\pm$ 2.3
Ionosphere	74.4 $\pm$ 4.0	74.9 $\pm$ 3.8	74.7 $\pm$ 5.7	75.1 $\pm$ 4.7	75.2 $\pm$ 5.9	76.7 $\pm$ 4.5
Iris	91.0 $\pm$ 4.0	91.0 $\pm$ 4.2	91.0 $\pm$ 3.5	91.8 $\pm$ 3.0	92.3 $\pm$ 3.2	92.9 $\pm$ 3.6
Wine	68.3 $\pm$ 5.2	69.0 $\pm$ 4.9	77.7 $\pm$ 8.4	77.0 $\pm$ 6.9	78.4 $\pm$ 8.5	78.6 $\pm$ 6.3
Waveform	72.9 $\pm$ 1.7	73.1 $\pm$ 1.5	73.3 $\pm$ 1.7	73.0 $\pm$ 2.3	74.5 $\pm$ 1.3	74.9 $\pm$ 2.8
Balance	63.4 $\pm$ 4.2	66.4 $\pm$ 4.7	77.8 $\pm$ 4.9	78.4 $\pm$ 3.7	78.3 $\pm$ 4.2	79.2 $\pm$ 3.2
SCCTS	80.0 $\pm$ 4.8	80.1 $\pm$ 5.0	80.5 $\pm$ 5.4	81.1 $\pm$ 6.2	81.2 $\pm$ 5.6	82.9 $\pm$ 6.7

TABLE 3: Time-domain statistical features.

Feature	Eq.	Feature	Eq.
Standard deviation	$x_{\text{std}} = \sqrt{\frac{\sum_{i=1}^N (x(i) - \bar{x})^2}{N}}$	Clearance factor	$\text{CLF} = \frac{x_p}{x_{\text{rms}}}$
Peak	$x_p = \max  x(i) $	Shape factor	$\text{SF} = \frac{x_{\text{rms}}}{(1/N) \sum_{i=1}^N  x(i) }$
Skewness	$x_{\text{ske}} = \frac{(1/N) \sum_{i=1}^N (x(i) - \bar{x})^3}{x_{\text{std}}^3}$	Impact factor	$\text{IF} = \frac{x_p}{(1/N) \sum_{i=1}^N  x(i) }$
Kurtosis	$x_{\text{kur}} = \frac{(1/N) \sum_{i=1}^N (x(i) - \bar{x})^4}{x_{\text{std}}^4}$	Square mean root	$x_{\text{smr}} = \left( \frac{1}{N} \sum_{i=1}^N \sqrt{ x(i) } \right)^2$
Crest factor	$\text{CF} = \frac{x_p}{x_{\text{rms}}}$		

where  $x(i)$  is a digital signal series,  $i = 1, 2, \dots, N$ ,  $N$  is the number of elements of the digital signal, and  $\bar{x} = \sum_{i=1}^N x(i)/N$  and  $x_{\text{rms}} = \sqrt{\sum_{i=1}^N x(i)^2/N}$  are the mean value and root-mean-square value of the digital signal series, respectively.

in the output reduced space to train one nearest neighborhood (INN) classifier for evaluating the classification rate of test set. To restrict the influence of random effects, the experiments of PCA, LDA, KPCA, KDA, TR-LDA, and TR-KDA-BIGA compared on each benchmark problem are independently performed for 20 runs. Table 2 compares the classification rate for benchmark problems of the proposed TR-KDA-BIGA with that of the PCA, LDA, KPCA, KDA, and TR-LDA. As seen in Table 2, the proposed TR-KDA-BIGA can perform better than all the compared methods, except in the case of Heart-statlog.

The results obtained demonstrate the ability of the proposed TR-KDA-BIGA in classifying different classes well. Thus, the proposed TR-KDA-BIGA may be effectively employed for fault diagnosis of rolling element bearings.

