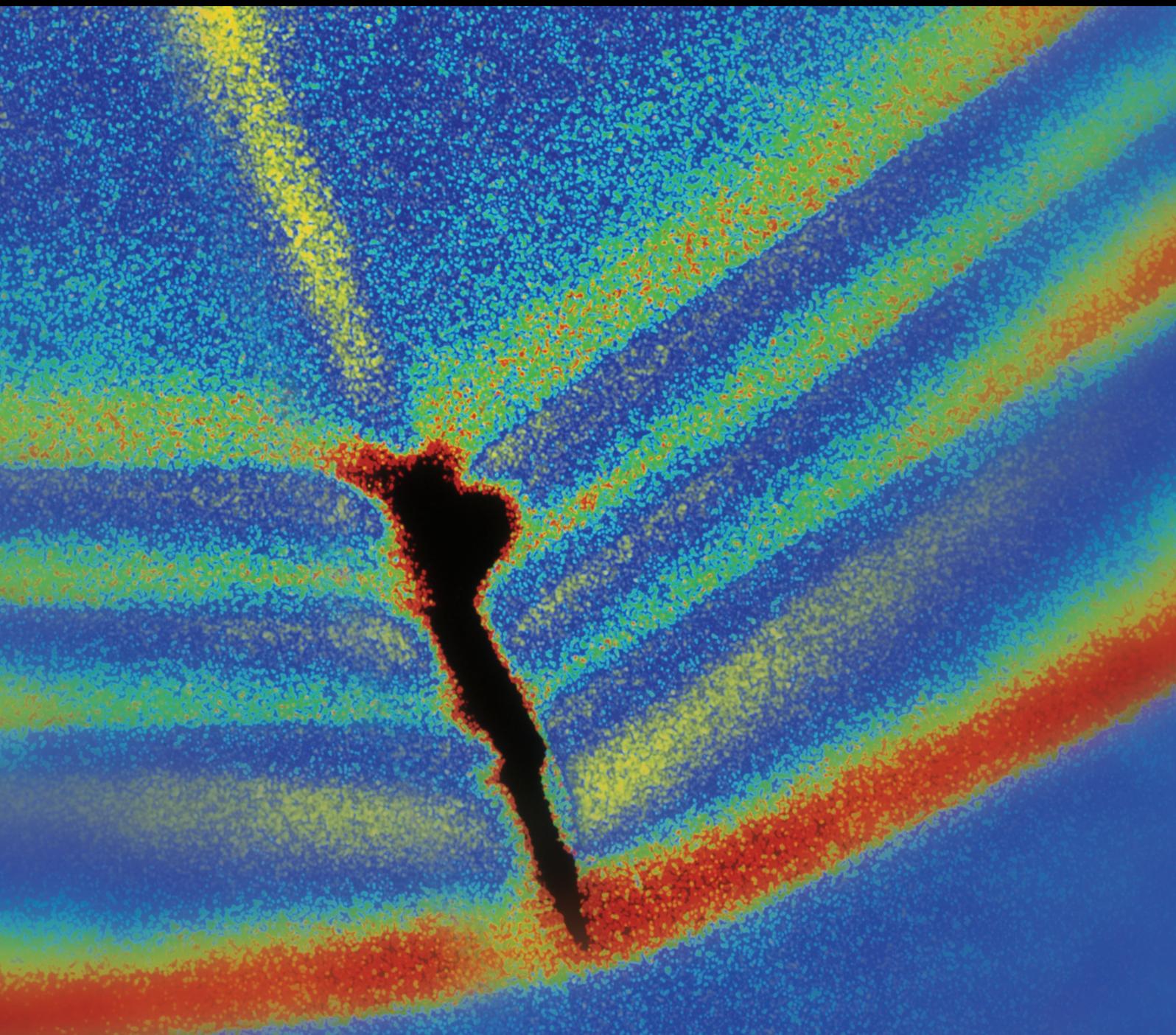


Shock and Vibration

Intelligent Fault Diagnosis Based on Vibration Signal Analysis

Guest Editors: Minvydas Ragulskis, Lu Chen, Ganging Song, and A. El Sinawi





Intelligent Fault Diagnosis Based on Vibration Signal Analysis

Shock and Vibration

Intelligent Fault Diagnosis Based on Vibration Signal Analysis

Guest Editors: Minvydas Ragulskis, Lu Chen, Ganging Song,
and A. El Sinawi



Copyright © 2017 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in "Shock and Vibration." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

Brij N. Agrawal, USA
Marco Alfano, Italy
Farbod Alijani, Canada
Sumeet S. Aphale, UK
Mariano Artés, Spain
Hassan Askari, Canada
Matteo Aureli, USA
Ranjan Banerjee, UK
Mahmoud Bayat, Iran
José A. Becerra Villanueva, Spain
Marco Belloli, Italy
Subhamoy Bhattacharya, UK
Ivo Caliò, Italy
Antonio Carcaterra, Italy
Dumitru I. Caruntu, USA
Noel Challamel, France
Athanasios Chasalevris, UK
Ashwin Chinnayya, France
Giorgio Dalpiaz, Italy
Farhang Daneshmand, Canada
Sergio De Rosa, Italy
Dario Di Maio, UK
Longjun Dong, China
Lorenzo Dozio, Italy
Mohamed El Badaoui, France
Mohammad Elahinia, USA
Fiorenzo A. Fazzolari, UK

Francesco Franco, Italy
Pedro Galvín, Spain
Alessandro Gasparetto, Italy
Gianluca Gatti, Italy
Anindya Ghoshal, USA
Nere Gil-Negrete, Spain
Hassan Haddadpour, Iran
M.I. Herreros, Spain
Hamid Hosseini, Japan
Reza Jazar, Australia
Sakdirat Kaewunruen, UK
Yuri S. Karinski, Israel
Jeong-Hoi Koo, USA
Georges Kouroussis, Belgium
Mickaël Lallart, France
Kenneth J. Loh, USA
Nuno M. Maia, Portugal
Giuseppe C. Marano, Italy
Laurent Mevel, France
Emiliano Mucchi, Italy
Tony Murmu, UK
Sundararajan Natarajan, India
Toshiaki Natsuki, Japan
Miguel Neves, Portugal
Coral Ortiz, Spain
Aleksandar Pavic, UK
Evgeny Petrov, UK

Antonina Pirrotta, Italy
Vitomir Racic, Italy
Carlo Rainieri, Italy
Didier Rémond, France
Francesco Ripamonti, Italy
Salvatore Russo, Italy
Edoardo Sabbioni, Italy
Jerzy T. Sawicki, USA
Onome E. Scott-Emuakpor, USA
Vadim V. Silberschmidt, UK
Kumar V. Singh, USA
Isabelle Sochet, France
Alba Sofi, Italy
Jussi Sopenan, Finland
Stefano Sorace, Italy
Narakorn Srinil, UK
Salvatore Strano, Italy
Chao Tao, China
Mario Terzo, Italy
Tai Thai, Australia
Carlo Trigona, Italy
Federica Tubino, Italy
Nerio Tullini, Italy
Marcello Vanali, Italy
Jörg Wallaschek, Germany
Zaili L. Yang, UK
Stana Živanović, UK

Contents

Intelligent Fault Diagnosis Based on Vibration Signal Analysis

Minvydas Ragulskis, Lu Chen, Ganging Song, and Ameen El Sinawi
Volume 2017, Article ID 9186989, 1 page

Output-Only Modal Parameter Recursive Estimation of Time-Varying Structures via a Kernel Ridge Regression FS-TARMA Approach

Zhi-Sai Ma, Li Liu, Si-Da Zhou, and Lei Yu
Volume 2017, Article ID 8176593, 14 pages

Electromagnetic and Mechanical Characteristics Analysis of a Flat-Type Vertical-Gap Passive Magnetic Levitation Vibration Isolator

Baoquan Kou, Yiheng Zhou, Xiaobao Yang, Feng Xing, and He Zhang
Volume 2016, Article ID 5327207, 12 pages

Automated Bearing Fault Diagnosis Using 2D Analysis of Vibration Acceleration Signals under Variable Speed Conditions

Sheraz Ali Khan and Jong-Myon Kim
Volume 2016, Article ID 8729572, 11 pages

Research on Dynamic Modeling and Application of Kinetic Contact Interface in Machine Tool

Dan Xu and Zhixin Feng
Volume 2016, Article ID 5658181, 9 pages

Control Performance and Robustness of Pounding Tuned Mass Damper for Vibration Reduction in SDOF Structure

Qichao Xue, Jingcai Zhang, Jian He, and Chunwei Zhang
Volume 2016, Article ID 8021690, 15 pages

Reliability Analysis with Multiple Dependent Features from a Vibration-Based Accelerated Degradation Test

Fuqiang Sun, Jingcheng Liu, Xiaoyang Li, and Haitao Liao
Volume 2016, Article ID 2315916, 15 pages

Research of Fault Diagnosis Based on Sensitive Intrinsic Mode Function Selection of EEMD and Adaptive Stochastic Resonance

Zhixing Li and Boqiang Shi
Volume 2016, Article ID 2841249, 12 pages

Prognostic and Remaining Life Prediction of Electronic Device under Vibration Condition Based on CPSD of MPI

Ying Chen, Ning Tang, and Zenghui Yuan
Volume 2016, Article ID 2650569, 11 pages

Distance and Density Similarity Based Enhanced k -NN Classifier for Improving Fault Diagnosis Performance of Bearings

Sharif Uddin, Md. Rashedul Islam, Sheraz Ali Khan, Jaeyoung Kim, Jong-Myon Kim, Seok-Man Sohn, and Byeong-Keun Choi
Volume 2016, Article ID 3843192, 11 pages

Improvement of Roller Bearing Diagnosis with Unlabeled Data Using Cut Edge Weight Confidence Based Tritraining

Wei-Li Qin, Wen-Jin Zhang, and Zhen-Ya Wang

Volume 2016, Article ID 1646898, 9 pages

Vibration Based Diagnosis for Planetary Gearboxes Using an Analytical Model

Liu Hong, Yongzhi Qu, Yuegang Tan, Mingyao Liu, and Zude Zhou

Volume 2016, Article ID 2620518, 11 pages

Adaptive Multiscale Noise Control Enhanced Stochastic Resonance Method Based on Modified EEMD with Its Application in Bearing Fault Diagnosis

Jimeng Li and Jinfeng Zhang

Volume 2016, Article ID 1485412, 13 pages

Bearing Performance Degradation Assessment Using Lifting Wavelet Packet Symbolic Entropy and SVDD

Jianmin Zhou, Huijuan Guo, Long Zhang, Qingyao Xu, and Hui Li

Volume 2016, Article ID 3086454, 10 pages

Stator Vibration Characteristic Identification of Turbogenerator among Single and Composite Faults Composed of Static Air-Gap Eccentricity and Rotor Interturn Short Circuit

Yu-Ling He, Wei-Qi Deng, Bo Peng, Meng-Qiang Ke, Gui-Ji Tang, Shu-Ting Wan, and Xiang-Yu Liu

Volume 2016, Article ID 5971081, 14 pages

Fault Diagnosis of a Hydraulic Pump Based on the CEEMD-STFT Time-Frequency Entropy Method and Multiclass SVM Classifier

Wanlin Zhao, Zili Wang, Jian Ma, and Lianfeng Li

Volume 2016, Article ID 2609856, 8 pages

An Approach to Fault Diagnosis for Gearbox Based on Image Processing

Yang Wang and Yujie Cheng

Volume 2016, Article ID 5898052, 10 pages

Intelligent Analysis Method of Gear Faults Based on FRWT and SVM

Hongfang Chen, Yanqiang Sun, Zhaoyao Shi, and Jiachun Lin

Volume 2016, Article ID 1582738, 9 pages

Investigations of Dynamic Behaviors of Face Gear Drives Associated with Pinion Dedendum Fatigue Cracks

Zhengminqing Li, Shuang Liu, Rupeng Zhu, and Xiaodong Xu

Volume 2016, Article ID 3173860, 10 pages

Vestas V90-3MW Wind Turbine Gearbox Health Assessment Using a Vibration-Based Condition Monitoring System

A. Romero, Y. Lage, S. Soua, B. Wang, and T.-H. Gan

Volume 2016, Article ID 6423587, 18 pages

Shannon Entropy and K -Means Method for Automatic Diagnosis of Broken Rotor Bars in Induction Motors Using Vibration Signals

David Camarena-Martinez, Martin Valtierra-Rodriguez, Juan P. Amezcua-Sanchez, David Granados-Lieberman, Rene J. Romero-Troncoso, and Arturo Garcia-Perez
Volume 2016, Article ID 4860309, 10 pages

Rolling Bearing Fault Diagnosis Based on STFT-Deep Learning and Sound Signals

Hongmei Liu, Lianfeng Li, and Jian Ma
Volume 2016, Article ID 6127479, 12 pages

Bearing Fault Diagnosis Based on Deep Belief Network and Multisensor Information Fusion

Jie Tao, Yilun Liu, and Dalian Yang
Volume 2016, Article ID 9306205, 9 pages

Fault Diagnosis for Rolling Bearing under Variable Conditions Based on Image Recognition

Bo Zhou and Yujie Cheng
Volume 2016, Article ID 1948029, 14 pages

Editorial

Intelligent Fault Diagnosis Based on Vibration Signal Analysis

Minvydas Ragulskis,¹ Lu Chen,² Ganging Song,³ and Ameen El Sinawi⁴

¹Department of Mathematical Modelling, Kaunas University of Technology, LT-51368 Kaunas, Lithuania

²School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China

³Department of Mechanical Engineering, University of Houston, Houston, TX 77204, USA

⁴Department of Mechanical Engineering, The Petroleum Institute, Abu Dhabi, UAE

Correspondence should be addressed to Minvydas Ragulskis; minvydas.ragulskis@ktu.lt

Received 30 January 2017; Accepted 30 January 2017; Published 19 February 2017

Copyright © 2017 Minvydas Ragulskis et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Intelligent fault diagnosis has attracted increasing attention during the past decades in various industrial fields, such as aerospace, shipbuilding, manufacturing, sustainable energy, infrastructure development, and transportation. Preventive fault diagnosis is expected to improve machinery operational reliability and safety for complicated systems or equipment, further reducing the cost of cycle life and avoiding system risk. Nowadays, many diagnostic applications are deployed based on the vibration signal analysis. Vibration signals can be conveniently acquired and contain abundant signature information that reflects the potential failures and performance degradation trends of the monitored system. This special issue aims to publish new progress with the state of the art in the various engineering applications and provides a platform for researchers and engineers to share their novel theories and methodologies so as to improve technologies in the research area of intelligent fault diagnosis.

This special issue on intelligent fault diagnosis based on vibration signal analysis offers a unique opportunity to present the latest trends and state-of-the-art research in this rapidly developing field of research. 9 of 23 papers are dedicated for developing new techniques for bearing fault diagnosis. Enhanced k -NN classifiers, sensitive selection of intrinsic model functions, adaptive stochastic resonance analysis algorithms, lifting wavelet packet symbolic entropy assessment, cut-edge weight confidence intervals based on tri-training, advanced image recognition algorithms in the EEMD space, STFT-based deep learning techniques, and deep belief networks coupled with multisensor information

fusion algorithms are just a few advanced approaches represented in this special issue in the area of rolling bearing diagnosis.

Other papers of the special issue are dedicated to other (but not less important) areas or application: intelligent fault diagnosis in gearboxes, wind turbines, hydraulic pumps, turbo generators, electronic devices, vibrations isolators, machine tools, and so forth. Recursive estimation of kernel ridge regression, advanced K -means and multiclass SVM classifiers, modified time-frequency Shannon entropy method, and advanced image analysis techniques in frequency spaces are just a few novel techniques and algorithms used for intelligent fault diagnosis in these demanding applications.

We hope that this special issue will be beneficial for specialists, graduate students, researchers, and scientists working in the rapidly developing area of intelligent fault diagnosis based on vibration signal analysis.

Acknowledgments

The editors of the special issue would like to thank authors of all submitted papers.

Minvydas Ragulskis
Lu Chen
Ganging Song
Ameen El Sinawi

Research Article

Output-Only Modal Parameter Recursive Estimation of Time-Varying Structures via a Kernel Ridge Regression FS-TARMA Approach

Zhi-Sai Ma, Li Liu, Si-Da Zhou, and Lei Yu

School of Aerospace Engineering, Beijing Institute of Technology, Zhongguancun South Street 5, Beijing 100081, China

Correspondence should be addressed to Si-Da Zhou; zhousida@bit.edu.cn

Received 21 July 2016; Revised 2 November 2016; Accepted 29 November 2016; Published 11 January 2017

Academic Editor: Lu Chen

Copyright © 2017 Zhi-Sai Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Modal parameter estimation plays an important role in vibration-based damage detection and is worth more attention and investigation, as changes in modal parameters are usually being used as damage indicators. This paper focuses on the problem of output-only modal parameter recursive estimation of time-varying structures based upon parameterized representations of the time-dependent autoregressive moving average (TARMA). A kernel ridge regression functional series TARMA (FS-TARMA) recursive identification scheme is proposed and subsequently employed for the modal parameter estimation of a numerical three-degree-of-freedom time-varying structural system and a laboratory time-varying structure consisting of a simply supported beam and a moving mass sliding on it. The proposed method is comparatively assessed against an existing recursive pseudolinear regression FS-TARMA approach via Monte Carlo experiments and shown to be capable of accurately tracking the time-varying dynamics in a recursive manner.

1. Introduction

The need for damage identification or fault diagnosis of a structure is pervasive throughout the aerospace, ship-building, manufacturing, infrastructure, and transportation engineering communities. In recent years, many efforts have been undertaken to detect, locate, and characterize damage in structural systems by examining changes in measured vibration responses, which may be collectively referred to as vibration-based damage identification or fault diagnosis [1–5]. Compared with visual or localized methods, vibration-based methods offer a number of potential advantages, as they require no visual inspection, are capable of working at a “system level” especially when the portion of the structure being inspected is not accessible, and can monitor a structure under its operating conditions [1–5]. The basic idea behind this technology is that changes in the physical properties of a structure will cause detectable changes in its modal parameters (natural frequencies, damping ratios, and mode shapes), as the modal parameters are functions of the physical properties of the structure (mass, damping, and stiffness) [1].

Because changes in modal parameters are usually being used as damage indicators, modal parameter estimation plays an important role in vibration-based damage detection of structures and is worth more attention and investigation.

Most structures in the real world are time-varying under a specific time scale and their intrinsic time-varying behavior is increasingly inevitable in industry. Typical examples include vibration absorbers with variable stiffness, bridges with crossing vehicles, launch vehicles with varying fuel mass, airplanes with varying additional aerodynamic effects in flight, deployable space structures, and rotating machinery. In contrast to time-invariant systems producing stationary responses with time-invariant statistical characteristics [6], time-varying systems exhibit time-varying/nonstationary characteristics, requiring time-dependent dynamic models and corresponding identification methods. Therefore, time-varying system identification becomes increasingly more important [7–12].

This paper focuses on the problem of output-only modal parameter recursive estimation of time-varying structures due to the following two reasons. Firstly, in many cases

controlled testing may not be feasible or the excitation may be unobservable under realistic operating conditions. Hence, there exists a need to estimate modal parameters of time-varying structures by exclusively using the available measured vibration responses (referred to as output-only modal parameter estimation). Secondly, in many practical cases in which the data observations are arriving in a continuous stream, the structure has to be identified before all of the response measurements are known. Hence, there also exists a need to estimate modal parameters of time-varying structures in a recursive instead of batch manner.

Many output-only identification methods have been developed based on parameterized representations of the time-dependent autoregressive moving average (TARMA) [10–14]. TARMA model-based identification methods are divided into three main classes, depending on the type of “structure” imposed upon the evolution of the time-varying model parameters [11, 12]: unstructured parameter evolution (UPE) methods, stochastic parameter evolution (SPE) methods, and deterministic parameter evolution (DPE) methods. Most of the output-only recursive identification methods are currently of the UPE type [9–13], which may be not suitable for fast varying structures as they impose no “structure” upon their model parameters [10–12]. These methods include the recursive maximum likelihood TARMA method [9–11] and the recursive pseudolinear regression TARMA (RPLR-TARMA) method [9, 10, 13]. In contrast with the UPE/SPE methods, the DPE methods are based on explicit models of the parameter variation. These models are achieved by approximating the parameter trajectory by a linear combination of known basis functions, belonging to specific functional subspaces. Such representations are generally referred to as functional series TARMA (FS-TARMA) representations/models [11–14]. Although different recursive DPE methods have been developed in recent years, their tracking capability of time-varying dynamics may be limited and more efforts are still needed [10–20]. From this point of view the FS-TARMA model-based recursive methods are promising and are therefore investigated in this work.

The remainder of the paper is organized as follows: Section 2 introduces the FS-TARMA model-based modal parameter estimation. The existing recursive pseudolinear regression FS-TARMA (RPLR-FS-TARMA) approach is reviewed in Section 3. Section 4 proposes a kernel ridge regression FS-TARMA (KRR-FS-TARMA) approach consisting of parameter matrix estimation and the corresponding model structure selection. The proposed method is, along with the existing RPLR-FS-TARMA method, numerically and experimentally validated in Sections 5 and 6, respectively. Section 7 summarizes the study.

2. FS-TARMA Model-Based Modal Parameter Estimation

2.1. TARMA Model. The TARMA representation resembles its stationary counterparts, with the significant difference being that it allows its parameters to depend upon time [10–12]. A TARMA (n_a, n_c) model, with n_a, n_c designating,

respectively, its autoregressive (AR) and moving average (MA) orders, is of the form [13]

$$\mathbf{x}[t] + \sum_{i=1}^{n_a} \mathbf{A}_i[t] \mathbf{x}[t-i] = \mathbf{e}[t] + \sum_{i=1}^{n_c} \mathbf{C}_i[t] \mathbf{e}[t-i], \quad (1)$$

$$\mathbf{e}[t] \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}[t])$$

with t designating normalized (by the sampling period) discrete time, $\mathbf{x}[t] \in \mathbb{R}^{k \times 1}$ the discrete-time nonstationary vibration signal to be modeled, and $\mathbf{e}[t] \in \mathbb{R}^{k \times 1}$ an unobservable uncorrelated innovation (residual) sequence with zero mean and time-varying nonsingular covariance matrix $\boldsymbol{\Sigma}[t] \in \mathbb{R}^{k \times k}$. $\mathbf{A}_i[t]$ and $\mathbf{C}_i[t]$ are, respectively, the model’s AR and MA parameter matrices. $\mathcal{N}(\cdot, \cdot)$ stands for normally independently distributed random variables with the indicated mean and covariance.

Once the time-dependent AR parameter matrix $\mathbf{A}_i[t]$ has been obtained, the system’s “frozen-time” poles can be derived by solving the following general eigenvalue problem [14]

$$\left(\mathbf{D}[t] - p_r[t] \mathbf{I}_{kn_a} \right) \mathbf{V}_r[t] = 0, \quad r = 1, 2, \dots, kn_a, \quad (2)$$

where \mathbf{I}_{kn_a} designates the identity matrix of order kn_a and $p_r[t]$ and $\mathbf{V}_r[t] = [p_r[t]^{-n_a} \mathbf{L}_r^T, \dots, p_r[t]^{-1} \mathbf{L}_r^T]^T$ are the r th pole and eigenvector of $\mathbf{D}[t]$ at instant of time t with the r th mode shape \mathbf{L}_r , respectively. The matrix $\mathbf{D}[t]$ is constructed from the AR parameter matrices as

$$\mathbf{D}[t] = \begin{bmatrix} \mathbf{0} & \mathbf{I}_k & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_k \\ -\mathbf{A}_{n_a}[t] & -\mathbf{A}_{n_a-1}[t] & \cdots & -\mathbf{A}_1[t] \end{bmatrix}. \quad (3)$$

The system’s “frozen-time” modes with natural frequencies and damping ratios are computed by [6]

$$f_r[t] = \frac{|\ln p_r[t]|}{2\pi\Delta t},$$

$$\zeta_r[t] = \frac{-\text{Re}(\ln p_r[t])}{|\ln p_r[t]|}, \quad (4)$$

$$r = 1, 2, \dots, kn_a$$

in which $|\cdot|$ denotes the absolute value and Δt the sampling period.

2.2. FS-TARMA Model. Any given function can be approximated by a set of selected basis functions with arbitrary accuracy, as long as a sufficient number of basis functions are used [21]. Hence, a time-varying parameter $v[t]$ can be expressed as

$$v[t] = \sum_{j=1}^{p_v} v^j \cdot G_{b_v(j)}[t], \quad (5)$$

where $G_{b_v(j)}[t]$ denotes a set of basis functions selected from a suitable functional subspace (such as trigonometric, Legendre, Chebyshev, wavelets, B-splines, and moving Kriging shape functions), v^j coefficient of projection, p_v the subspace dimensionality, and the index $b_v(i)$ ($i = 1, \dots, p_v$) the specific basis functions of the selected functional subspace that are included in the basis.

