

Motor Bearing Fault Diagnosis Using Trace Ratio Linear Discriminant Analysis

Xiaohang Jin, Mingbo Zhao, Tommy W. S. Chow, *Senior Member, IEEE*, and Michael Pecht, *Fellow, IEEE*

Abstract—Bearings are critical components in induction motors and brushless direct current motors. Bearing failure is the most common failure mode in these motors. By implementing health monitoring and fault diagnosis of bearings, unscheduled maintenance and economic losses caused by bearing failures can be avoided. This paper introduces trace ratio linear discriminant analysis (TR-LDA) to deal with high-dimensional non-Gaussian fault data for dimension reduction and fault classification. Motor bearing data with single-point faults and generalized-roughness faults are used to validate the effectiveness of the proposed method for fault diagnosis. Comparisons with other conventional methods, such as principal component analysis, local preserving projection, canonical correction analysis, maximum margin criterion, LDA, and marginal Fisher analysis, show the superiority of TR-LDA in fault diagnosis.

Index Terms—Bearing, fault diagnosis, linear discriminant analysis (LDA), pattern recognition, trace ratio (TR) criterion, vibrations.

NOMENCLATURE

BLDC	Brushless DC.
BF	Ball fault.
CCA	Canonical correction analysis.
DC	Direct current.
FMMEA	Failure modes, mechanisms, and effects analysis.
hp	Horsepower.
IRF	Inner race fault.
ITR	Iterative trace ratio.
LDA	Linear discriminant analysis.
LLE	Locally linear embedding.
LPP	Local preserving projection.
MFA	Marginal Fisher analysis.

MMC	Maximum margin criterion.
NN	Nearest neighbor.
OLDA	Orthogonal LDA.
ORF	Outer race fault.
PCA	Principal component analysis.
PMM	Power ratio of maximal defective frequency to mean.
SLPP	Supervised LPP.
TR	Trace ratio.

I. INTRODUCTION

MOTORS are widely used in modern society. They consume over 50% of the electrical energy in industrial applications in the United States [1]. Bearings are important components in motors. According to reports [2]–[4], bearing failure is the most common failure mode of induction motors and BLDC motor fans. Bearing failure can lead to motor breakdown, loss of production and income, and even human casualties [5]. Therefore, health monitoring and fault diagnosis for bearings are important and play a key role in the reliable operation of motors. Although signals such as vibration, current, temperature, acoustic emission, and sound pressure can be used for bearing health monitoring and fault diagnosis, vibration signal analysis, used in this paper, is the most reliable, effective, standardized, and popular method [1]–[3], [6]–[19].

Bearing faults that are widespread in the industry can be categorized into two types, namely, single-point faults and generalized-roughness faults [10], [11]. A single-point fault is an obvious defect (e.g., pit and spall) localized on the bearing surface. It is usually caused by overloading during operation, which leads to a fatigue crack in the bearing surface until a piece of metal drops off. Generalized-roughness faults are faults that considerably degrade, roughen, or even deform the bearing surface. Some common causes of this type of fault are contamination, lack or loss of lubricant, and misalignment. In this paper, both of these two types of bearing faults are studied.

Normally, bearing fault diagnosis methods are based on time domain analysis [12]–[14], frequency domain analysis [15], [16], or time–frequency domain analysis [17]. In this paper, a set of features was constructed from vibration signals. In that way, bearing fault diagnosis is transformed into a pattern recognition problem with a high-dimensional and non-Gaussian data set. However, dealing with high-dimensional data has always been a major problem for pattern recognition. Hence, dimensionality reduction techniques can be used to reduce the complexity of the original data and embed high-dimensional

Manuscript received May 16, 2012; revised November 17, 2012 and March 11, 2013; accepted June 6, 2013. Date of publication August 21, 2013; date of current version October 18, 2013. This work was supported in part by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China, under Grant CityU8/CRF/09 and in part by the City University Strategic Research Grants under Grant 7002914.

X. Jin and T. W. S. Chow are with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong, and also with the Centre for Prognostics and System Health Management, City University of Hong Kong, Kowloon, Hong Kong (e-mail: xiaohangjin2-c@my.cityu.edu.hk; xiaohangjin@yahoo.com; eetchow@cityu.edu.hk).

M. Zhao is with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong (e-mail: mzhao4@cityu.edu.hk).

M. Pecht was with the Centre for Prognostics and System Health Management, City University of Hong Kong, Kowloon, Hong Kong. He is now with the Center for Advanced Life Cycle Engineering, University of Maryland, College Park, MD 20742 USA.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIE.2013.2273471

data into low-dimensional data while keeping most of the desired intrinsic information [20], [21]. Such intrinsic information can be geometrical [20]–[24] or discriminative [25], [26]. Several methods have been proposed for solving high-dimensional data in fault diagnosis [27]–[33].

Over the past decades, many dimensionality reduction methods have been proposed, including ISOMAP [20], LLE [21], LPP [26], PCA [34], and LDA [34], [35]. Recently, Yan *et al.* have proposed a new graph embedding framework to unify these existing methods (e.g., PCA, LDA, ISOMAP, LLE, and LPP), in which the statistical and geometrical properties of the data are encoded as graph relationships [36]. Among the dimensionality reduction methods, PCA [34] and LDA [34], [35] are the most popular methods. PCA pursues the direction of maximum variance for optimal reconstruction, whereas LDA is to find the optimal low-dimensional representation of the original data set by maximizing the between-class scatter matrix, i.e., S_b while minimizing the within-class scatter matrix, i.e., S_w . Due to the utilization of the labeled information, LDA can achieve better classification results than those obtained by PCA if sufficient labeled samples are provided.

Several variants of LDA have been proposed during the past decades, and TR-LDA is one of the most common [37], [38]. TR-LDA is based on the TR criterion, which can directly reflect Euclidean distances between data points of inter- and intraclass. In addition, the optimal projection obtained by TR-LDA is orthogonal. When evaluating the similarities between data points based on Euclidean distance, the orthogonal projection can preserve such similarities without any change [37]. Thus, TR-LDA tends to perform empirically better than the classical LDA in many problems. Recently, another type of LDA variants has been proposed by bringing the idea from the mentioned unsupervised manifold methods. These methods start from the local structure of the labeled sample and then aim to preserve the geometric information provided by both the data sample and the labeled information. Typical methods include MFA [36], locality sensitive discriminant analysis [39], local Fisher discriminant analysis [40], [41] and robust linearly optimized discriminant analysis [42].

The contributions of this paper can be summarized here.

- 1) TR-LDA, an orthogonal variant of LDA, is utilized for bearing fault diagnosis. It is also extended to deal with the non-Gaussian data sets confronted in many real-world fault diagnosis problems.
- 2) Real motor bearing data, including single-point fault bearing data and generalized-roughness fault bearing data, are used to validate the effectiveness of the proposed method for fault diagnosis. Results show the superiority of the TR-LDA method to other methods.

The rest of this paper is organized as follows: The basic concepts of LDA and TR-LDA are described in Section II. Section III introduces preprocessing and feature construction from bearing vibration signals. In Section IV, motor bearing data are used to test the effectiveness of the TR-LDA-based fault diagnosis, and the performance of the proposed method is also compared with other methods. Finally, conclusions are drawn in Section V.

