Bearing defect classification based on individual trained wavelet kernel local fisher discriminant analysis with particle swarm optimization

Published in:
IEEE Transactions on Industrial Informatics

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
Copyright 2015 IEEE. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.
Bearing Defect Classification based on Individual Wavelet Local Fisher Discriminant Analysis with Particle Swarm Optimization

Mien Van, Hee-Jun Kang

Abstract—In order to enhance the performance of bearing defect classification, feature extraction and dimensionality reduction have become important. In order to extract the effective features, wavelet kernel local fisher discriminant analysis (WKLFDA) is first proposed; herein, a new wavelet kernel function is proposed to construct the kernel function of LFDA. In order to automatically select the parameters of WKLFDA, a particle swarm optimization (PSO) algorithm is employed, yielding a new PSO-WKLFDA. When compared to the other state-of-the-art methods, the proposed PSO-WKLFDA yields better performance. However, the use of a single global transformation of PSO-WKLFDA for the multiclass task does not provide excellent classification accuracy due to the fact that the projected data still significantly overlap with each other in the projected subspace. In order to enhance the performance of bearing defect classification, a novel method is then proposed by transforming the multiclass task into all possible binary classification tasks using a one-against-one (OAO) strategy. Then, individual PSO-WKLFDA (I-PSO-WKLFDA) is used for extracting effective features of each binary class. The extracted effective features of each binary class are inputted to a support vector machine (SVM) classifier. Finally, a decision fusion mechanism is employed to merge the classification results from each SVM classifier in order to identify the bearing condition. Simulation results using synthetic data and experimental results using different bearing fault types show that the proposed PSO-WKLFDA outperforms the other state-of-the-art methods, the proposed PSO-WKLFDA yields better performance. However, the use of a single global transformation of PSO-WKLFDA for the multiclass task does not provide excellent classification accuracy due to the fact that the projected data still significantly overlap with each other in the projected subspace. In order to enhance the performance of bearing defect classification, a novel method is then proposed by transforming the multiclass task into all possible binary classification tasks using a one-against-one (OAO) strategy. Then, individual PSO-WKLFDA (I-PSO-WKLFDA) is used for extracting effective features of each binary class. The extracted effective features of each binary class are inputted to a support vector machine (SVM) classifier. Finally, a decision fusion mechanism is employed to merge the classification results from each SVM classifier in order to identify the bearing condition. Simulation results using synthetic data and experimental results using different bearing fault types show that the proposed method is well suited and effective for bearing defect classification.

Index Terms—Local fisher discriminant analysis, bearing defect classification, wavelet kernel, feature extraction, dimensional reduction, pattern recognition.

I. INTRODUCTION

Rotating machines are widely used in the manufacturing industry. Bearings play a significant role in modern rotating machinery, and their carrying capacity and reliability are crucial for overall machine performance. Since bearings frequently fail and can consequently lead to catastrophic system failure, it is important to be able to automatically detect and rapidly and accurately diagnose the existence and severity of bearing faults.

During the past decades, signal analysis-based fault diagnosis methods have been widely used to identify multiple bearing defects. This tool extracts fault features and then identifies fault patterns. In this field, two major approaches have been widely developed for bearing fault diagnosis: acoustic signal analysis and vibration signal analysis [1]. Among these, vibration signal-based diagnosis has become the most popular monitoring technique because of its ease of measurement. However, the difficulty of defect detection and classification in bearings from vibration data is that the signature of a defective bearing presents a wide frequency band and can be overwhelmed by noise [2].

In literature, there are two approaches to diagnose the bearing faults. The first approach is based on identifying the bearing characteristic frequency (BCF) of vibration spectra [3]. This approach, however, is not effective when the signal is in heavy noise. It means that the absence of clear BCF due to noise should not be informed as a completely healthy condition of the bearing [10]. In the second approach, fault diagnosis can be regarded as a problem concerning pattern recognition, which mainly includes three important stages: feature calculation, feature extraction and dimensionality reduction, and pattern classification [4-5]. In order to acquire more fault information and improve the accuracy of the diagnosis system, many feature calculation methods have been developed based on three domains: time domain analysis, frequency domain analysis, and time-frequency domain analysis [6-7]. Recently, time-frequency analysis based on empirical mode decomposition (EMD) has been developed as an efficient method to extract bearing characteristics [8]. However, EMD is sensitive to measurement noise, which decreases the performance of feature extraction. To increase the performance of EMD for feature extraction, a hybrid nonlocal means (NLM) de-noising was developed in our previous work [9]. Based on three domains, many features can be calculated from a vibration signal to generate a rich feature set, which may contain many aspects to represent the bearing defect condition. In this way, the subjects of bearing defect identification can be transformed into a pattern recognition problem with high-dimensional data. However, dealing with high-dimensional data has always been a major problem in pattern recognition. High-dimensionality data processing...
suffers many difficulties in real applications such as need to massive computational resources and storage capacity. Meanwhile, the constructed features are often correlated, and not all of them are useful for the specific classification task [10]. Therefore, before employing classification tasks, a second step, dimensionality reduction (DR), should be employed to increase the performance of the classification system. There are two approaches for dimensional reduction: feature selection [11-12] and feature transformation [13-14]. Feature selection aims to find the most important features from a given set of features, while feature transformation aim to obtain low-dimensional feature representation with enhanced discrimination power. Conventional dimensionality-reduction techniques based on feature transformation include unsupervised approaches, such as principle component analysis (PCA), and supervised approaches, such as fisher discriminant analysis (FDA) [13]. Since FDA aims to identify projections with the most discriminant information, whereas PCA-based methods identify projections with minimal reconstruction errors, FDA-based methods generally perform better than PCA-based methods [13]. In [14], an orthogonal variant of FDA, which is called trace ratio linear discriminant analysis (TRLDA), has been developed in order to eliminate redundant information from the scatter matrices in LDA. However, the FDA and its variant TRLDA tend to produce undesired results if data samples in a class are multimodal. According to our experiment, measurement noise and feature extraction in different domain are the main sources generated multimodal of bearing vibrations signal. In [15], multimodal approach has been developed for gearbox fault diagnosis using deep learning. The idea of this approach is to classify the time domain, frequency domain, and time-frequency domain as separate modal feature representations, and then use the deep learning method to learn the modal feature representation. However, this approach also tends to produce undesired results if data samples in a class of a modal feature representation are multimodal. Another approach is to use local fisher discriminant analysis (LFDA) [16], which combines FDA and locality preserving projection (LPP). LFDA considers the local structure of the data samples so that the multimodal data can be appropriately embedded. However, LFDA is a linear method, which makes it difficult to use to describe a complex nonlinear system. To overcome this limitation, the kernel trick-based nonlinear LFDA extensions, called kernel LFDA (KLFDA), have attracted a great deal of attention. The basic idea is to nonlinearly map the input data from the input space to a higher dimensional feature space, where the complex distribution is expected to become linearly separable in the feature space, and then perform LFDA in the feature space. Linear, polynomial, and radial basis function (RBF) kernels are generally used in many applications, among which, RBF kernels are most commonly used [17].