By using (5), the time-dependent AR and MA parameter matrices $\mathbf{A}_i[t]$ and $\mathbf{C}_i[t]$ can be, respectively, expressed as [11–14]

$$\mathbf{A}_i[t] \approx \begin{bmatrix} \sum_{j=1}^{p_a} a_{1,1}^{i,j} \cdot G_{b_a(j)}[t] & \cdots & \sum_{j=1}^{p_a} a_{1,k}^{i,j} \cdot G_{b_a(j)}[t] \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^{p_a} a_{k,1}^{i,j} \cdot G_{b_a(j)}[t] & \cdots & \sum_{j=1}^{p_a} a_{k,k}^{i,j} \cdot G_{b_a(j)}[t] \end{bmatrix}, \quad (6)$$

$$\mathbf{C}_i[t] \approx \begin{bmatrix} \sum_{j=1}^{p_c} c_{1,1}^{i,j} \cdot G_{b_c(j)}[t] & \cdots & \sum_{j=1}^{p_c} c_{1,k}^{i,j} \cdot G_{b_c(j)}[t] \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^{p_c} c_{k,1}^{i,j} \cdot G_{b_c(j)}[t] & \cdots & \sum_{j=1}^{p_c} c_{k,k}^{i,j} \cdot G_{b_c(j)}[t] \end{bmatrix}.$$

Substituting (6) into (1), the TARMA model can be rewritten into the following FS-TARMA model:

$$\mathbf{x}[t] = -\sum_{i=1}^{n_a} \mathbf{A}_i[t] \mathbf{x}[t-i] + \sum_{i=1}^{n_c} \mathbf{C}_i[t] \mathbf{e}[t-i] + \mathbf{e}[t] \implies \quad (7)$$

$$\mathbf{x}[t] = \boldsymbol{\theta}^T \boldsymbol{\phi}[t] + \mathbf{e}[t]$$

with

$$\boldsymbol{\theta} = \begin{bmatrix} a_{1,1}^{1,1} & \cdots & a_{1,1}^{1,p_a} & \cdots & c_{1,k}^{n_c,1} & \cdots & c_{1,k}^{n_c,p_c} \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ a_{k,1}^{1,1} & \cdots & a_{k,1}^{1,p_a} & \cdots & c_{k,k}^{n_c,1} & \cdots & c_{k,k}^{n_c,p_c} \end{bmatrix}^T \quad (8)$$

$$\in \mathbb{R}^{k(n_a p_a + n_c p_c) \times k}$$

being the time-invariant parameter matrix, and

$$\boldsymbol{\phi}[t] = \begin{bmatrix} -\mathbf{x}[t-1] \otimes \mathbf{g}_{b_a}[t] \\ \vdots \\ -\mathbf{x}[t-n_a] \otimes \mathbf{g}_{b_a}[t] \\ \mathbf{e}[t-1] \otimes \mathbf{g}_{b_c}[t] \\ \vdots \\ \mathbf{e}[t-n_c] \otimes \mathbf{g}_{b_c}[t] \end{bmatrix} \in \mathbb{R}^{k(n_a p_a + n_c p_c) \times 1} \quad (9)$$

the corresponding regression vector, where \otimes is the Kronecker product [22] and $\mathbf{g}_{b_a}[t]$ and $\mathbf{g}_{b_c}[t]$ are, respectively, given by

$$\mathbf{g}_{b_a}[t] = [G_{b_a(1)}[t], G_{b_a(2)}[t], \dots, G_{b_a(p_a)}[t]]^T \quad (10)$$

$$\in \mathbb{R}^{p_a \times 1},$$

$$\mathbf{g}_{b_c}[t] = [G_{b_c(1)}[t], G_{b_c(2)}[t], \dots, G_{b_c(p_c)}[t]]^T \quad (11)$$

$$\in \mathbb{R}^{p_c \times 1}.$$

Unlike the TARMA model of (1), the parameter matrix of the FS-TARMA model of (7) is time-invariant; that is, $\boldsymbol{\theta}$ is a time-invariant matrix. In other words, a time-varying TARMA model can be transformed into a time-invariant FS-TARMA model by expanding the time-dependent model parameters onto the selected functional subspaces. The parameter matrix $\boldsymbol{\theta}$ may be thus estimated by means of time-invariant system identification techniques, while the time-dependent AR and MA parameter matrices $\mathbf{A}_i[t]$ and $\mathbf{C}_i[t]$ may be subsequently estimated by using (6), after substituting $\boldsymbol{\theta}$ by the obtained estimate $\hat{\boldsymbol{\theta}}$.

3. Recursive Pseudolinear Regression FS-TARMA Approach

So far, we have presented that modal parameters can be extracted from the FS-TARMA model parameter matrix estimate $\hat{\boldsymbol{\theta}}$. The problem of FS-TARMA model parameter matrix estimation based on the existing recursive pseudolinear regression scheme is reviewed in this section.

3.1. RPLR-FS-TARMA Method. The FS-TARMA model is time-invariant, as shown in (7), and the well-known recursive pseudolinear regression scheme [9, 10] can be utilized to accomplish the estimation of $\boldsymbol{\theta}$. Given the regularization parameter $\gamma \geq 0$, the regularized least squares cost function is defined as

$$\min_{\boldsymbol{\theta}} \left[\frac{1}{2} \sum_{\tau=1}^t \|\mathbf{x}[\tau] - \boldsymbol{\theta}^T \boldsymbol{\phi}[\tau]\|^2 + \frac{1}{2} \gamma \|\boldsymbol{\theta}\|^2 \right], \quad (12)$$

where $\|\cdot\|$ denotes the Euclidean norm. The RPLR-FS-TARMA method can be obtained by minimizing the cost function of (12) and is summarized in the following [11]:

$$\text{Estimator update: } \hat{\boldsymbol{\theta}}[t]^T = \hat{\boldsymbol{\theta}}[t-1]^T + \hat{\mathbf{e}}[t | t-1] \mathbf{K}[t]^T,$$

$$\begin{aligned}
\text{Prediction error: } \widehat{\mathbf{e}}[t | t-1] &= \mathbf{x}[t] - \widehat{\mathbf{x}}[t | t-1] = \mathbf{x}[t] - \widehat{\boldsymbol{\theta}}[t-1]^T \boldsymbol{\phi}[t], \\
\text{Gain: } \mathbf{K}[t] &= \mathbf{P}[t-1] \boldsymbol{\phi}[t] \left(1 + \boldsymbol{\phi}[t]^T \mathbf{P}[t-1] \boldsymbol{\phi}[t]\right)^{-1}, \\
\text{Covariance update: } \mathbf{P}[t] &= \mathbf{P}[t-1] - \mathbf{K}[t] \boldsymbol{\phi}[t]^T \mathbf{P}[t-1].
\end{aligned} \tag{13}$$

By recursively applying (13) to the data record $\mathbf{x}[t]$ ($t = 1, \dots, L$), the final parameter matrix estimate can be obtained as $\widehat{\boldsymbol{\theta}}[L]^T$, where L is the length of the data record. For the initialization of the method it is customary to set $\widehat{\boldsymbol{\theta}}[0]^T = \mathbf{0}$, $\mathbf{P}[0] = \gamma^{-1} \mathbf{I}_{k(n_a p_a + n_c p_c)}$. It should be further stressed that the recursive scheme may yield biased parameter estimates even if parameters can be described exactly as weighted sums of basis functions, as the regression matrix $\boldsymbol{\phi}[t]$ contains entries that are constructed from data using past models.

Generally, the capability of the RPLR-FS-TARMA method is limited due to the following four reasons. Firstly, the estimate $\widehat{\boldsymbol{\theta}}[t]^T$ is recursively updated at each time step, but the final estimate $\widehat{\boldsymbol{\theta}}[L]^T$ involves processing the entire data record. Hence, the RPLR-FS-TARMA method is a recursive, but not a real-time, time-varying system identification method. Secondly, the RPLR-FS-TARMA method, as a member of running-basis estimators [10], will encounter numerical problems in case that the basis sequences are not bounded. Hence, this method may be limited to the cases in which L is known and the basis functions can be scaled in advance. Thirdly, we have to consider the problem of data saturation [23], which means we may reach a point in our analysis of data that sampling more data will not lead to more accurate estimation of $\widehat{\boldsymbol{\theta}}[L]^T$. Hence, this method may be limited to the cases in which L is not so large or

data saturation can be avoided. Finally, with the increase in L , the difficulty of expanding the time-dependent model parameters onto specific functional subspaces increases as well. For example, the selection of a suitable functional subspace could be very difficult, and the subspace dimensionalities p_a and/or p_c could be quite large. Hence, the superior achievable accuracy, lower computational complexity, and enhanced tracking capability of the RPLR-FS-TARMA method will not be guaranteed any more.

3.2. Exponentially Weighted RPLR-FS-TARMA Method. In order to avoid some of those limitations of the RPLR-FS-TARMA method, the exponentially weighted RPLR-FS-TARMA method is presented by including an exponentially weighted mechanism in the RPLR-FS-TARMA method [9, 11, 13].

Given the forgetting factor $0 < \lambda \leq 1$ (a positive scalar, usually close to one) and the regularization parameter $\gamma \geq 0$, the exponentially weighted regularized least squares cost function is defined as

$$\min_{\boldsymbol{\theta}[t]} \left[\frac{1}{2} \sum_{\tau=1}^t \lambda^{t-\tau} \|\mathbf{x}[\tau] - \boldsymbol{\theta}[\tau]^T \boldsymbol{\phi}[\tau]\|^2 + \frac{1}{2} \lambda^t \gamma \|\boldsymbol{\theta}[t]\|^2 \right]. \tag{14}$$

The exponentially weighted RPLR-FS-TARMA method can be obtained by minimizing the cost function of (14) and is summarized in the following [9, 11, 13]:

$$\begin{aligned}
\text{Estimator update: } \widehat{\boldsymbol{\theta}}[t]^T &= \widehat{\boldsymbol{\theta}}[t-1]^T + \widehat{\mathbf{e}}[t | t-1] \mathbf{K}[t]^T, \\
\text{Prediction error: } \widehat{\mathbf{e}}[t | t-1] &= \mathbf{x}[t] - \widehat{\mathbf{x}}[t | t-1] = \mathbf{x}[t] - \widehat{\boldsymbol{\theta}}[t-1]^T \boldsymbol{\phi}[t], \\
\text{Gain: } \mathbf{K}[t] &= \mathbf{P}[t-1] \boldsymbol{\phi}[t] \left(\lambda + \boldsymbol{\phi}[t]^T \mathbf{P}[t-1] \boldsymbol{\phi}[t]\right)^{-1}, \\
\text{Covariance update: } \mathbf{P}[t] &= \lambda^{-1} \left(\mathbf{P}[t-1] - \mathbf{K}[t] \boldsymbol{\phi}[t]^T \mathbf{P}[t-1]\right).
\end{aligned} \tag{15}$$

By recursively applying (15) to the data record $\mathbf{x}[t]$ ($t = 1, \dots, L$), the parameter matrix estimate at each time step can be obtained as $\widehat{\boldsymbol{\theta}}[t]^T$. Similarly, for the initialization of the method it is customary to set $\widehat{\boldsymbol{\theta}}[0]^T = \mathbf{0}$, $\mathbf{P}[0] = \gamma^{-1} \mathbf{I}_{k(n_a p_a + n_c p_c)}$. Obviously, the exponentially weighted RPLR-FS-TARMA method reduces to the RPLR-FS-TARMA method of (13), when $\lambda = 1$. Besides, the exponentially weighted RPLR-FS-TARMA method reduces to the exponentially weighted RPLR-TARMA method [9, 13], when $p_a = p_c = 1$, $G_{b_a(1)}[t] = G_{b_c(1)}[t] = 1$.

Some limitations of the RPLR-FS-TARMA method are relaxed by using the exponentially weighted mechanism.

Firstly, the estimate $\widehat{\boldsymbol{\theta}}[t]^T$ is recursively estimated at each time step, allowing the exponentially weighted RPLR-FS-TARMA method to operate in real-time. Secondly, even if L is unknown, the exponentially weighted RPLR-FS-TARMA method can be still implemented by selecting a suitable functional subspace with bounded basis sequences (which is the case when a trigonometric basis is used). Thirdly, the problem of data saturation is successfully overcome by putting more emphasis on recent data and deemphasizing data from remote past.

Unfortunately, the exponentially weighted RPLR-FS-TARMA method may be an impractical method due to its

fatal drawback. The FS-TARMA model structure calls for a constant parameter matrix. Therefore, if the estimate $\hat{\boldsymbol{\theta}}[t]^T$ is used as the projection coefficients of the time-dependent AR and MA parameter matrices of (6) at each time step, the structured parameter evolution will be partially violated. In other words, the exponentially weighted RPLR-FS-TARMA method loses those advantages of the DPE methods and sometimes even underperforms its UPE counterpart, that is, the exponentially weighted RPLR-TARMA method [9, 13].

4. Kernel Ridge Regression FS-TARMA Approach

In the previous section we have reviewed the RPLR-FS-TARMA approach which suffers from many limitations. This section proposes a kernel ridge regression FS-TARMA approach which can outperform the RPLR-FS-TARMA approach in terms of achievable accuracy and tracking capability.

4.1. Parameter Matrix Estimation. The FS-TARMA model may suffer from limited expressiveness, while this problem can be overcome by using kernel methods [24–26]. The main idea of kernel methods can be summarized as follows: project the inputs $\boldsymbol{\phi}[t]$ from the input space \mathbb{U} into a high-dimensional feature space $\boldsymbol{\varphi}(\boldsymbol{\phi}[t]) \in \mathbb{H}$ via a reproducing kernel and then apply the linear model in this feature space instead of directly on the inputs themselves. Based on this idea, we propose to rewrite the FS-TARMA model of (7) into the following form:

$$\begin{aligned} \mathbf{x}[t] &= \boldsymbol{\theta}^T \boldsymbol{\phi}[t] + \mathbf{e}[t] \implies \\ \mathbf{x}[t] &= \boldsymbol{\vartheta}^T \boldsymbol{\Theta}[t] + \mathbf{e}[t] \end{aligned} \quad (16)$$

with $\boldsymbol{\vartheta} \in \mathbb{R}^{\mathcal{D} \times k}$ being the time-invariant parameter matrix and $\boldsymbol{\Theta}[t] \triangleq \boldsymbol{\varphi}(\boldsymbol{\phi}[t]) \in \mathbb{R}^{\mathcal{D} \times 1}$ the corresponding regression vector. \mathcal{D} denotes the dimensionality of the feature space.

Given the regularization parameter $\gamma \geq 0$, the regularized least squares cost function is defined as

$$\min_{\boldsymbol{\vartheta}} \left[\frac{1}{2} \sum_{\tau=1}^t \|\mathbf{x}[\tau] - \boldsymbol{\vartheta}^T \boldsymbol{\Theta}[\tau]\|^2 + \frac{1}{2} \gamma \|\boldsymbol{\vartheta}\|^2 \right]. \quad (17)$$

By using kernel ridge regression [27, 28], the solution to this problem is given by

$$\hat{\boldsymbol{\vartheta}}[t] = (\gamma \mathbf{I}_{\mathcal{D}} + \tilde{\boldsymbol{\Theta}}[t] \tilde{\boldsymbol{\Theta}}[t]^T)^{-1} \tilde{\boldsymbol{\Theta}}[t] \tilde{\mathbf{x}}[t] \quad (18)$$

with $\tilde{\boldsymbol{\Theta}}[t] = [\boldsymbol{\Theta}[1], \boldsymbol{\Theta}[2], \dots, \boldsymbol{\Theta}[t]]$, $\tilde{\mathbf{x}}[t] = [\mathbf{x}[1], \mathbf{x}[2], \dots, \mathbf{x}[t]]^T$.

As the dimensionality of the feature space \mathcal{D} can be very high (even infinite), (18) may be not computable in practice. By using the matrix inversion lemma [9, 26]

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{C}^{-1} + \mathbf{DA}^{-1} \mathbf{B})^{-1} \mathbf{DA}^{-1} \quad (19)$$

with the identifications $\gamma \mathbf{I}_{\mathcal{D}} \rightarrow \mathbf{A}$, $\tilde{\boldsymbol{\Theta}}[t] \rightarrow \mathbf{B}$, $\mathbf{I}_t \rightarrow \mathbf{C}$, and $\tilde{\boldsymbol{\Theta}}[t]^T \rightarrow \mathbf{D}$, we can rewrite (18) into the following form:

$$\hat{\boldsymbol{\vartheta}}[t] = \tilde{\boldsymbol{\Theta}}[t] (\gamma \mathbf{I}_t + \tilde{\boldsymbol{\Theta}}[t]^T \tilde{\boldsymbol{\Theta}}[t])^{-1} \tilde{\mathbf{x}}[t], \quad (20)$$

where $\tilde{\boldsymbol{\Theta}}[t]^T \tilde{\boldsymbol{\Theta}}[t]$ is computable by using the ‘‘kernel trick’’ $\kappa(\boldsymbol{\phi}[i], \boldsymbol{\phi}[j]) = \boldsymbol{\Theta}[i]^T \boldsymbol{\Theta}[j]$ [24–28], as follows:

$$\tilde{\boldsymbol{\Theta}}[t]^T \tilde{\boldsymbol{\Theta}}[t] = \begin{bmatrix} \kappa(\boldsymbol{\phi}[1], \boldsymbol{\phi}[1]) & \cdots & \kappa(\boldsymbol{\phi}[1], \boldsymbol{\phi}[t]) \\ \vdots & \ddots & \vdots \\ \kappa(\boldsymbol{\phi}[t], \boldsymbol{\phi}[1]) & \cdots & \kappa(\boldsymbol{\phi}[t], \boldsymbol{\phi}[t]) \end{bmatrix} \quad (21)$$

with $\kappa(\cdot, \cdot)$ is called a reproducing kernel of \mathbb{H} [29], and its form can be finalized after the selection of kernel functions.

Denoting $\mathbf{Q}[t] = \gamma \mathbf{I}_t + \tilde{\boldsymbol{\Theta}}[t]^T \tilde{\boldsymbol{\Theta}}[t]$ and $\mathbf{k}[t] = \tilde{\boldsymbol{\Theta}}[t-1]^T \boldsymbol{\Theta}[t]$, we have

$$\mathbf{Q}[t] = \begin{bmatrix} \mathbf{Q}[t-1] & \mathbf{k}[t] \\ \mathbf{k}[t]^T & \gamma + \kappa(\boldsymbol{\phi}[t], \boldsymbol{\phi}[t]) \end{bmatrix}. \quad (22)$$

Obviously, the dimensionality of $\mathbf{Q}[t]$ increases as the number of samples increases. If the dimensionality of $\mathbf{Q}[t]$ is smaller than a predefined sliding-window length N , by using the block matrix inversion identity [26]

$$\begin{aligned} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} \\ = \begin{bmatrix} (\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1} & (\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1} \end{bmatrix} \end{aligned} \quad (23)$$

the inverse matrix of $\mathbf{Q}[t]$ can be computed by

$$\begin{aligned} \mathbf{Q}[t]^{-1} \\ = \begin{bmatrix} \mathbf{Q}[t-1]^{-1} + \mathbf{z}[t] \mathbf{z}[t]^T r[t]^{-1} & -\mathbf{z}[t] r[t]^{-1} \\ -\mathbf{z}[t]^T r[t]^{-1} & r[t]^{-1} \end{bmatrix} \end{aligned} \quad (24)$$

with $\mathbf{z}[t] = \mathbf{Q}[t-1]^{-1} \mathbf{k}[t]$ and $r[t] = \gamma + \kappa(\boldsymbol{\phi}[t], \boldsymbol{\phi}[t]) - \mathbf{z}[t]^T \mathbf{k}[t]$.

Assume that $N_1 \geq 1$ and $N_2 \geq 1$ denote, respectively, the rate of data forgetting and the degree of data overlap, and $N = N_1 + N_2$. When the dimensionality of $\mathbf{Q}[t]$ equals the sliding-window length N , we need to downsize the matrix $\mathbf{Q}[t]$ and obtain the corresponding downsized matrix $\mathbf{Q}^\# [t] \in \mathbb{R}^{N_2 \times N_2}$ by removing the first N_1 rows and columns of $\mathbf{Q}[t]$. The inverse matrix of $\mathbf{Q}^\# [t]$ can be computed by using the following downsized matrix inversion formula. Assume that the previously obtained $\mathbf{Q}[t]^{-1}$ is of the form

$$\mathbf{Q}[t]^{-1} = \begin{bmatrix} *_{N_1 \times N_1} & *_{N_1 \times N_2} \\ *_{N_2 \times N_1} & \mathbf{Q}^\# [t] \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}; \quad (25)$$

we have

$$\mathbf{Q}^\# [t]^{-1} = \mathbf{D} - \mathbf{CA}^{-1} \mathbf{B}. \quad (26)$$

Downsize $\tilde{\boldsymbol{\Theta}}[t] \leftarrow [\boldsymbol{\Theta}[t-N_2+1], \boldsymbol{\Theta}[t-N_2+2], \dots, \boldsymbol{\Theta}[t]]$, $\tilde{\mathbf{x}}[t] \leftarrow [\mathbf{x}[t-N_2+1], \mathbf{x}[t-N_2+2], \dots, \mathbf{x}[t]]^T$ and redefine

$\mathbf{Q}[t] \leftarrow \mathbf{Q}^\# [t]$; we have the newly defined matrix $\mathbf{Q}[t] \in \mathbb{R}^{N_2 \times N_2}$, whose dimensionality N_2 is smaller than N . Once the next data observation has been seen, $\mathbf{Q}[t + 1]$ can be upsized by (22) and its inverse matrix $\mathbf{Q}[t + 1]^{-1}$ can be computed by (24). In case the dimensionality of $\mathbf{Q}[t + N_1]$ equals N , first compute $\mathbf{Q}^\# [t + N_1]^{-1}$ by using the downsized matrix inversion formulas (25) and (26), and then redefine $\mathbf{Q}[t + N_1] \leftarrow \mathbf{Q}^\# [t + N_1]$.

Substituting $\mathbf{Q}[t]^{-1}$ into (20), the parameter matrix $\hat{\boldsymbol{\theta}}$ at each time step can be obtained by

$$\hat{\boldsymbol{\theta}}[t] = \tilde{\boldsymbol{\Theta}}[t] \mathbf{Q}[t]^{-1} \tilde{\mathbf{x}}[t]. \quad (27)$$

In order to avoid violating the structured parameter evolution of the FS-TARMA model, the parameter matrix $\hat{\boldsymbol{\theta}}[t]$ is updated once every N_1 data observations. In other words, the parameter matrix $\hat{\boldsymbol{\theta}}[t]$ is updated by (27) only when the dimensionality of $\mathbf{Q}[t]$ is equal to N . It should be noted that each $\hat{\boldsymbol{\theta}}[t]$ is estimated based on the data observations within $[t - N + 1, t]$, and each $\hat{\boldsymbol{\theta}}[t]$ can be subsequently used to compute the FS-TARMA model parameter matrix $\hat{\boldsymbol{\theta}}$ within $[t - N + 1, t]$. Furthermore, both of the adjacent parameter matrix estimates $\hat{\boldsymbol{\theta}}[t]$ and $\hat{\boldsymbol{\theta}}[t + N_1]$ can be used to compute $\hat{\boldsymbol{\theta}}$ at each arbitrary time step within $[t - N_2 + 1, t]$, and that is why N_2 is called the degree of data overlap.

Finally, the FS-TARMA model parameter matrix $\hat{\boldsymbol{\theta}}$ can be obtained by substituting (27) into (16):

$$\begin{aligned} \hat{\boldsymbol{\theta}}[t]^T &= \hat{\boldsymbol{\theta}}[t]^T \boldsymbol{\varphi}(\mathbf{I}_{k(n_a p_a + n_c p_c)}) \\ &= \tilde{\mathbf{x}}[t]^T \mathbf{Q}[t]^{-1} \begin{bmatrix} \kappa(\boldsymbol{\phi}[t - N + 1], \mathbf{I}_{k(n_a p_a + n_c p_c)}) \\ \kappa(\boldsymbol{\phi}[t - N + 2], \mathbf{I}_{k(n_a p_a + n_c p_c)}) \\ \vdots \\ \kappa(\boldsymbol{\phi}[t], \mathbf{I}_{k(n_a p_a + n_c p_c)}) \end{bmatrix}. \end{aligned} \quad (28)$$

4.2. Model Structure Selection. The model structure selection for the KRR-FS-TARMA method includes the selection of basis functions (a suitable functional subspace with proper dimensionalities p_a , p_c and indices $b_a(i)$, $b_c(i)$), the kernel function $\kappa(\cdot, \cdot)$, the regularization parameter γ , the sliding-window length N and the data forgetting rate N_1 , and the AR and MA orders n_a and n_c . The Akaike information criterion and the Bayesian information criterion may not be formally used in connection with methods that recursively estimate the AR and MA parameter matrices at each time step [11]. Therefore, once the functional subspaces, p_a , p_c , $b_a(i)$, $b_c(i)$, $\kappa(\cdot, \cdot)$, γ , N , and N_1 , have been initially selected, model orders including AR and MA orders will be selected based on the residual sum of squares (RSS) criterion [9, 11–13] as

$$\text{RSS} = \sum_{\tau=1}^L \hat{\mathbf{e}}[\tau]^T \cdot \hat{\mathbf{e}}[\tau] \quad (29)$$

with $\hat{\mathbf{e}}[\tau]$ being the estimated innovations (residual) sequence vector.

Model structure selection may be viewed as an iterative optimization problem that can be tackled via following three steps.

Step 1. Select a suitable functional subspace with proper dimensionalities p_a , p_c and indices $b_a(i)$, $b_c(i)$, a suitable kernel function $\kappa(\cdot, \cdot)$, the initial values of γ , N and N_1 based upon prior knowledge, physical understanding, and expected form of model parameter variation.

Step 2. Select model orders including AR and MA orders n_a and n_c via optimization procedure [14] based upon the minimization of the RSS used as ‘‘fitness’’ function:

- (a) AR and MA orders ($n_a = n_c$) within the search space are sequentially selected, and the best fit AR order \hat{n}_a can be determined when the decrease of RSS values is quite subtle.
- (b) MA orders within the search space $[0, \hat{n}_a]$ are sequentially selected under the previously obtained AR order \hat{n}_a , and the best fit MA order \hat{n}_c can be determined in the same way as that of AR order.

Step 3. Refine the candidate model structure obtained in Step 2 by further optimizing those parameters including p_a , p_c , $b_a(i)$, $b_c(i)$, γ , N , and N_1 .

The workflow of the KRR-FS-TARMA approach is summarized as follows:

Model structure selection: basis functions including p_a , p_c , $b_a(i)$, and $b_c(i)$, kernel function $\kappa(\cdot, \cdot)$, AR and MA orders n_a and n_c , regularization parameter γ , sliding-window length N , forgetting rate N_1 , and overlap degree $N_2 = N - N_1$

Initialization ($t = 1$):

$$\mathbf{Q}[1]^{-1} = (\gamma + \kappa(\boldsymbol{\phi}[1], \boldsymbol{\phi}[1]))^{-1},$$

$$\tilde{\boldsymbol{\Theta}}[1] = \boldsymbol{\Theta}[1], \quad (30)$$

$$\tilde{\mathbf{x}}[1] = \mathbf{x}[1]^T$$

Computation ($t > 1$):

if $\dim(\mathbf{Q}[t]) < N$

Upsizing: $\tilde{\boldsymbol{\Theta}}[t] \leftarrow [\tilde{\boldsymbol{\Theta}}[t - 1], \boldsymbol{\Theta}[t]]$, $\tilde{\mathbf{x}}[t] \leftarrow [\tilde{\mathbf{x}}[t - 1], \mathbf{x}[t]^T]$

Upsizing: obtain $\mathbf{Q}[t]$ according to Eq. (22)

Compute $\mathbf{Q}[t]^{-1}$ according to Eq. (24)

elseif $\dim(\mathbf{Q}[t]) = N$

Compute the FS-TARMA model parameter matrix $\hat{\boldsymbol{\theta}}[t]^T$ according to Eq. (28)

Compute the time-dependent AR parameter matrix $\hat{\mathbf{A}}_i[t]$ according to Eq. (6)

Compute the modal parameters according to Eq. (2), (3) and (4)

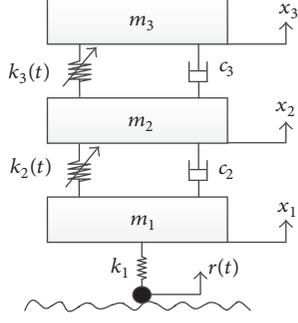


FIGURE 1: A time-varying structural system.