II. TR-LDA

LDA uses the within-class scatter matrix, i.e., S_w , to evaluate the compactness within each class and the between-class scatter matrix, i.e., S_b , to evaluate the separability of different classes. The goal of LDA is to find a linear transformation matrix, i.e., $W \in R^{D \times d}$, mapping the original D -dimensional space onto a reduced d -dimensional feature space with $d \ll D$, for which the between-class scatter matrix is maximized, whereas the within-class scatter matrix is minimized. Let $X = \{x_1, x_2, \dots, x_l\} \in R^{D \times l}$ be the training set; each x_i belongs to a class $c_i = \{1, 2, \dots, c\}$. Let l_i be the number of data points in the i th class and l be the number of data points in all classes. Then, the between-class scatter matrix, i.e., S_b , the within-class scatter matrix, i.e., S_w , and the total-class scatter matrix, i.e., S_t , are defined as

$$S_b = \sum_{i=1}^c l_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (1)$$

$$S_w = \sum_{i=1}^c \sum_{x_i \in c_i} (x_i - \mu_i)(x_i - \mu_i)^T \quad (2)$$

$$S_t = \sum_{i=1}^l (x_i - \mu)(x_i - \mu)^T \quad (3)$$

where $\mu_i = 1/l_i \sum_{x_i \in c_i} x_i$ is the mean of the data points in the i th class, and $\mu = 1/l \sum_{i=1}^l x_i$ is the mean of the data points in all classes. The original formulation of LDA, called the Fisher LDA [35], can only deal with binary classification. Two optimization criteria can be used to extend the Fisher LDA to solve the multiclass classification problem. For the first criterion, the optimization of LDA is to find the optimal projection matrix \widehat{W} satisfying

$$\widehat{W} = \arg \max_W \text{Tr}((W^T S_w W)^{-1} (W^T S_b W)). \quad (4)$$

To distinguish this from the TR problem introduced later, the above optimization is called the ratio trace problem in this paper. If we assume that S_w is nonsingular, the optimization problem in (4) can be solved by generalized eigenvalue decomposition as [34], i.e.,

$$S_b w_k = \tau_k S_w w_k \quad (5)$$

where $w_k \in R^D$ is the eigenvector corresponding to the k th largest eigenvalue τ_k . We then form \widehat{W} by the w_k corresponding to the d largest eigenvalues. Finally, the original high-dimensional set X can be projected into a low-dimensional set $Y \in R^{d \times l}$ by $Y = \widehat{W}^T X$.

Another reasonable optimization criterion of LDA is to maximize $\text{Tr}(W^T S_b W)$ while minimizing $\text{Tr}(W^T S_w W)$. The optimization problem can be given by

$$\widehat{W} = \arg \max_{W^T W = I} \frac{\text{Tr}(W^T S_b W)}{\text{Tr}(W^T S_w W)}. \quad (6)$$

TABLE I
ITR ALGORITHM FOR SOLVING TR PROBLEM

- (1) Initialize the projection matrix W_0 as an arbitrary column-orthogonal matrix.
- (2) Compute the TR value $\lambda_t = \text{Tr}(W_t^T S_b W_t) / \text{Tr}(W_t^T S_w W_t)$.
- (3) Compute the eigen-decomposition of $S_b - \lambda_t S_w$ as $(S_b - \lambda_t S_w)w_i = \tau_i w_i$, where $w_i (i = 1, 2, \dots, D)$ is the eigenvector of $S_b - \lambda_t S_w$.
- (4) Choose the d eigenvectors w_i having the d largest eigenvalues τ_i to form W_t .
- (5) Iterate steps (2-4) until $\|W_{t+1} - W_t\| < \varepsilon$. Output $\hat{W} = W_{t+1}$.

The above problem is called the TR problem in this paper. In fact, TR-LDA is a new variant of LDA, which is based on the TR criterion. Let the i th column of projection matrix W as w_i , i.e., $W = \{w_1, w_2, \dots, w_d\}$, the goal of TR-LDA is to find $W^* = \arg \max_W (\sum_{i=1}^d w_i^T S_b w_i) / (\sum_{i=1}^d w_i^T S_w w_i)$. On the other hand, the ratio trace-based LDA is to iteratively find the i th column w_i that maximizes $w_i^T S_b w_i / w_i^T S_w w_i$, which is to use the greedy algorithm that optimizes $W^* = \arg \max_W \sum_{i=1}^d (w_i^T S_b w_i / w_i^T S_w w_i)$. From the above analysis, we can see that the optimal projection matrix obtained by the ratio trace-based LDA is only an approximated solution to that of TR-LDA. Hence, the TR-LDA is theoretically better than the ratio trace LDA.

A. Efficient Algorithm for Solving TR Problem

Solving the TR problem has never been a straightforward issue because it does not have a closed-form solution [38]. Fortunately, a recent study in [38] has rigorously proven that the TR problem can be solved equivalently to searching the optimal TR value $\lambda^* = \max_W W^T (S_b - \lambda^* S_w) W$ by using iterative procedures [37], [38], [43]. Based on this point, Wang *et al.* [37] proposed an efficient method, called the ITR algorithm, to solve the TR problem. The basic steps of the algorithm are shown in Table I.

Though the ITR algorithm works well for solving the TR problem, it has its drawbacks. First, the initialized orthogonal matrix W_0 is arbitrary and hard to choose. In some cases, when W_0 is well chosen, the algorithm is able to converge relatively quickly, while in most cases, an inappropriate W_0 dramatically increases the number of iterations. On the other hand, initializing λ_0 seems much easier for any λ_0 satisfying $0 \leq \lambda_0 < +\infty$ [38], [43]. Hence, λ_0 can simply be set as 0 in practice. Second, as shown in Step 4 in Table I, the ITR algorithm method has chosen d eigenvectors corresponding to the d largest eigenvalues of $S_b - \lambda_t S_w$ to form W_t . These eigenvectors can only maximize the trace difference value $\text{Tr}[W^T (S_b - \lambda_t S_w) W]$; however, they cannot maximize the TR value $\text{Tr}(W^T S_b W) / \text{Tr}(W^T S_w W)$. Thus, how to find d eigenvectors to maximize the TR value is an important question. Motivated by this issue, an improved ITR algorithm, called the ITR-Score algorithm, is proposed in the preliminary work of [25], [44] to solve this problem. For all the eigenvectors of $S_b - \lambda_t S_w$, the ITR-Score algorithm computes a score $s_i = w_i^T S_b w_i / w_i^T S_w w_i$ for each eigenvector w_i . Then, d eigenvectors with the largest scores form W_t . This procedure can be viewed as a greedy algorithm that optimizes the

TABLE II
ITR-SCORE ALGORITHM FOR SOLVING TR PROBLEM

- (1) Initialize $\lambda_0 = 0$.
- (2) Compute the eigen-decomposition of $S_b - \lambda_t S_w$ as $(S_b - \lambda_t S_w)w_i = \tau_i w_i$, where $w_i (i = 1, 2, \dots, D)$ is the eigenvector of $S_b - \lambda_t S_w$.
- (3) Compute the score $s_i = w_i^T S_b w_i / w_i^T S_w w_i$ for each eigenvector w_i .
- (4) Choose the top d eigenvectors w_i having the d largest eigenvalues s_i to form W_t .
- (5) Update $\lambda_{t+1} = \text{Tr}(W_t^T S_b W_t) / \text{Tr}(W_t^T S_w W_t)$.
- (6) Iterate the steps (2-5) until $|\lambda_{t+1} - \lambda_t| < \varepsilon$. Output $\hat{W} = W_{t+1}$.