## 6. The Proposed TR-KDA Using BIGA for Fault Diagnosis of Rolling Element Bearings

In this section, the proposed TR-KDA-BIGA is applied to fault diagnosis of rolling element bearings. Vibration signals resulting from rolling element bearings are first filtered by using a low-pass filter. Then, the filtered vibration signals are divided into sections of equal window length. One set of relevant features obtained from each window is used for characterizing to some extent the health status of the rolling element bearings. Most of the faults occurring in rolling element

bearings will introduce the increased friction and impulsive forces when bearings are rotating, which generally lead the vibration signals in time-domain, frequency-domain, and/or time-frequency-domain to vary (become different) from the normal ones. In this study, 9 time-domain statistical features (Table 3) are extracted from the vibration signal. All of these 9 time-domain statistical features reflect the characteristics of time series data in the time-domain. Moreover, 6 time-frequency-domain wavelet features about the percentages of energy corresponding to wavelet coefficients are extracted from the vibration signal by using Daubechies-4 (db4) wavelet to decompose the vibration signal into five levels [32]. Wavelet features extracted in such a way can to the greatest extent reflect the vibration energy distribution in the time-frequency-domain. Thus, 9 time-domain statistical features together with 6 time-frequency-domain wavelet features are used to represent each window's vibration signals.

*6.1. Experimental Setup.* In order to demonstrate the performance of the proposed TR-KDA-BIGA, rolling element bearing data obtained from the Bearing Data Centre, Case Western Reserve University [36], are used. The test rolling element bearings were SKF 6205 JEM, a type of deep groove ball bearing. Single-point faults were seeded into the drive end ball bearing using electrodischarge machining. Faults occurring in rolling element bearings introduced impact-like vibration signals when bearings were rotating. An accelerometer was mounted on the drive end of the motor housing

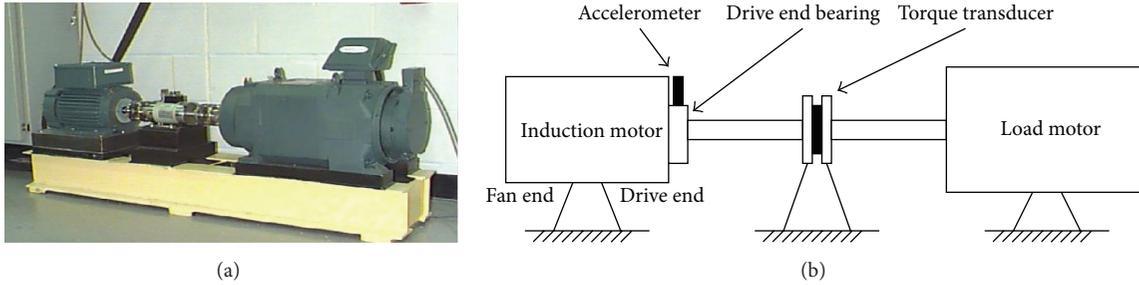


FIGURE 1: Experimental setup: (a) test rig; (b) schematic description of the test rig.

to detect such impacts that behaved like damped oscillations. Vibration signals were captured from four different health statuses of bearing, that is, normal bearings (Normal), inner race fault (IR), ball fault (BA), and outer race fault (OR). For each of the three abnormal statuses (IR, BA, and OR), there are three different levels of severity with fault diameters (0.007 inches, 0.014 inches, and 0.021 inches). All the experiments were done for three different load conditions (1 HP, 2 HP, and 3 HP). Figure 1 illustrates the experimental setup. Experimental data were collected from the drive end ball bearing of an induction motor (Reliance Electric 2 HP IQPreAlert) driven test rig. Table 4 gives a short description of rolling element bearing data.

## 6.2. Experiment Results

**6.2.1. Visualization of Bearing Data.** Visualization performances of the proposed TR-KDA-BIGA are compared with those of PCA, LDA, KPCA, KDA, and TR-LDA using simulations, where KPCA and KDA are the kernel extensions to PCA and LDA, respectively. The two-dimensional visualization results of bearing data for three different load conditions (1, 2, and 3 HP) obtained with PCA, LDA, KPCA, KDA, TR-LDA, and the proposed TR-KDA-BIGA are summarized in Figures 2, 3, and 4, respectively. As seen in Figures 2, 3, and 4, the proposed TR-KDA-BIGA outperforms all the compared methods in not only closely conglomerating bearing data belonging to the same class but also clearly separating bearing data belonging to different classes of three different load conditions (1, 2, and 3 HP). Compared with the unsupervised methods (i.e., PCA and KPCA), the supervised methods (i.e., LDA, KDA, TR-LDA, and TR-KDA-BIGA) can preserve more discriminative information embedded in bearing data and obtain clearer and less overlapped boundaries. It can also be concluded from Figures 2, 3, and 4 that the methods using kernel trick (i.e., KPCA, KDA, and TR-KDA-BIGA) performed better than the methods without using kernel trick (i.e., PCA, LDA, and TR-LDA) in separating the discriminative property—samples from different classes in the learned subspace.