TABLE 1: Time-varying system parameters.

Property	Value
Mass (kg)	$m_1 = 0.5, m_2 = 1.5, m_3 = 1.0$
Damping (N/(m/s))	$c_2 = 0.5, c_3 = 0.3$
Stiffness (N/m)	$k_1 = 300$ $k_2(t) = 100 + 60 \sin\left(\frac{2\pi t}{170}\right) + 20 \sin\left(\frac{2\pi t}{85}\right)$ $k_3(t) = 120 + 72 \sin\left(\frac{2\pi t}{141.67}\right) + 24 \sin\left(\frac{2\pi t}{94.44}\right)$

Downsizing: obtain $\mathbf{Q}^\# [t]$ by removing the first N_1 rows and columns of $\mathbf{Q} [t]$
 Compute $\mathbf{Q}^\# [t]^{-1}$ according to Eq. (25) and (26)
 Downsizing:

$$\begin{aligned} \tilde{\Theta} [t] &\leftarrow [\Theta [t - N_2 + 1], \Theta [t - N_2 + 2], \dots, \Theta [t]] \\ \tilde{\mathbf{x}} [t] &\leftarrow [\mathbf{x} [t - N_2 + 1], \mathbf{x} [t - N_2 + 2], \dots, \mathbf{x} [t]]^T \end{aligned} \quad (31)$$

Redefining: $\mathbf{Q} [t] \leftarrow \mathbf{Q}^\# [t]$
 end if

5. Numerical Validation

5.1. The Time-Varying Structural System. For validating the proposed method for output-only modal parameter recursive estimation of time-varying structures, the three-degree-of-freedom structural system given by [11, 30] is used in this section, as shown in Figure 1. This system represents a simplified model of one-quarter of a truck vehicle [11, 30], which is subject to a unobservable, stationary, Gaussian, zero mean, and uncorrelated road excitation $\mathbf{r}(t)$. The numerical values of system parameters are given in Table 1.

Given a particular excitation realization, the vibration displacement responses for $t \in [-16.67, 850]$ s are obtained via the Runge-Kutta 4/5 method characterized by variable integration step and subsequently recorded at a sampling frequency 12 Hz [11, 30]. The initial 200 samples of the obtained responses for $t \in [-16.67, 0]$ s are discarded to minimize initialization effects. The rest of the obtained responses for $t \in [0, 850]$ s are corrupted by Gaussian white noise

with the signal-to-noise ratio of 20 dB. In a word, each noise-corrupted response realization, sampled at 12 Hz, is 850 seconds long, producing 10201 sample-long versions of the nonstationary vibration signals. This procedure is repeated for each Monte Carlo run (number of runs: 30), with a different excitation realization used each time.

5.2. Modal Parameter Estimation Results. The proposed KRR-FS-TARMA method is, along with the existing RPLR-FS-TARMA method, applied to estimate modal parameters of the time-varying numerical system. The proposed method is assessed against the theoretical modal parameters, and further comparisons with the RPLR-FS-TARMA method are also made via Monte Carlo experiments.

5.2.1. The RPLR-FS-TARMA Method Based Estimation. The functional basis spanned by trigonometric functions of the form [11, 30]

$$\begin{aligned} G_0 [t] &= 1, \\ G_{2k-1} [t] &= \sin \left[\frac{k\pi (t-1)}{L-1} \right], \\ G_{2k} [t] &= \cos \left[\frac{k\pi (t-1)}{L-1} \right], \end{aligned} \quad (32)$$

$k = 1, 2, \dots,$

is selected. The corresponding dimensionalities and indices are, respectively, $p_a = 40$, $b_a = [0 : 39]$, and $p_c = 6$, $b_c = [0 : 5]$. The value of $\gamma = 10^{-4}$ is selected. Model order selection is achieved via the RSS criterion, as the RPLR-FS-TARMA method is based on the class of recursive TARMA representations. Figure 2(a) depicts the RSS normalized by the sum of squares of the signal samples (series sum of squares, SSS) of the obtained RPLR-FS-TARMA (p, q) ($p = q = 1, 2, 3, 4$), which suggests AR order $n_a = 2$. Figure 2(b) depicts the RSS normalized by the SSS of the obtained RPLR-TARMA (p, q) ($p = 2, q = 0, 1, 2$), which suggests MA order $n_c = 2$. Figure 3 depicts the “frozen-time” RPLR-FS-TARMA based modal parameter estimates (with damping ratio $0 < \zeta_i [t] < 0.1$) from the 30 Monte Carlo tests, along with their theoretical counterparts. The blue dots notate the modal parameter estimates from the 30 tests, and the green dashed lines notate the theoretical modal parameters.

5.2.2. The Exponentially Weighted RPLR-TARMA Method Based Estimation. The functional basis given by (32) is selected with $p_a = p_c = 1$, $b_a = b_c = 0$. In other words, the exponentially weighted RPLR-FS-TARMA method reduces to its UPE counterpart, that is, the exponentially weighted RPLR-TARMA method. The values of $\gamma = 10^{-4}$ and $\lambda = 0.988$ are selected. Figure 4(a) depicts the RSS normalized by the SSS of the obtained RPLR-TARMA (p, q) ($p = q = 1, 2, 3, 4$), which suggests AR order $n_a = 2$. Figure 4(b) depicts the RSS normalized by the SSS of the obtained RPLR-TARMA (p, q) ($p = 2, q = 0, 1, 2$), which suggests MA order $n_c = 2$. Figure 5 depicts the “frozen-time” RPLR-TARMA based modal parameter estimates (with damping ratio $0 <$

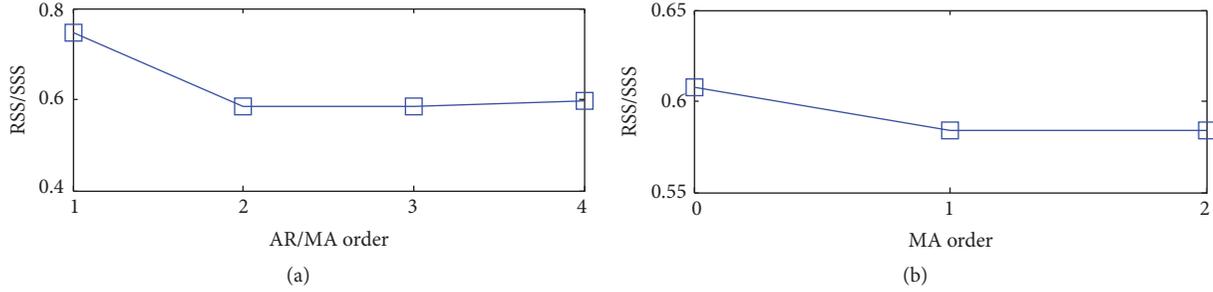


FIGURE 2: RPLR-FS-TARMA model order selection: (a) AR order and (b) MA order.

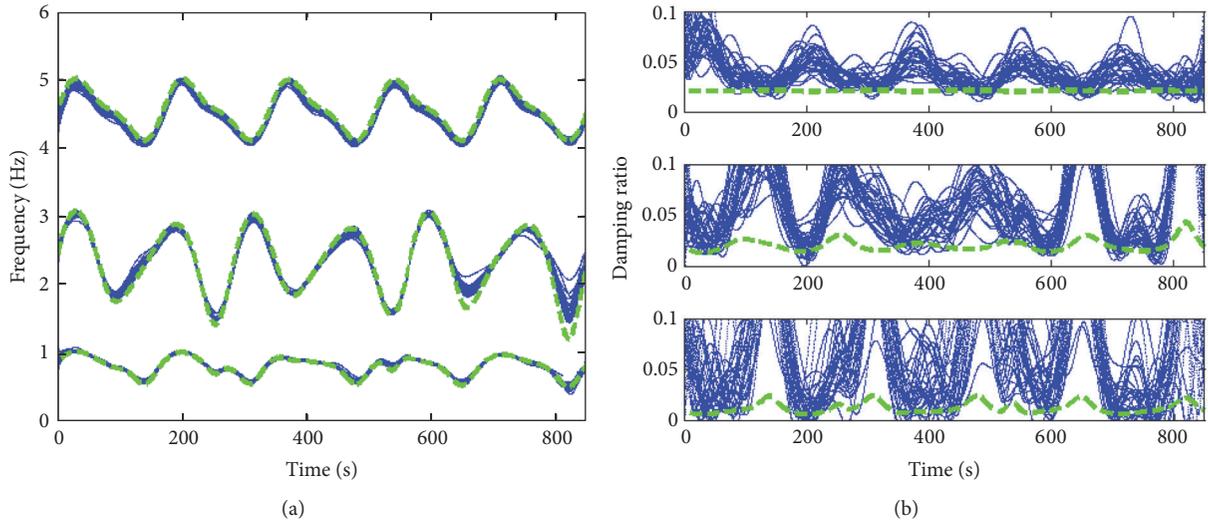


FIGURE 3: RPLR-FS-TARMA based modal parameter Monte Carlo estimates (30 realizations): (a) natural frequency estimates and (b) damping ratio estimates.

$\zeta_i[t] < 0.1$) from the 30 Monte Carlo tests, along with their theoretical counterparts.

5.2.3. The KRR-FS-TARMA Method Based Estimation. The functional basis given by (32) is selected with $p_a = 20$, $b_a = [0 : 19]$ and $p_c = 3$, $b_c = [0 : 2]$. The simplest linear kernel $\kappa(\phi[i], \phi[j]) = \phi[i]^T \phi[j]$ is here selected. The values of $\gamma = 10^{-3}$, $N = 1020$, and $N_1 = 510$ are selected. Figure 6(a) depicts the RSS normalized by the SSS of the obtained KRR-FS-TARMA (p, q) ($p = q = 1, 2, 3, 4$), which suggests AR order $n_a = 2$. Figure 6(b) depicts the RSS normalized by the SSS of the obtained KRR-FS-TARMA (p, q) ($p = 2, q = 0, 1, 2$), which suggests MA order $n_c = 2$. Figure 7 depicts the “frozen-time” KRR-FS-TARMA based modal parameter estimates (with damping ratio $0 < \zeta_i[t] < 0.1$) from the 30 Monte Carlo tests, along with their theoretical counterparts.

5.2.4. Comparisons. From the natural frequency estimates shown in Figures 3 and 5 it may be observed that the RPLR-FS-TARMA based natural frequency estimates exhibit better accuracy and tracking than its UPE counterpart obtained by the exponentially weighted RPLR-TARMA method. From the natural frequency estimates shown in Figures 3, 5, and 7 it may be further observed that the KRR-FS-TARMA based

natural frequency estimates exhibit best accuracy and tracking than its RPLR-FS-TARMA and exponentially weighted RPLR-TARMA counterparts. The damping ratios are filtered by $0 < \zeta_i[t] < 0.1$, and their estimates are generally much less accurate, which is common for almost all methods.

In order to better compare these methods by using Monte Carlo natural frequency estimates, a mean absolute error (MAE) is introduced as

$$\text{MAE} = \frac{1}{R} \sum_{i=1}^R \left(\frac{1}{L} \sum_{t=1}^L |f(t) - \hat{f}_i(t)| \right) \quad (33)$$

with $f(t)$ designating the baseline value of the natural frequency of interest, $\hat{f}_i(t)$ its estimated value based on the i th Monte Carlo run, L the length of samples, and R the number of Monte Carlo runs. Figure 8 illustrates the MAE values obtained by each method. Evidently, the identified KRR-FS-TARMA model attains best overall performance compared to its RPLR-FS-TARMA and exponentially weighted RPLR-TARMA counterparts.

6. Experimental Validation

6.1. The Laboratory Time-Varying Structure. A coupled moving mass and simply supported beam time-varying structure

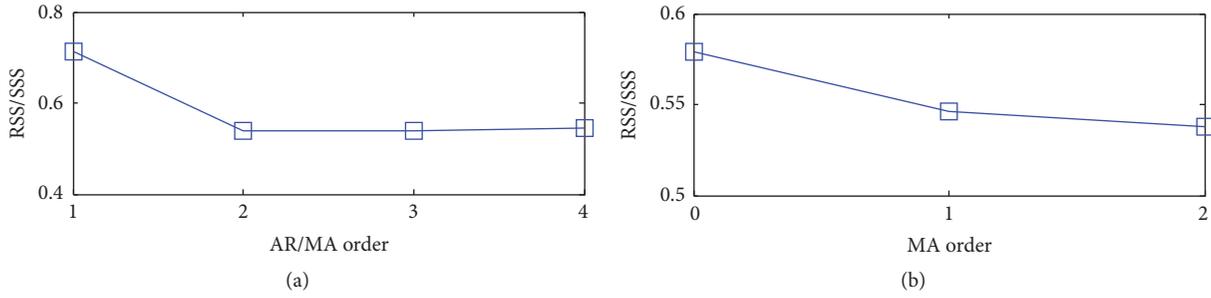


FIGURE 4: Exponentially weighted RPLR-TARMA model order selection: (a) AR order and (b) MA order.

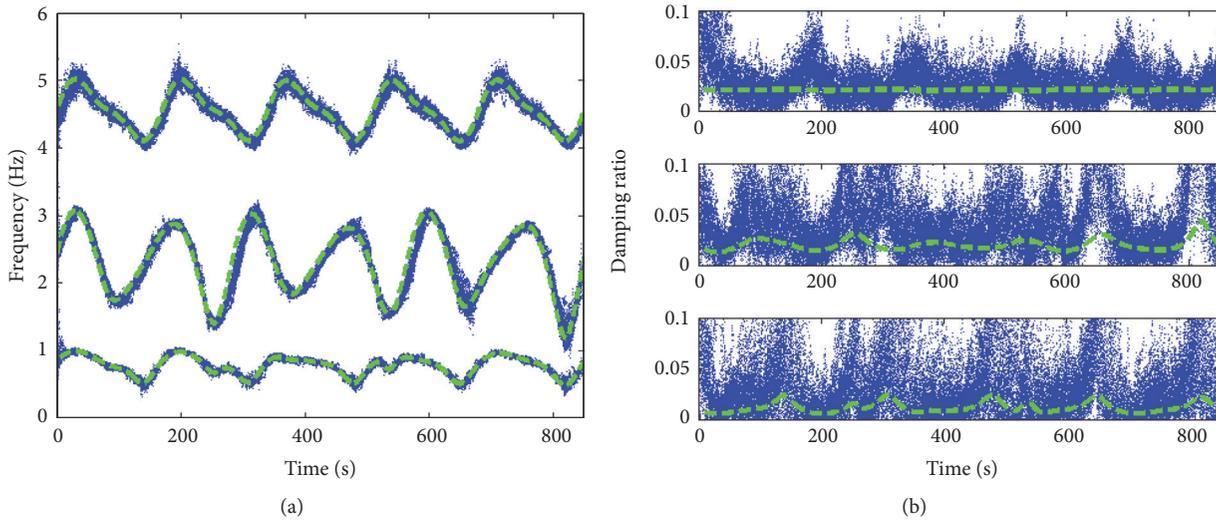


FIGURE 5: Exponentially weighted RPLR-TARMA based modal parameter Monte Carlo estimates (30 realizations): (a) natural frequency estimates and (b) damping ratio estimates.

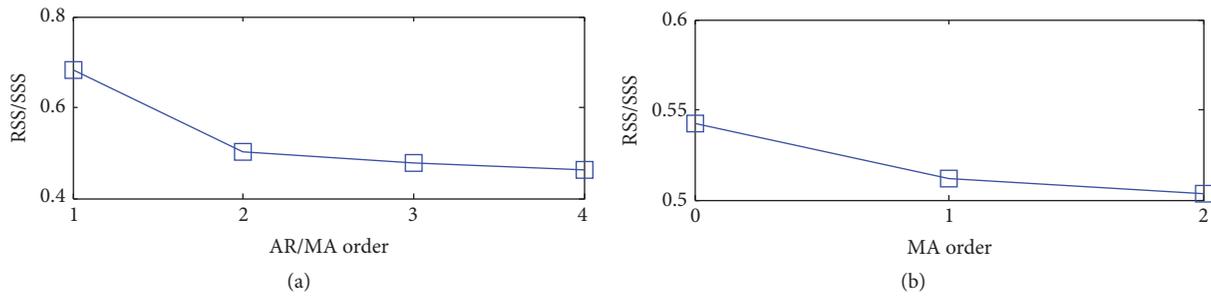


FIGURE 6: KRR-FS-TARMA model order selection: (a) AR order and (b) MA order.

[31] is used for experimental validation of the proposed KRR-FS-TARMA approach. The experimental system is composed of the test structure, an exciter system, force and motion transducers, measurement and analysis systems, control systems, and boundary conditions. Figure 9 shows the schematic diagram of the experimental system and its set-up.

A nonstationary experiment is conducted with the moving mass sliding on the beam with a uniform velocity of 0.2 m/s. Figure 10(a) shows its motion form, that is, the position over time. In the experiment, no external excitation is generated to excite the coupled time-varying system (the

shaker shown in Figure 9 is not in use), and the system is only excited by the motion of the moving mass. Fifteen accelerometers are used to measure the acceleration responses of the simply supported beam at fifteen uniformly distributed positions along the axial direction of the beam from left to right, as shown in Figure 9. The vertical acceleration signals, sampled at 256 Hz, are 32 seconds long, producing 8192 sample-long versions of the nonstationary vibration signals. Figure 10(b) shows the first four baseline natural frequencies of the coupled time-varying system, obtained by using the “frozen-time” technique [31].

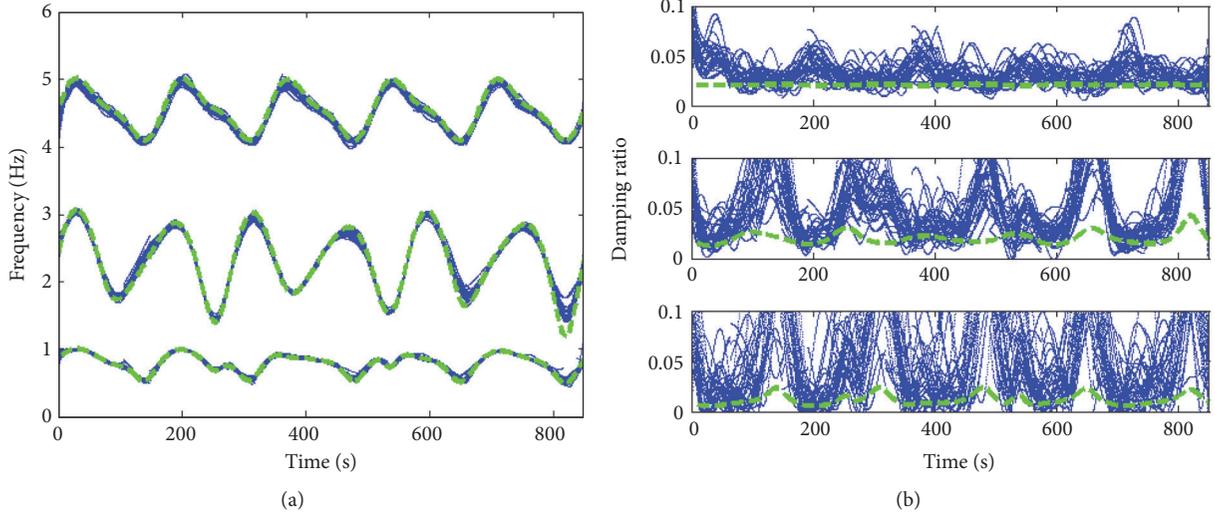


FIGURE 7: KRR-FS-TARMA based modal parameter Monte Carlo estimates (30 realizations): (a) natural frequency estimates and (b) damping ratio estimates.

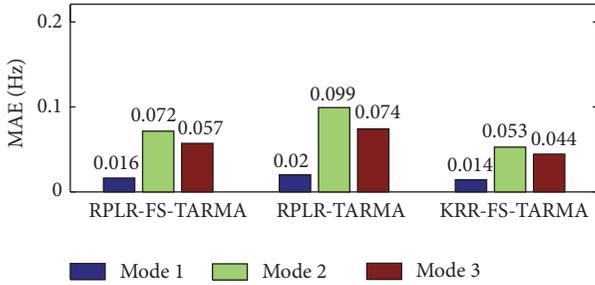


FIGURE 8: Monte Carlo comparison of different methods: MAE of natural frequency estimates.

6.2. Modal Parameter Estimation Results. In order to effectively and reliably assess and compare various methods, the study in this section is also carried out based upon Monte Carlo runs (number of runs: 30). In the same manner as previously introduced, 30 tests are repeatedly carried out, and the coupled time-varying system undergoes the same variation in each test. The modal parameter estimation of the experimental time-varying structure is subsequently considered based on three methods including the RPLR-FS-TARMA method, the exponentially weighted RPLR-TARMA method, and the proposed KRR-FS-TARMA method.

6.2.1. The RPLR-FS-TARMA Method Based Estimation. The functional basis given by (32) is selected with $p_a = 50$, $b_a = [0 : 49]$ and $p_c = 1$, $b_c = 0$. The value of $\gamma = 10^{-4}$ is selected. Figure 11(a) depicts the RSS normalized by the SSS of the obtained RPLR-FS-TARMA (p, q) ($p = q = 1, 2, 3, 4$), which suggests AR order $n_a = 2$. Figure 11(b) depicts the RSS normalized by the SSS of the obtained RPLR-FS-TARMA (p, q) ($p = 2$, $q = 0, 1, 2$), which suggests MA order $n_c = 1$. Figure 12 depicts the “frozen-time” RPLR-FS-TARMA based modal parameter estimates (with damping ratio $0 < \zeta_i[t] < 0.1$) from the 30 Monte Carlo tests, along with their baseline counterparts.

6.2.2. The Exponentially Weighted RPLR-TARMA Method Based Estimation. The exponentially weighted RPLR-TARMA method is employed for modal parameter estimation of the laboratory structure with $\gamma = 10^{-4}$ and $\lambda = 0.964$. Figure 13(a) depicts the RSS normalized by the SSS of the obtained RPLR-TARMA (p, q) ($p = q = 1, 2, 3, 4$), which suggests AR order $n_a = 2$. Figure 13(b) depicts the RSS normalized by the SSS of the obtained RPLR-TARMA (p, q) ($p = 2$, $q = 0, 1, 2$), which suggests MA order $n_c = 0$. Figure 14 depicts the “frozen-time” RPLR-TARMA based modal parameter estimates (with damping ratio $0 < \zeta_i[t] < 0.1$) from the 30 Monte Carlo tests, along with their baseline counterparts.

6.2.3. The KRR-FS-TARMA Method Based Estimation. The functional basis given by (32) is selected with $p_a = 40$, $b_a = [0 : 39]$ and $p_c = 1$, $b_c = 0$. The linear kernel functions, $\gamma = 10^{-4}$, $N = 1024$, and $N_1 = 512$ are selected. Figure 15(a) depicts the RSS normalized by the SSS of the obtained KRR-FS-TARMA (p, q) ($p = q = 1, 2, 3, 4$), which suggests AR order $n_a = 2$. Figure 16(b) depicts the RSS normalized by the SSS of the obtained KRR-FS-TARMA (p, q) ($p = 2$, $q = 0, 1, 2$), which suggests MA order $n_c = 2$. Figure 16 depicts the “frozen-time” KRR-FS-TARMA based modal parameter estimates (with damping ratio $0 < \zeta_i[t] < 0.1$) from the 30 Monte Carlo tests, along with their baseline counterparts.

6.2.4. Comparisons. From the natural frequency estimates shown in Figure 14 it may be observed that significant scatter is evidently associated with the estimated modes, and a number of false modes appear, especially in the frequency range near the third and fourth modes. Compared with the exponentially weighted RPLR-TARMA based estimates in Figure 14, the RPLR-FS-TARMA based natural frequency estimates appear more smooth and less false modes are exhibited, as shown in Figure 12. Compared with the RPLR-FS-TARMA based estimates in Figure 12, the KRR-FS-TARMA based

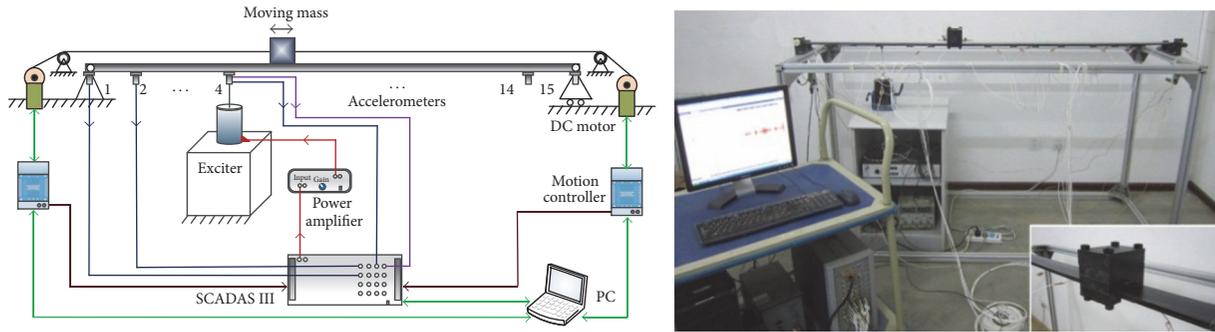


FIGURE 9: Schematic diagram of the experimental system and photo of its set-up.