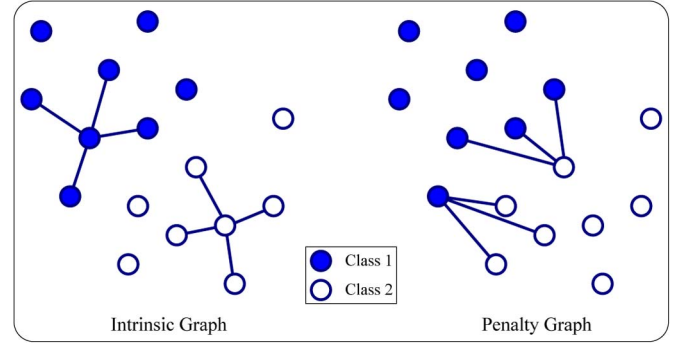


Fig. 1. Intrinsic graph and penalty graph. In the intrinsic graph, each sample is connected to its k NNs within the same class, whereas in the penalty graph, each sample is connected to its k' NNs in different classes.

TR problem, causing it to be faster than the ITR algorithm [25], [44]. The basic steps of this algorithm are listed in Table II. For more details about the convergent analysis of ITR-Score algorithm, please refer to our preliminary work in [25] and [44].

B. Application for Non-Gaussian Data Sets

The above algorithm was developed with the assumption that the samples in each class follow a Gaussian distribution. However, in many fault diagnosis problems, samples in a data set may follow a non-Gaussian distribution that cannot satisfy the above assumption. Without this assumption, the separability of different classes may not be well characterized by the scatter matrices, causing the classification results to be degraded [34]. To solve this problem, some researchers have developed new scatter matrices of S_w and S_b to characterize the intraclass compactness and interclass separability [36], [37]. In this paper, inspired by the work in [36], S_w and S_b are constructed using the particularly designed intrinsic graph, i.e., W , and penalty graph, i.e., W^+ (see Fig. 1 for details). Specifically, let $N_k(x_i)$ be the set of k NNs of x_i in the same class, and $N_{k'}^+(x_i)$ be the set of k' NNs of x_i in different classes. The intrinsic graph W and penalty graph W^+ can be defined by

$$W_{ij} = \begin{cases} 1 & \text{if } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i) \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

$$W_{ij}^+ = \begin{cases} 1 & \text{if } x_i \in N_{k'}^+(x_j) \text{ or } x_j \in N_{k'}^+(x_i) \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

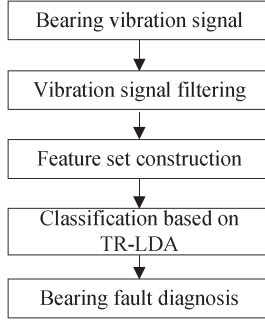


Fig. 2. TR-LDA-based fault diagnosis.

The intraclass compactness S_c from the intrinsic graph and interclass separability S_p from the penalty graph can then be characterized by

$$\begin{aligned}
 S_c &= \sum_{i,j: x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i)} \|W^T x_i - W^T x_j\|^2 \\
 &= 2Tr(W^T X(D - W)X^T W) \\
 &= 2Tr(W^T S'_w W) \quad (9)
 \end{aligned}$$

$$\begin{aligned}
 S_p &= \sum_{i,j: x_i \in N_{k'}^+(x_j) \text{ or } x_j \in N_{k'}^+(x_i)} \|W^T x_i - W^T x_j\|^2 \\
 &= 2Tr(W^T X(D^+ - W^+)X^T W) \\
 &= 2Tr(W^T S'_b W) \quad (10)
 \end{aligned}$$

where D and D^+ are two diagonal matrices satisfying $D_{ii} = \sum_j W_{ij}$ and $D_{ii}^+ = \sum_j W_{ij}^+$, and $S'_w = X(D - W)X^T$ and $S'_b = X(D^+ - W^+)X^T$ are the newly designed scatter matrices. The above strategy for constructing the scatter matrices can avoid the drawback that samples in each class must follow a Gaussian distribution [36]. Thus, for a non-Gaussian data set, it is better to represent the separability of different classes than the intraclass covariance, such as in LDA. In this paper, we calculate the scatter matrices in the proposed TR-LDA method using the above strategy. Hence, the objective function of TR-LDA for a non-Gaussian data set can be rewritten by notation substitutions in (6), i.e., $S_b \rightarrow S'_b$ and $S_w \rightarrow S'_w$. Thus,

$$\widehat{W} = \arg \max_{W^T W = I} \frac{Tr(W^T S'_b W)}{Tr(W^T S'_w W)}. \quad (11)$$

III. VIBRATION SIGNAL PREPROCESSING AND FEATURE SET CONSTRUCTION

For fault diagnosis, it is necessary to separate all the possible fault types and the normal condition. The TR-LDA, which maximizes the between-class measure while minimizing the within-class measure, was used for fault diagnosis of the motor bearings, as illustrated in Fig. 2. A low-pass filter filtered the vibration signals first. Then, the filtered vibration signals were divided into sections of equal window lengths, as shown in Fig. 3. A set of features was then constructed from each window to represent the characteristics of the vibration signals.

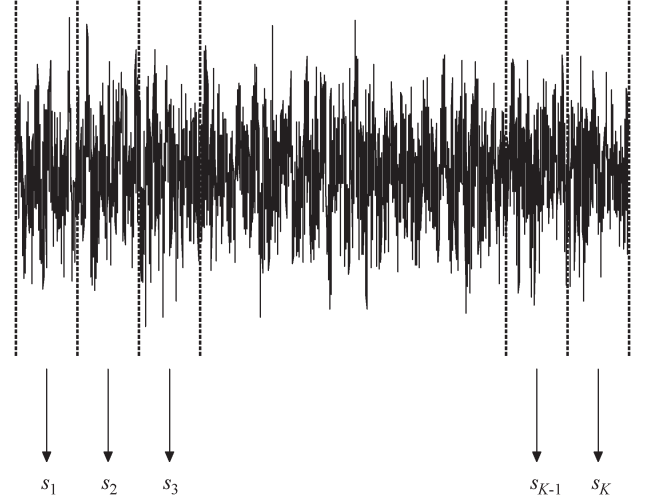


Fig. 3. Division of signal with equal window lengths.

TABLE III
NINE TIME-DOMAIN STATISTICAL PARAMETERS

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

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 $x_{rms} = \sqrt{\sum_{i=1}^N x(i)^2}/N$ are the mean value and root-mean-square value of the digital signal series respectively.

TABLE IV
MOTOR BEARING DATA DESCRIPTIONS

Data set	#Samples	#Dim	#Class	#Samples per class
Single-point faults	800	15	10	80
Generalized-roughness faults	600	15	6	100

When faults occur in bearings, the increased friction and impulsive forces cause both the time-domain parameters and the time-frequency-domain parameters to vary from the normal