Since the wavelet technique emerged as a powerful tool for nonlinear signal approximation [19], it has been employed as a kernel function (called a wavelet kernel) for support vector machine (SVM) classifiers [18, 20]. The experimental results shown that the wavelet kernel can provide a better approximation than does the use of the RBF kernel when incorporated into the SVM classifier. This is a valuable motivation for us to study the application of a wavelet kernel for LFDA. However, when applied to a wavelet kernel for LFDA, a problem arises in terms of selection of wavelet kernel parameters so that the KLFDA can generate the highest performance. In addition, in practical applications of LFDA and KLFDA, one of the most important issues is selection of the nearest neighbor parameter  

The main contributions of this paper are summarized as follows: 1) This paper claims that the conventional frequency domain approach fail to identify the bearing fault when the signal is immersed in heavy noise. 2) This paper shows that the data, which contains the features extracted from the bearing vibration signal, is a type of multimodal data. Thus, the multimodal dimensional reduction would be a good choice to enhance the classification accuracy.
3) A new multimodal dimensional reduction, namely I-
PSO-WKLFDA (OAA-PSO-WKLFDA and OAO-PSO-
WKLFDA), is proposed to enhance the diagnosis performance.
4) Experiments results for measured bearing vibration data
show that the proposed I-PSO-WKLFDA outperforms
conventional frequency domain approach, feature selection
approach, and other state-of-the-art multimodal DR methods.

The rest of this paper is organized as follows. Section II
describes the proposed WKLFD and PSO-WKLFDA
methods. The OAA and OAO strategies and the proposed I-
PSO-WKLFDA method are described in Section III. Section
IV presents the construction of a feature set based on time,
frequency, and time-frequency domains and the proposed
bearing defect identification. Section V provides experimental
data to verify the effectiveness of the proposed WKLFDA,
PSO-WKLFDA, and I-PSO-WKLFDA algorithms. Finally,
concluding remarks are given in Section VI.

II. WAVELET KERNEL LOCAL FISHER DISCRIMINANT ANALYSIS

A. Local Fisher Discriminant Analysis (LFDA)

LFDA [16] is a recent extension of FDA that can
effectively handle the multimodal problem. By combining the
properties of FDA and an unsupervised manifold technique,
I.e., Locality Preserving Projection (LPP), LFDA has the
ability to simultaneously preserve both between-class
separation and within-class local structure.

Consider a data set with a training sample with
p features
\{x_i^p\}_{i=1}^n, x_i \in \mathbb{R}^p
\text{ and class labels } \{y_i\}_{i=1}^c, y_i \in \{1, 2, ..., c\},
where c is the number of classes, and n is the total number of
training samples. Let n_i be the number of training samples
available for the ith class, and \sum n_i = n. We define
\begin{align}
A_{i,j} & = \left[ 1 \right] \\
\text{ as the “affinity” between } x_i \text{ and } x_j \text{ given by the } \\
A_{i,j} & = \exp \left( \frac{-\|x_i - x_j\|^2}{\gamma_{i,j}} \right) \\
\end{align}

where \gamma_i is the local scaling around x_i, defined by
\gamma_i = \left\| x_i - x_k^i \right\|, \text{ and } x_k^i \text{ is the k-th nearest neighbor of } x_i.

A_{i,j} \text{ is large if } x_i \text{ and } x_j \text{ are close to each other in the}
feature space, otherwise it is small. The parameter k-th is a
tuning factor and is a function of the embedding space. A
heuristic choice of k=7 has been shown to be effective [23].
However, this value is not applicable for general embedding
spaces. In this paper, we use PSO to effectively select this
parameter, as described later. In LFDA, the local between-
class \( S_{(lb)} \) and within-class \( S_{(lw)} \) scatter matrices are
respectively defined as
\begin{align}
S_{(lb)} & = \frac{1}{2} \sum_{i,j=1}^n W_{i,j}^{lb} \left( x_i - x_j \right) \left( x_i - x_j \right)^T \\
\text{where } W_{i,j}^{lb} \text{ and } W_{i,j}^{lw} \text{ are } n \times n \text{ matrices respectively defined as } \\
w_{i,j}^{lb} & = \begin{cases} \frac{1}{n} & \text{if } y_i = y_j = l \\
1/n & \text{if } y_i \neq y_j \end{cases} \\
w_{i,j}^{lw} & = \begin{cases} \frac{1}{n_i} & \text{if } y_i = y_j = l \\
0 & \text{if } y_i \neq y_j \end{cases}
\end{align}
The transformation matrix \( W_{LFDA} \) can then be computed by
maximizing the local Fisher’s ratio \( W^{T} S_{(lb)} W^{-1} W^{T} S_{(lw)} W \) as
\begin{align}
W_{LFDA} & = \underset{W}{\text{argmax}} \frac{W^{T} S_{(lb)} W}{W^{T} S_{(lw)} W} \\
The above optimization problem can be equivalently solved
by the generalized eigenvalue decomposition \( S_{(lb)} = \lambda S_{(lw)} \),
where \( \lambda \) is the generalized eigenvalue, and V denotes
the eigenvector that corresponds to a Fisher discriminant
direction. Assuming that the generalized eigenvectors
\( v_1 \geq v_2 \geq ... \geq v_p \) are arranged in descending order, the
local generalized eigenvectors \( v_1 \geq v_2 \geq ... \geq v_p \) are
the localized Fisher discriminant directions of decreasing class
severability.

B. Kernel Local Fisher Discriminant Analysis (KLFDA)

Since LFDA is a linear technique of dimensionality
reduction and feature extraction, it tends to provide inaccurate
results for complex nonlinear systems. In order to overcome
this limitation, KLFDA [17, 18] is presented in this section.
KLFDA has been shown to be a very effective feature
reduction algorithm in Reproducing Kernel Hilbert Space
(RKHS). The kernel trick for LFDA can be explained as follows.