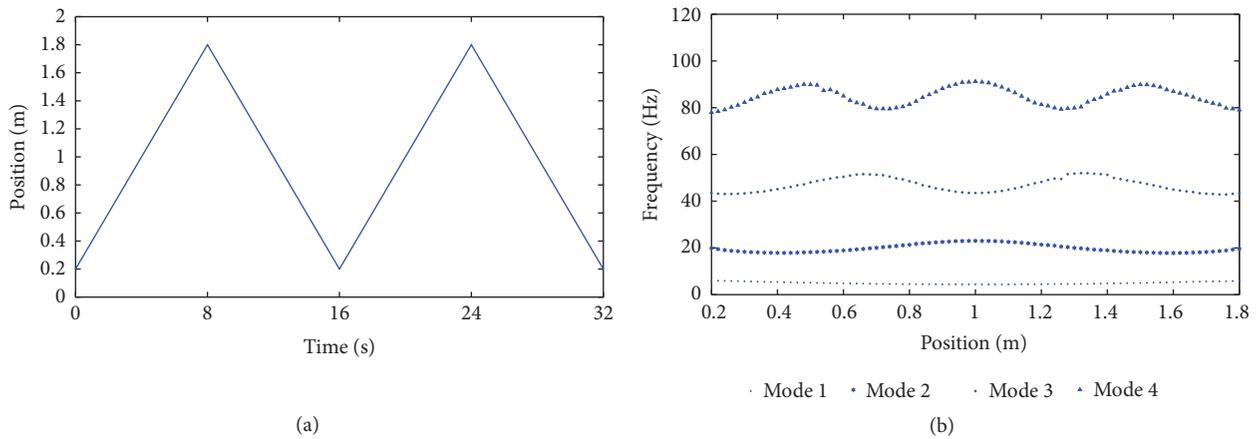


FIGURE 10: The motion form of the moving mass and the corresponding first four baseline natural frequencies of the experimental system: (a) motion form and (b) baseline natural frequencies.

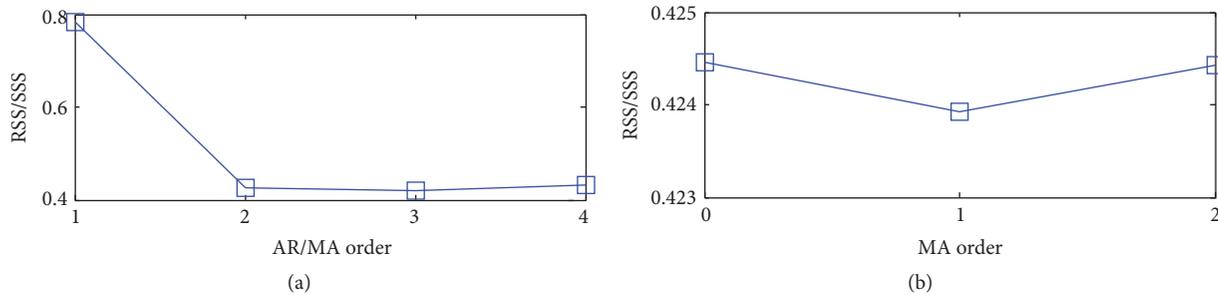


FIGURE 11: RPLR-FS-TARMA model order selection: (a) AR order and (b) MA order.

natural frequency estimates shown in Figure 16 exhibit better accuracy and tracking, especially for the third and fourth modes. The damping ratios are filtered by $0 < \zeta_i[t] < 0.1$, and their estimates are generally much less accurate. Therefore quantitative analysis is carried out based on the natural frequency estimates instead of the damping ratio estimates. Figure 17 illustrates the MAE values obtained by each method. Compared with the exponentially weighted RPLR-TARMA, its DPE counterparts including the RPLR-FS-TARMA method and the KRR-FS-TARMA method can accurately track the third and fourth modes, but not for the first mode. However, the identified KRR-FS-TARMA model

attains better overall performance than its RPLR-FS-TARMA counterpart.

7. Conclusions

This work focuses on the output-only modal parameter estimation of time-varying structures by using recursive FS-TARMA approach. A KRR-FS-TARMA approach was proposed by exclusively using the available response measurements. The proposed method was numerically and experimentally validated and further assessed against the existing RPLR-FS-TARMA method via Monte Carlo experiments.

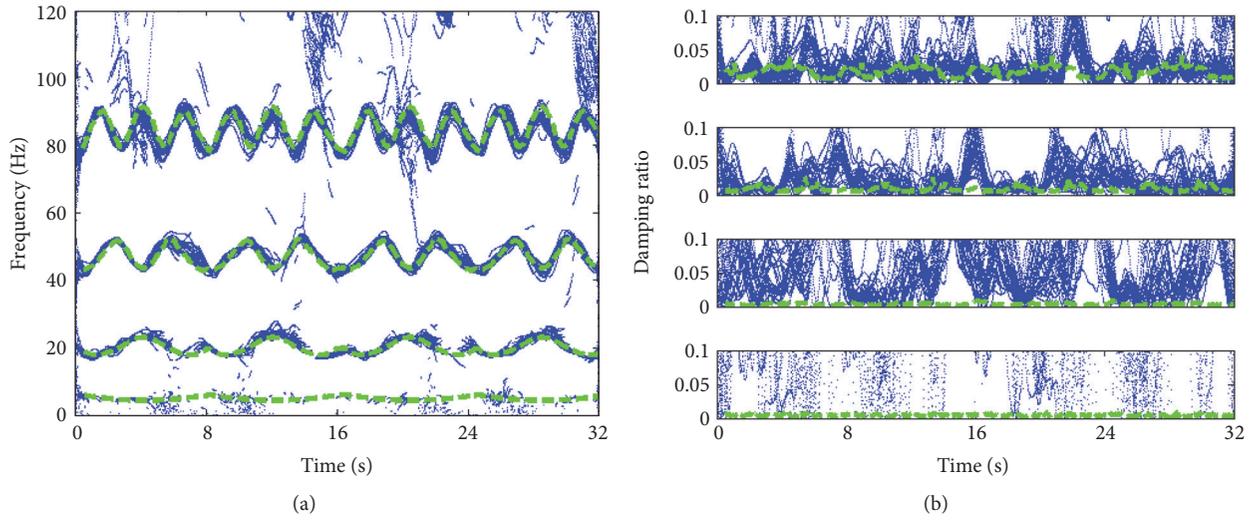


FIGURE 12: RPLR-FS-TARMA based modal parameter Monte Carlo estimates (30 realizations): (a) natural frequency estimates and (b) damping ratio estimates.

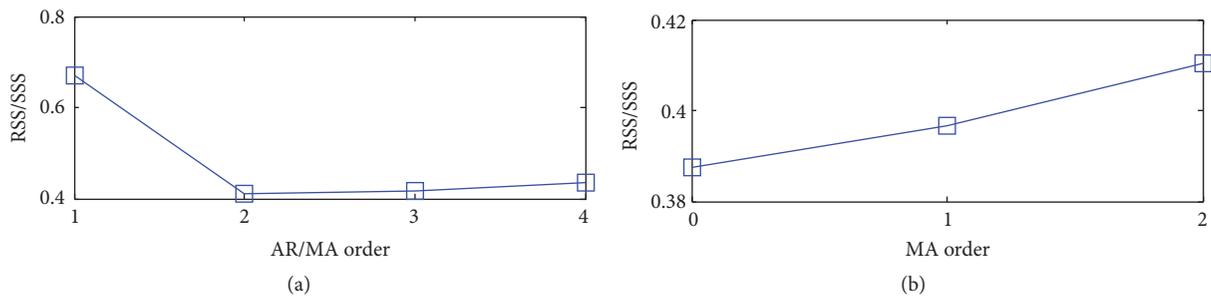


FIGURE 13: Exponentially weighted RPLR-TARMA model order selection: (a) AR order and (b) MA order.

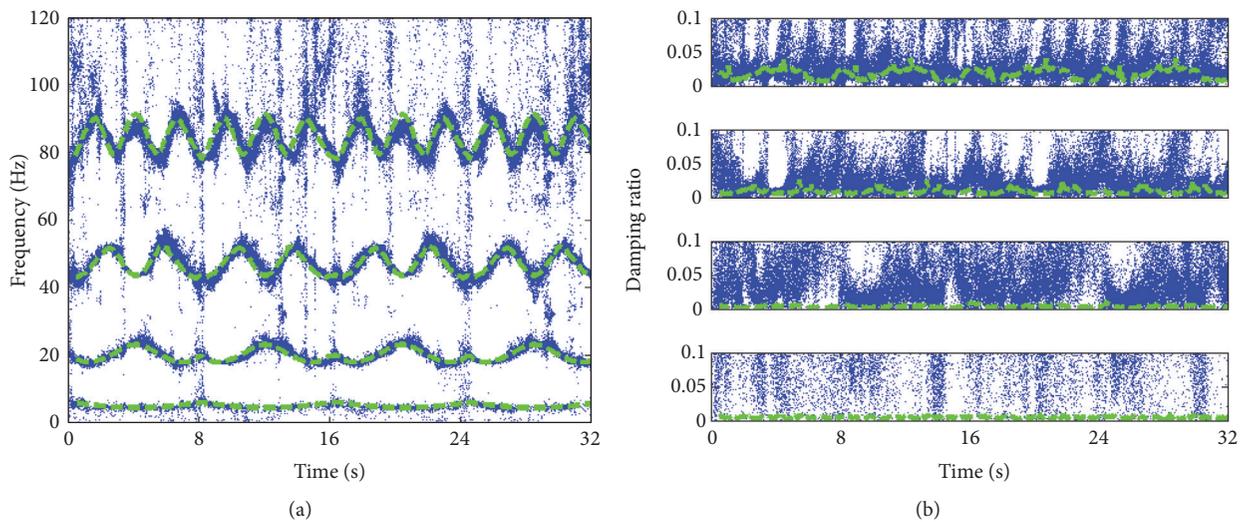


FIGURE 14: Exponentially weighted RPLR-TARMA based modal parameter Monte Carlo estimates (30 realizations): (a) natural frequency estimates and (b) damping ratio estimates.

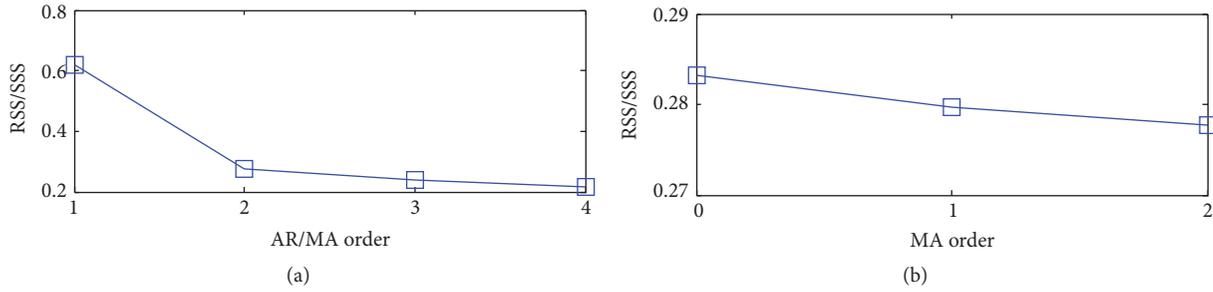


FIGURE 15: KRR-FS-TARMA model order selection: (a) AR order and (b) MA order.

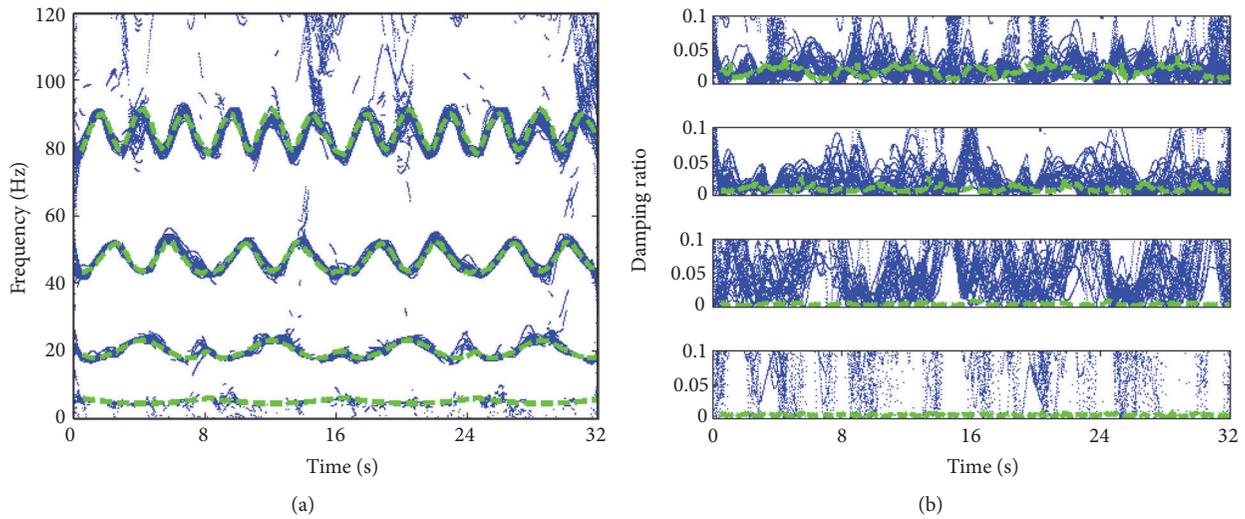


FIGURE 16: KRR-FS-TARMA based modal parameter Monte Carlo estimates (30 realizations): (a) natural frequency estimates and (b) damping ratio estimates.

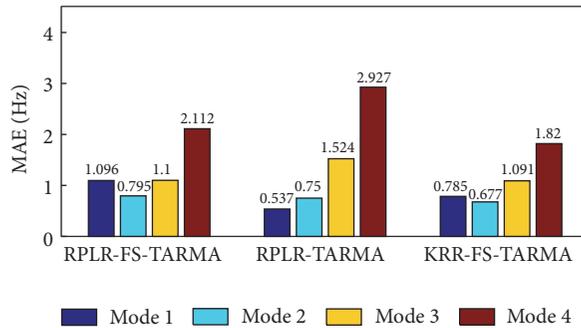


FIGURE 17: Monte Carlo comparison of different methods: MAE of natural frequency estimates.

The comparative assessments demonstrate the proposed KRR-FS-TARMA approach has superior achievable accuracy and enhanced tracking capability that are typical of the DPE methods but, at the same time, has improved flexibility of recursive estimation, which is typical of the UPE methods.

Abbreviations

AR: Autoregressive
 MA: Moving average
 TARMA: Time-dependent autoregressive moving average (model)

FS-TARMA: Functional series TARMA (model)
 RPLR-TARMA: Recursive pseudolinear regression TARMA (method)
 RPLR-FS-TARMA: Recursive pseudolinear regression FS-TARMA (method)
 KRR-FS-TARMA: Kernel ridge regression FS-TARMA (method)
 UPE: Unstructured parameter evolution
 SPE: Stochastic parameter evolution
 DPE: Deterministic parameter evolution
 RSS: Residual sum of squares

SSS: Series sum of squares

MAE: Mean absolute error.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The authors acknowledge the support from the China Scholarship Council and National Natural Science Foundation of China (Grant no. 11402022).

References

- [1] S. W. Doebling, C. R. Farrar, and M. B. Prime, "A summary review of vibration-based damage identification methods," *Shock and Vibration Digest*, vol. 30, no. 2, pp. 91–105, 1998.
- [2] E. P. Carden and P. Fanning, "Vibration based condition monitoring: a review," *Structural Health Monitoring*, vol. 3, no. 4, pp. 355–377, 2004.
- [3] S. D. Fassois and J. S. Sakellariou, "Time-series methods for fault detection and identification in vibrating structures," *Mathematic, Physical & Engineering Sciences*, vol. 365, pp. 411–448, 2007.
- [4] W. Fan and P. Qiao, "Vibration-based damage identification methods: A Review And Comparative Study," *Structural Health Monitoring*, vol. 10, no. 1, pp. 83–111, 2011.
- [5] M. D. Spiridonakos and S. D. Fassois, "An FS-TAR based method for vibration-response-based fault diagnosis in stochastic time-varying structures: experimental application to a pick-and-place mechanism," *Mechanical Systems and Signal Processing*, vol. 38, no. 1, pp. 206–222, 2013.
- [6] W. Heylen, S. Lammens, and P. Sas, *Modal Analysis Theory and Testing*, Katholieke Universiteit Leuven, Leuven, Belgium, 2007.
- [7] H. D'Angelo, *Linear Time-Varying Systems: Analysis and Synthesis*, Allyn & Bacon, Boston, Mass, USA, 1970.
- [8] Y. S. Shmaliy, *Continuous-Time Systems*, Springer, Dordrecht, The Netherlands, 2007.
- [9] L. Ljung, *System Identification—Theory for the User*, Prentice Hall, Upper Saddle River, NJ, USA, 2nd edition, 1999.
- [10] M. Niedzwiecki, *Identification of Time-Varying Processes*, John Wiley & Sons, New York, NY, USA, 2000.
- [11] A. G. Poulimenos and S. D. Fassois, "Parametric time-domain methods for non-stationary random vibration modelling and analysis—a critical survey and comparison," *Mechanical Systems and Signal Processing*, vol. 20, no. 4, pp. 763–816, 2006.
- [12] M. D. Spiridonakos and S. D. Fassois, "Non-stationary random vibration modelling and analysis via functional series time-dependent ARMA (FS-TARMA) models—a critical survey," *Mechanical Systems and Signal Processing*, vol. 47, no. 1–2, pp. 175–224, 2014.
- [13] M. D. Spiridonakos and S. D. Fassois, "Parametric identification of a time-varying structure based on vector vibration response measurements," *Mechanical Systems and Signal Processing*, vol. 23, no. 6, pp. 2029–2048, 2009.
- [14] W. Yang, L. Liu, S.-D. Zhou, and Z.-S. Ma, "Moving Kriging shape function modeling of vector TARMA models for modal identification of linear time-varying structural systems," *Journal of Sound and Vibration*, vol. 354, pp. 254–277, 2015.
- [15] X. Y. Xie and R. J. Evans, "Discrete time stochastic adaptive control for time-varying systems," *IEEE Transactions on Automatic Control*, vol. 29, no. 7, pp. 638–640, 1984.
- [16] Z. Li, "Discrete-time adaptive control of deterministic fast time-varying systems," *IEEE Transactions on Automatic Control*, vol. 32, no. 5, pp. 444–447, 1987.
- [17] M. Niedzwiecki, "Recursive functional series modeling estimators for identification of time-varying plants—more bad news than good?" *IEEE Transactions on Automatic Control*, vol. 35, no. 5, pp. 610–616, 1990.
- [18] M. K. Tsatsanis and G. B. Giannakis, "Modelling and equalization of rapidly fading channels," *International Journal of Adaptive Control and Signal Processing*, vol. 10, no. 2–3, pp. 159–176, 1996.
- [19] A. Joensen, H. Madsen, H. A. Nielsen, and T. S. Nielsen, "Tracking time-varying parameters with local regression," *Automatica*, vol. 36, no. 8, pp. 1199–1204, 2000.
- [20] M. Niedzwiecki and T. Klaput, "Fast recursive basis function estimators for identification of time-varying processes," *IEEE Transactions on Signal Processing*, vol. 50, no. 8, pp. 1925–1934, 2002.
- [21] G. G. Walter, *Wavelets and Other Orthogonal Systems with Applications*, Taylor & Francis, Boca Raton, Fla, USA, 1994.
- [22] H. Lütkepohl, *Handbook of Matrices*, John Wiley & Sons, New York, NY, USA, 1996.
- [23] S. Yang, Y. Wu, and J. Xuan, *Time Series Analysis in Engineering Application*, Huazhong University of Science and Technology Press, Wuhan, China, 2nd edition, 2007.
- [24] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, USA, 2006.
- [25] C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning*, Adaptive Computation and Machine Learning, MIT Press, Cambridge, Massachusetts, USA, 2006.
- [26] W. Liu, J. C. Principe, and S. Haykin, *Kernel Adaptive Filtering: A Comprehensive Introduction*, John Wiley & Sons, New Jersey, NJ, USA, 2010.
- [27] S. Van Vaerenbergh, J. Via, and I. Santamaría, "Nonlinear system identification using a new sliding-window Kernel RLS algorithm," *Journal of Communications*, vol. 2, no. 3, pp. 1–8, 2007.
- [28] J. A. K. Suykens, M. Signoretto, and A. Argyriou, *Regularization, Optimization, Kernels, and Support Vector Machines*, CRC Press, Boca Raton, Fla, USA, 2014.
- [29] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, pp. 337–404, 1950.
- [30] A. Poulimenos, M. Spiridonakos, and S. Fassois, "Parametric time-domain methods for non-stationary random vibration identification and analysis: an overview and comparison," in *Proceedings of the International Conference on Noise and Vibration Engineering (ISMA '06)*, pp. 2885–2905, Leuven, Belgium, 2006.
- [31] Z.-S. Ma, L. Liu, S.-D. Zhou, and W. Yang, "Modal parameter estimation of the coupled moving-mass and beam time-varying system," in *Proceedings of the International Conference on Noise and Vibration Engineering (ISMA '14)*, pp. 587–596, Leuven, Belgium, 2014.

Research Article

Electromagnetic and Mechanical Characteristics Analysis of a Flat-Type Vertical-Gap Passive Magnetic Levitation Vibration Isolator

Baoquan Kou, Yiheng Zhou, Xiaobao Yang, Feng Xing, and He Zhang

School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, China

Correspondence should be addressed to Baoquan Kou; koubq@hit.edu.cn

Received 23 August 2016; Revised 22 November 2016; Accepted 5 December 2016

Academic Editor: Lu Chen

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

In this paper, we describe a flat-type vertical-gap passive magnetic levitation vibration isolator (FVPMLVI) for active vibration isolation system (AVIS). A dual-stator scheme and a special stator magnet array are adopted in the proposed FVPMLVI, which has the effect of decreasing its natural frequency, and this enhances the vibration isolation capability of the FVPMLVI. The structure, operating principle, analytical model, and electromagnetic and mechanical characteristics of the FVPMLVI are investigated. The relationship between the force characteristics (levitation force, horizontal force, force ripple, and force density) and major structural parameters (width and thickness of stator and mover magnets) is analyzed by finite element method. The experiment result is in good agreement with the theoretical analysis.

1. Introduction

Active vibration isolation system (AVIS), which integrates actuators with passive gravity compensation devices, can effectively improve the accuracy of measuring and machining equipment. Therefore, active vibration isolation systems have been widely used in many advanced industrial applications such as microscopy and lithography. As an important component of active vibration isolation system, the passive gravity compensation device plays the role of supporting and vibration isolation, and such characteristics like high force density and low natural frequency are required. More importantly, lower natural frequency leads to wider vibration isolation bandwidth and lower vibration transmissibility. Owing to the advantage of low natural frequency, air springs are widely adopted as passive gravity compensation devices in many active vibration isolation systems. However, some ultraprecision equipment must operate in a moderate vacuum, for example, extreme ultraviolet lithography [1]. It is difficult for air springs to be applied in vacuum environment, because air springs need compressed gas [2]. To solve the problem, passive magnetic levitation vibration isolator is used as a substitute for air spring.

Passive magnetic levitation vibration isolators generate levitation force by interaction between magnets and have drew increased interest in recent years due to their feature of vacuum compatibility. Puppini and Fratello proposed a vibration isolation apparatus composed of four passive magnetic levitation vibration isolators in 2002; its natural frequency is 6.1 Hz at a load of 4 kg and 5.8 Hz at a load of 15 kg [2]. Zhu et al. proposed a passive magnetic levitation vibration isolator composed of ring-shaped permanent magnets; they studied its axial force and stiffness characteristics, and its natural frequency is about 6 Hz [3]. Robertson et al. proposed a multipole array passive magnetic levitation vibration isolator [4] and studied the design method of passive magnetic levitation vibration isolators [5, 6]. Lomonova et al. also studied the design method of passive magnetic levitation vibration isolators [7–9]. Zhu et al. proposed a vibration isolator composed of permanent magnets and rubber ligaments, they reduced the natural frequency of the vibration isolator by about 50%, and its lowest natural frequency is 2.75 Hz [10]. Xu et al. designed a vibration isolation system composed of permanent magnets and a coil spring [11]. Wu et al. designed a vibration isolator composed of three cuboidal magnets and a coil spring; its natural frequency is reduced from 10.45 Hz to

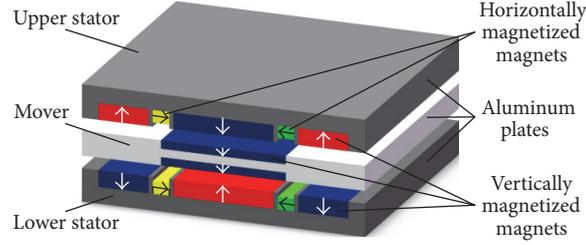


FIGURE 1: The structure of flat-type vertical-gap passive magnetic levitation vibration isolator.

4.96 Hz at a load of 2.29 kg [12]. Shin analyzed the maximum vibration transmissibility of a vibration isolator composed of four magnets and two coil springs [13]. Zheng et al. designed a vibration isolator composed of ring-shaped magnets and a coil spring; its natural frequency is reduced from 9.0 Hz to 5.8 Hz [14]. As mentioned above, most published works about this kind of device show higher natural frequency than air springs, which is adverse for vibration isolation.

Herein, we propose a flat-type vertical-gap passive magnetic levitation vibration isolator (FVPMLVI) that features vacuum compatibility, low natural frequency, and no mechanical contact. A dual-stator scheme is adopted in the proposed FVPMLVI to decrease its natural frequency. This paper focuses on the operation principle and characteristics analysis of the FVPMLVI and provides useful advices for its application. This paper is organized as follows. In Section 2, the structure and operation principle of the FVPMLVI are introduced. In Section 3, the analytical model of the FVPMLVI is established. In Section 4, the electromagnetic and mechanical characteristics of the FVPMLVI are studied in detail by finite element method, and the force characteristics experiment is carried out. A summary is included in Section 5.

2. Structure and Operating Principle

The proposed flat-type vertical-gap passive magnetic levitation vibration isolator consists of three components, that is, upper stator, mover, and lower stator, as shown in Figure 1. The mover consists of an aluminum plate and two vertically magnetized magnets. Both the upper and the lower stators consist of an aluminum plate, three vertically magnetized magnets, and two horizontally magnetized magnets. The magnetization direction of the magnets is shown as arrows in Figures 1 and 2. The major structural parameters of flat-type vertical-gap passive magnetic levitation vibration isolator are shown in Figure 2.

The flat-type vertical-gap passive magnetic levitation vibration isolator generates levitation force by the attraction and repulsion between stator and mover magnets. The mover is suspended above the stator by the magnetic force between stator and mover magnets. When the mover moves relative to the stators along the vertical direction, the upper stator and mover generate attractive force with negative stiffness and the lower stator and mover generate repulsive force with positive stiffness; thus the total levitation force is near-constant. When

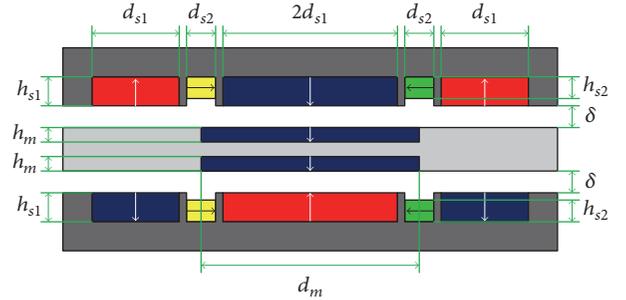


FIGURE 2: The major structural parameters of flat-type vertical-gap passive magnetic levitation vibration isolator and the magnetization direction of magnets.

the mover moves relative to the stators along the horizontal direction, the total levitation force generated by magnets on the left side and the right side is also near-constant. Through adopting the dual-stator structure and inserting horizontally magnetized magnets, the flat-type vertical-gap passive magnetic levitation vibration isolator can generate near-constant levitation force and near-zero stiffness in six-degree of freedom stroke, which is useful for improving the vibration isolation performance of active vibration isolation system.