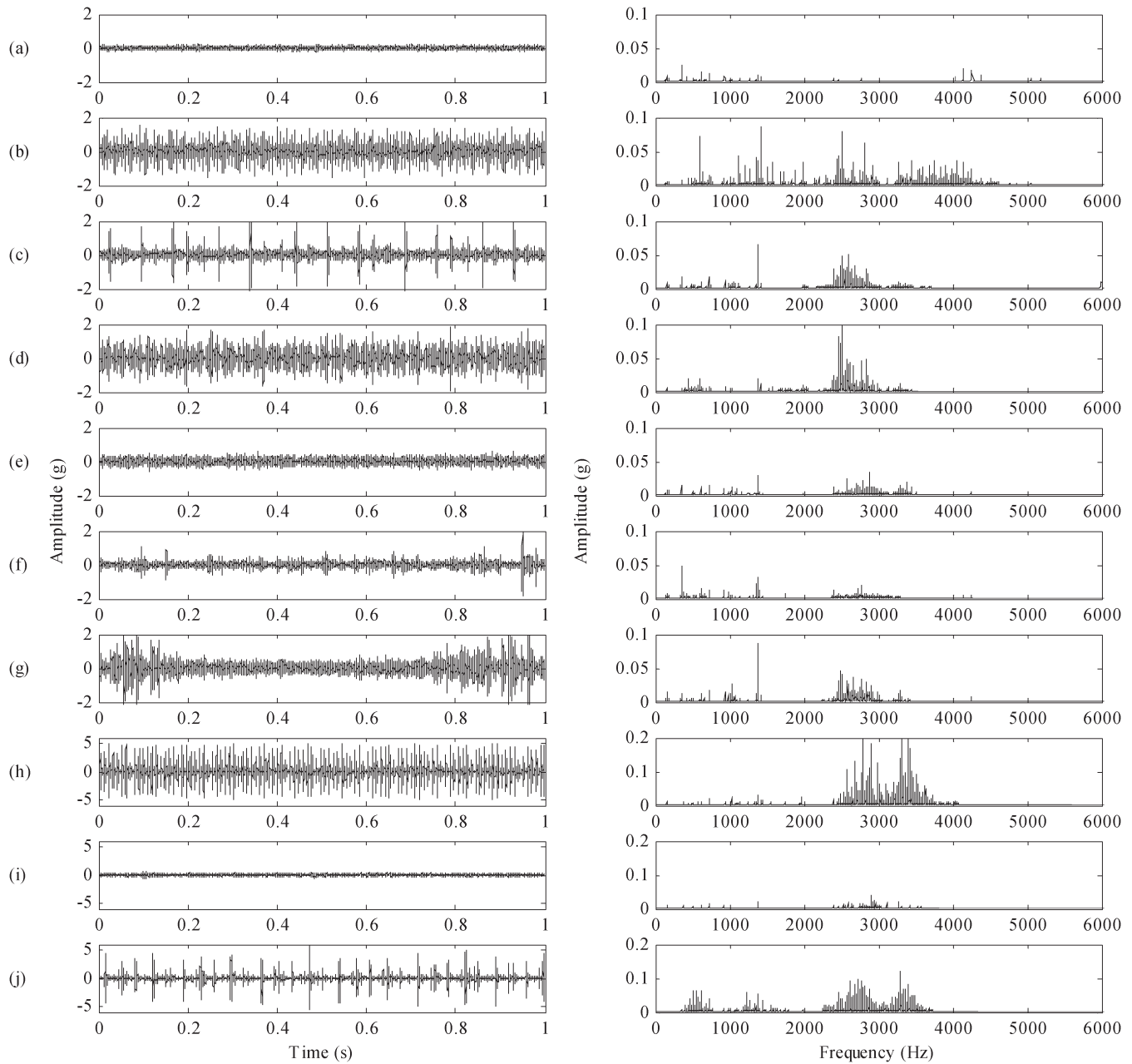


Fig. 4. (Left) Raw vibration signals and (right) their corresponding spectra. (a) Normal. (b) 0.007-in IRF. (c) 0.014-in IRF. (d) 0.021-in IRF. (e) 0.007-in BF. (f) 0.014-in BF. (g) 0.021-in BF. (h) 0.007-in ORF. (i) 0.014-in ORF. (j) 0.021-in ORF.

ones. Normal bearings and bearings with different faults usually have different data distributions in their signals. In this paper, nine time-domain statistical parameters, as described in Table III, are considered [27], [30], [45]. More specifically, such as peak reflects the maximum vibration amplitude in the time domain, standard deviation is a measure of how spread out the amplitudes are, skewness is a asymmetry indicator, kurtosis reflects the sharpness of the peak, and crest factor defines the ratio of peak value to the RMS value. All of these nine parameters reflect the characteristics of time series data in the time domain. Furthermore, six parameters about the percentage of energy corresponding to wavelet coefficients are also considered by decomposing the vibration signal by a wavelet transform using “db4” at level 5 [46], since they

reflect the vibration energy distribution in the time-frequency domain [47]. Thus, nine time-domain statistical parameters together with six time–frequency domain parameters are used to represent each window’s vibration signals. This 15-feature data set constructed from the bearing vibration signals is the input of the proposed method. It should be noted that our proposed method is not limited by these 15 features; other features, such as PMM [30], spectrum peak ratio [47], and mean frequency [47], can be also used. Detailed discussions of which types of features are more useful than others for bearing fault diagnosis are beyond the scope of this paper. Based on the experimental motor bearing data, the analysis results in Section IV are drawn to verify the effectiveness of these 15 features for bearing fault diagnosis.