**6.2.2. Classification of Bearing Data.** Classification performances of the proposed TR-KDA-BIGA are compared with those of PCA, LDA, KPCA, KDA, and TR-LDA. In order to show the robustness of the proposed TR-KDA-BIGA, we

TABLE 4: Description of rolling element bearing data.

#samples	#dimension	#class	#samples per class
800	15	10	80

perform 4 independent experiments for each load condition in terms of 4 different data partitions. In this study, 10, 20, 30, and 40 samples per class in bearing data set are randomly selected from each class in bearing data as the training set and the remaining samples as the test set. Then, each method uses the training set to train a INN classifier in order to classify different health status in test set. Tables 5, 6, and 7 summarize the average classification results of PCA, LDA, KPCA, KDA, TR-LDA, and the proposed TR-KDA-BIGA with various numbers of training samples for 1HP, 2HP, and 3HP load conditions, respectively. It can be observed that the overall average performance of the classification of health status is fairly good. Tables 5, 6, and 7 demonstrate that the proposed TR-KDA-BIGA performs remarkably better than the compared methods (PCA, LDA, KPCA, KDA, and TR-LDA). It should be noted that the proposed model can also provide the capability of real-time visualization that is very useful for the practitioners to monitor the health status of rolling element bearings. Tables 5, 6, and 7 also demonstrate that the number of training samples does significantly affect the classification accuracy for bearing health status.

## 7. Conclusions

The rolling element bearing is a core component of many systems, and their failure can lead to reduced capability, downtime, and even catastrophic breakdowns. Effective and efficient fault diagnosis of rolling element bearings plays an extremely important role in the safe and reliable operation of their host systems. In the current study, fault diagnosis of rolling element bearings is done in a pattern recognition way by calculating a high-dimensional feature data set from vibration signals, which represents the different status of bearings. Specifically, the TR-KDA is presented for fault diagnosis of rolling element bearings and the BIGA is employed to solve the trace ratio problem in TR-KDA. The numerical results obtained using extensive simulation indicate that the proposed TR-KDA-BIGA can effectively