The proposed flat-type vertical-gap passive magnetic levitation vibration isolator has two advantages: (1) using parallel magnetized cubic magnets; the vertically and horizontally magnetized magnets used in FVPMLVI are parallel magnetized cubic magnets; compared with radially magnetized magnets used in many vibration isolators, the parallel magnetized cubic magnets have advantages of simple structure, good manufacturability, and low cost; and (2) low natural frequency. The natural frequency of FVPMLVI can be decreased effectively by adopting dual-stator scheme and special stator magnet array.

3. Analytical Model

The force generated by flat-type vertical-gap passive magnetic levitation vibration isolator can be calculated by superposition of magnetic force between each two magnets (one of the two magnets is on the stator and another is on the mover). The magnetic force between magnets can be calculated by the equivalent charge model. The equivalent charge model of flat-type vertical-gap passive magnetic levitation vibration

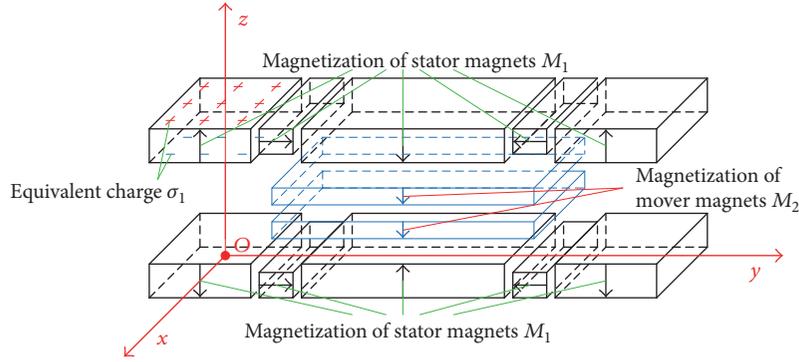


FIGURE 3: The equivalent charge model of flat-type vertical-gap passive magnetic levitation vibration isolator.

isolator is shown in Figure 3. The cuboidal magnets are equivalent to a series of charged rectangular surface.

According to the work of Allag et al. [15], the magnetic force between two magnets with parallel and perpendicular magnetization can be expressed as F_{parallel} and $F_{\text{perpendicular}}$, separately. Since there are 6 vertically magnetized magnets on the stators and 2 vertically magnetized magnets on the mover, thus there are 6×2 pairs of magnets with parallel magnetization. Similarly, there are 4 horizontally magnetized magnets on the stators; thus there are 4×2 pairs of magnets with perpendicular magnetization. Then, the force of the flat-type vertical-gap passive magnetic levitation vibration isolator is given by

$$\begin{aligned}
 F &= \sum_{m=1}^{6 \times 2} F_{\text{parallel}} + \sum_{n=1}^{4 \times 2} F_{\text{perpendicular}} \\
 &= \sum_{m=1}^{12} F_{\text{parallel}} + \sum_{n=1}^8 F_{\text{perpendicular}}
 \end{aligned} \quad (1)$$

where F_{parallel} is the magnetic force between two magnets with parallel magnetization and $F_{\text{perpendicular}}$ is the magnetic force between two magnets with perpendicular magnetization.

Thus the stiffness k and natural frequency f of the flat-type vertical-gap passive magnetic levitation vibration isolator can be expressed as (2) and (3).

$$k = \frac{dF}{ds}, \quad (2)$$

where k is the stiffness, dF is the variation of force, and ds is the displacement.

$$f = \frac{1}{2\pi} \sqrt{\frac{k}{m}} = \frac{\sqrt{g}}{2\pi} \sqrt{\frac{dF/F}{ds}}, \quad (3)$$

where f is the natural frequency, m is the total mass of mover, and load g is the gravitational acceleration.

As shown in (3), compared to stiffness k , the force ripple within unit displacement $dF/F/ds$ is a better indicator which can reflect the vibration isolator's vibration isolation performance, because the natural frequency f is directly related to vibration isolation ability. Therefore, force ripple is analyzed instead of stiffness in the next section.

4. Electromagnetic and Mechanical Characteristics Analysis

In order to provide useful advice for application, the electromagnetic and mechanical characteristics of FVPMLVI such as levitation force, horizontal force, force ripple, and force density are analyzed in this section. Firstly, the force characteristics of FVPMLVI which varied with mover position are studied, and the principle of reducing force ripple is analyzed. Then, the influence of major structural parameters on the device performance like levitation force, force density, and force ripple is studied. At last, the force characteristics of a FVPMLVI prototype is tested to validate the analysis.

4.1. Force Characteristics Varied with Mover Position and Principle of Reducing Force Ripple. When the mover moves relative to the stators, the magnetic force generated by interaction between magnets will change inevitably. The finite element method is used here to analyze the vibration of the force. The mesh result and magnetic field distribution are shown in Figure 4.

When the mover moves relative to the stators along the vertical direction, the levitation force characteristic which varied with vertical motion is shown in Figure 5(a). For a certain horizontal position y , with the increase in vertical position z , the levitation force first decreases and then increases. The minimum levitation force occurs when z is about 0.4 mm. When the mover moves relative to the stators along the horizontal direction, the levitation force characteristic which varied with horizontal motion is shown in Figure 5(b). For a certain vertical position z , with the increase in horizontal position y , the levitation force first increases and then decreases when z is larger than -0.6 mm. The maximum levitation force occurs when x is 0.0 mm. It should be noted that the vertical position $z = 0$ mm means the position 1 mm below the center of FVPMLVI.

When the mover moves relative to the stators along the vertical direction, the horizontal force characteristic which varied with vertical motion is shown in Figure 6(a). For a certain horizontal position y , with the increase in vertical position z , the horizontal force decreases. When the mover moves relative to the stators along the horizontal direction, the horizontal force characteristic which varied with horizontal motion is shown in Figure 6(b). For a certain vertical

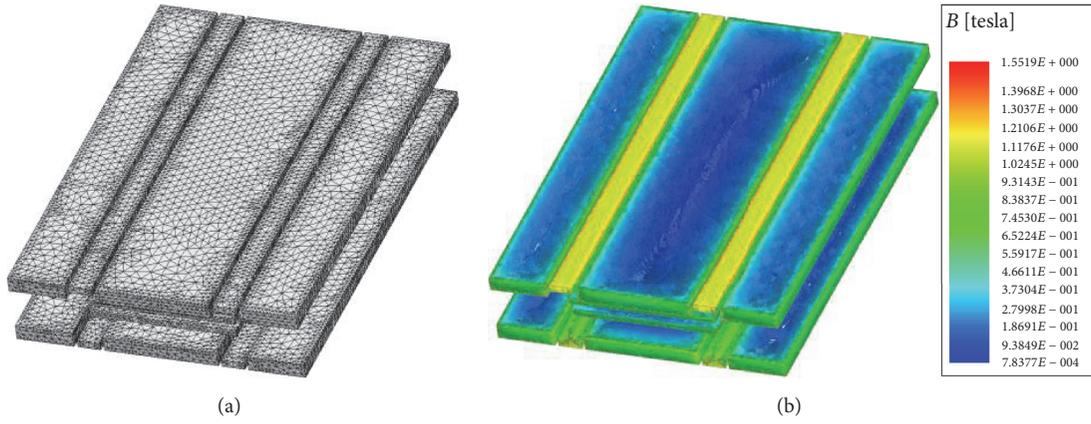


FIGURE 4: The finite element model of FVPMLVI: (a) mesh result; (b) magnetic field distribution.

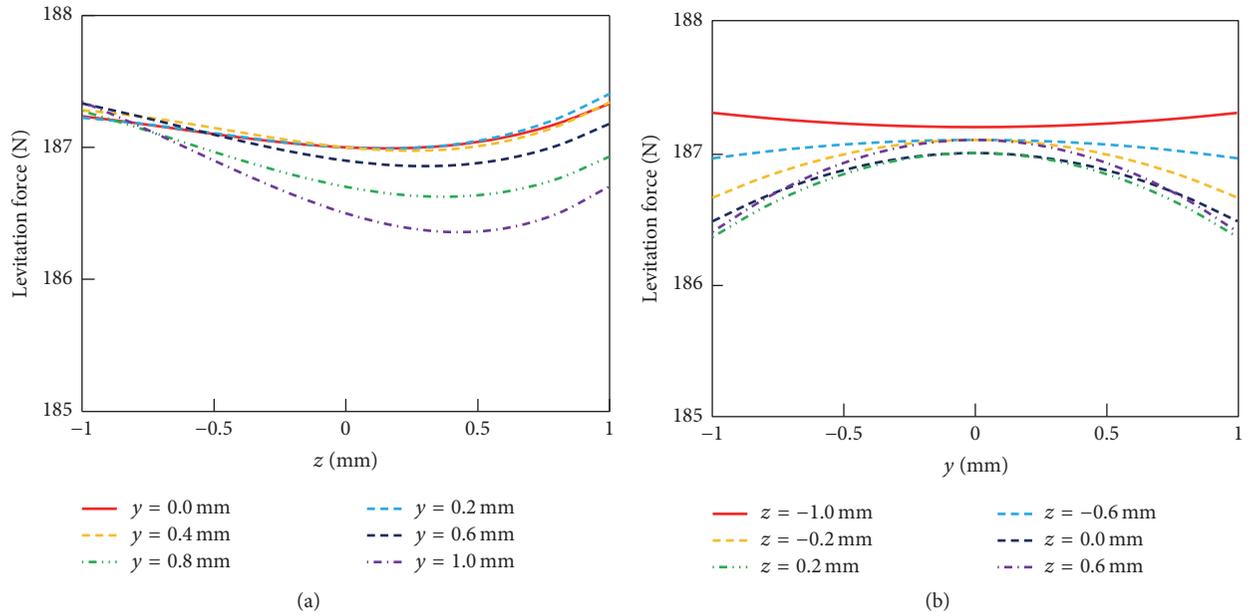


FIGURE 5: Levitation force characteristics versus motion of mover: (a) vertical motion z ; (b) horizontal motion y .

position z , with the increase in horizontal position y , the levitation force is approximately proportional to horizontal position x . As shown in Figures 5 and 6, the vibration of levitation force generated by FVPMLVI is not more than 1 N in the ± 1 mm stroke, and the vibration of horizontal force is not more than 4 N.

The following is the analysis about the role of special stator magnet array and dual-stator scheme. Firstly, the role of special stator magnet array is analyzed. As shown in Figure 7, the stator magnets are divided into three magnet arrays, that is, magnet array A, magnet array B, and magnet array C. The force generated by FVPMLVI is the superposition of magnetic force between these three magnet arrays and mover magnets.

The levitation force characteristics of these three magnet arrays are shown in Figure 8. When the mover moves relative to the stators along the vertical direction, the stiffness of

levitation force generated by magnet array A is near-zero, the stiffness of levitation force generated by magnet array B is negative, and the stiffness of levitation force generated by magnet array C is positive. When the mover moves relative to the stators along the horizontal direction, the levitation force generated by magnet array A and magnet array B increases when $|y|$ increases, and the levitation force generated by magnet array C decreases when $|y|$ increases. Thus, a near-constant levitation force can be obtained by appropriate combination.

The horizontal force characteristics of these three magnet arrays are shown in Figure 9. When the mover moves relative to the stators along the vertical direction, the horizontal force generated by magnet array A and magnet array B decreases when z increases, and the horizontal force generated by magnet array C increases when z increases. When the mover moves relative to the stators along the horizontal direction,

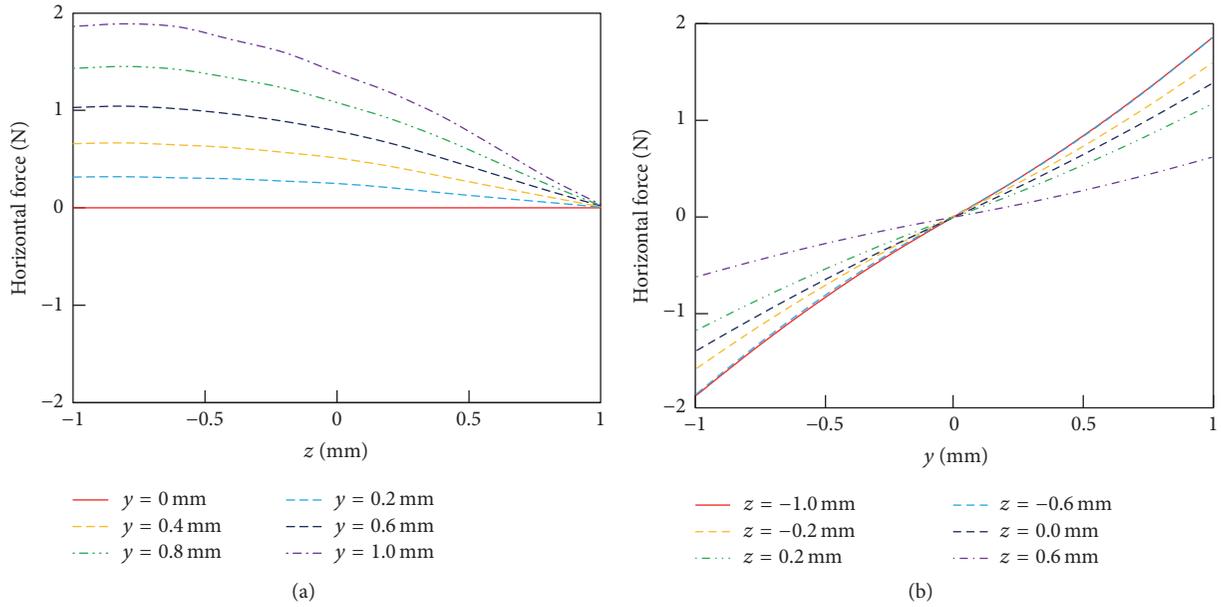


FIGURE 6: Horizontal force characteristics versus motion of mover: (a) vertical motion z ; (b) horizontal motion y .

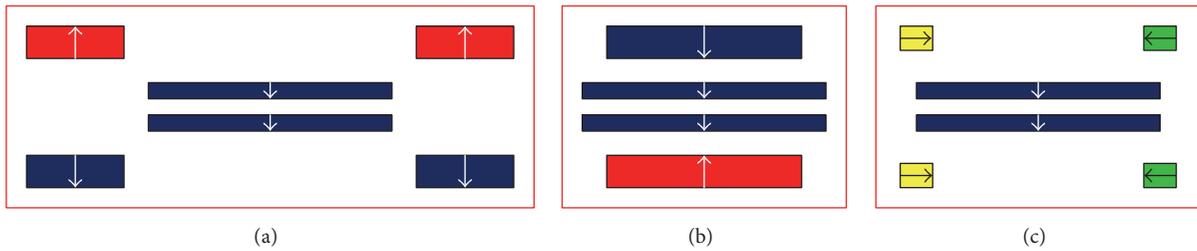


FIGURE 7: The structure of three magnet arrays: (a) magnet array A; (b) magnet array B; (c) magnet array C.

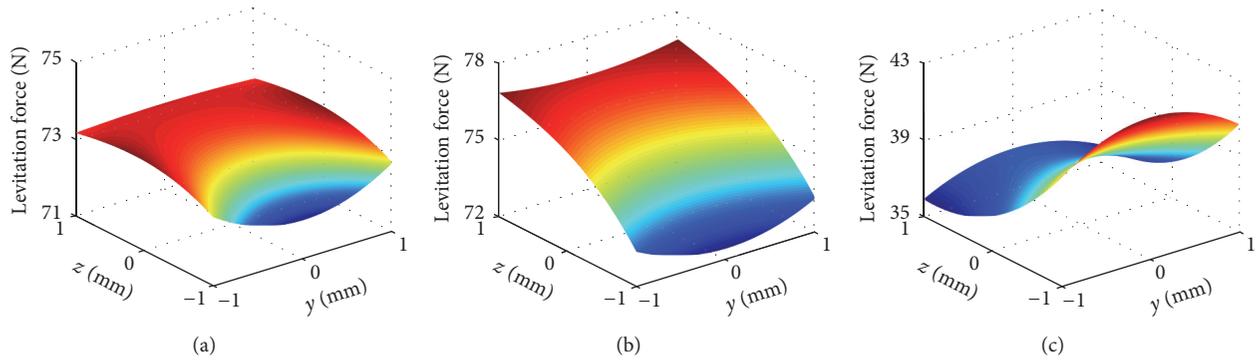


FIGURE 8: Levitation force characteristics of three magnet arrays: (a) array A; (b) array B; (c) array C.

the stiffness of horizontal force generated by magnet array A and magnet array B is positive, and the stiffness of levitation force generated by magnet array C is negative. Thus, a near-zero vertical stiffness can be obtained by appropriate combination. Thus, the special stator magnet array, which is composed of three vertically magnetized magnets and two horizontally magnetized magnets, can effectively decrease the force ripple of FVPMLVI.

Similarly, the role of dual-stator scheme is analyzed. As shown in Figure 10, the stator magnets are divided into two magnet arrays, that is, magnet array D and magnet array E. The force generated by FVPMLVI is the superposition of magnetic force between these two magnet arrays and mover magnets.

The levitation force characteristics of these two magnet arrays are shown in Figure 11. When the mover moves relative

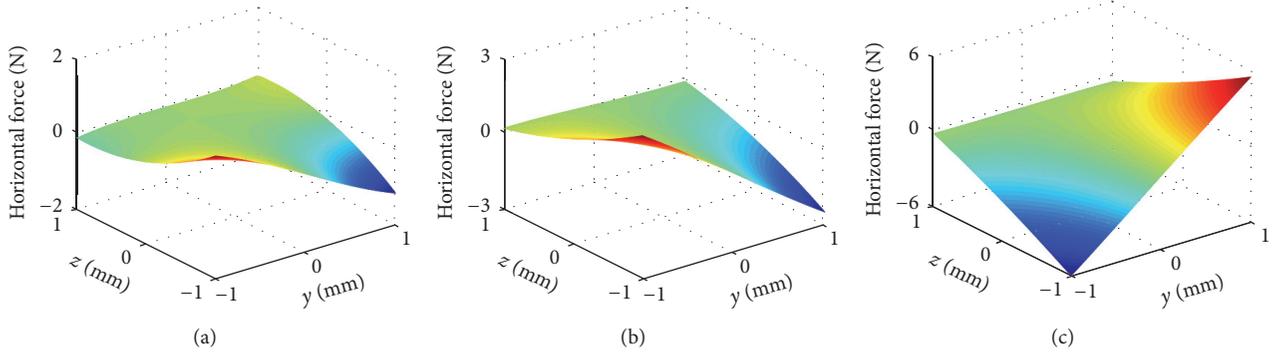


FIGURE 9: Horizontal force characteristics of three magnet arrays: (a) array A; (b) array B; (c) array C.

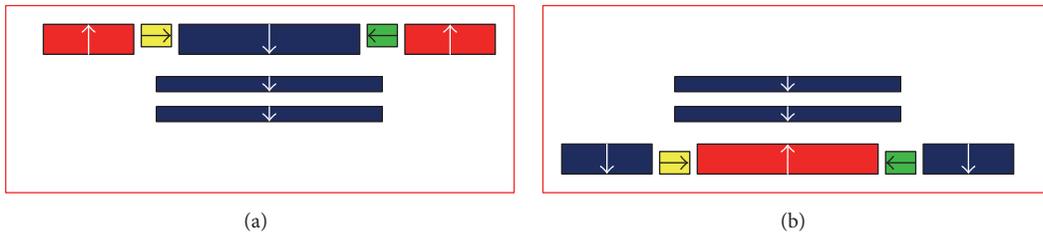


FIGURE 10: The structure of two magnet arrays: (a) magnetic array D; (b) magnetic array E.

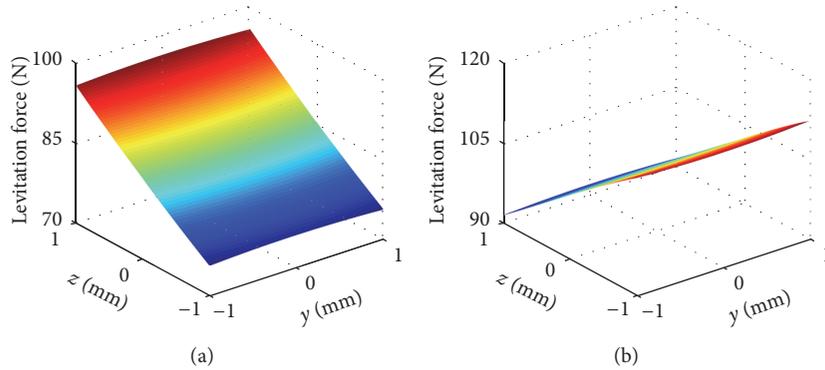


FIGURE 11: Levitation force characteristics of two magnet arrays: (a) magnetic array D; (b) magnetic array E.

to the stators along the vertical direction, the stiffness of levitation force generated by magnet array D is negative, and the stiffness of levitation force generated by magnet array E is positive. Thus, a near-constant levitation force can be obtained by appropriate combination.

The horizontal force characteristics of these three magnet arrays are shown in Figure 12. When the mover moves relative to the stators along the horizontal direction, the stiffness of horizontal force generated by magnet array D is positive, and the stiffness of levitation force generated by magnet array E is negative. Thus, a near-zero vertical stiffness can be obtained by appropriate combination.

In conclusion, the natural frequency of FVPMLVI can be decreased effectively by adopting the dual-stator scheme and special stator magnet array, and the structural parameters of magnets should be carefully designed because the force

characteristics of FVPMLVI depend on the combination of force characteristics of several magnet arrays.

4.2. Influence of Major Structural Parameters on Force Ripple and Force Density. In order to provide useful advice for application, the influence of major structural parameters on force ripple and force density should be analyzed in order to find a useful optimized method for future design process.

The major structural parameters are defined as the structural parameters which can determine the FVPMLVI's electromagnetic and mechanical characteristics. As shown in Figure 2, the major structural parameters include width of stator magnets d_{s1} and d_{s2} , thickness of stator magnets h_{s1} and h_{s2} , and width and thickness of mover magnets d_m and h_m . The force ripple within unit displacement σ and force density ρ are calculated by (4) and (5), separately.

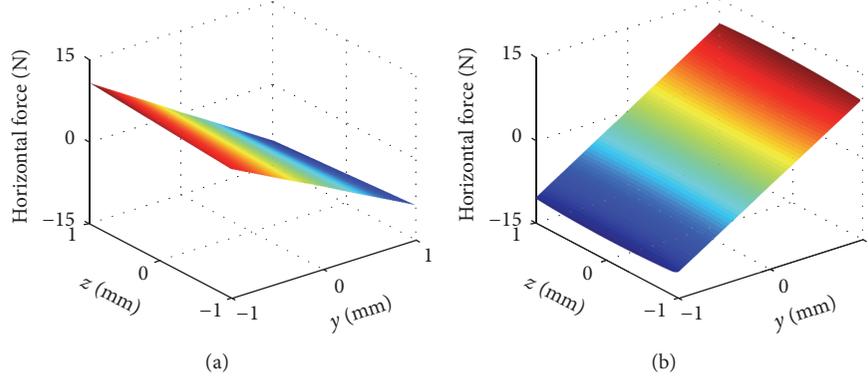


FIGURE 12: Horizontal force characteristics of two magnet arrays: (a) magnetic array D; (b) magnetic array E.

$$\sigma = \frac{dF/F}{ds} \times 100\% \quad (4)$$

$$\rho = \frac{F}{V} = \frac{F}{[(4d_{s1} + 2d_{s2} + 4t_s + 2t_{ap}) \times (2h_{s1} + 2h_m + 2\delta + t_m + 2t_{ap}) \times l_{ap}]}, \quad (5)$$

where F is the levitation force of FVPMLVI, V is the volume of FVPMLVI, t_s is the spacing between stator magnets, t_{ap} is the thickness of stator aluminum plates, δ is the thickness of air gap, t_m is the spacing between mover magnets, and l_{ap} is the length of FVPMLVI in x -axis.

The influence of major structural parameters on force ripple and force density is analyzed by finite element method. The major structural parameters are studied with some auxiliary structural parameters fixed; that is, the length of FVPMLVI in x -axis l_{ap} is 100 mm; the spacing between stator magnets t_s and between mover magnets t_m is 1 mm and 2 mm, respectively. Considering the stroke and manufacturability, the air gap length δ is chosen to be 3 mm.

Firstly, from Figures 8 and 9, the stiffness of levitation force generated by magnet array B is opposite to that of magnet array C. Thus, there is a reasonable range of d_{s1} and d_{s2} , where the total stiffness of magnet array B and C is near-zero; then the force ripple can be decreased significantly. Simultaneously, the levitation force will increase when d_{s1} and d_{s2} are increased. However, from (5), the volume of PMLVI will also increase with the increase of d_{s1} and d_{s2} , which brings a negative influence on the force density. Thus, there is a reasonable range of d_{s1} and d_{s2} , where force density obtains the maximum.

The force ripple which varied with d_{s1} and d_{s2} is shown in Figure 13. From Figure 13, (1) the force ripple caused by vertical motion is decreased significantly when d_{s1} and d_{s2} take appropriate values; (2) the minimum of force ripple caused by vertical motion occurs when d_{s2} is in the range of 6.2 mm~7.2 mm; (3) the force ripple caused by horizontal motion is decreased significantly when d_{s1} and d_{s2} take appropriate values; (4) the minimum force ripple caused by

vertical motion occurs when d_{s2} is in the range of 8.4 mm~12.2 mm.

The levitation force and force density which varied with d_{s1} and d_{s2} are shown in Figure 14. From Figure 14, (1) the levitation force increases with the increase of d_{s1} and d_{s2} ; (2) the maximum of force density occurs when d_{s1} is in the range of 4 mm~6 mm, and d_{s2} is expected to be as small as possible. This shows that although force ripple is decreased by adopting special stator magnet array, the force density is also decreased.