Let \( S_{(lm)} \) be the local mixture scatter matrix defined by
\( S_{(lm)} = S_{(lb)} + S_{(lw)} \). From (2)-(5), \( S_{(lm)} \) can be expressed as
\begin{align}
S_{(lm)} & = \frac{1}{2} \sum_{i,j=1}^n W_{i,j} \left( x_i - x_j \right) \left( x_i - x_j \right)^T \\
\text{where } W_{i,j} \text{ is the n-dimensional matrix with the (i,j)-th } \\
element being \\
w_{i,j} & = \begin{cases} \frac{1}{n} & \text{if } y_i = y_j = l \\
1/n & \text{if } y_i \neq y_j \end{cases} \\
\end{align}

From (7), \( S_{(lm)} \) can be expressed as
\begin{align}
S_{(lm)} & = \frac{1}{2} \sum_{i,j=1}^n W_{i,j} \left( x_i x_i^T + x_j x_j^T - x_i x_j^T - x_j x_i^T \right) \\
& = \frac{1}{2} \sum_{i=1}^c \sum_{j=1}^c W_{i,j} \left( x_i x_i^T - \sum_{i,j=1}^c W_{i,j} x_i x_j^T \right) \\
\end{align}

which can be expressed in matrix form as
\( S_{(lm)} = XX^T \)
where $l^{lm} = D^{lm} - W^{lm}$, and $p^{lm}$ is the $n$-dimensional matrix with the $i$-th diagonal element of $D^{lm}_{i,j} = \sum_{j=1}^{n} W_{i,j}^{lm}$. Similarly, $s^{lb}$ can be expressed in the form $s^{lb} = XL^{lb}X^{T}$, where $l^{lb} = L^{lb} - W^{lb}$, and $p^{lb}$ is an $n$-dimensional matrix with the $i$-th diagonal element of $D^{lb}_{i,j} = \sum_{j=1}^{n} W_{i,j}^{lb}$. Thus, the eigenvector problem $s^{lb}v = \lambda s^{lb}v$ can be expressed as $XL^{lb}X^{T}V = \lambda XL^{lb}X^{T}V$ (11)

where $l^{lb} = l^{lm} - l^{lw}$. Since $X^{T}V$ is in the range of $X^{T}$, it can be expressed using some vector $a \in \mathbb{R}^{n}$ as $X^{T}V = XX^{T}a = Ka$

where $K$ is the $n$-dimensional matrix with the $(i,j)$-th element $K_{i,j} = x^{T}_{i}x_{j}$

Substituting this into (11) and multiplying the resulting equation by $X^{T}$, we obtain $KL^{lb}Ka = \lambda KL^{lb}Ka$ (12)

This implies that $(x_{i})^{n}_{i=1}$ appear only in terms of their inner products. Thus, we obtain a nonlinear variant of LFDA using the kernel trick.

We further define a nonlinear mapping $\Phi(\cdot)$ from input space $\mathbb{R}^{p}$ to a higher-dimensional RKHS $\Omega$ as follows $\Phi : \mathbb{R}^{p} \rightarrow \Omega, x \rightarrow \Phi(x)$ (13)

Then, a kernel function $K$ is defined as $K(x_{i}, x_{j}) = \langle \Phi(x_{i}), \Phi(x_{j}) \rangle$ (14)

where $\langle \cdot, \cdot \rangle$ is the inner product of two vectors.

Commonly used kernel functions include linear, polynomial, and RBF kernels. Among these, the RBF kernel is the most popular and is defined as

$$K(x_{i}, x_{j}) = \exp \left( -\frac{\|x_{i} - x_{j}\|^{2}}{2\sigma^{2}} \right)$$ (15)

where $\sigma$ is the dilation parameter, which is determined by users.

**C. Wavelet Kernel Local Fisher Discriminant Analysis (WKLFDA)**

The principle of wavelet analysis is to express or approximate a signal or function with a family of functions generated by dilations and translations of a mother wavelet

$$h_{a,c}(x) = \left| \frac{1}{\sqrt{a}} \right| h \left( \frac{x - c}{a} \right)$$ (16)

where $x, a, c \in \mathbb{R}$, $a$ is a dilation factor, $c$ is a translation factor, and $h(x)$ is the mother wavelet, which satisfies the following condition [19]

$$W_{h} = \int_{0}^{\infty} |F(\omega)|^{2} < \infty$$ (17)

where $F(\omega)$ is the Fourier transform of $h(x)$. The wavelet transform of a function, $g(x)$, can be expressed as

$$w_{a,c}(g) = \langle g(x), h_{a,c}(x) \rangle$$ (18)

On the right-hand side of (18), $\langle \cdot, \cdot \rangle$ denotes the dot product. Equation (18) represents the decomposition of a function, $g(x)$, on a wavelet basis, $h_{a,c}(x)$.

A wavelet function can be written in the following form:

$$h(x) = \sum_{i=1}^{n} h(x_{i})$$ (19)

where $x = [x_{1}, x_{2},...,x_{N}]^{T} \in \mathbb{R}^{N}$. Then, if $x, x' \in \mathbb{R}^{N}$, the dot-product wavelet kernels can be expressed as

$$K(x, x') = \sum_{i=1}^{n} h(x_{i} - x_{i}')\sum_{i=1}^{n} h(x_{i} - x_{i}')$$ (20)

and the translation-invariant wavelet kernels are expressed as

$$K(x, x') = \sum_{i=1}^{N} h\left( \frac{x_{i} - x_{i}'}{a} \right)$$ (21)

Equation (21) represents a multidimensional wavelet function.

Based on the wavelet function in [19], without loss of generality, we propose a Morlet wavelet function as the translation invariant wavelet kernel function

$$h(x) = \cos(\lambda x) \cdot \exp\left(-\frac{x^{2}}{2}\right)$$ (22)

Considering the mother wavelet defined in (22) and the dilation $a$, the wavelet kernel of this mother wavelet is

$$K(x, x') = \sum_{i=1}^{N} h\left( \frac{x_{i} - x_{i}'}{a} \right)$$ (23)

where $a$ is a parameter of the RBF kernel, and $\lambda$ is a new parameter that controls the kernel shape and must be suitably selected. It can be seen that when $\lambda = 0$, $K(x, x')$ represents a RBF kernel. When $\lambda = 1.5$, $K(x, x')$ approximates the Mexhat kernel in the range of $[-1,1]$.

**D. Particle Swarm Optimization (PSO) for Parameter Selection of WKLFDA (PSO-WKLFDA)**

The PSO algorithm was first developed by Eberhart and Kennedy in 1995. It is a powerful tool for dealing with global optimization problems [21]. PSO possesses several advantages compared with other heuristic optimization techniques such as simplicity, ease of implementation, robustness to control parameters, and computational efficiency.