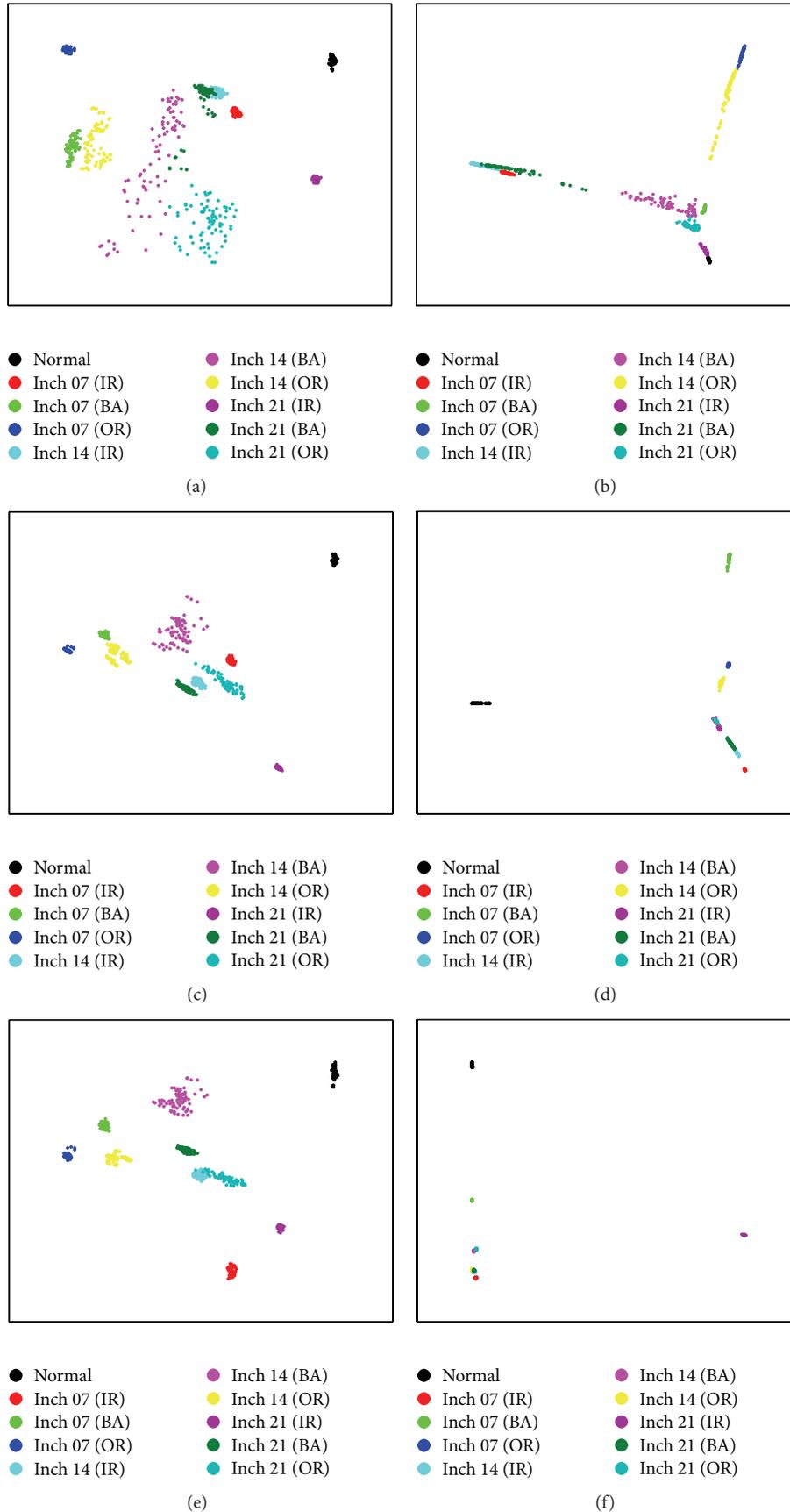


FIGURE 2: Two-dimensional representation of bearing data under 1 HP motor load by each method: (a) PCA; (b) KPCA; (c) LDA; (d) KDA; (e) TR-LDA; (f) TR-KDA-BIGA.