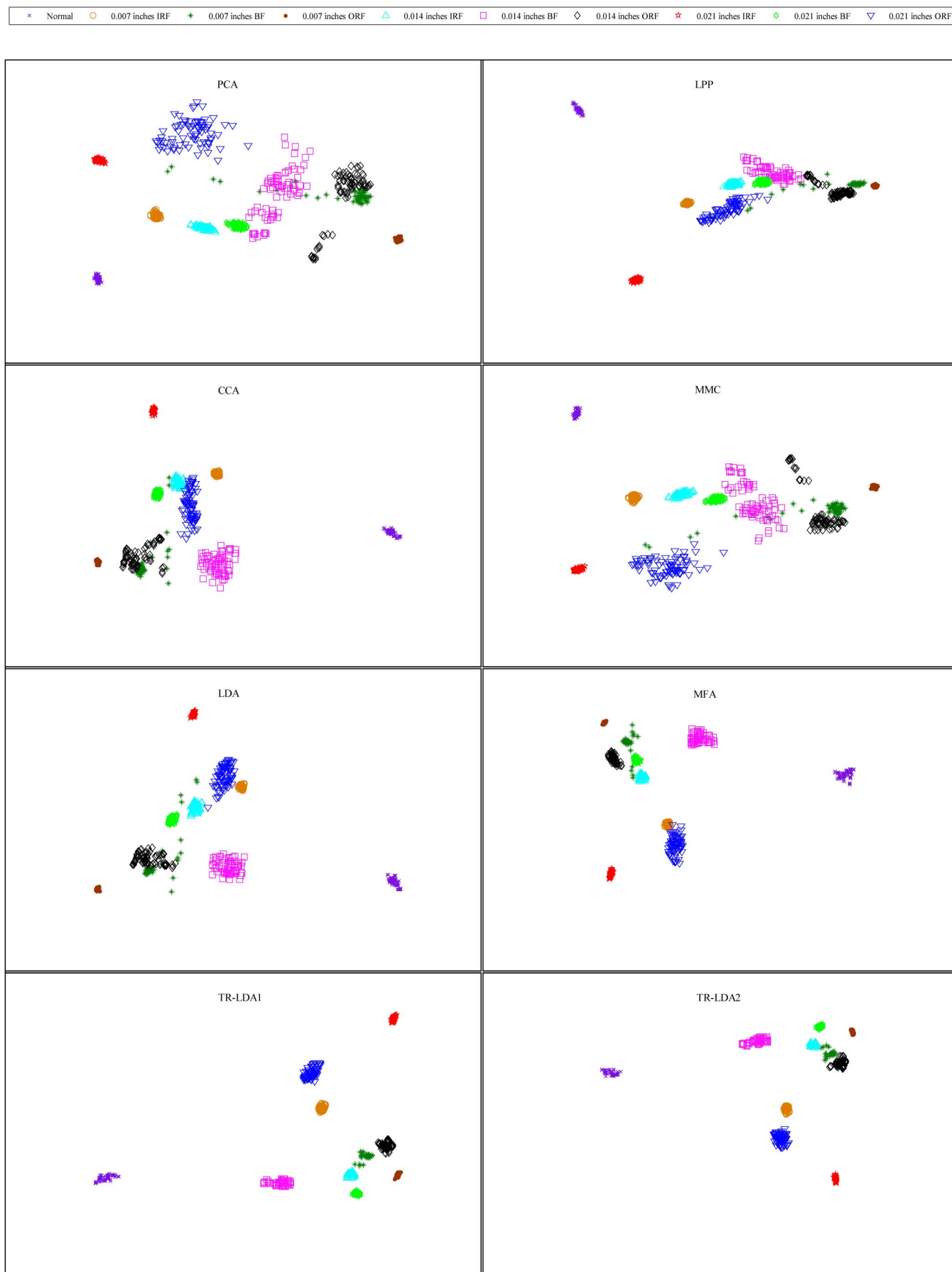


Fig. 5. Two-dimensional representation of single-point fault bearing data by each method.

IV. EXPERIMENT SETUP AND FAULT DIAGNOSIS

To validate the effectiveness of the proposed approach for fault diagnosis, both bearing data with single-point faults and bearing data with generalized-roughness faults were studied. These two types of bearing fault data were each measured from a different type of motor bearing under unique experimental conditions. The details of the two data sets are shown in Table IV. The visualization and classification of bearing fault data using TR-LDA are reported and compared with other methods, including unsupervised methods, such as PCA and LPP [26], and supervised methods, such as CCA [48], MMC [49], LDA, and MFA [36].

A. Single-Point Faults

Motor bearing data with single-point faults obtained from Case Western Reserve University [50] are analyzed in this paper. Experimental data were collected from the drive end ball bearing of an induction motor (2 hp Reliance Electric Motor) driven test rig. The test bearings were SKF 6205 JEM, a type of deep groove ball bearing. The accelerometer was mounted on the drive end of the motor housing. Single-point faults were seeded into the drive end ball bearing using electrodischarge machining. Vibration signals were collected for four different conditions, namely: 1) normal; 2) ORF; 3) IRF; and 4) BF. For ORF, IRF, and BF cases, vibration signals for three different severity levels (0.007-, 0.014-, and 0.021-in diameter) were separately collected. All the experiments were done for one load condition (3 hp), whereas the rotation speed was 1730 r/min. Therefore, experimental data (see Fig. 4) consisted of one vibration signal for the normal condition and three vibration signals for the ORF, IRF, and BF conditions. The sampling rate was 48 kHz.

The cutoff frequency for the low-pass filter was set to 6 kHz. The equal window length was set to 12 000. Fig. 5 shows the 2-D visualization results of single-point fault bearing data by PCA, LPP, CCA, MMC, LDA, MFA, and TR-LDA, in which we chose all samples in the data set as the training set. Here, TR-LDA1 is to use the scatter matrices as LDA that are defined in (1) and (2), whereas TR-LDA2 is to use the scatter matrices as MFA defined in (9) and (10). From the results in Fig. 5, the unsupervised methods, PCA and LPP, show the different classes of bearing data, but the boundaries among different classes are heavily overlapped and unclear. Therefore, both PCA and LPP cannot preserve the discriminative properties of the data. On the other hand, by providing the discriminative information based on labeled data, the boundaries learned by CCA, MMC, LDA, MFA, and the proposed TR-LDA algorithm are clearer and less overlapped. That is to say, supervised methods are able to preserve more discriminative information embedded in bearing data than unsupervised methods. Fig. 5 also demonstrates that the proposed method, TR-LDA, outperforms the other supervised methods, MMC, CCA LDA, and MFA, so that the same class of bearing data is closely conglomerated, whereas those belonging to different classes are clearly separated. In the following simulation, we randomly chose 8, 16, 24, and 40 samples from each class as the training set and the remaining samples as the test set to evaluate the classification accuracy

TABLE V
CLASSIFICATION ACCURACY OF MOTOR BEARING DATA WITH SINGLE-POINT FAULTS (1-NN CLASSIFIER)

Method	Training size			
	10%	20%	30%	50%
PCA	89.00	92.55	94.45	96.75
LPP	89.00	92.70	94.50	96.85
CCA	91.35	94.25	96.00	98.95
MMC	90.00	93.80	95.85	98.55
LDA	91.35	94.25	96.00	98.95
OLDA	91.50	94.50	96.00	99.00
SLPP	90.00	93.50	95.00	97.55
MFA	91.35	94.55	96.00	99.00
TR-LDA1	92.00	95.05	97.10	100.00
TR-LDA2	92.50	95.35	97.40	100.00



Fig. 6. Fan and accelerometer.

for different methods. Table V shows the test accuracies for an average of 20 random trials. Here, we also include with another orthogonal variant of LDA, i.e., OLDA [51], and SLPP [26] for comparisons. From Table V, we can observe that TR-LDA achieves competitive performance compared with other methods. These results verify the effectiveness of the proposed method for diagnosing single-point faults in bearings.

B. Generalized-Roughness Faults

The motor bearing data with generalized-roughness faults analyzed in this paper were collected from a BLDC motor fan with no lubricant in the ball bearings. Jin *et al.* [4] conducted FMMEA on fans and concluded that lubricant degradation was the most common cause of bearing failure. Moreover, the bearing loads in fans are very light. Therefore, lubricant plays a vital role in the reliability of fan bearings. In this paper, the loss of lubricant was simulated by not adding any lubricant to the ball bearings. The test bearings were NSK 693 ZZ, a type of miniature ball bearing. These customized ball bearings without lubricant were used in a BLDC motor fan. The fan was from Sanyo Denki Co., Ltd., and its size was 80 mm × 80 mm × 25 mm. Although no lubricant was added to the ball bearings, all other bearing components (including the inner race, outer race, retainer, and balls) were in good condition with protecting oil. The protecting oil acted as a temporary lubricant before it evaporated during the experiment. In other words, the nonlubricated fan acted like a normal fan before its protecting