In PSO, the population is referred to as a swarm, and the individuals are classed as particles. The $i$th particle is characterized by its current position vector, $x_{i}(t) = (x_{i1}(t), x_{i2}(t),...,x_{ID}(t))$, in the search space, where $D$ is the dimensionality of the search space, and the velocity vector is $v_{i}(t) = (v_{i1}(t), v_{i2}(t),...,v_{ID}(t))$. Each particle maintains a record of its personal best position, $p_{best,i}(t) = (p_{best,i1}(t), p_{best,i2}(t),...,p_{best,ID}(t))$, and the whole swarm of particles maintains a record of the global best
position, \( G_{\text{best}}(t) = (g_{\text{best},1}(t), g_{\text{best},2}(t), ..., g_{\text{best},d}(t)) \). Particles move in the search space in order to search for the optimal solution. During the movement, each particle updates its position and velocity according to the distance to its personal best position and the distance to the global best position with the following equations:

\[
\begin{align*}
\omega_{k}^{t+1} & = \omega_{k}^{t} - \omega_{\text{max}} - \omega_{\text{min}} \times \text{iter} \\
\dot{x}_{i}^{t+1} & = \omega_{k}^{t} + c_{1} \cdot r_{1} \cdot (p_{\text{best},i}^{t} - x_{i}^{t}) + c_{2} \cdot r_{2} \cdot (g_{\text{best},i}^{t} - x_{i}^{t}) \\
x_{i}^{t+1} & = x_{i}^{t} + \dot{x}_{i}^{t+1}
\end{align*}
\]

(24) (25) (26)

where \( \omega_{\text{min}} \) is the minimal inertia weight, and \( \omega_{\text{max}} \) is the maximal inertia weight. \( \text{iter} \) is the current iteration number, \( \text{iter}_{\text{max}} \) is the maximum iteration number. \( t \) represents the \( t \)th iteration in the evolutionary process, while \( d \in D \) represents the \( d \)th dimension in the search space. \( c_{1} \) and \( c_{2} \) are acceleration constants, and \( r_{1} \) and \( r_{2} \) are random values uniformly distributed in \([0, 1]\).

The WKLFDA feature extraction constructed using the wavelet kernel function defined in (23) has three determined parameters, a nearest neighbor parameter \( \lambda \)-th and two wavelet kernel parameters \( \lambda \) and \( a \). In this study, we employ a PSO-based SVM classifier; herein, the classification accuracy of an SVM classifier is used as a fitness function to automatically select the parameters of the WKLFDA feature extraction. The step-by-step implementation details are described below:

**Step 1:** Initialization  The upper and lower limits of the position and the velocity of the particles and parameters such as \( c_{1}, c_{2}, \text{iter}_{\text{max}}, \omega_{\text{max}}, \omega_{\text{min}} \) are initialized. Further, particle swarm with population \( N \) is initialized, as is the position and velocity of each particle \((k, \lambda, a - \text{parameters for WKLFDA})\).

**Step 2:** Evaluation of initial population  All the particles are evaluated for fitness based on a cost function. In this step, the following work is performed to obtain the fitness function. The training and validation data are projected onto the modeling WKLFDA subspace to obtain training and validation features. Then, the training and validation features are input into the SVM classifier to obtain classification accuracy. The classification accuracy is used as the fitness function and is defined as follows:

\[
\text{fitness function} = \frac{N_{T}}{N_{T} + N_{F}}
\]

(27)

where \( N_{T} \) and \( N_{F} \) denote the number of true and false classifications, respectively. A particle with the high classification accuracy value produces a high fitness value. The \( p_{\text{best}} \) of the individual particles and \( g_{\text{best}} \) of the population are identified.

**Step 3:** Updating  For each particle, update particle velocity according to Eq. (25) and particle position according to Eq. (26) in order to generate a new swarm.

**Step 4:** Evaluation of updated population  Evaluate the fitness value of the newly updated particles in a process similar to Step 2, and then update the \( p_{\text{best}}, \) and \( G_{\text{best}} \) of the swarm. For an individual particle, if the newly updated fitness value is greater than the historical local best value, the local best position, \( p_{\text{best},i}, \) will be replaced by the current position. For the swarm of particles, if the current fitness value is greater than the global best, the global best position, \( G_{\text{best}}, \) will be replaced by the current position.

**Step 5:** If the maximum number of iterations is not yet reached, return to Step 3. Otherwise, go to Step 6.

**Step 6:** Select the global best position, \( G_{\text{best}}, \) in the swarm as the ultimate solution. The value encoded from the global best position, \( G_{\text{best}}, \) is assigned as the optimal value for the parameters of the WKLFDA algorithm.

In order to increase the robustness, a \( k \)-fold CV method [24] is employed to compute the fitness function in Step 2. In the \( k \)-fold CV estimation procedure, the training data is randomly divided into \( k \) equal size subsamples. Of the \( k \) subsamples, the \( k-1 \) subsamples are used as the training data, and the remaining subsample is retained as the validation data for testing the model. We repeat this procedure until each of the subsamples is used as a validation set. The \( k \) results are then averaged to produce a single estimation. The averaged classification accuracy is used to gauge the fitness of the PSO algorithm. In this study, \( k \) is set to 5. This is a reasonable compromise considering the computational complexity and modeling robustness.

---

**Fig. 1 Structure diagram of individual PSO-WKLFDA.**
III. INDIVIDUAL PSO-WKLFD

A. One-Against-All (OAA) Strategy

The OAA strategy [22] has been applied widely for SVM classification, in which an \( m \) class classification problem is divided into \( m \) binary class classification problems. Each problem involves a binary classifier, which is responsible for distinguishing one of the classes from all other classes combined. Using a Winner-Takes-All strategy, each binary classifier is trained considering the samples from one of the classes as positive and the samples from all other classes as negative. The decision of the multiclassifier is chosen from the class whose binary classifier provides the greatest output. Formally, given a vector \( y = f(x) \) with the outputs of the binary classifiers, the multiclassifier generates a vector \( L = (l_1, ..., l_m) \) in the following way

\[
i^* = \arg \max_{i=1,...,m} f_j(x)
\]

\( L_i = \begin{cases} +1 & \text{if } i = i^*, i = 1, ..., m \\ -1 & \text{otherwise} \end{cases} \)

(28)

B. One-Against-One (OAO) Strategy

The OAO strategy divides an \( m \) class classification problem into \( m(m-1)/2 \) binary class classification problems. Each problem involves a binary classifier, which is responsible for distinguishing the samples of one class from the samples of another class. Each binary classifier is trained considering the samples from one of the classes as positive and the other class as negative. Formally, each classifier considers two information classes \( l \) and \( s \) \( (l, s \in \{1, 2, ..., C\}, l \neq s) \) via a decision function \( g_{l,s}(x) \). The global decision function is

\[
N_l(x) = \sum_{i=l}^{C} sgn(g_{l,i}(x))
\]