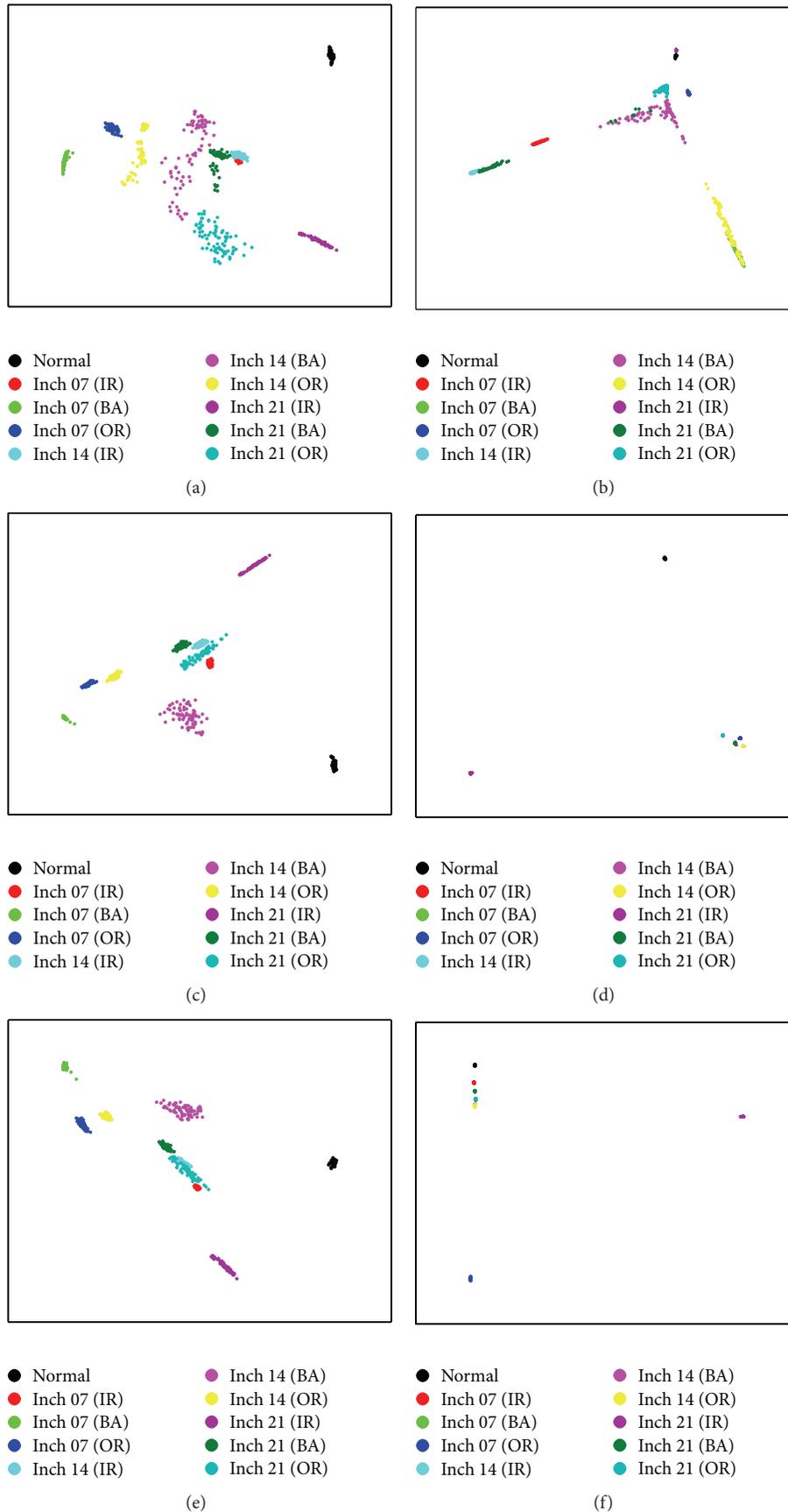


FIGURE 3: Two-dimensional representation of bearing data under 2 HP motor load by each method: (a) PCA; (b) KPCA; (c) LDA; (d) KDA; (e) TR-LDA; (f) TR-KDA-BIGA.