Similar to the analysis of d_{s1} and d_{s2} , there is a reasonable range of h_{s1} and h_{s2} , where force ripple can be decreased significantly and force density reaches the maximum. It should be noted that h_{s1} is designed to be larger than h_{s2} , because the air gap length is fixed, and a larger h_{s2} will lead to a shorter stroke. Through changing the thickness of stator magnets h_{s1} and h_{s2} , the force ripple is calculated, as shown in Figure 15. From Figure 15, (1) the force ripple caused by vertical motion is decreased significantly when d_{s1} and d_{s2} take appropriate values; (2) the minimum of force ripple caused by vertical motion occurs when $h_{s1} = -0.1214h_{s2}^2 + 2.179h_{s2} + 0.14$; (3) the force ripple caused by horizontal motion is decreased significantly when h_{s1} and h_{s2} take appropriate values; the minimum of force ripple caused by horizontal motion occurs when h_{s1} and h_{s2} are about 4 mm and 1 mm, respectively.

The levitation force and force density which varied with h_{s1} and h_{s2} are shown in Figure 16. From Figure 16, (1) the levitation force increases with the increase of h_{s1} and h_{s2} ; (2) the maximum of force density occurs when h_{s2} takes larger values and $h_{s1} = h_{s2}$; the force density can be further increased if h_{s2} takes values in a wider range; however, the force ripple will increase.

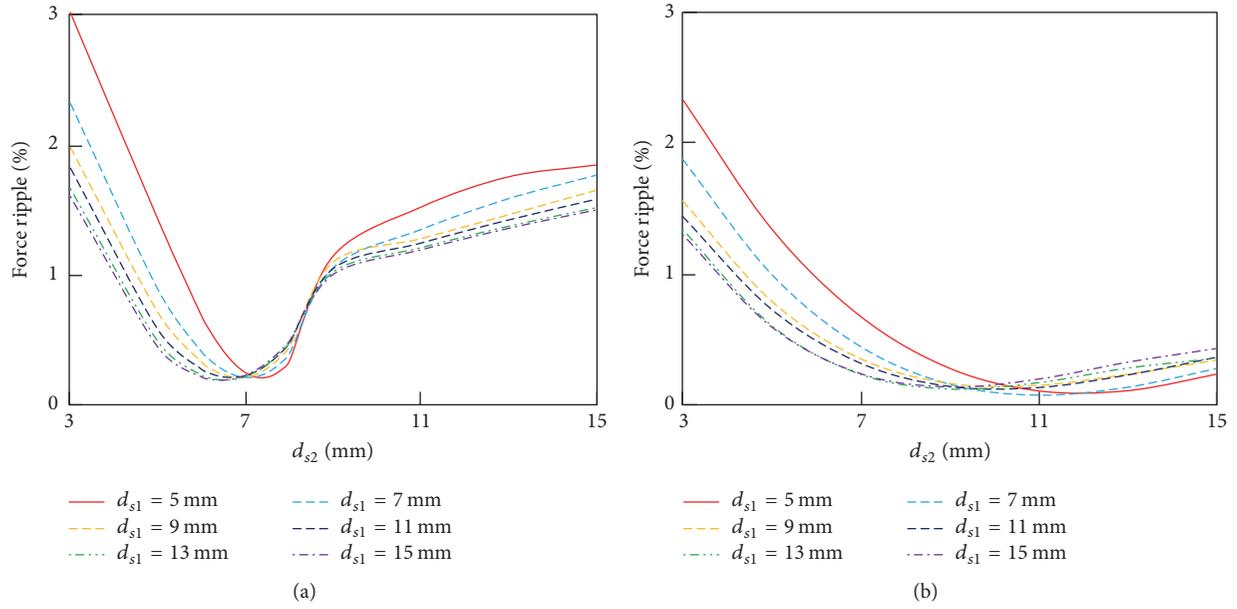


FIGURE 13: Force characteristics versus width of stator magnets d_{s1} and d_{s2} : (a) force ripple caused by vertical motion; (b) force ripple caused by horizontal motion.

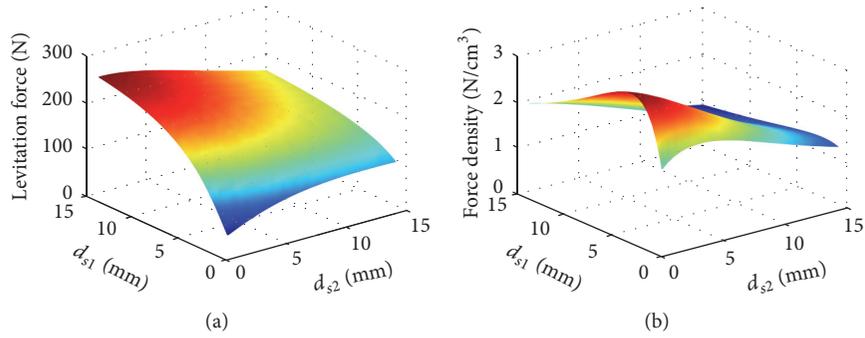


FIGURE 14: Force characteristics versus width of stator magnets d_{s1} and d_{s2} : (a) levitation force; (b) force density.

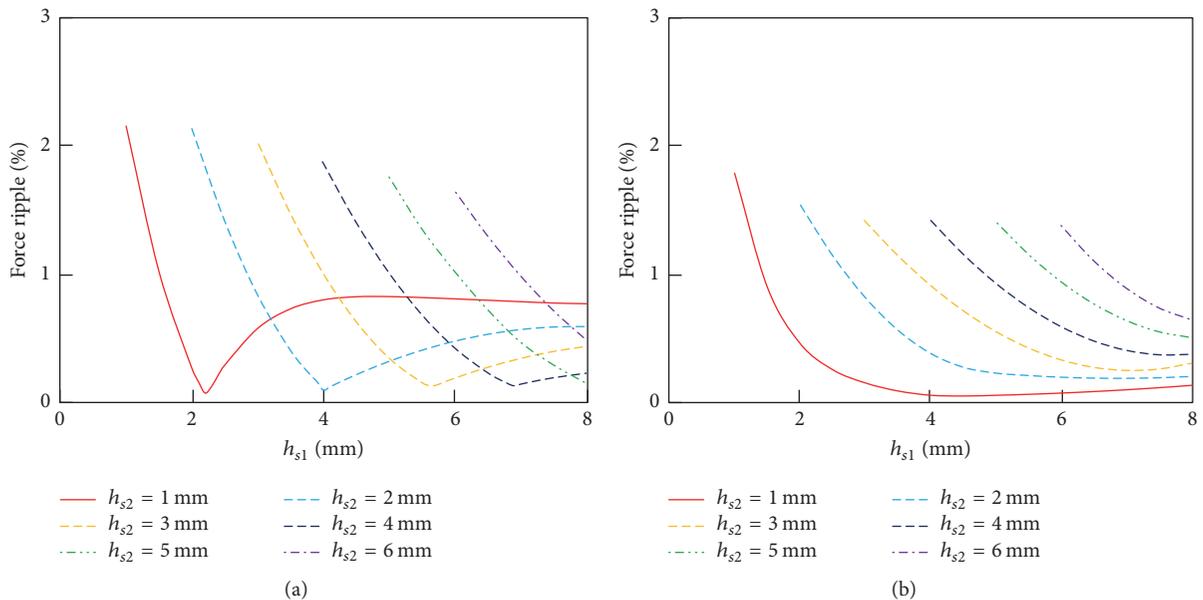


FIGURE 15: Force characteristics versus thickness of stator magnets h_{s1} and h_{s2} : (a) force ripple caused by vertical motion; (b) force ripple caused by horizontal motion.

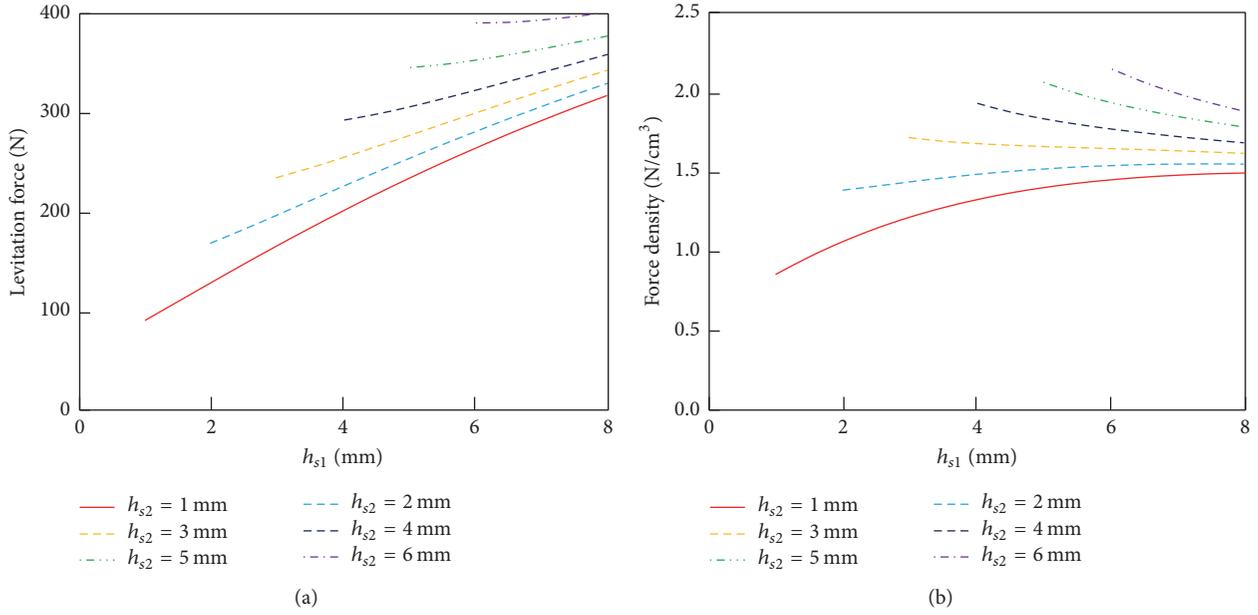


FIGURE 16: Force characteristics versus thickness of stator magnets h_{s1} and h_{s2} : (a) levitation force; (b) force density.

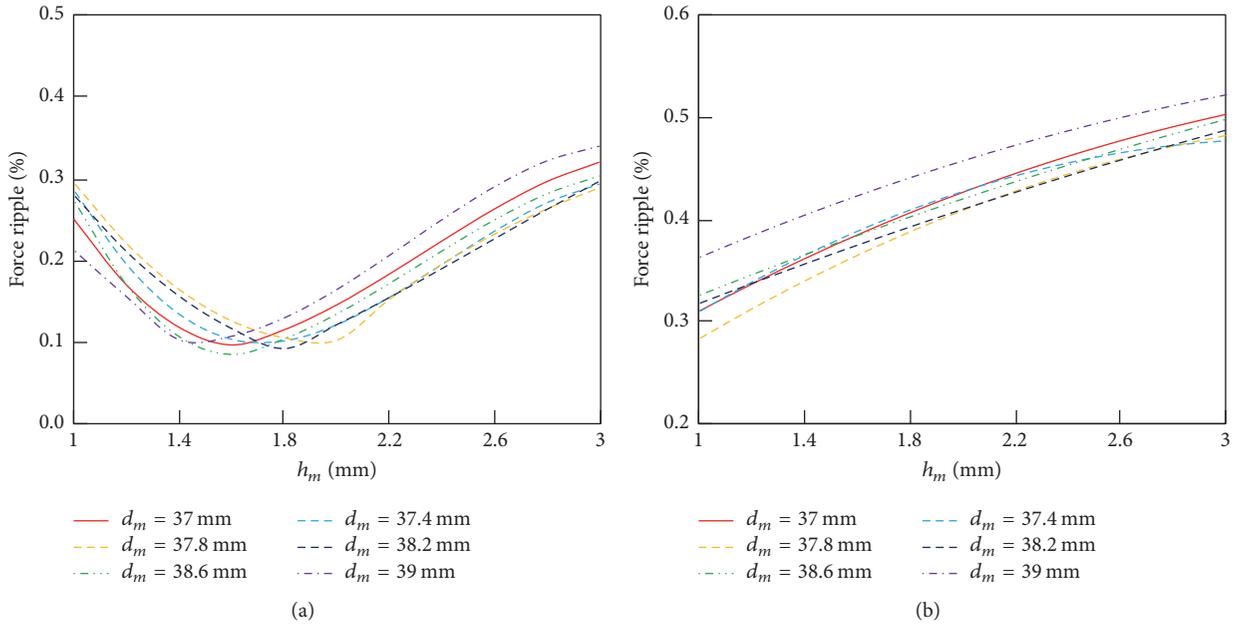


FIGURE 17: Force characteristics versus width and thickness of mover magnets d_m and h_m : (a) force ripple caused by vertical motion; (b) force ripple caused by horizontal motion.

The force characteristics of PMLVI could be also adjusted by the structural parameters of mover. Through changing the width and thickness of mover magnets d_m and h_m in a small range, the force ripple is calculated as shown in Figure 17. From Figure 17, (1) the force ripple caused by vertical motion is decreased significantly when d_m and h_m take appropriate values; the minimum of force ripple caused by vertical motion occurs when $h_m = -0.4138d_m^2 + 31.33d_m - 591.1$; (3) the force ripple caused by horizontal motion is decreased significantly when h_m decreases.

The levitation force and force density which varied with d_m and h_m are shown in Figure 18. From Figure 18, (1) the levitation force increases significantly with the increase of h_m ; the levitation force is nearly proportional to the thickness of mover magnets h_m ; (2) from (5), with the increase of h_m , the increase magnitude in levitation force is larger than the increase magnitude of its volume; thus the force density also increases significantly.

Based on the analysis results above, an optimized FVPM-LVI is designed and analyzed by experiment. It should be

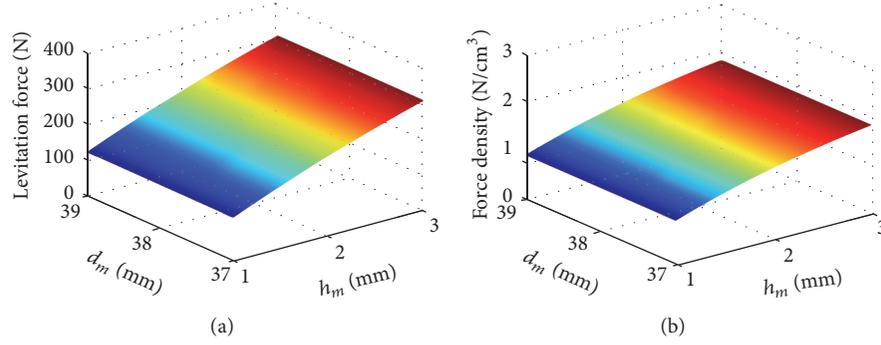


FIGURE 18: Force characteristics versus width and thickness of mover magnets d_m and h_m : (a) levitation force; (b) force density.

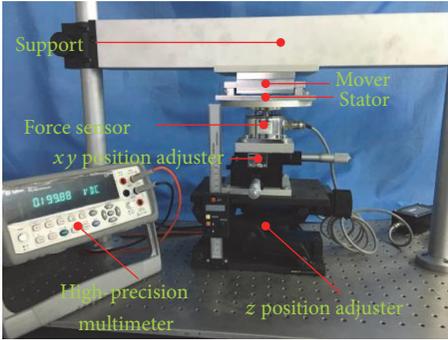


FIGURE 19: The test platform for measuring force characteristics of the FVPMLVI.

mentioned that the parameters of the optimized model in this section are preliminary optimization results based on the forgoing FEM analysis. A more accurate optimization result can be obtained by using some optimization methods in the design process, for example, numerical optimization algorithms or approximate models [16, 17].

4.3. Experiment. To verify the analysis presented above, a prototype of the FVPMLVI is manufactured. The structural parameters of prototype are listed in Table 1. The test platform for measuring force characteristics of the FVPMLVI is shown in Figure 19. The test platform is composed of a three-degree of freedom position adjuster, a force sensor, a high-precision multimeter, and a support. The three-degree of freedom position adjuster is composed of a xy position adjuster and a z position adjuster. The three-degree of freedom position adjuster and the support are fixed on a passive vibration isolation platform separately, and the force sensor is fixed on the three-degree of freedom position adjuster. The mover of FVPMLVI is fixed on the support, and the stator of FVPMLVI is fixed on the force sensor. Therefore the relative position between stator and mover can be adjusted by the three-degree of freedom position adjuster. The relationship between force and mover position can be measured.

Figure 20 shows the levitation force distribution with different vertical displacement (z) and horizontal displacement (x and y). The measured nominal levitation force is

TABLE 1: Structural parameters of PMLVI prototype.

Parameter	Data
Width of stator magnet d_{s1}	15 mm
Thickness of stator magnet h_{s1}	4 mm
Width of stator magnet d_{s2}	6 mm
Thickness of stator magnet h_{s2}	2 mm
Length of stator magnet l_s	108 mm
Stator magnet spacing t_s	1 mm
Width of mover magnet d_m	38.6 mm
Thickness of stator magnet h_m	1.6 m
Length of mover magnet l_m	100 mm
Mover magnet spacing t_m	2 mm
Length of air gap δ	3 mm
Aluminum plate thickness t_{ap}	4 mm

197.64 N, which matches well with the finite element model. The variation of the levitation force is quite small for the whole stroke.

The stiffness is obtained from the levitation force characteristics by using (6). Figure 21 shows the vertical stiffness distribution with different vertical displacement (z) and horizontal displacement (x and y). It can be seen that the vertical stiffness can be as low as zero in some areas. The maximum positive vertical stiffness is 1008.26 N/m, and the natural frequency near this point calculated by (3) is 1.13 Hz. Compared with passive magnetic levitation vibration isolators with natural frequency more than 4 Hz [2, 3, 10–14], the proposed passive maglev vibration isolation unit features advantage in lower natural frequency.

$$K_z = \frac{\partial F_z}{\partial z}. \quad (6)$$

5. Conclusions

A flat-type vertical-gap passive magnetic levitation vibration isolator has been investigated for active vibration isolation system. The dual-stator scheme and special stator magnet array can decrease FVPMLVI's natural frequency effectively, and this enhances the vibration isolation capability of the FVPMLVI. The magnets used in FVPMLVI are parallel

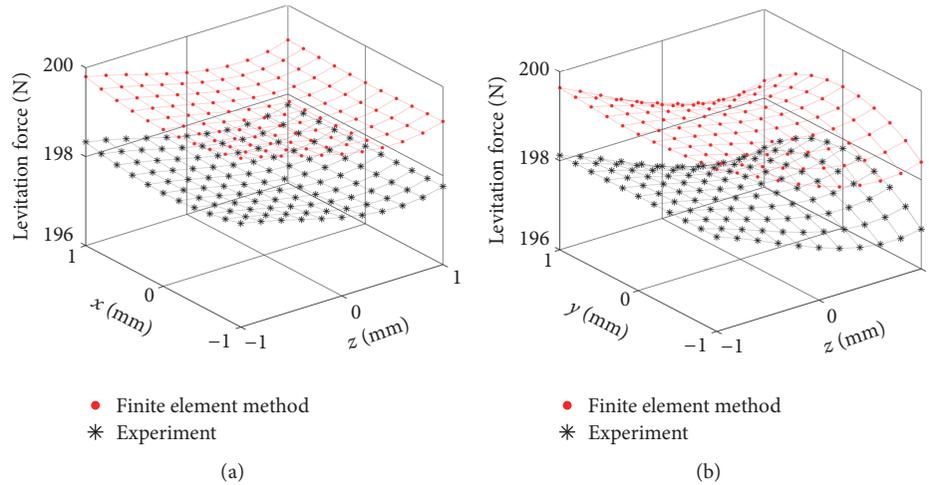


FIGURE 20: The levitation force characteristics of FVPMLVI.

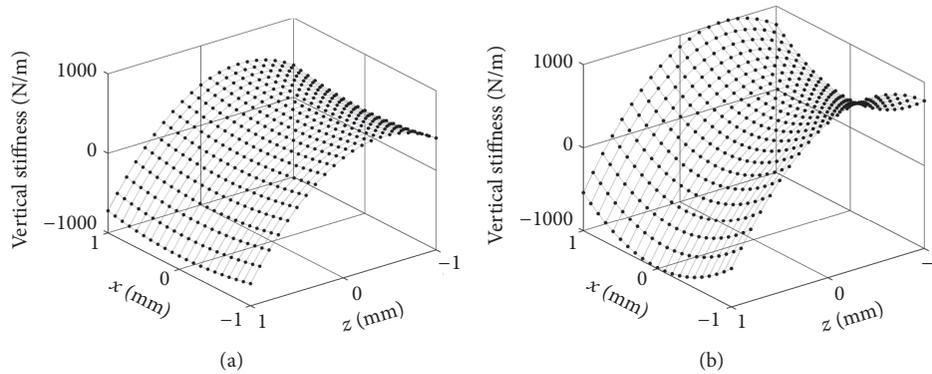


FIGURE 21: The stiffness characteristics of FVPMLVI.

magnetized cubic magnets, which have advantages of simple structure, good manufacturability, and low cost. The analytical expression of magnetic force is deduced by equivalent charge model. The influence of major structural parameters and motion on levitation force, horizontal force, force density, and force ripple are studied. The dimensions of both stator magnets and mover magnets have significant impact on levitation force, horizontal force, force density, and force ripple. In comparison with conventional passive magnetic levitation vibration isolators, the proposed flat-type vertical-gap passive magnetic levitation vibration isolator shows advantage in lower natural frequency, wider vibration isolation bandwidth, and lower vibration transmissibility.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (51507034).

References

- [1] R. H. Stulen and D. W. Sweeney, "Extreme ultraviolet lithography," *IEEE Journal of Quantum Electronics*, vol. 35, no. 5, pp. 694–699, 1999.
- [2] E. Puppini and V. Fratello, "Vibration isolation with magnet springs," *Review of Scientific Instruments*, vol. 73, no. 11, p. 4034, 2002.
- [3] Y. Zhu, Q. Li, D. Xu, and M. Zhang, "Modeling of axial magnetic force and stiffness of ring-shaped permanent-magnet passive vibration isolator and its vibration isolating experiment," *IEEE Transactions on Magnetics*, vol. 48, no. 7, pp. 2228–2238, 2012.
- [4] W. Robertson, B. Cazzolato, and A. Zander, "A multipole array magnetic spring," *IEEE Transactions on Magnetics*, vol. 41, no. 10, pp. 3826–3828, 2005.
- [5] W. S. Robertson, M. R. F. Kidner, B. S. Cazzolato, and A. C. Zander, "Theoretical design parameters for a quasi-zero stiffness magnetic spring for vibration isolation," *Journal of Sound and Vibration*, vol. 326, no. 1-2, pp. 88–103, 2009.
- [6] W. Robertson, B. Cazzolato, and A. Zander, "Theoretical analysis of a non-contact spring with inclined permanent magnets for load-independent resonance frequency," *Journal of Sound and Vibration*, vol. 331, no. 6, pp. 1331–1341, 2012.

- [7] J. L. G. Janssen, J. J. H. Paulides, E. A. Lomonova, B. Delinchant, and J.-P. Yonnet, "Design study on a magnetic gravity compensator with unequal magnet arrays," *Mechatronics*, vol. 23, no. 2, pp. 197–203, 2013.
- [8] D. T. E. H. van Casteren, J. J. H. Paulides, J. L. G. Janssen, and E. A. Lomonova, "Analytical force, stiffness, and resonance frequency calculations of a magnetic vibration isolator for a microbalance," *IEEE Transactions on Industry Applications*, vol. 51, no. 1, pp. 204–210, 2015.
- [9] D. T. E. H. van Casteren, J. J. H. Paulides, and E. A. Lomonova, "A comparison between cylindrical and cross-shaped magnetic vibration isolators: ideal and practical," *Archives of Electrical Engineering*, vol. 64, no. 4, pp. 593–604, 2015.
- [10] Y. Zhu, Q. Li, D. Xu, C. Hu, and M. Zhang, "Modeling and analysis of a negative stiffness magnetic suspension vibration isolator with experimental investigations," *Review of Scientific Instruments*, vol. 83, no. 9, Article ID 095108, 2012.
- [11] D. Xu, Q. Yu, J. Zhou, and S. R. Bishop, "Theoretical and experimental analyses of a nonlinear magnetic vibration isolator with quasi-zero-stiffness characteristic," *Journal of Sound and Vibration*, vol. 332, no. 14, pp. 3377–3389, 2013.
- [12] W. Wu, X. Chen, and Y. Shan, "Analysis and experiment of a vibration isolator using a novel magnetic spring with negative stiffness," *Journal of Sound and Vibration*, vol. 333, no. 13, pp. 2958–2970, 2014.
- [13] K. Shin, "On the performance of a single degree-of-freedom high-static-low-dynamic stiffness magnetic vibration isolator," *International Journal of Precision Engineering and Manufacturing*, vol. 15, no. 3, pp. 439–445, 2014.
- [14] Y. Zheng, X. Zhang, Y. Luo, B. Yan, and C. Ma, "Design and experiment of a high-static-low-dynamic stiffness isolator using a negative stiffness magnetic spring," *Journal of Sound and Vibration*, vol. 360, pp. 31–52, 2016.
- [15] H. Allag, J.-P. Yonnet, and M. E. H. Latreche, "3D analytical calculation of forces between linear halbach-type permanent-magnet arrays," in *Proceedings of the 8th International Symposium on Advanced Electromechanical Motion Systems and Electric Drives Joint Symposium (ELECTROMOTION '09)*, July 2009.
- [16] G. Lei, T. Wang, Y. Guo, J. Zhu, and S. Wang, "System-level design optimization methods for electrical drive systems: deterministic approach," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 12, pp. 6591–6602, 2014.
- [17] G. Lei, T. Wang, J. Zhu, Y. Guo, and S. Wang, "System-level design optimization method for electrical drive systems—robust approach," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 8, pp. 4702–4713, 2015.

Research Article

Automated Bearing Fault Diagnosis Using 2D Analysis of Vibration Acceleration Signals under Variable Speed Conditions

Sheraz Ali Khan and Jong-Myon Kim

School of Electrical, Electronics and Computer Engineering, University of Ulsan, Building No. 7, Room No. 308, 93 Daehak-ro, Nam-gu, Ulsan 680-749, Republic of Korea

Correspondence should be addressed to Jong-Myon Kim; jongmyon.kim@gmail.com

Received 2 September 2016; Accepted 17 November 2016

Academic Editor: Minvydas Ragulskis

Copyright © 2016 S. A. Khan and J.-M. Kim. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Traditional fault diagnosis methods of bearings detect characteristic defect frequencies in the envelope power spectrum of the vibration signal. These defect frequencies depend upon the inherently nonstationary shaft speed. Time-frequency and subband signal analysis of vibration signals has been used to deal with random variations in speed, whereas design variations require retraining a new instance of the classifier for each operating speed. This paper presents an automated approach for fault diagnosis in bearings based upon the 2D analysis of vibration acceleration signals under variable speed conditions. Images created from the vibration signals exhibit unique textures for each fault, which show minimal variation with shaft speed. Microtexture analysis of these images is used to generate distinctive fault signatures for each fault type, which can be used to detect those faults at different speeds. A k -nearest neighbor classifier trained using fault signatures generated for one operating speed is used to detect faults at all the other operating speeds. The proposed approach is tested on the bearing fault dataset of Case Western Reserve University, and the results are compared with those of a spectrum imaging-based approach.