(29)

where \( N_l(x) \) is the number of times that class \( i \) is assigned for the testing vector \( x \). The final class label is assigned according to

\[
y^* = \arg \max_{i=1,...,m} N_l(x)
\]

(30)

C. Individual PSO-WKLFD

Although the proposed WKLFDA and PSO-WKLFD algorithms yield better performance than previous state-of-the-art approaches, the use of a single global transformation for the multiclass task does not provide excellent classification accuracy due to the fact that the projected data still overlap with each other in the projected subspace. In this paper, we consider the OAA and OAO strategies to reduce the overlapped data in the projected subspace. In the following, we present the proposed method based on OA frame work (the method based on OAO frame work can be executed in a similar way). Fig. 1 illustrates the proposed classification framework based on OAA strategy. In this way, the \( m \) class classification subject is first transformed into an \( m \) binary class classification subject. For each binary class, an individual PSO-WKLFD is used to extract the effective feature. In contrast to the conventional approach, which uses FDA, LFDA, or WKLFDA for the \( m \) class classification problem to obtain a single global transformation, this paper use \( m \) individual PSO-WKLFDAs for the \( m \) binary class classification problem in order to obtain \( m \) single transformations. The parameters of each WKLFDA are individually selected by PSO so that the corresponding binary classifier yields the best classification accuracy. The effective features extracted by each WKLFDAs are used as the input to

<table>
<thead>
<tr>
<th>Feature</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>( x_{std} = \frac{\sum_{i=1}^{N}(x(n) - \bar{x})^2}{N} )</td>
</tr>
<tr>
<td>Peak</td>
<td>( x_P = \max</td>
</tr>
<tr>
<td>Skewness</td>
<td>( x_{skew} = \frac{\sum_{i=1}^{N}(x(n) - \bar{x})^3}{(N-1)x_{std}} )</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>( x_{kurt} = \frac{\sum_{i=1}^{N}(x(n) - \bar{x})^4}{(N-1)x_{std}^4} )</td>
</tr>
<tr>
<td>Crest factor</td>
<td>( CF = \frac{x_P}{x_{rms}} )</td>
</tr>
<tr>
<td>Root mean square</td>
<td>( x_{rms} = \sqrt{\frac{\sum_{i=1}^{N}(x(n))^2}{N}} )</td>
</tr>
<tr>
<td>Clearance factor</td>
<td>( CLF = \frac{x_P}{\sqrt{\frac{\sum_{i=1}^{N}(x(n))^2}{N}}} )</td>
</tr>
<tr>
<td>Shape factor</td>
<td>( SF = \frac{x_{rms}}{\sqrt{\frac{\sum_{i=1}^{N}(x(n))^2}{N}}} )</td>
</tr>
<tr>
<td>Impulse factor</td>
<td>( IF = \frac{\sum_{i=1}^{N}(x(n))}{\sqrt{\frac{\sum_{i=1}^{N}(x(n))^2}{N}}} )</td>
</tr>
</tbody>
</table>

where \( x(n) \) is a signal series for \( n = 1, 2, ..., N \), and \( N \) is the number of data points.

IV. PROPOSED DIAGNOSIS METHODOLOGY

A. Feature Calculation

In order to measure the changes in vibration signal due to diverse bearing defects, a vibration sensor is generally attached to a non-rotating part of the machinery, (e.g., bearing housing). Although the bearing housing is the closest element on which to place a vibration sensor, the distance from the source of bearing failures causes the vibration signal to be overwhelmed by noise due to the effects of other components. Thus, the vibration signal should be effectively de-noised.
before being used for analysis. In this paper, we employed the hybrid NLM and EMD method, which had been effectively developed in our previous work [9], to extract the effective features. In this method, the raw vibration signal is first preprocessed using the NLM algorithm to eliminate or reduce the measurement noise; the de-noised signal is then decomposed using EMD to obtain a number of reliable intrinsic mode functions (IMFs). Based on the EMD decomposed using EMD to obtain a number of reliable intrinsic mode functions (IMFs). Based on the EMD algorithm, the de-noised signal, \( x(t) \), can be decomposed into a number of IMFs, \( C(t) \):

\[
x(t) = \sum_{j=1}^{n} C_j(t) + r_n(t)
\]

The first IMF, \( C_1(t) \), contains mostly high-frequency components. The IMFs \( C_2(t), C_3(t), \ldots, C_n(t) \) include different frequency bands ranging from high to low, and \( r_n(t) \) usually does not contain any signal information. Since the diverse bearing failures exists primarily in mid-and high frequency bands [12], we choose the first four IMFs as the most components for calculating fault signature because they represent the mid- and high-frequency components.

When a rolling bearing with different faults is used in the operation, different resonance frequency components are produced in the vibration signals. The energy of the fault vibration signal depends on the frequency band. To illustrate this change, IMF energy features are introduced in this paper.

For the first four IMF components, the total signal energy \( E_j(t) \) is expressed as \( E_j(t) \), so

\[
E_j = \int |C_j(t)|^2 dt
\]

To encompass the large value of \( E_j(t) \), the normalized energy of the IMF is defined as

\[
T_j = E_j / T
\]

where \( T \) is defined as

\[
T = \left( \sum_{j=1}^{36} |C_j(t)|^2 \right)^{1/2}
\]

When faults occur in rotating machinery, the vibration signals may change. The amplitude and distribution of the time-domain signals may be different from those of normal bearings. In this paper, the nine time-domain dimensionless parameters including standard deviation, peak, skewness, kurtosis, crest factor, root mean square, clearance factor, shape factor and impulsive factor, described in Table I are extracted.

\[
T = \left( \sum_{j=1}^{36} |C_j(t)|^2 \right)^{1/2}
\]

A. Support Vector Machine (SVM) Classifier

The SVM binary classifier locates a hyperplane between the two categories with the largest margin in the feature space. This hyperplane is used to classify test samples into one of the two categories.