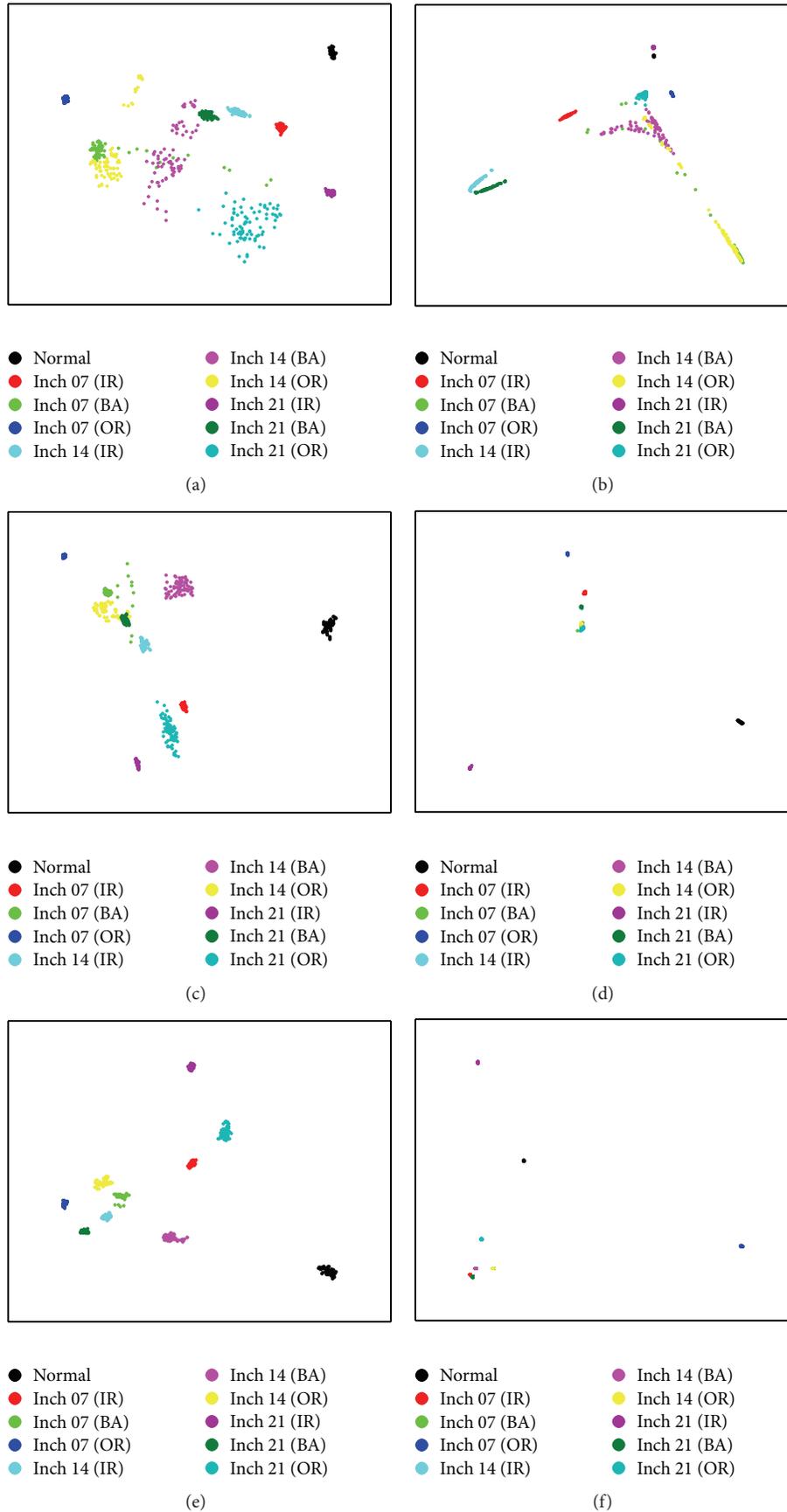


FIGURE 4: Two-dimensional representation of bearing data under 3 HP motor load by each method: (a) PCA; (b) KPCA; (c) LDA; (d) KDA; (e) TR-LDA; (f) TR-KDA-BIGA.

TABLE 5: Classification accuracy of bearing data under 1 HP motor load.

Method	Number of training samples				Average
	10	20	30	40	
PCA	94.7793	95.9016	96.7564	97.7785	96.30395
KPCA	94.9203	95.9007	96.5873	98.4867	96.47375
LDA	96.2741	97.1929	98.0017	98.8438	97.57813
KDA	96.7026	97.1957	98.1504	99.1511	97.79995
TR-LDA	97.1914	97.9510	99.0587	99.7006	98.47543
TR-KDA-BIGA	97.5793	98.8971	99.5416	99.8068	98.95620

TABLE 6: Classification accuracy of bearing data under 2 HP motor load.

Method	Number of training samples				Average
	10	20	30	40	
PCA	94.4654	95.4452	96.3670	97.6571	95.98368
KPCA	94.2155	96.5207	97.3084	98.7273	96.69298
LDA	96.0477	96.9640	97.6314	99.0035	97.41165
KDA	96.3333	97.5962	97.2495	99.4105	97.64738
TR-LDA	97.3188	98.0867	98.7036	99.7777	98.47170
TR-KDA-BIGA	98.9123	98.6209	99.7737	99.9311	99.30950

TABLE 7: Classification accuracy of bearing data under 3 HP motor load.

Method	Number of training samples				Average
	10	20	30	40	
PCA	96.1337	97.3051	97.6719	97.9738	97.27113
KPCA	96.7442	97.0369	97.6968	98.2854	97.44083
LDA	97.0421	98.5900	99.2875	99.4996	98.60480
KDA	97.4820	98.8408	99.1516	99.6082	98.77065
TR-LDA	98.8961	99.3457	99.9190	99.6082	99.44225
TR-KDA-BIGA	99.2918	99.9153	99.9103	99.9291	99.76163

classify different classes of rolling element bearing data, while also providing the capability of real-time visualization that is very useful for the practitioners to monitor the health status of rolling element bearings. Empirical comparisons show that the proposed TR-KDA-BIGA performs better than existing methods in classifying different rolling element bearing data. The proposed TR-KDA-BIGA may be a promising tool for fault diagnosis of rolling element bearings.

Three research directions are worth pursuing. First, although this study considers the specific fault diagnosis of rolling element bearings, the proposed method can be modified and extended to address the fault diagnosis of gearboxes [37, 38] and cutting tools [39, 40]. Second, frequency-domain information can be utilized for fault diagnosis of rolling element bearings [41, 42]; it would thus be interesting to integrate frequency-domain features to time-domain and time-frequency-domain features. Third, empirical mode decomposition is a very powerful tool for nonlinear and nonstationary signal processing [43–45]; it would be also interesting to employ the empirical mode decomposition to extract periodic components and random transient components from the bearing vibration signal mixture, which may be very helpful for extraction of fault signatures from a collected bearing vibration signal.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The research is funded partially by the National Science Foundation of China (51405239), National Defense Basic Scientific Research Program of China (A2620132010, A2520110003), Jiangsu Provincial Natural Science Foundation of China (BK20150745, BK20140727), Jiangsu Province Science and Technology Support Program (BE2014134), Fundamental Research Funds for the Central Universities (1005-YAH15055), and Jiangsu Postdoctoral Science Foundation of China (1501024C). The authors would like to express sincere appreciation to Professor KA Loparo and Case Western Reserve University for their efforts to make bearing data set available and permission to use data set.