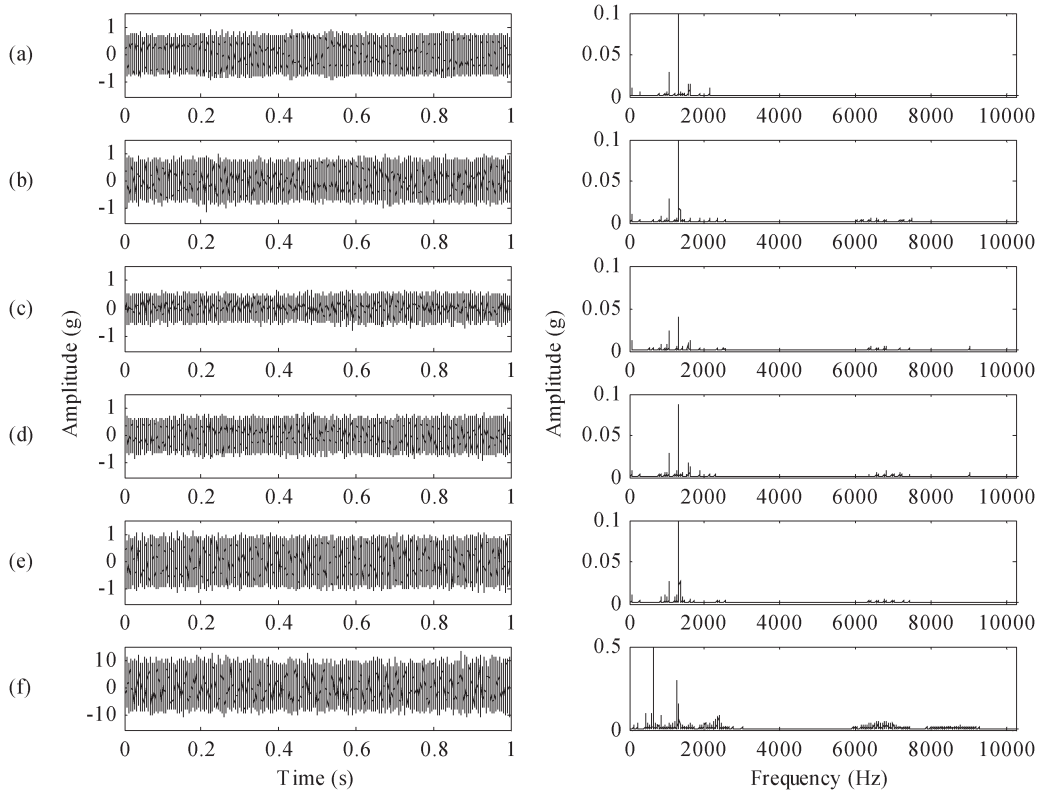


Fig. 7. (Left) Raw vibration signals and (right) their corresponding spectrums. (a) 0 h. (b) 8 h. (c) 16 h. (d) 24 h. (e) 48 h. (f) 72 h.

oil evaporated. Therefore, it was regarded as being in good condition before the stress test.

The high-temperature (70 °C) stress test was conducted on the tailor-made BLDC motor fan when it was running under free airflow conditions. The degradation of ball bearings in the BLDC motor fan was expected due to the loss of lubricant during the high-temperature stress test. The lack of lubricant increases friction and wear between the balls and races. Such friction and wear cause generalized-roughness faults in the ball bearing.

The fan was powered by a 12-V DC power supply and controlled by a pulsewidth modulation input signal (duty cycle was set at 74%) to simulate fan operation in actual application conditions. The rotation speed was 4000 r/min. An accelerometer was attached to the fan housing, and the fan was mounted on a test plenum for the acquisition of the vibration signal, as shown in Fig. 6. After the fan underwent 0, 8, 16, 24, 48, and 72 h of high-temperature stress tests, its vibration signal was measured, as shown in Fig. 7. When bearings degraded due to the lack of lubricant, the increased friction damaged the surfaces of the bearings and resulted in small pieces of metal dropping off. The dropped metal and the defective bearing surface were smoothly grinded by the continuous operation of the cooling fan bearings. Hence, the vibration signals could become weaker [see Fig. 7(c)]. However, as the defects grew, the vibration became stronger. National Instruments' LabVIEW was used for the data collection. The data sampling rate was 102.4 kHz. The high-temperature stress test was stopped when the fan's sound pressure level increased 3 dBA from the initial value, which is one of the failure criteria defined in IPC-9591 [52]. Fig. 8 shows a failed bearing after it underwent 72 h of the



Fig. 8. (a) Failed bearing. (b) Close up of (a). Arrows indicate tiny metal particles detached from bearing surfaces.

high-temperature stress test; tiny metal particles detached from the bearing surfaces were observed.

The cutoff frequency for the low-pass filter was set to 10.24 kHz. The equal window length was set to 20480. Fig. 9 shows the 2-D visualization results for the generalized-roughness fault bearing data using different methods. Here, we also choose all samples in the data set as the training set. From the results, the following phenomena can be observed.

- 1) The supervised methods, namely, MMC, CCA, LDA, MFA, and TR-LDA, outperformed the unsupervised methods, namely, PCA and LPP, as they preserved more discriminative information embedded in the bearing data.
- 2) Compared with the other supervised methods—MMC, CCA, LDA, and MFA—TR-LDA2 can achieve the best performance in a way that the submanifolds of different classes are more separated and less overlapped, and the submanifolds in each class are smoothly preserved.

For other supervised methods, there are some overlapped parts among the boundaries of different classes. The main

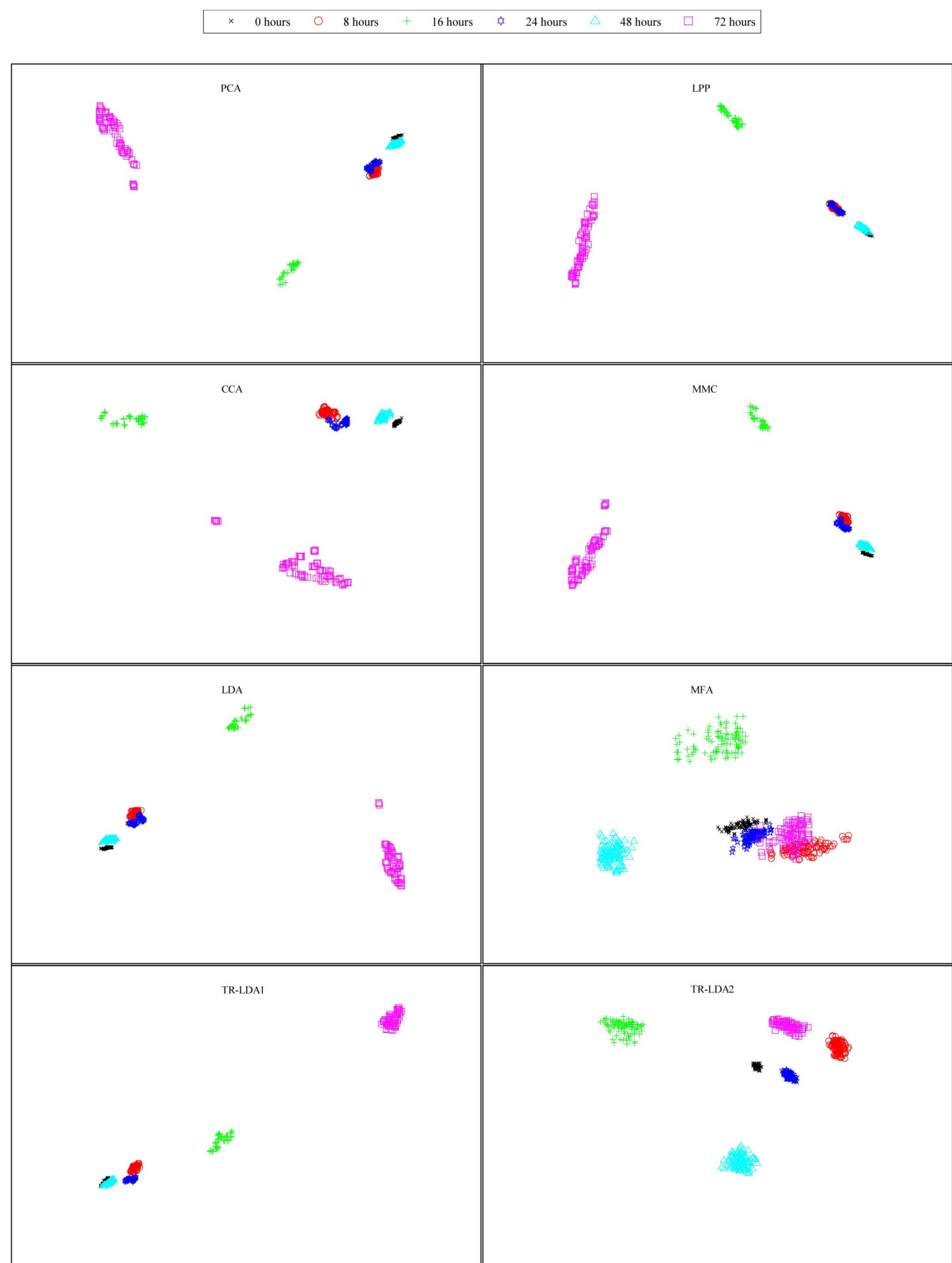


Fig. 9. Two-dimensional representation of generalized-roughness fault bearing data by each method.