1. Introduction

In modern industries, motion is mostly powered by electromechanical systems (e.g., induction motors), which account for nearly 70% of the gross energy consumption in industrialized economies [1]. Induction motors and other industrial machines that undergo rotatory motion use bearings to reduce friction. The reduction in friction conserves energy that would otherwise be lost in overcoming it. It also increases the useful life of a machine by reducing its wear. Nevertheless, adverse operating conditions and cyclic loading can lead to material fatigue in bearings, which manifests itself in the form of surface cracks and spalls [2]. These cracks and spalls, if allowed to go undetected, can lead to costly and unexpected shutdowns, which is detrimental to economic productivity.

Bearings are at the heart of condition monitoring techniques since they are the most common cause of failure in induction motors (i.e., in more than 50% of the cases) [3].

In addition, their failure can cause prolonged downtimes. According to a recent study, gearbox bearings cause the longest downtime per failure in wind turbines [4, 5]. Hence, condition monitoring techniques for bearings are widely used in almost all forms of rotary machinery (e.g., gearboxes, wind turbines, helicopters, and even rotary microelectromechanical systems or MEMS) [6–10]. As bearing degradation is accompanied by increased levels of noise and vibration, vibration condition monitoring of bearings is a standard practice in the industry and an essential component in any predictive maintenance strategy. The vibration levels of machine components, especially bearings, are measured using accelerometers and analyzed to determine underlying faults [11, 12].

A detailed survey of fault diagnosis and fault-tolerant techniques, which have been developed in different domains, is provided in [13, 14]. These techniques have been broadly categorized into *model-based*, *signal-based*, *knowledge-based*, and *hybrid/active* approaches. However, fault diagnosis in

bearings has mostly been done through signal-based techniques. These techniques usually involve three major steps: (1) measurement of the signal that will be used for fault diagnosis (different types of signals, such as structural vibration [10, 15–18], stator current for bearings in induction motors [7, 8], acoustic emissions [19–25], temperature [6], and more recently rotor speed [26], have been used); (2) processing the signal to extract features that are characteristic of anomalous conditions; (3) using different classifiers such as k -NN, support vector machines (SVMs), or artificial neural networks (ANNs) for classifying normal and faulty signals.

Signal-based approaches detect localized defects in bearings, mostly by extracting their associated characteristic frequencies from the modulated fault signal through envelope analysis [27, 28]. These characteristic frequencies are the ball pass frequency for the outer raceway (BPFO), which is associated with faults on the outer raceway of a bearing, the ball pass frequency for the inner race way (BPFI), which is associated with faults on the inner raceway, and the first even harmonic of the ball spin frequency ($2xBSF$), which is associated with roller faults [29, 30]. Performance of the envelope analysis is improved by using it in conjunction with time-frequency analysis tools such as the discrete wavelet transform (DWT) [17, 31, 32], short time Fourier transform (STFT) [33, 34], empirical mode decomposition (EMD) [35–38], and discrete wavelet packet transform (DWPT) [19–23, 39]. These tools are primarily used to filter frequency bands near the carrier frequency, where the signal components corresponding to defects are modulated.

Nevertheless, these characteristic or defect frequencies are kinematic quantities that depend on the *shaft speed* and bearing geometry. The shaft speed and load angle from the radial plane are subject to random variations, which make the bearing signal inherently nonstationary and cause variations in the fundamental defect frequencies [29]. The detection of defect frequencies that are subject to random variations is challenging and hence requires tedious methods, which are difficult to implement. In [21, 22, 39], Kang et al. reduced the effects of nonstationarity by using subband analysis of fault signals through filter banks. They proposed measures like the mean-peak ratio and Gaussian mixture model-based residual component-to-defect component ratios to select the most informative subband. After selecting the most effective subband, features such as the relative wavelet packet energy and wavelet packet node kurtosis were extracted and then used for fault diagnosis. In [16], Amar et al. used binary vibration spectrum images and artificial neural networks for bearing fault diagnosis. The classification performance of this approach is highly susceptible to the quality of the binary spectral images. This method therefore hinges on the appropriate selection of the grayscale threshold value, which is used to generate those binary spectral images. Despite their complexity, these methods are dependent on the shaft speed and their performance is affected by its random variations. Moreover, these methods cover the design variations in shaft speed by dividing the vibration acceleration or acoustic emission signals into different datasets on the basis of shaft speed (revolutions per minute or rpm). A unique instance of the classifier is trained for each operating speed, which can only be used to classify

the test samples for that operating speed alone. For a different operating speed, a new instance of the classifier needs to be retrained on a new set of features. In [40], a mechanism for feature extraction was proposed, which can be used to diagnose bearing faults under gear interference and variable speed conditions. However, this approach is very tedious and computationally complex since it extracts a feature, called the instantaneous dominant meshing multiplying, using STFT, and then resamples the original signal using this feature, decomposes the resampled signal into intrinsic mode functions (IMF) using EMD, and finally carries out the envelope demodulation of the IMF, with the highest kurtosis value, to determine bearing fault. Moreover, the output is not processed automatically; rather it requires manual interpretation to diagnose bearing faults.

The proposed approach addresses the three primary limitations of existing techniques: (1) They are tedious and complex processes, and hence practical solutions based on these methods are difficult to implement and more likely to be costly; (2) they require retraining the classifier each time if there is a change in the motor rpm to achieve given operating conditions; and (3) these techniques are not fully automated and require manual interpretation of the output. In contrast with the conventional approach, the proposed scheme is automated and simple to implement, uses pure pattern analysis, and requires the training of only a single instance of the classifier for all the four operating speeds considered in this study. The classifier is trained using features extracted from any of the four datasets and it can effectively diagnose faults at all the operating speeds as demonstrated in Section 4. In this study, the operating speed varies by approximately $\pm 5\%$ across all the datasets.

The main contribution of this study is that it proposes a method for fault diagnosis of bearings that is impervious to both random and design variations in shaft speed. In this method, the time domain vibration signal is converted into grayscale images. The dimensions of these images are determined experimentally, to ensure minimal variation in textures across different shaft speeds. The proposed approach, which is discussed in detail in Section 3, is validated using the publicly available benchmark dataset from [41]. A comparison of the proposed approach with vibration spectrum imaging [16] is provided.

The rest of the paper is organized as follows: Section 2 describes the seeded fault test data used to validate the proposed approach. Section 3 provides a detailed discussion of the proposed fault diagnosis scheme. Section 4 provides the discussion and analysis of experimental results, and Section 5 concludes the paper.

2. Experimental Setup and Vibration Fault Data

The proposed approach is tested on the publicly available seeded fault test data of Case Western Reserve University [41]. The data was collected using a 2-horsepower (hp) motor with a torque transducer and a dynamometer. The dynamometer is used to apply different loads on the bearing (i.e., 0 hp, 1 hp, 2 hp, and 3 hp). In this study, the vibration acceleration signals

TABLE 1: Specifications of bearings.

Attribute	Value
Model	JEM SKF 6205-2RS
Location	Drive end
Outside diameter	2.0472 Inches
Inside diameter	0.9843 Inches
Thickness	0.5906 Inches
Ball diameter	0.3126 Inches
Pitch diameter	1.537 Inches

TABLE 2: Specifications of fault types.

Fault type	Fault location	Fault diameter (Inches)	Fault depth (Inches)
Inner raceway fault (IRF)	Inner raceway	0.007	0.011
Outer raceway fault (ORF)	Outer raceway	0.007	0.011
Ball fault (BF)	Ball	0.007	0.011
Normal	Nil	Nil	Nil

that were recorded for the drive end bearings are considered for analysis. The specifications of the drive end bearings are given in Table 1.

The test bearings are seeded with single point localized defects on the rollers, the inner raceways, and the outer raceways. The dimensions of the seeded faults are given in Table 2. The vibration data used in this analysis was measured using accelerometers placed at the 12 o'clock position on the bearing housing. The signal was recorded at a sampling rate of 12,000 Hz, using a 16-channel encoder.

As mentioned in Table 2, a total of four fault conditions including the normal or fault-free condition, an inner raceway fault, a ball fault, and an outer raceway fault are considered in this study. The snapshots of vibration acceleration signals for each of these four conditions are given in Figure 1.

For each fault condition, the vibration acceleration signals are available at four different shaft speeds (i.e., 1796 revolutions per minute (rpm), 1772 rpm, 1748 rpm, and 1722 rpm). Therefore, a total of sixteen vibration acceleration signals are analyzed in this study. These vibration signals are divided into four datasets, one for each shaft speed including 1796 rpm, 1772 rpm, 1748 rpm, and 1722 rpm. For each dataset, the measured shaft speed is assumed as constant and taken as the value given in [41]. In other words, any inevitable random variations in the measured speed are not considered and the proposed method is not affected by either these random or design variations to achieve specified operating conditions. The details of these datasets are given in Table 3.

The length of vibration acceleration signals, in terms of the number of cycles of available data, varies across the datasets. In the proposed approach, the rectified vibration signal is divided into cycle length slices, as discussed in Section 3. The length of each slice, therefore, is different and varies with the shaft speed (i.e., the cycle length for the 1796 rpm dataset is ~400 samples; for the 1772 rpm dataset, it

TABLE 3: Datasets for the proposed fault diagnosis scheme.

Dataset	Fault type	Shaft speed (rpm)	Motor load (hp)	Number of cycles
1	Inner raceway	1796	0	~302
	Outer raceway	1796	0	~304
	Ball	1796	0	~305
	Normal	1796	0	~608
2	Inner raceway	1772	1	~300
	Outer raceway	1772	1	~301
	Ball	1772	1	~298
	Normal	1772	1	~1190
3	Inner raceway	1748	2	~296
	Outer raceway	1748	2	~295
	Ball	1748	2	~294
	Normal	1748	2	~1177
4	Inner raceway	1722	3	~293
	Outer raceway	1722	3	~293
	Ball	1722	3	~290
	Normal	1722	3	~1615

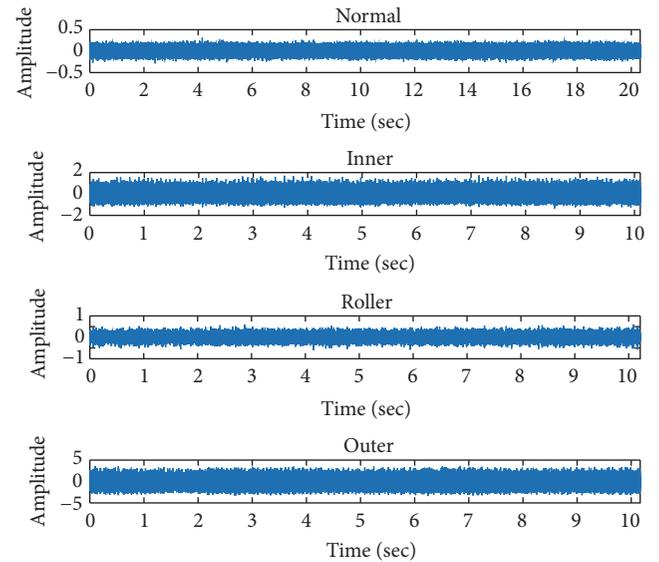


FIGURE 1: The time domain vibration acceleration signals for normal, inner race, ball/roller, and outer race faults.

is ~406 samples; for the 1748 rpm dataset, it is ~412 samples; and for the 1722 rpm dataset, it is ~418 samples). These slices are then stacked over each other to construct the grayscale vibration images. The heights of these images correspond to the number of slices that are stacked during their construction. The details of this process and the motivation for it are discussed in detail in Section 3. The raw vibration acceleration signals are first converted into grayscale images, which are then used for extracting microtexture information using the local binary operator.

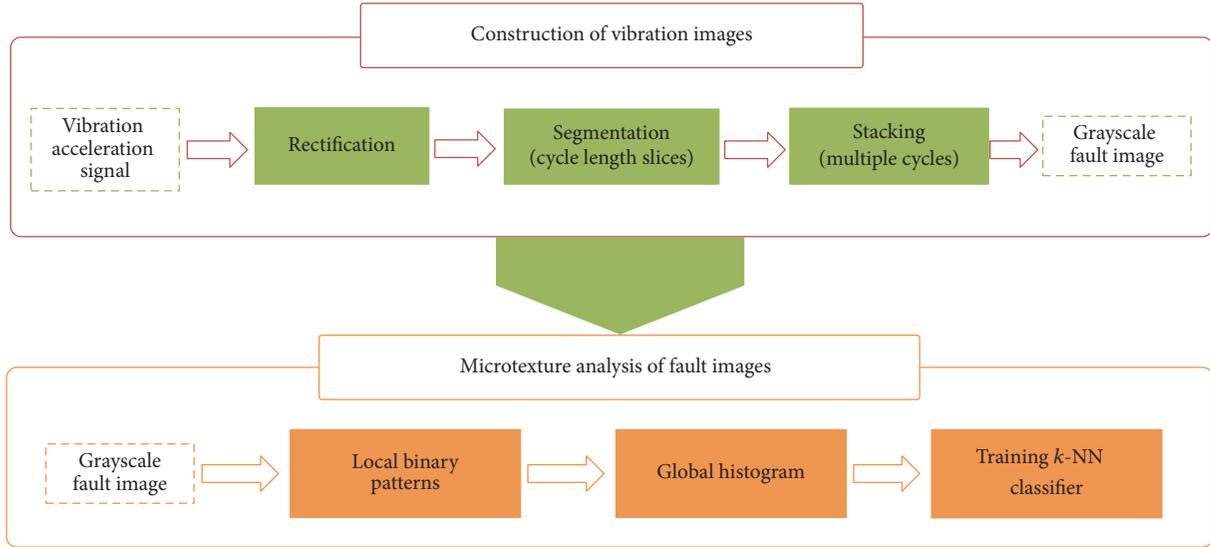


FIGURE 2: The proposed bearing fault diagnosis scheme based on microtexture analysis of the grayscale vibration acceleration fault images.

3. The Proposed Fault Diagnosis Scheme

3.1. Vibration Image Construction. The proposed fault diagnosis scheme, based on the microtexture analysis of vibration images using local binary patterns, is shown in Figure 2. The scheme works on two-dimensional (2D) images that are constructed from the time domain vibration acceleration signals. A vibration acceleration signal $x(t)$ is first rectified to get rid of the negative values. The resultant signal $|x(t)|$ is segmented into equal length slices. The length of each slice, l_s , is equal to the number of samples in one full shaft cycle or revolution of the bearing and is calculated using

$$l_s = \text{round}\left(\frac{60f_s}{\omega}, 3\right), \quad (1)$$

where f_s is the sampling frequency and ω is the shaft speed in revolutions per minute (rpm). The value of the quotient in (1) is rounded off to the nearest multiple of 3 because the local binary pattern (LBP) operator used in this study works on 3×3 pixel blocks. A total of H slices, each of length l_s , are used to construct the vibration fault images. Both dimensions including H and l_s of the grayscale vibration fault images are chosen so that they are integer multiples of 3. It also ensures an integer number of 3×3 -pixel blocks in each vibration image. This eliminates the chance of any overlap or loss of vibration data during segmentation and stacking. The process of image construction does not change or omit any samples from the original data. It simply projects the original vibration acceleration signals into a 2D grayscale intensity space, where the instantaneous acceleration values can be viewed as pixels. The intuition behind this method is to observe the behavior of the time domain vibration signal in intervals of H cycles and discover unique patterns in that behavior for each fault condition. The number of cycles H is determined experimentally such that the selected value of H gives the highest classification accuracy, as discussed in Section 4. These H

cycles of the vibration signal are stacked so that the width of the resulting image is equal to l_s pixels, while its height is equal to H lines. Experiments with images of different dimensions, which are discussed in Section 4, show that images with widths corresponding to cycle lengths display marginal variation in texture due to changes in the shaft frequency or speed.

3.2. Local Binary Patterns. LBPs were first proposed for texture classification [42] as a simple, effective, and computationally efficient technique that is invariant to rotation and changes in illumination. Since then, they have been extensively used in texture classification, image indexing, and facial recognition applications [43]. The use of local binary patterns has also been explored for diagnosing rotor unbalance, broken rotor bars, eccentricity, stator faults, and bowed rotor faults in induction motors [44].

LBP, as the name suggests, looks for micropatterns in small neighborhoods and then constructs a global frequency distribution of those micropatterns across the entire image in the form of a histogram. It is a nonlinear operator, which encodes microtexture information across small neighborhoods into n -bit texture descriptors or codes. The histogram of these texture descriptors is used to uniquely identify an image. These micropatterns or texture primitives, as they are called, can be an edge, a corner, a line end, a spot, or a flat area. The LBP operator can be applied to both circular and noncircular neighborhoods. In this study, this approach is applied to 3×3 noncircular neighborhoods, whereas in the literature it has been applied to both circular and noncircular neighborhoods of 3×3 and 4×4 pixels and even larger dimensions. The performance of the LBP operator is invariant under changes of illumination. The LBP operator is computationally simple, yet highly effective [43]. The illumination invariance property is especially useful in dealing with noisy vibration signals because the noise mostly affects the illumination level of the grayscale image [44].

The LBP operator constructs the texture descriptor by thresholding each pixel of the 3×3 neighborhood with respect to its central pixel. Simply, the outer eight pixels of the 3×3 neighborhood are compared with the central pixel (the 9th pixel) and replaced by a “1,” if it is greater than or equal to the central pixel and “0” otherwise. This thresholding reduces the neighboring pixels to a binary value of either 1 or 0, and these values are then used to construct an 8-bit texture descriptor for that neighborhood. The 8-bit texture descriptor, which has a decimal value within the range 0 to 255, encodes the texture information for that particular neighborhood. This process is repeated for all the neighborhoods in the entire image, and the texture information for each neighborhood is encoded into 8-bit texture descriptors. The total number of neighborhoods T_N in each image is determined using

$$T_N = \frac{Hl_s}{N}, \quad (2)$$

where H is the height of the image, l_s is its width, and N is the size of each neighborhood. In this study, N has been set to nine pixels, as we are considering only 3×3 neighborhoods. These local texture descriptors are then used to construct a global histogram that can be used to uniquely identify the entire image. Mathematically, for a grayscale image $G(x, y)$, if we let i_p denote the gray value of any arbitrary pixel $p(x, y)$ in a given neighborhood and let i_c denote the gray value of the central pixel in that neighborhood, then the texture descriptor t_N for a neighborhood of size N is given as follows [42]:

$$t_N = \sum_{n=0}^{N-1} s(i_p - i_c) 2^n, \quad (3)$$

where $s(i_p - i_c)$ is the thresholding step function, which is defined as

$$s(i_p - i_c) = \begin{cases} 1, & i_p - i_c \geq 0 \\ 0, & i_p - i_c < 0. \end{cases} \quad (4)$$

The texture descriptor in (2) can have 2^{N-1} unique values. Therefore, the global histogram for the entire image would require 2^{N-1} bins. However, a previous study [42] suggests that there are certain micropatterns, called the uniform patterns, which occur more frequently than others and are more discriminative. Uniform patterns have a uniformity measure of at most two, which is calculated by counting the number of binary transitions (i.e., 0 to 1 or 1 to 0 in the texture descriptor). Thus, a texture descriptor that has two or less than two binary transitions is considered uniform. Among the 256 (2^{N-1} for $N = 9$) possible texture descriptors, only 58 are uniform, while the rest are nonuniform. Hence, our global LBP histogram has 59 bins (i.e., 58 bins for the 58 uniform texture descriptors and one for the remaining 198 nonuniform texture descriptors). This global LBP histogram is used to uniquely identify each vibration fault image and the histograms of images for the same fault conditions would be similar to each other. In this study, we use Euclidean distance based similarity measure for the LBP histograms.

3.3. Classification Using k -NN. The duration of vibration signals for each fault condition, considered in this study, spans across hundreds of cycles as mentioned in Table 3. Each signal is converted into multiple grayscale images, which are then encoded into LBP histograms. These histograms are used by the classifier to uniquely identify each fault. In this study, the k -nearest neighbor classifier (k -NN) is used for classification [45]. It uses the Euclidean distance between the histogram of an unknown fault image and histograms of the training dataset images to classify the unknown image. The Euclidean distance $d(x, y)$ between the histograms of two fault images x and y is calculated using

$$d(x, y) = \sqrt{\sum_{j=1}^L (x_j - y_j)^2}, \quad (5)$$

where L is the number of bins in the histogram ($L = 59$ in this study). In this study, k , which is the number of training set samples or nearest neighbors considered in determining the class of an unknown sample, is assigned the value of 3. The diagnostic performance of the classifier is evaluated using average classification accuracy, sensitivity, and specificity, which are calculated using (6), (7), and (8), respectively:

$$ACA = \frac{\sum_{N_{\text{classes}}} N_{\text{TP}}}{N_{\text{images}}} \times 100 (\%), \quad (6)$$

$$\text{Sensitivity} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}, \quad (7)$$

$$\text{Specificity} = \frac{N_{\text{TN}}}{N_{\text{TN}} + N_{\text{FN}}}, \quad (8)$$

where N_{TP} is the number of images in class c that are correctly classified as class c , N_{FN} is the number of images in class c that are not classified as class c , N_{TN} is the number of images that are not in class c and are classified as not in class c , N_{FN} is the number of images that are in class c but classified as not in class c , N_{images} is the total number of images for all classes combined, and N_{classes} is the number of fault types or classes in the study.

4. Experimental Results and Analysis

4.1. Determining the Optimal Image Dimensions. Prior to the conversion of the vibration signal into 2D grayscale images, the image dimensions, which would yield a more uniform and robust texture, are determined. It is observed that variation in the image dimensions causes stark variations in texture. This is because stacking different segments of the vibration signal over each other would give rise to a more consistent texture if those segments have good correlation. For a given shaft speed, the images do show a uniform texture, but it may change as we change the shaft speed depending upon the selected image dimension. This is clearly evident in Figure 3, which shows the images for an inner race fault. It is observed that the texture spreads across the image as the shaft speed decreases and hence they look different for each shaft speed. Textures like these, as shown in Figure 3, cannot achieve shaft

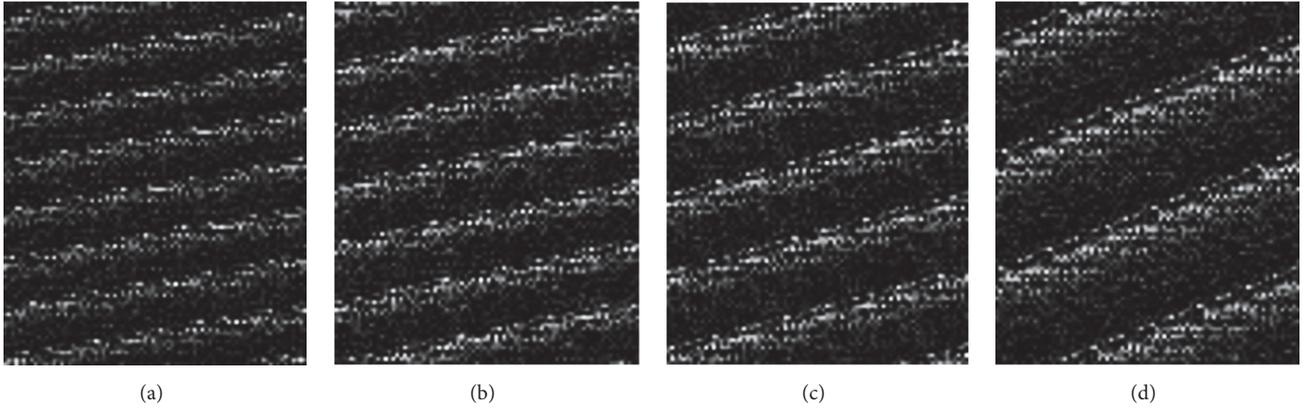


FIGURE 3: Vibration acceleration images for an inner race fault at (a) 1796 rpm, (b) 1772 rpm, (c) 1748 rpm, and (d) 1722 rpm.

speed invariance in fault diagnosis, because they are different from one another and a classifier trained on the LBP histogram of one cannot be used to accurately detect the other.

On the contrary, in Figure 4(a), the texture changes little with changes in shaft speed for the same fault condition. The only notable variation occurs in the illumination and global distribution of micropatterns or texture elements across the image. The LBP operator, as discussed earlier, is invariant to changes in illumination and therefore does not suffer any loss in performance. Similarly, the variation of the global spatial distribution of micropatterns does not adversely affect the performance of the LBP operator, as it only considers the frequency of occurrence of micropatterns, contrary to their spatial distribution. In Figure 4, the fault images have widths equal to l_s , which is calculated using (1). It can also be observed that the images in Figure 4 have higher aspect ratios as compared to those in Figure 3.

In general, it is observed that images with higher aspect ratios result in better classification accuracy, as shown in Figure 5. The higher aspect ratios prevent the texture from smearing across the image when there are changes in shaft speed. Therefore, in this study, vibration images with higher aspect ratios are used, as shown in Figure 4.

The width of these images is set to l_s , while their height is usually set to 15 pixels. The height of these images corresponds to the number of cycles of data used in constructing them; that is, one cycle of the vibration acceleration signal contributes one line of pixels to the fault image. It is clearly evident from Figure 6 that the classification accuracy improves as the height of the image (the number of cycles of vibration data used in constructing it) is increased, which is understandable as more cycles of vibration data result in a more distinct pattern for each fault. However, it reaches its asymptote when the image height is 15; any increase in image height beyond that point does not improve the classification accuracy.