## References

- [1] X. Jin, E. W. M. Ma, L. L. Cheng, and M. Pecht, "Health monitoring of cooling fans based on mahalanobis distance with mRMR feature selection," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 8, pp. 2222–2229, 2012.

- [2] X. Jin and T. W. S. Chow, "Anomaly detection of cooling fan and fault classification of induction motor using Mahalanobis-Taguchi system," *Expert Systems with Applications*, vol. 40, no. 15, pp. 5787–5795, 2013.
- [3] J. Zarei, "Induction motors bearing fault detection using pattern recognition techniques," *Expert Systems with Applications*, vol. 39, no. 1, pp. 68–73, 2012.
- [4] J.-B. Yu, "Bearing performance degradation assessment using locality preserving projections," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7440–7450, 2011.
- [5] D. Wang, P. W. Tse, and Y. L. Tse, "A morphogram with the optimal selection of parameters used in morphological analysis for enhancing the ability in bearing fault diagnosis," *Measurement Science and Technology*, vol. 23, no. 6, Article ID 065001, 2012.
- [6] W. Wang and M. Pecht, "Economic analysis of canary-based prognostics and health management," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 7, pp. 3077–3089, 2011.
- [7] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—a tutorial," *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 485–520, 2011.
- [8] Y. Yang, Y. Liao, G. Meng, and J. Lee, "A hybrid feature selection scheme for unsupervised learning and its application in bearing fault diagnosis," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11311–11320, 2011.
- [9] J. Rafiee, M. A. Rafiee, and P. W. Tse, "Application of mother wavelet functions for automatic gear and bearing fault diagnosis," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4568–4579, 2010.
- [10] W. He, Z.-N. Jiang, and K. Feng, "Bearing fault detection based on optimal wavelet filter and sparse code shrinkage," *Measurement*, vol. 42, no. 7, pp. 1092–1102, 2009.
- [11] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [12] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [13] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, "Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3398–3407, 2013.
- [14] M. B. Zhao, Z. Zhang, and T. W. S. Chow, "Trace ratio criterion based generalized discriminative learning for semi-supervised dimensionality reduction," *Pattern Recognition*, vol. 45, no. 4, pp. 1482–1499, 2012.
- [15] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.
- [16] K. Feng, Z. Jiang, W. He, and B. Ma, "A recognition and novelty detection approach based on Curvelet transform, nonlinear PCA and SVM with application to indicator diagram diagnosis," *Expert Systems with Applications*, vol. 38, no. 10, pp. 12721–12729, 2011.
- [17] Q. Jiang, M. Jia, J. Hu, and F. Xu, "Machinery fault diagnosis using supervised manifold learning," *Mechanical Systems and Signal Processing*, vol. 23, no. 7, pp. 2301–2311, 2009.
- [18] E. G. Strangas, S. Aviyente, and S. S. H. Zaidi, "Time-frequency analysis for efficient fault diagnosis and failure prognosis for interior permanent-magnet AC motors," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 12, pp. 4191–4199, 2008.
- [19] Y. Wang, E. W. M. Ma, T. W. S. Chow, and K.-L. Tsui, "A two-step parametric method for failure prediction in hard disk drives," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 419–430, 2014.
- [20] J. Yu, "Local and nonlocal preserving projection for bearing defect classification and performance assessment," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 5, pp. 2363–2376, 2012.
- [21] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, 1990.
- [22] H. Wang, S. Yan, D. Xu, X. Tang, and T. Huang, "Trace ratio vs. ratio trace for dimensionality reduction," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, Minneapolis, Minn, USA, June 2007.
- [23] M. B. Zhao, R. H. M. Chan, P. Tang, T. W. S. Chow, and S. W. H. Wong, "Trace ratio linear discriminant analysis for medical diagnosis: a case study of dementia," *IEEE Signal Processing Letters*, vol. 20, no. 5, pp. 431–434, 2013.
- [24] L. Zhou, L. Wang, and C. H. Shen, "Feature selection with redundancy-constrained class separability," *IEEE Transactions on Neural Networks*, vol. 21, no. 5, pp. 853–858, 2010.
- [25] T. Sun and S. Chen, "Class label versus sample label-based CCA," *Applied Mathematics and Computation*, vol. 185, no. 1, pp. 272–283, 2007.
- [26] Y.-F. Guo, S.-J. Li, J.-Y. Yang, T.-T. Shu, and L.-D. Wu, "A generalized Foley-Sammon transform based on generalized fisher discriminant criterion and its application to face recognition," *Pattern Recognition Letters*, vol. 24, no. 1–3, pp. 147–158, 2003.
- [27] M. B. Zhao, X. H. Jin, Z. Zhang, and B. Li, "Fault diagnosis of rolling element bearings via discriminative subspace learning: visualization and classification," *Expert Systems with Applications*, vol. 41, no. 7, pp. 3391–3401, 2014.
- [28] X. H. Jin, M. B. Zhao, T. W. S. Chow, and M. S. Pecht, "Motor bearing fault diagnosis using trace ratio linear discriminant analysis," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2441–2451, 2014.
- [29] Y. Q. Jia, F. P. Nie, and C. S. Zhang, "Trace ratio problem revisited," *IEEE Transactions on Neural Networks*, vol. 20, no. 4, pp. 729–735, 2009.
- [30] J. Yang, A. F. Frangi, J.-Y. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 230–244, 2005.
- [31] C. Zhang, F. Nie, and S. Xiang, "A general kernelization framework for learning algorithms based on kernel PCA," *Neurocomputing*, vol. 73, no. 4–6, pp. 959–967, 2010.
- [32] S. Ji and J. Ye, "Kernel uncorrelated and regularized discriminant analysis: a theoretical and computational study," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 10, pp. 1311–1321, 2008.
- [33] L. C. Jiao and L. Wang, "A novel genetic algorithm based on immunity," *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, vol. 30, no. 5, pp. 552–561, 2000.
- [34] F. R. K. Chung, *Spectral Graph Theory*, CBMS Regional Conference Series in Mathematics, No. 92, American Mathematical Society, 1997.