TABLE VI
CLASSIFICATION ACCURACY OF MOTOR BEARING DATA WITH
GENERALIZED-ROUGHNESS FAULTS (1-NN CLASSIFIER)

Method	Training size			
	10%	20%	30%	50%
PCA	96.80	96.90	97.45	97.90
LPP	96.85	97.00	97.50	98.00
CCA	97.80	98.50	99.00	99.45
MMC	97.70	98.55	98.90	99.55
LDA	97.80	98.50	99.00	99.45
OLDA	97.85	98.80	99.00	99.45
SLPP	97.50	97.95	98.50	99.00
MFA	97.85	98.80	99.00	99.45
TR-LDA1	98.85	100.00	100.00	100.00
TR-LDA2	99.00	100.00	100.00	100.00

reason is that TR-LDA2 is based on TR criterion, which can directly reflect the distances between data samples of inter- and intraclass. It also includes more local discriminative information by using the local scatter matrices, which is good for classification. Table VI shows the classification performance using the different methods. In the following simulation, we randomly choose 10, 20, 30, and 50 samples from each class as the training set and the remaining samples as the test set to evaluate the classification accuracy for different methods. Table VI shows the test accuracies for an average of 20 random trials. From Table VI, we can see that TR-LDA achieved a better performance than the other methods. These results verify the effectiveness of the proposed method for diagnosing generalized-roughness faults in bearings.

V. CONCLUSION

In this paper, TR-LDA, a new efficient dimensionality reduction method, has been used and has been extended to deal with the non-Gaussian data sets confronted in the real-world fault diagnosis. Different from other dimensionality reduction methods, TR-LDA can directly reflect Euclidean distances between data points of inter- and intraclass. Hence, the compactness within each class can be minimized, whereas the separability of different classes can be maximized, which is suitable for fault diagnosis. In addition, motor bearing data, including single-point fault bearing data and generalized-roughness fault bearing data, are used to validate the effectiveness of the approach. Results showing the 2-D visualization and classification performance of the data demonstrate that the extended TR-LDA method achieves better performance than other algorithms, such as PCA, LPP, MMC, CCA LDA, and MFA, in fault diagnosis.

REFERENCES

- [1] B. Li, M. Y. Chow, Y. Tipsuwan, and J. C. Hung, "Neural-network-based motor rolling bearing fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 47, no. 5, pp. 1060–1068, Oct. 2000.
- [2] A. Bellini, F. Filippetti, C. Tassoni, and G. A. Capolino, "Advances in diagnostic techniques for induction machines," *IEEE Trans. Ind. Electron.*, vol. 55, no. 12, pp. 4109–4126, Dec. 2008.
- [3] S. Nandi, H. A. Toliyat, and X. Li, "Condition monitoring and fault diagnosis of electrical motors—A review," *IEEE Trans. Energy Convers.*, vol. 20, no. 4, pp. 719–729, Dec. 2005.
- [4] X. Jin, M. H. Azarian, C. Lau, L. L. Cheng, and M. Pecht, "Physics-of-failure analysis of cooling fans," in *Proc. Prognost. Syst. Health Manage. Conf.*, Shenzhen, China, 2011, pp. 1–8.
- [5] W. Wang and M. Pecht, "Economic analysis of canary-based prognostics and health management," *IEEE Trans. Ind. Electron.*, vol. 58, no. 7, pp. 3077–3089, Jul. 2011.
- [6] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—A tutorial," *Mech. Syst. Signal Process.*, vol. 25, no. 2, pp. 485–520, Feb. 2011.
- [7] T. W. S. Chow and S. Hai, "Induction machine fault diagnostic analysis with wavelet technique," *IEEE Trans. Ind. Electron.*, vol. 51, no. 3, pp. 558–565, Jun. 2004.
- [8] W. Zhou, T. Habetler, and R. Harley, "Bearing fault detection via stator current noise cancellation and statistical control," *IEEE Trans. Ind. Electron.*, vol. 55, no. 12, pp. 4260–4269, Dec. 2008.
- [9] J. Cusido, L. Romeral, J. A. Ortega, J. A. Rosero, and A. Garcia Espinosa, "Fault detection in induction machines using power spectral density in wavelet decomposition," *IEEE Trans. Ind. Electron.*, vol. 55, no. 2, pp. 633–643, Feb. 2008.
- [10] F. Immovilli, M. Cocconcelli, A. Bellini, and R. Rubini, "Detection of generalized-roughness bearing fault by spectral-kurtosis energy of vibration or current signals," *IEEE Trans. Ind. Electron.*, vol. 56, no. 11, pp. 4710–4717, Nov. 2009.
- [11] J. R. Stack, R. G. Harley, and T. G. Habetler, "An amplitude modulation detector for fault diagnosis in rolling element bearings," *IEEE Trans. Ind. Electron.*, vol. 51, no. 5, pp. 1097–1102, Oct. 2004.
- [12] T. W. S. Chow and H. Z. Tan, "HOS-based nonparametric and parametric methodologies for machine fault detection," *IEEE Trans. Ind. Electron.*, vol. 47, no. 5, pp. 1051–1059, Oct. 2000.
- [13] H. R. Martin and F. Honarvar, "Application of statistical moments to bearing failure detection," *Appl. Acoust.*, vol. 44, no. 1, pp. 67–77, Jan. 1995.
- [14] R. B. W. Heng and M. J. M. Nor, "Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition," *Appl. Acoust.*, vol. 53, no. 1–3, pp. 211–226, Jan. 1998.
- [15] J. Courrech, "Envelope analysis for effective rolling-element fault detection—fact or fiction?" *UPTIME Mag.*, vol. 8, no. 1, pp. 14–17, 2000.
- [16] A. Ibrahim, M. El Badaoui, F. Guillet, and F. Bonnardot, "A new bearing fault detection method in induction machine based on instantaneous power factor," *IEEE Trans. Ind. Electron.*, vol. 55, no. 12, pp. 4252–4259, Dec. 2008.
- [17] P. W. Tse, Y. H. Peng, and R. Yam, "Wavelet analysis and envelope detection for rolling element bearing fault diagnosis—their effectiveness and flexibilities," *J. Vib. Acoust.*, vol. 123, no. 3, pp. 303–310, Mar. 2001.
- [18] 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 Trans. Instrum. Meas.*, vol. 61, no. 8, pp. 2222–2229, Aug. 2012.
- [19] X. Jin and T. W. S. Chow, "Anomaly detection of cooling fan and fault classification of induction motor using Mahalanobis-Taguchi system," *Exp. Syst. Appl.*, vol. 40, no. 15, pp. 5787–5795, Nov. 2013.
- [20] 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, Dec. 2000.
- [21] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
- [22] P. Demartines and J. Hérault, "Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 148–154, Jan. 1997.
- [23] M. Delgado, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, "Bearing faults detection by a novel condition monitoring scheme based on statistical-time features and neural networks," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3398–3407, Aug. 2013.
- [24] N. Kambhatla and T. K. Leen, "Dimension reduction by local principal component analysis," *Neural Comput.*, vol. 9, no. 7, pp. 1493–1516, Oct. 1997.
- [25] M. Zhao, Z. Zhang, and T. W. S. Chow, "Trace ratio criterion based generalized discriminative learning for semi-supervised dimensionality reduction," *Pattern Recognit.*, vol. 45, no. 4, pp. 1482–1499, Apr. 2012.
- [26] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