Given a training set of instances and class label pairs \((x_i, y_i), \ for \ i = 1, \ldots, J\), where \( x_i \in \mathbb{R}^n \) and \( y_i \in \{1, -1\}^J \), the following minimization optimal problem can be solved to find the optimal hyperplane to separate the two categories:

\[
\min_{w, \xi, \xi^T} \frac{1}{2} w^T w + C \left( \sum_{i=1}^{J} \xi_i \right)
\]

subject to \( y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i \) and \( \xi_i \geq 0 \)

where \( w \) is a normal vector to the hyperplane, \( b \) is a constant such that \( \|w\| \) represents the Euclidean distance between the
hyperplane and the original of the feature space, \( \phi \) is a nonlinear function to map the original feature space into the high-dimensional nonlinear feature space. Parameter \( C > 0 \) is the penalty factor of the error term and may be seen as a factor that controls the tradeoff between separation margin and training errors, and \( \xi \) are slack variables that measure the degree of misclassification.

The minimization optimal problem defined in (35) can be written in dual form by applying Lagrange optimization as follows:

\[
\max \left\{ \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j \phi(x_i)^T \phi(x_j) \right\}
\]

subject to \( \sum_{i=1}^{l} \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, \forall i = 1, 2, \ldots, l, \)

where the \( \alpha_i \)'s are Lagrange multipliers, and \( x_i \) and \( x_j \) are any two different samples in the training dataset. Furthermore, \( K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \) represents the kernel function. Kernels, which are often selected based on the data structure and type of the boundaries between classes, can take many forms such as linear, polynomial, radial basis function, and sigmoid. In this paper, the RBF kernel is used:

\[
K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}
\]

where \( \gamma \) is an adjustable parameter to be carefully tuned.

C. Bearing Defect Classification based on Individual PSO-WKLFDA and SVM Classifier

Based on the above presented feature extraction method and proposed individual PSO-WKLFDA approach, a new bearing fault diagnosis approach is proposed to classify the data mixed with both normal and multiple types of faulty data. The bearing defect classification is based on NLM and EMD-based feature calculation, individual PSO-WKLFDA-based feature extraction and dimensionality reduction, and the SVM classifier. The proposed procedure is described in Fig. 2 and includes the following steps.

Step 1: Vibration signals acquired from the machine are preprocessed using NLM algorithm and EMD in order to obtain IMFs effectively.

Step 2: Extract the energy- and time-domain dimensionless feature parameters from the de-noised signal and the first four IMFs in order to obtain a combined feature set.

Step 3: Building data set and partition into a training data set and testing data set.

Step 4: Convert multiclass classification problem into a number of binary class classification problem based on OAA (or OAO) strategy.

Step 5: **Offline Training:** this step is done to obtain the optimal PSO-WKLFDA model for each binary class. The training data is used at this step. For each binary class, select the wavelet kernel function for WKLFDA feature extraction as in (23) and then select its parameters using the PSO- based SVM classifier in order to establish an optimal WKLFDA model for each binary class. After this step, we will obtain an \( m \) optimized PSO-WKLFDA model for corresponding \( m \) binary class, and then use it for the online fault diagnosis.

**Step 6:** **Online Fault Diagnosis:**

i) For each binary class, project the training and testing data sets onto the optimized WKLFDA to obtain testing features and training features for each binary class.

ii) The training and testing features are inputted into the SVM classifiers.

iii) Finally, a decision fusion mechanism is employed to merge the classification results from each SVM classifier to identify the bearing condition (labels).

V. EXPERIMENTAL RESULTS

A. Training and Test Data Configuration

To validate the proposed method for bearing defect recognition, measured bearing vibration data is used as an example. The vibration data used for analysis are taken from the Case Western Reserve University Bearing Data Center (2014) [25]. The test stand, shown in Fig. 3, consists of a 2hp motor (left), a torque transducer/encoder (center), a dynamometer (right), and control electronics (not show). Test bearings support the motor shaft. Single-point faults with diameters of 7, 14, and 21 mil were introduced in the test

![Fig. 3. Experimental setup for vibration monitoring in the Case Western Reserve University Bearing Data Center [25].](image-url)
bearings using electro-discharge machining (1 mil=0.001 inches). Vibration data is collected using accelerometers, which are attached to the housing with magnetic bases. Accelerometers were placed at the three o’clock position on the drive end of the motor housing. Vibration signals were collected using a 16-channel operating conditions: (1) normal condition, (2) outer race fault (ORF), (3) inner race fault (IRF), and (4) ball fault (BF). Each fault condition includes three different sizes, 0.007, 0.014, and 0.021 in. Hence, there are ten conditions (10 classes) that need to be identified in this experiment. All the experiments were done for one load condition (3 hp), where the rotation speed was 1730 r/min. The sampling rate was 12,000 Hz. The collected raw vibration signals were divided into sections of equal window lengths. Each window contains 1024 point. For each window, NLM and EMD are employed to obtain IMF components. Fig. 4a) shows the time domain of a window after NLM de-noising, and Fig. 4b) shows the de-noised signal IMF components after EMD decomposition. From this figure, we can see that the first four IMFs represent the mid- and high frequency components of the original signal. Hence, we used the first four IMFs for analysis as discussed above. Then, a set of features (49 features) as defined in Section III was then constructed from each de-noised signal and the first four IMFs of each window were used to represent the characteristic of the vibration signals. For each condition, 100 samples were used, and therefore the whole dataset corresponding to the ten signal conditions consists of 1000 samples.

B. Performance Evaluation

In order verify the four main contributions of this paper, as described in the introduction part, the experiment set is divided into four subsets: 1) the performance of the conventional frequency domain analysis is analyzed, 2) verifies that the bearing fault data is a type of multimodal data, 3) the performance of the proposed method is compared with other state-of-the-art multimodal dimensional reduction methods, and 4) we compare the performance of the proposed method with the used of lower dimensional data and feature selection methods.

Firstly, we analyze the performance of the conventional frequency domain analysis to see its performance. According to [9], after NLM and EMD decomposition, the first IMF is passed through the envelope to identify the BCFs, which can be used to detect and isolate the fault. First, we consider the inner race faults that are 0.007, 0.014, and 0.021 inches thick. The results of the FFT plots of the first IMF are shown in Fig. 5a), b), c), respectively. From these figures, we can see that the peaks corresponding to the inner race fault frequency can be easily identified in Fig. 5a) and Fig. 5c), but fail to identify in Fig. 5b). Similar analysis is executed for outer race faults of 0.007, 0.014, and 0.021 inches thick. The FFT plots are shown in Fig. 5d), e) and f), respectively. From the results, we can see that the characteristic frequency of the outer race fault and its harmonic are obvious in Figs. 5d) and 5f), but unseen in Fig. 5e). From the results, we can conclude that the conventional frequency domain analysis is sometime fail to extract BCFs for the noisy vibration signal.

Secondly, the distribution of the samples of the bearing data is analyzed. Fig. 6 shows the sample distribution of three example classes. It is obvious to see that the data, which contains the extracted features of the bearing fault signal, is a type of multimodal data. Thus, multimodal dimensional reduction technique would be a good choice to enhance the performance of the bearing fault diagnosis.