- [35] J. S. Zhang, Z. B. Xu, and Y. Liang, "The whole annealing genetic algorithms and their sufficient and necessary conditions of convergence," *Science of China*, vol. 27, no. 2, pp. 154–164, 1997.
- [36] K. A. Loparo, "Bearings vibration data set," Case Western Reserve University, <http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website>.
- [37] D. Wang, Q. Miao, and R. Kang, "Robust health evaluation of gearbox subject to tooth failure with wavelet decomposition," *Journal of Sound and Vibration*, vol. 324, no. 3–5, pp. 1141–1157, 2009.
- [38] D. Wang, P. W. Tse, W. Guo, and Q. Miao, "Support vector data description for fusion of multiple health indicators for enhancing gearbox fault diagnosis and prognosis," *Measurement Science and Technology*, vol. 22, no. 2, Article ID 025102, 2011.
- [39] F. J. Alonso and D. R. Salgado, "Analysis of the structure of vibration signals for tool wear detection," *Mechanical Systems and Signal Processing*, vol. 22, no. 3, pp. 735–748, 2008.
- [40] K. P. Zhu, G. S. Hong, and Y. S. Wong, "A comparative study of feature selection for hidden Markov model-based micro-milling tool wear monitoring," *Machining Science and Technology*, vol. 12, no. 3, pp. 348–369, 2008.
- [41] D. Wang, Q. Miao, X. F. Fan, and H.-Z. Huang, "Rolling element bearing fault detection using an improved combination of Hilbert and wavelet transforms," *Journal of Mechanical Science and Technology*, vol. 23, no. 12, pp. 3292–3301, 2010.
- [42] Y. G. Lei, M. J. Zuo, Z. J. He, and Y. Y. Zi, "A multidimensional hybrid intelligent method for gear fault diagnosis," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1419–1430, 2010.
- [43] D. Wang, W. Guo, and P. W. Tse, "An enhanced empirical mode decomposition method for blind component separation of a single-channel vibration signal mixture," *Journal of Vibration and Control*, 2014.
- [44] Y. G. Lei, M. J. Zuo, and M. R. Hoseini, "The use of ensemble empirical mode decomposition to improve bispectral analysis for fault detection in rotating machinery," *Proceedings of the Institution of Mechanical Engineers Part C: Journal of Mechanical Engineering Science*, vol. 224, no. 8, pp. 1759–1769, 2010.
- [45] Y. G. Lei, Z. J. He, and Y. Y. Zi, "Application of the EEMD method to rotor fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1327–1338, 2009.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