- [27] Q. Jiang, M. Jia, J. Hu, and F. Xu, "Machinery fault diagnosis using supervised manifold learning," *Mech. Syst. Signal Process.*, vol. 23, no. 7, pp. 2301–2311, Oct. 2009.
- [28] 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 Trans. Ind. Electron.*, vol. 55, no. 12, pp. 4191–4199, Dec. 2008.
- [29] B. Ayhan, M. Y. Chow, and M. H. Song, "Multiple discriminant analysis and neural-network-based monolith and partition fault-detection schemes for broken rotor bar in induction motors," *IEEE Trans. Ind. Electron.*, vol. 53, no. 4, pp. 1298–1308, Jun. 2006.
- [30] J. Yu, "Local and nonlocal preserving projection for bearing defect classification and performance assessment," *IEEE Trans. Ind. Electron.*, vol. 59, no. 5, pp. 2363–2376, May 2012.
- [31] Y. Wang, Q. Miao, E. W. M. Ma, K.-L. Tsui, and M. G. Pecht, "Online anomaly detection for hard disk drives based on Mahalanobis distance," *IEEE Trans. Rel.*, vol. 62, no. 1, pp. 136–145, Mar. 2013.
- [32] Y. Wang, E. W. M. Ma, T. W. S. Chow, and K. Tsui, "A two-step parametric method for failure prediction in hard disk drives," *IEEE Trans. Ind. Informat.*, to be published, DOI: 10.1109/TII.2013.2264060.
- [33] A. S. S. Vasan, B. Long, and M. Pecht, "Diagnostics and prognostics method for analog electronic circuits," *IEEE Trans. Ind. Electron.*, vol. 60, no. 11, pp. 5277–5291, Nov. 2013.
- [34] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York, NY, USA: Academic, 1990.
- [35] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, Sep. 1936.
- [36] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.
- [37] H. Wang, S. Yan, D. Xu, X. Tang, and T. Huang, "Trace ratio vs. ratio trace for dimensionality reduction," in *Proc. IEEE CVPR*, 2007, pp. 1–8.
- [38] Y. Guo, S. Li, J. Yang, T. Shu, and L. Wu, "A generalized Foley-Sammon transform based on generalized Fisher discriminant criterion and its application to face recognition," *Pattern Recognit. Lett.*, vol. 24, no. 1–3, pp. 147–158, Jan. 2003.
- [39] D. Cai, X. He, K. Zhou, J. Han, and H. Bao, "Locality sensitive discriminant analysis," in *Prof. IJCAI*, Hyderabad, India, 2007, pp. 708–713.
- [40] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, pp. 1027–1061, May 2007.
- [41] Z. Zhang and T. W. S. Chow, "Tensor locally linear discriminative analysis," *IEEE Signal Process. Lett.*, vol. 18, no. 11, pp. 643–646, Nov. 2011.
- [42] Z. Zhang and T. W. S. Chow, "Robust linear optimized discriminant analysis," *Neurocomputing*, vol. 79, pp. 140–157, Mar. 2012.
- [43] Y. Jia, F. Nie, and C. Zhang, "Trace ratio problem revisited," *IEEE Trans. Neural Netw.*, vol. 20, no. 4, pp. 729–735, Apr. 2009.
- [44] M. Zhao, Z. Zhang, and T. W. S. Chow, "ITR-Score algorithm: An efficient trace ratio criterion based algorithm for supervised dimensionality reduction," in *Proc. IJCNN*, 2011, pp. 145–152.
- [45] Y. Lei, Z. He, Y. Zi, and X. Chen, "New clustering algorithm-based fault diagnosis using compensation distance evaluation technique," *Mech. Syst. Signal Process.*, vol. 22, no. 2, pp. 419–435, Feb. 2008.
- [46] J. Daubechies, "Orthonormal bases of compact supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, no. 7, pp. 909–996, Oct. 1988.
- [47] R. Yan and R. X. Gao, "Energy-based feature extraction for defect diagnosis in rotary machines," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 9, pp. 3130–3139, Sep. 2009.
- [48] D. R. Hardoon, S. Szedmak, and J. S. Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Comput.*, vol. 16, no. 12, pp. 2639–2664, Dec. 2004.
- [49] H. Li, T. Jiang, and K. Zhang, "Efficient and robust feature extraction by maximum margin criterion," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 157–165, Jan. 2006.
- [50] K. A. Loparo, Bearings vibration data set, Cleveland, OH, USA. [Online]. Available: <http://www.eecs.cwru.edu/laboratory/bearing/download.htm>
- [51] J. Ye, "Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems," *J. Mach. Learn. Res.*, vol. 6, no. 4, pp. 483–502, Apr. 2005.
- [52] *Performance Parameters (Mechanical, Electrical, Environmental and Quality/Reliability) for Air Moving Devices, First Edition*, IPC-9591, Apr. 2006.



management.



machine learning, pattern recognition, and fault diagnosis.

Prof. Chow was the recipient of the Best Paper Award at the 2002 IEEE Industrial Electronics Society Annual meeting in Seville, Spain.



has authored more than 20 books on electronic product development, use, and supply chain management and over 500 technical articles. He consults for 22 major international electronics companies, providing expertise in strategic planning, design, test, intellectual property, and risk assessment of electronic products and systems.

Xiaohang Jin received the B.Sc. degree from Zhejiang University of Technology, Hangzhou, China, and the M.Sc. degree from Shanghai Jiao Tong University, Shanghai, China. He is currently working toward the Ph.D. degree in the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong.

He was a Certified Reliability Engineer with the American Society for Quality. His research interests include reliability evaluation and prediction, fault diagnosis, and prognostics and system health

Mingbo Zhao received the B.Sc. and M.S. degrees from the Department of Electronic Engineering, Shanxi University, Shanxi, China, in 2005 and 2008, respectively, and the Ph.D. degree in the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong, in 2013.

He is currently a Research Staff member with the Department of Electronic Engineering, City University of Hong Kong. His research interests include machine learning, data mining, and pattern recognition and its applications.

Tommy W. S. Chow (M'93–SM'03) received the B.Sc. degree (with honors) and the Ph.D. degree from the Department of Electrical and Electronic Engineering, University of Sunderland, Sunderland, U.K.

He is currently a Professor with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong. He has authored or coauthored over 200 technical articles related to his research, five book chapters, and one book. His main research has been in the area of neural networks,

Michael Pecht (M'83–SM'90–F'92) received the M.S. degree in electrical engineering and the M.S. and Ph.D. degrees in engineering mechanics from the University of Wisconsin-Madison, WI, USA.

He is the Founder of the Center for Advanced Life Cycle Engineering, University of Maryland, College Park, MD, USA, which is funded by over 150 of the world's leading electronics companies at more than U.S. \$6 million/year. He is also a Chair Professor of mechanical engineering and a Professor of applied mathematics with the University of Maryland. He