Thirdly, we compare the performance of the proposed methods with other state-of-the-art multimodal dimensional reduction methods. In this paper, we developed three multimodal dimensional reduction methods, WKLFDA, PSO-WKLFDA and I-PSO-WKLFDA (OAA-PSO-WKLFDA and OAO-PSO-WKLFDA), in which the I-PSO-WKLFDA algorithm is particularly concerned. To validate the effectiveness of the proposed feature extraction algorithms, we compared classification performance of the proposed method with seven other state-of-the-art feature analysis approaches, including global FDA, global TRLDA, global LFDA, global KLFDA (RBF kernel), global WKLFDA, global PSO-KLFDA (PSO is used to identify the sigma parameter of RBF kernel) and global PSO-WKLFDA algorithms. For LFDA, KLFDA and WKLFDA, the nearest neighbor parameter k(-th) were chosen as k(-th) = 7.

To evaluate the performance of the methods, the extracted feature vectors are used as input for the SVM classifier to obtain classification accuracy. The classification accuracy in this study is computed as follows:

\[
\text{Accuracy} = \frac{\text{Number of correctly classified samples}}{\text{Total number of samples}}
\]
\[ C_{\text{accuracy}} = \frac{\sum_{i} N_{TP}}{N_{\text{samples}}} \times 100(\%) \]  \hspace{1cm} (36)

where \( L \) is the number of classes (L=10) in this study, \( N_{TP} \) is the number of true positives (TP), defined as the total number of faults in class \( i \) that are correctly classified as class \( i \), and \( N_{\text{samples}} \) is the total number of samples used to evaluate the performance of the proposed bearing failure diagnosis.

To estimate the generalized classification accuracy, \( l \)-fold CV, described in section II.D, is also employed in this validation step. In the \( l \)-fold CV estimation procedure, the training data is randomly divided into \( l \) equal size folds. The classification performance is measured using one fold for training, and the remaining folds for testing. In this experiment, \( l \) is set to 4. This is a reasonable compromise considering the computational complexity and modeling robustness. More specifically, 100 samples of created feature vectors are divided into four mutual folds (each fold includes 25 randomly samples for each bearing condition), and one fold (25 samples for each bearing condition) are used for the training set. The remaining folds (75 samples for each bearing condition) are used as the testing set. To obtain precise classification results, \( l \)-fold CV is performed ten times in this study. To select the WKLFDA parameters using PSO, the training set is used. In our experiments, the PSO had a population size set at 20 particles, with \( c_1 \) and \( c_2 \) both set to 2.0. The inertial weight decreased linearly from 0.9 to 0.4 over the number of iterations, as defined in (24). To select the parameters for each binary SVM classifier, grid search algorithm is employed. In this paper, the searching range of parameters are \( \gamma \in \{2^{-3}, 2^{-2}, 2^{-1}, 2, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6\} \) and \( \xi \in \{2^{-3}, 2^{-2}, 2^{-1}, 1, 2, 2^2, 2^3, 2^4, 2^5, 2^6\} \). Finally, the grid search algorithm gives the best combination of pair \( \{C, \gamma\} \), yielding the highest classification accuracy.

Fig. 7 show the test accuracies of the dimensional reduction method in the \( d \)-dimensional subspace, where \( d=1,2,...,20 \). From Fig. 7 we can see that the performance of each method varies with reduced dimensions. The mean and best results of each method are also reported in Table II for easy in comparison. The subspaces according to the best records are assigned as the optimal subspaces. Observing from Fig. 7 and Table II, we conclude that the I-PSO-WKLFDA (OAA-PSO-WKLFDA and OAO-PSO-WKLFDA) method yields the highest accuracy, in which the OAO-PSO-WKLFDA provides better performance than OAA-PSO-WKLFDA. The experimental results coincided with theoretical analysis and can be explained as follows. Due to the limitation in the data, in which samples in a class are multimodal or samples between classes are nonlinearly separated in the input space, FDA generates the worst accuracy, 81.12%. Using the trace ratio criteria instead of the ratio trace criteria to obtain transformation, the TRLDA (82.78%) perform betters than the\( L \)-fold CV estimation procedure, the training data is randomly divided into \( l \) equal size folds. The classification performance is measured using one fold for training, and the remaining folds for testing. In this experiment, \( l \) is set to 4. This is a reasonable compromise considering the computational complexity and modeling robustness. More specifically, 100 samples of created feature vectors are divided into four mutual folds (each fold includes 25 randomly samples for each bearing condition), and one fold (25 samples for each bearing condition) are used for the training set. The remaining folds (75 samples for each bearing condition) are used as the testing set. To obtain precise classification results, \( l \)-fold CV is performed ten times in this study. To select the WKLFDA parameters using PSO, the training set is used. In our experiments, the PSO had a population size set at 20 particles, with \( c_1 \) and \( c_2 \) both set to 2.0. The inertial weight decreased linearly from 0.9 to 0.4 over the number of iterations, as defined in (24). To select the parameters for each binary SVM classifier, grid search algorithm is employed. In this paper, the searching range of parameters are \( C \in \{2^{-3}, 2^{-2}, 2^{-1}, 2, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6\} \) and \( \gamma \in \{2^{-3}, 2^{-2}, 2^{-1}, 1, 2, 2^2, 2^3, 2^4, 2^5, 2^6\} \). Finally, the grid search algorithm gives the best combination of pair \( \{C, \gamma\} \), yielding the highest classification accuracy.

Fig. 7 show the test accuracies of the dimensional reduction method in the \( d \)-dimensional subspace, where \( d=1,2,...,20 \). From Fig. 7 we can see that the performance of each method varies with reduced dimensions. The mean and best results of each method are also reported in Table II for easy in comparison. The subspaces according to the best records are assigned as the optimal subspaces. Observing from Fig. 7 and Table II, we conclude that the I-PSO-WKLFDA (OAA-PSO-WKLFDA and OAO-PSO-WKLFDA) method yields the highest accuracy, in which the OAO-PSO-WKLFDA provides better performance than OAA-PSO-WKLFDA. The experimental results coincided with theoretical analysis and can be explained as follows. Due to the limitation in the data, in which samples in a class are multimodal or samples between classes are nonlinearly separated in the input space, FDA generates the worst accuracy, 81.12%. Using the trace ratio criteria instead of the ratio trace criteria to obtain transformation, the TRLDA (82.78%) perform betters than the

![Fig. 7. Recognition accuracy versus reduced dimensionality using 25 training samples.](image-url)

![Table III](table-url)

<table>
<thead>
<tr>
<th>NM</th>
<th>ORF1</th>
<th>IRF1</th>
<th>BF1</th>
<th>ORF2</th>
<th>IRF2</th>
<th>BF2</th>
<th>ORF3</th>
<th>IRF3</th>
<th>BF3</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM</td>
<td>2916</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>39</td>
<td>0</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>ORF1</td>
<td>15</td>
<td>2950</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>ORF2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2983</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>ORF3</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>ORF4</td>
<td>24</td>
<td>23</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>ORF5</td>
<td>25</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>262</td>
</tr>
</tbody>
</table>

| Sensitivity(%) | 97.20  | 98.33  | 99.43  | 99.5   | 99.63  | 99.87  | 97.63  | 98.20  | 99.47  | 97.83  |
| Specificity(%) | 99.70  | 99.80  | 99.92  | 99.95  | 99.96  | 99.97  | 99.98  | 99.80  | 99.94  | 98.73  |

![Table II](table-url)

<table>
<thead>
<tr>
<th>Methods</th>
<th>SVM</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDA</td>
<td>81.12</td>
<td>94.67</td>
</tr>
<tr>
<td>TRLDA</td>
<td>82.78</td>
<td>95.00</td>
</tr>
<tr>
<td>LFDA</td>
<td>83.02</td>
<td>95.33</td>
</tr>
<tr>
<td>KLFDA</td>
<td>84.57</td>
<td>96.00</td>
</tr>
<tr>
<td>WKLFDA</td>
<td>85.61</td>
<td>96.16</td>
</tr>
<tr>
<td>PSO-KLFDA</td>
<td>86.20</td>
<td>96.70</td>
</tr>
<tr>
<td>PSO-WKLFDA</td>
<td>89.34</td>
<td>97.20</td>
</tr>
<tr>
<td>OAA-PSO-WKLFDA</td>
<td>91.07</td>
<td>97.80</td>
</tr>
<tr>
<td>OAO-PSO-WKLFDA</td>
<td>93.38</td>
<td>98.80</td>
</tr>
</tbody>
</table>
FDA. By preserving the information within a class, the LFDA (83.01%) provides better performance compared to FDA. By employing a kernel trick to map the data into higher dimension in which the nonlinear data possibly becomes linear separation, the KLFDA (84.57%), WKLFDA (86.51%), PSO-KLFDA (86.20%) and PSO-WKLFDA (89.34%) have much better performances. However, due to the capability of the wavelet kernel over the RBF kernel to approximate the nonlinear function, WKLFDA and PSO-WKLFDA yields better performance compared to KLFDA and PSO-KLFDA, respectively. Meanwhile, by using PSO to select the optimized WKLFDA parameters corresponding to the reduced dimension, the proposed PSO-WKLFDA algorithm yields better performance than WKLFDA. However, due to the capability of the wavelet kernel over the RBF kernel to approximate the nonlinear function, WKLFDA and PSO-WKLFDA yields better performance compared to KL FDA and PSO-KLFDA, respectively. Meanwhile, by using PSO to select the optimized WKLFDA parameters corresponding to the reduced dimension, the proposed PSO-WKLFD A algorithm yields better performance than WKLFDA. However, due to the capability of the wavelet kernel over the RBF kernel to approximate the nonlinear function, WKLFDA and PSO-WKLFDA yields better performance compared to KLFDA and PSO-KLFDA, respectively. Meanwhile, by using PSO to select the optimized WKLFDA parameters corresponding to the reduced dimension, the proposed PSO-WKLFDA algorithm yields better performance than WKLFDA.

<table>
<thead>
<tr>
<th>Fault diagnosis methods</th>
<th>Classification accuracy(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three selected features (kurtosis, crest factor, skewness) of the raw vibration signal</td>
<td>76.50</td>
</tr>
<tr>
<td>Nine features of the raw vibration signal</td>
<td>89.10</td>
</tr>
<tr>
<td>All 49 features</td>
<td>82.83</td>
</tr>
<tr>
<td>DET-Feature selection</td>
<td>94.10</td>
</tr>
<tr>
<td>OAO-DET-Feature selection</td>
<td>95.56</td>
</tr>
<tr>
<td>PSO-FS</td>
<td>96.40</td>
</tr>
<tr>
<td>OAO-PSO-FS</td>
<td>98.17</td>
</tr>
<tr>
<td>PSO-WKLFDA</td>
<td>97.20</td>
</tr>
<tr>
<td>OAO-PSO-WKLFDA</td>
<td>98.80</td>
</tr>
</tbody>
</table>

VI. CONCLUSION

This paper started by analyzing the performance of the conventional frequency domain analysis for different vibration signals. The results show that the technique fails to identify the BCF for some noisy vibration signals. Then, the distribution of the bearing data was analyzed. The results show that the distribution of the bearing data is a type of multimodal data. Thus, to enhance defect classification performance, a new multimodal dimensional reduction method, namely I-PSO-WKLFDA, was proposed. First, fault features based on NLM and EMD methods, was computed to represent diverse symptoms of bearing defects. The multiclass bearing defect classification problem was then converted to a multi-binary class bearing defect classification problem based on an OAO strategy. The I-PSO-WKLFDA was then applied for each binary class to extract the corresponding effective features. The effective features were used as the input to the SVM classifier. Finally, a decision fusion mechanism was employed to merge the classification results from each SVM classifier to identify the bearing condition. Experimental results indicated that the proposed fault diagnosis methodology achieves excellent classification accuracy.
the other hand, the main contribution of this paper is to propose a new effective dimensionality reduction method, thus the proposed methods can be applied for other pattern recognition problems.

REFERENCES


Mien Van received his B.S. degree in Electrical Engineering from Danang University of Technology, DaNang city, Vietnam, in 2009. He received his Ph.D. degree in the School of Electrical Engineering, University of Ulsan, Korea in 2015.

He is currently a Post-Doctoral Research Fellow with the Advanced Robotic Centre, Faculty Engineering, National University of Singapore, Singapore. His research interests include automatic control, fault diagnosis and fault tolerant, machine learning, robot-vision system and sensing and perception.

Hee-Jun Kang received his B.S. degree in Mechanical Engineering from Seoul National University, Korea in 1985. He received his M.S and Ph.D. degrees in Mechanical Engineering from The University of Texas at Austin, USA in 1988 and 1991, respectively. Since March in 1992, he has been a Professor of Electrical Engineering in University of Ulsan. His current research interests are sensor based robotic application, robot calibration, haptics, robot fault diagnosis and mechanism analysis.