4.2. Diagnostic Performance of the Proposed Method. In order to validate the proposed method, the available data is divided into four datasets, as shown in Table 3. The speed invariance of the proposed approach is verified by considering four

different scenarios. In the first scenario, dataset 1 is used for training, whereas datasets 2, 3, and 4 are used for testing. That is, the classifier is trained using the vibration acceleration signals for 1796 rpm only and then used to classify the unknown fault signals for 1772, 1748, and 1722 rpm. In the second scenario, dataset 2 is used for training, whereas datasets 1, 3, and 4 are used for testing. That is, the classifier is trained using the vibration acceleration signals for 1772 rpm only and then used to classify the unknown fault signals for 1796, 1748, and 1722 rpm. Similarly, in the third and fourth scenarios, datasets 3 and 4 are used as training datasets, respectively, whereas the remaining datasets are used for testing in each case. As mentioned in Table 3, each of these datasets corresponds to vibration signals recorded at a different shaft speed or frequency. Thus we verify the proposed method by training our classifier on a dataset for one shaft speed while testing it on datasets for three different shaft speeds or frequencies and we repeat this process for each of the four shaft frequencies considered in this study.

The diagnostic performance of the proposed approach is given in Table 4, which clearly indicates that the proposed method is effective in diagnosing bearing faults independent of variations in shaft speed, both random and planned. As explained earlier, this is due to the uniformity of textures in the vibration images, which are constructed after determining their optimal dimensions. It is observed from the results given in Table 4 that the diagnostic performance of the proposed method in terms of its classification accuracy, sensitivity, and specificity generally improves with increasing height of the fault images, which is equal to the number of cycles of the vibration data used to construct these images. The average classification accuracy reaches a maximum value of 99.74% for $H = 18$. In each scenario, the test dataset is three times larger than the training dataset, which shows that the proposed approach achieves very good generalization compared to existing techniques, which are generally validated using 3-fold cross validation, where 2 datasets out of 3 datasets are used for training and the remaining one is used for testing.

4.3. Comparison with Vibration Spectrum Imaging-Based Fault Diagnosis. The proposed method is compared with a

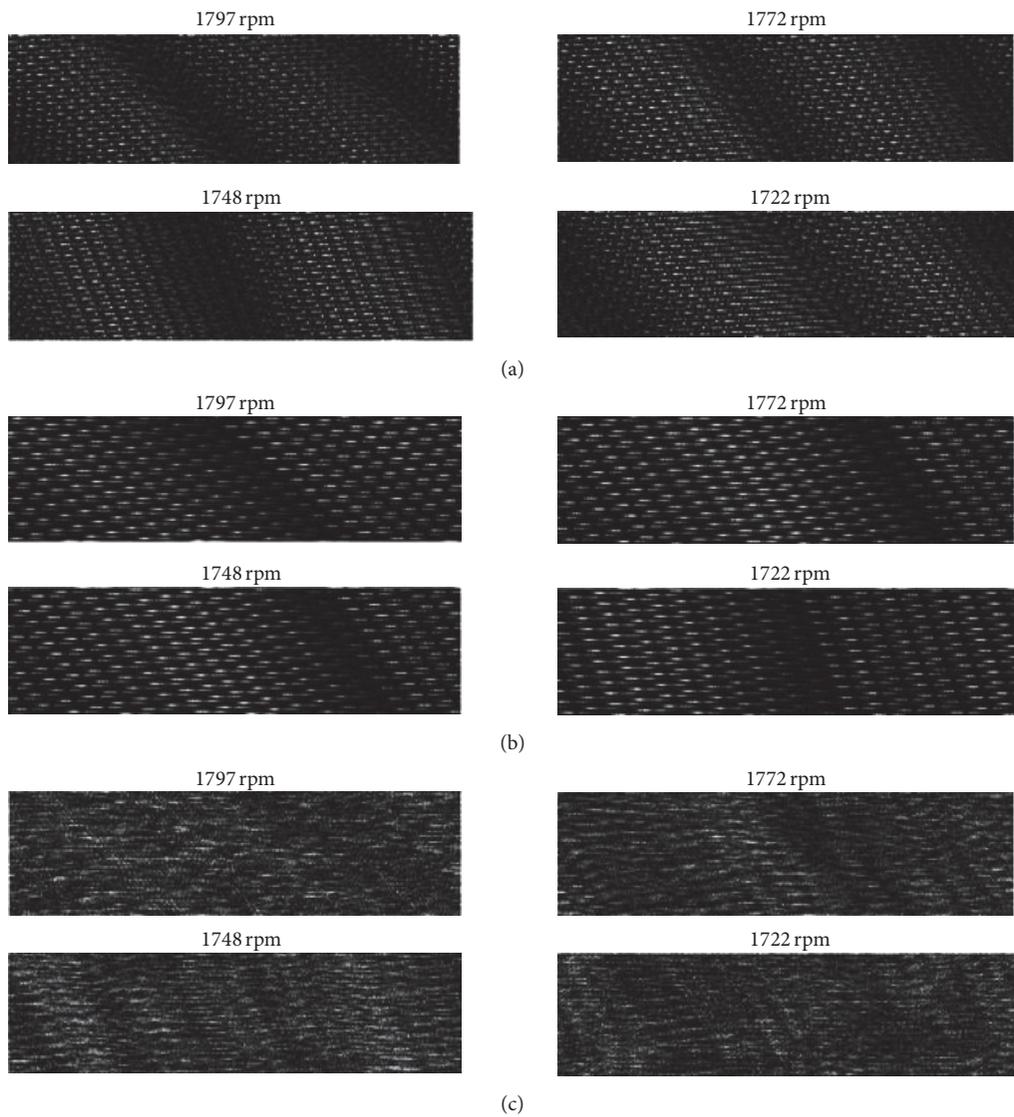


FIGURE 4: The vibration images for (a) the inner race fault, (b) the outer race fault, and (c) the ball fault.

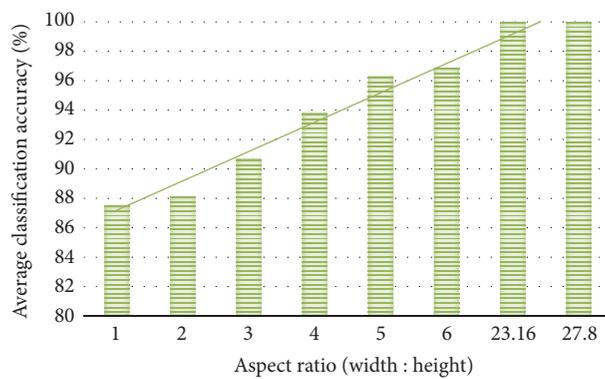


FIGURE 5: Effect of the vibration image aspect ratio on average classification accuracy.

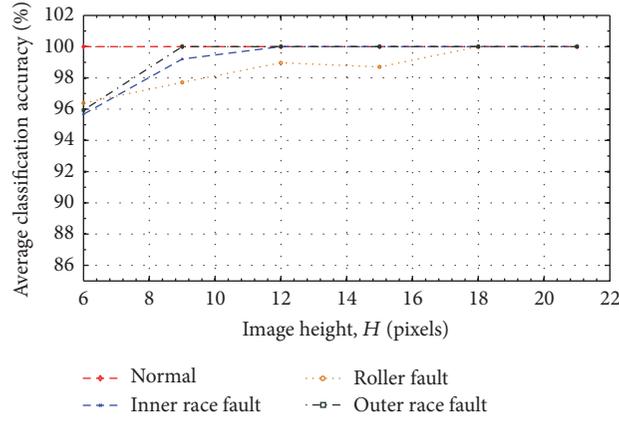


FIGURE 6: Effect of image height (H) on the average classification accuracy (ACA) for normal, inner, ball or roller, and outer raceway faults.

TABLE 4: Diagnostic performance of the proposed scheme under different scenarios.

Training dataset (H , number of training samples)	Testing datasets (number of test samples)	Classification accuracy (%)				Average classification accuracy (%) (sensitivity, specificity)
		Normal	Inner race fault (IRF)	Ball fault (BF)	Outer race fault (ORF)	
Dataset 1 (6, 144)	Datasets 2, 3, 4 (432)	90.28	86.81	86.81	97.92	90.45 (0.90, 1.00)
Dataset 2 (6, 144)	Datasets 1, 3, 4 (432)	88.19	88.19	96.53	94.44	91.84 (0.88, 1.00)
Dataset 3 (6, 144)	Datasets 1, 2, 4 (432)	99.31	89.52	98.61	95.83	95.83 (0.99, 1.00)
Dataset 4 (6, 144)	Datasets 1, 2, 3 (432)	93.75	85.42	100.0	97.22	94.10 (0.94, 1.00)
Dataset 1 (9, 96)	Datasets 2, 3, 4 (288)	100.0	91.67	100.0	97.92	97.40 (0.93, 1.00)
Dataset 2 (9, 96)	Datasets 1, 3, 4 (288)	93.75	94.79	100.0	98.96	96.88 (0.94, 1.00)
Dataset 3 (9, 96)	Datasets 1, 2, 4 (288)	100.0	91.67	100.0	97.92	97.40 (1.00, 1.00)
Dataset 4 (9, 96)	Datasets 1, 2, 3 (288)	98.96	94.79	98.96	98.96	97.92 (0.99, 1.00)
Dataset 1 (12, 72)	Datasets 2, 3, 4 (216)	94.44	95.83	98.61	97.22	96.53 (0.94, 1.00)
Dataset 2 (12, 72)	Datasets 1, 3, 4 (216)	100.0	98.61	98.61	95.83	98.26 (1.00, 1.00)
Dataset 3 (12, 72)	Datasets 1, 2, 4 (216)	100.0	100.0	100.0	98.61	99.65 (1.00, 1.00)
Dataset 4 (12, 72)	Datasets 1, 2, 3 (216)	100.0	97.22	100.0	100.0	99.31 (1.00, 1.00)
Dataset 1 (15, 57)	Datasets 2, 3, 4 (171)	91.23	98.25	98.25	100.0	96.93 (0.91, 1.00)
Dataset 2 (15, 57)	Datasets 1, 3, 4 (171)	98.25	100.0	100.0	100.0	99.56 (0.98, 1.00)
Dataset 3 (15, 57)	Datasets 1, 2, 4 (171)	100.0	98.25	100.0	96.49	98.68 (1.00, 1.00)
Dataset 4 (15, 57)	Datasets 1, 2, 3 (171)	100.0	98.25	100.0	100.0	99.56 (1.00, 1.00)
Dataset 1 (18, 48)	Datasets 2, 3, 4 (144)	97.92	100.0	100.0	100.0	99.48 (1.00, 1.00)
Dataset 2 (18, 48)	Datasets 1, 3, 4 (144)	100.0	100.0	100.0	100.0	100.0 (1.00, 1.00)
Dataset 3 (18, 48)	Datasets 1, 2, 4 (144)	100.0	100.0	100.0	100.0	100.0 (1.00, 1.00)
Dataset 4 (18, 48)	Datasets 1, 2, 3 (144)	100.0	97.92	100.0	100.0	99.48 (1.00, 1.00)

TABLE 5: Comparison of classification accuracy between existing BSI and the proposed method.

Method	Classification accuracy (%)				Average classification accuracy (%)
	Normal	Inner race fault	Outer race fault	Ball fault	
Vibration spectrum imaging [16]	78.20	77.50	79.00	75.80	77.62
Proposed	100.0	100.0	100.0	100.0	100.0

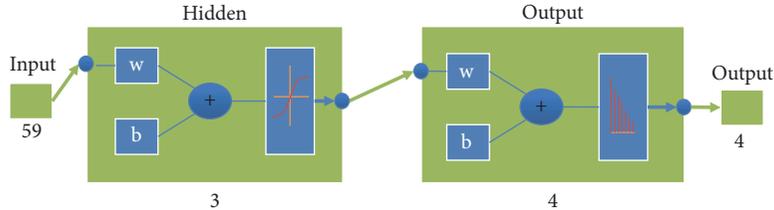


FIGURE 7: The artificial neural network used for fault diagnosis in bearings using the LBP histograms.

recent study [16], which uses vibration spectrum imaging (VSI) and artificial neural networks (ANN) to diagnose bearing faults. The VSI based approach uses data from Case Western Reserve University but considers only dataset 3 (i.e., the dataset for 1748 rpm in Table 3). It takes eight windows of the time domain vibration signal, each 1024 samples in length, and then applies a 513-point fast Fourier transform (FFT) to each window. The resultant spectral information for these eight windows is stacked on top of each other to create a 513×8 pixel grayscale vibration spectrum image. This image is smoothed using a 2D averaging filter (of size 8×4), and then the filtered grayscale image is converted into a binary image using thresholding. The threshold is determined by optimizing a certain cost function, and the optimum value of 0.7 is used for thresholding. The performance of the VSI based method is highly susceptible to the value of the threshold because it governs the quality of the input vectors to the ANN and hence its classification accuracy. The binary image with 513×8 (4104 binary spectral components as inputs to the neural network) is used as an input to the ANN. The data is divided into training (70%), validation (15%), and testing (15%) sets. In order to compare the proposed approach with VSI, the same network architecture is used as shown in Figure 7, except for the input layer (the proposed approach uses an input vector of length 59).

The datasets in Table 3 are merged to create one dataset, with four fault conditions (i.e., normal, inner race fault, roller fault, and outer race fault) at four different shaft speeds (i.e., 1796 rpm, 1772 rpm, 1748 rpm, and 1722 rpm). The diagnostic performance of VSI and the proposed approach is shown in Table 5. It is clearly evident that the proposed approach delivers superior diagnostic performance on datasets with variations in shaft speed compared to VSI, which uses the shaft speed dependent spectral information for fault diagnosis. Under variable speed conditions the diagnostic performance of methods that use spectral information is bound to deteriorate, as speed variations drastically change the spectral content of the vibration signals.

5. Conclusion

This paper investigates a new dimension in bearing fault diagnosis and presents a new method that is invariant to both random and premeditated variations in shaft speed. This is a very important aspect of fault diagnosis in bearings since traditional approaches diagnose bearing defects generally by detecting their fundamental defect frequencies. Though highly effective, these techniques have certain caveats. The defect frequencies depend on the nonstationary shaft speed, and variations in shaft speed cause inevitable variations in these fundamental defect frequencies. These variations can be small random variations that are usually tackled by time-frequency and subband analysis of the vibration signals, which makes these approaches tedious and computationally expensive, with costly and difficult practical implementations. In the case of large premeditated variations in shaft speed, the fault data is divided into different datasets depending upon the speed, and each dataset is separately processed for fault diagnosis. Thus, premeditated variations in shaft speed entail the recalculation of features and the training of a new instance of the classifier for every dataset. These crucial limitations justified investigation into a simple, automated, approach for fault diagnosis in bearings that is effective under variable speed conditions. The proposed scheme transforms the time domain vibration acceleration signal into grayscale fault images of appropriate dimensions and then classifies those images based upon their unique textures. The image textures are encoded using the local binary pattern operator, which is a highly effective texture descriptor. This study validated the proposed scheme by using fault images for four different operating speeds. A k -NN classifier was trained using images for one operating speed, and then its classification performance was measured by testing it with the fault images for the remaining three operating speeds. This was repeated for the fault images in all the datasets. The classifier yields an average classification accuracy of 99.74%, which shows that the proposed approach is invariant to variations in shaft speed. A comparison with

a recent technique based upon vibration spectrum imaging shows that the proposed method gives better diagnostic performance. Despite its advantages, there are certain aspects of the proposed approach that need further investigation, such as its performance at relatively low shaft speeds.

Competing Interests

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

Acknowledgments

This work was supported by Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (no. 20162220100050, no. 20161120100350), in part by the Leading Human Resource Training Program of Regional Neo Industry through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2016H1D5A1910564), in part by Business for Cooperative R&D between Industry, Academy, and Research Institute funded by Korea Small and Medium Business Administration in 2016 (Grants nos. C0395147, S2381631), in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A3B03931927), and in part by the Development of a Basic Fusion Technology in Electric Power Industry (Ministry of Trade, Industry & Energy, 201301010170D).

References

- [1] R. Saidur, "A review on electrical motors energy use and energy savings," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 3, pp. 877–898, 2010.
- [2] I. El-Thalji and E. Jantunen, "Dynamic modelling of wear evolution in rolling bearings," *Tribology International*, vol. 84, pp. 90–99, 2015.
- [3] O. V. Thorsen and M. Dalva, "Failure identification and analysis for high-voltage induction motors in the petrochemical industry," *IEEE Transactions on Industry Applications*, vol. 35, no. 4, pp. 810–818, 1999.
- [4] S. Sheng, *Report on Wind Turbine Subsystem Reliability—A Survey of Various Databases*, National Renewable Energy Laboratory, 2013.
- [5] W. Qiao and D. Lu, "A survey on wind turbine condition monitoring and fault diagnosis—part I: components and sub-systems," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 10, pp. 6536–6545, 2015.
- [6] P. Bangalore and L. B. Tjernberg, "An artificial neural network approach for early fault detection of gearbox bearings," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 980–987, 2015.
- [7] X. Gong and W. Qiao, "Bearing fault diagnosis for direct-drive wind turbines via current-demodulated signals," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3419–3428, 2013.
- [8] X. Gong and W. Qiao, "Current-based mechanical fault detection for direct-drive wind turbines via synchronous sampling and impulse detection," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 3, pp. 1693–1702, 2015.
- [9] D. Siegel, C. Ly, and J. Lee, "Methodology and framework for predicting helicopter rolling element bearing failure," *IEEE Transactions on Reliability*, vol. 61, no. 4, pp. 846–857, 2012.
- [10] J. Feldman, B. M. Hanrahan, S. Misra et al., "Vibration-based diagnostics for rotary MEMS," *Journal of Microelectromechanical Systems*, vol. 24, no. 2, pp. 289–299, 2015.
- [11] D. Hansen and A. H. Olsson, *ISO Standard 13373-1: 2002—Condition Monitoring and Diagnostics of Machines—Vibration Condition Monitoring—Part 1: General Procedures*, International Standards Organization, 2002.
- [12] D. Hansen and A. H. Olsson, "ISO Standard 13373-2: 2005: condition monitoring and diagnostics of machines-vibration condition monitoring-part 2: processing, analysis and presentation of vibration data," *International Standards Organization*, 2009.
- [13] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part II: fault diagnosis with knowledge-based and hybrid/active approaches," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3768–3774, 2015.
- [14] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—part I: fault diagnosis with model-based and signal-based approaches," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3757–3767, 2015.
- [15] T. W. Rauber, F. de Assis Boldt, and F. M. Varejão, "Heterogeneous feature models and feature selection applied to bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 637–646, 2015.
- [16] M. Amar, I. Gondal, and C. Wilson, "Vibration spectrum imaging: a novel bearing fault classification approach," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 494–502, 2015.
- [17] J. Seshadrinath, B. Singh, and B. K. Panigrahi, "Investigation of vibration signatures for multiple fault diagnosis in variable frequency drives using complex wavelets," *IEEE Transactions on Power Electronics*, vol. 29, no. 2, pp. 936–945, 2014.
- [18] B. Eftekharnajad, M. R. Carrasco, B. Charnley, and D. Mba, "The application of spectral kurtosis on Acoustic Emission and vibrations from a defective bearing," *Mechanical Systems and Signal Processing*, vol. 25, no. 1, pp. 266–284, 2011.
- [19] M. Kang, J. Kim, B. Choi, and J.-M. Kim, "Envelope analysis with a genetic algorithm-based adaptive filter bank for bearing fault detection," *The Journal of the Acoustical Society of America*, vol. 138, no. 1, pp. EL65–EL70, 2015.
- [20] M. Kang, J. Kim, and J.-M. Kim, "Reliable fault diagnosis for incipient low-speed bearings using fault feature analysis based on a binary bat algorithm," *Information Sciences*, vol. 294, pp. 423–438, 2015.
- [21] M. Kang, J. Kim, and J.-M. Kim, "An FPGA-based multicore system for real-time bearing fault diagnosis using ultrasampling rate AE signals," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 4, pp. 2319–2329, 2015.
- [22] M. Kang, J. Kim, and J.-M. Kim, "High-performance and energy-efficient fault diagnosis using effective envelope analysis and denoising on a general-purpose graphics processing unit," *IEEE Transactions on Power Electronics*, vol. 30, no. 5, pp. 2763–2776, 2015.
- [23] M. Kang, J. Kim, J.-M. Kim, A. C. C. Tan, E. Y. Kim, and B.-K. Choi, "Reliable fault diagnosis for low-speed bearings

- using individually trained support vector machines with kernel discriminative feature analysis," *IEEE Transactions on Power Electronics*, vol. 30, no. 5, pp. 2786–2797, 2015.
- [24] R. Islam, S. A. Khan, and J.-M. Kim, "Discriminant feature distribution analysis-based hybrid feature selection for online bearing fault diagnosis in induction motors," *Journal of Sensors*, vol. 2016, Article ID 7145715, 16 pages, 2016.
- [25] P. H. Nguyen and J.-M. Kim, "Multifault diagnosis of rolling element bearings using a wavelet kurtogram and vector median-based feature analysis," *Shock and Vibration*, vol. 2015, Article ID 320508, 14 pages, 2015.
- [26] M. Hamadache, D. Lee, and K. C. Veluvolu, "Rotor speed-based bearing fault diagnosis (RSB-BFD) under variable speed and constant load," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 10, pp. 6486–6495, 2015.
- [27] D. Wang, P. W. Tse, and K. L. Tsui, "An enhanced Kurtogram method for fault diagnosis of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 176–199, 2013.
- [28] M. Pineda-Sanchez, J. Perez-Cruz, J. Roger-Folch, M. Riera-Guasp, A. Sapena-Bano, and R. Puche-Panadero, "Diagnosis of induction motor faults using a DSP and advanced demodulation techniques," in *Proceedings of the 9th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives (SDEMPED '13)*, pp. 69–76, Valencia, Spain, August 2013.
- [29] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—a tutorial," *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 485–520, 2011.
- [30] I. Bediaga, X. Mendizabal, A. Arnaiz, and J. Munoa, "Ball bearing damage detection using traditional signal processing algorithms," *IEEE Instrumentation and Measurement Magazine*, vol. 16, no. 2, pp. 20–25, 2013.
- [31] P. K. Kankar, S. C. Sharma, and S. P. Harsha, "Fault diagnosis of rolling element bearing using cyclic autocorrelation and wavelet transform," *Neurocomputing*, vol. 110, pp. 9–17, 2013.
- [32] J. Rafiee, M. A. Rafiee, and P. W. Tse, "Application of mother wavelet functions for automatic gear and bearing fault diagnosis," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4568–4579, 2010.
- [33] E. Cabal-Yepez, A. G. Garcia-Ramirez, R. J. Romero-Troncoso, A. Garcia-Perez, and R. A. Osornio-Rios, "Reconfigurable monitoring system for time-frequency analysis on industrial equipment through STFT and DWT," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 2, pp. 760–771, 2013.
- [34] S. Nandi, T. C. Ilamparithi, S. B. Lee, and D. Hyun, "Detection of eccentricity faults in induction machines based on nameplate parameters," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 1673–1683, 2011.
- [35] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 108–126, 2013.
- [36] J. Zheng, J. Cheng, and Y. Yang, "Generalized empirical mode decomposition and its applications to rolling element bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 40, no. 1, pp. 136–153, 2013.
- [37] X. Zhang and J. Zhou, "Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines," *Mechanical Systems and Signal Processing*, vol. 41, no. 1-2, pp. 127–140, 2013.
- [38] G. F. Bin, J. J. Gao, X. J. Li, and B. S. Dhillon, "Early fault diagnosis of rotating machinery based on wavelet packets—empirical mode decomposition feature extraction and neural network," *Mechanical Systems and Signal Processing*, vol. 27, pp. 696–711, 2012.
- [39] M. Kang, J. Kim, L. M. Wills, and J.-M. Kim, "Time-varying and multiresolution envelope analysis and discriminative feature analysis for bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 12, pp. 7749–7761, 2015.
- [40] D. Zhao, J. Li, and W. Cheng, "Feature extraction of faulty rolling element bearing under variable rotational speed and gear interferences conditions," *Shock and Vibration*, vol. 2015, Article ID 425989, 9 pages, 2015.
- [41] Case Western Reserve University, "Seeded Fault Test Data," <http://csegroups.case.edu/bearingdatacenter/home>.
- [42] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [43] S. Brahmam, L. C. Jain, L. Nanni, and A. Lumini, *Local Binary Patterns: New Variants and Applications*, Springer, 2014.
- [44] M. R. Shahriar, T. Ahsan, and U. Chong, "Fault diagnosis of induction motors utilizing local binary pattern-based texture analysis," *EURASIP Journal on Image and Video Processing*, vol. 2013, article 29, 2013.
- [45] D. H. Pandya, S. H. Upadhyay, and S. P. Harsha, "Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN," *Expert Systems with Applications*, vol. 40, no. 10, pp. 4137–4145, 2013.

Research Article

Research on Dynamic Modeling and Application of Kinetic Contact Interface in Machine Tool

Dan Xu^{1,2} and Zhixin Feng^{1,2}

¹School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China

²Science & Technology on Reliability & Environmental Engineering Laboratory, Beihang University, Beijing 100191, China

Correspondence should be addressed to Dan Xu; xudan@buaa.edu.cn

Received 22 July 2016; Accepted 16 October 2016

Academic Editor: A. El Sinawi

Copyright © 2016 D. Xu and Z. Feng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A method is presented which is a kind of combining theoretic analysis and experiment to obtain the equivalent dynamic parameters of linear guideway through four steps in detail. From statics analysis, vibration model analysis, dynamic experiment, and parameter identification, the dynamic modeling of linear guideway is synthetically studied. Based on contact mechanics and elastic mechanics, the mathematic vibration model and the expressions of basic mode frequency are deduced. Then, equivalent stiffness and damping of guideway are obtained in virtue of single-freedom-degree mode fitting method. Moreover, the investigation above is applied in a certain gantry-type machining center; and through comparing with simulation model and experiment results, both availability and correctness are validated.

1. Introduction

The dynamic analysis and simulation technology of numerical control machine tool are a very important research direction in modern advanced manufacturing technology and equipment technology. In a numerical control machine tool, the linear rolling guide is not only a significant functional component but also an important kinetic contact interface, meaning its characteristic has a direct effect on the machining precision and performance. Therefore, the modeling of contact interface, which takes the dynamic property of the NC machine tool into consideration, is a prerequisite research for establishing the dynamic overall model. Without this, it is impossible to get a practical conclusion.

For the dynamic analysis and prediction of machine tool, the modeling of contact interface and accurate identifying of the parameters are the main difficulty. In recent years, the domestic and overseas research of contact interface is basic on three aspects: the mechanism of contact interface, modeling, and parameter identification. Some of the researchers, who are represented by Wen et al. [1, 2], put forward the scale-independent stiffness fractal modeling of the normal and tangential contact. At the same time, they reveal the nonlinear relationship between the contact stiffness and each interface

parameter. The researchers, represented by Zhang et al. [3] and Mao et al. [4], build the fundamental property model of contact interface and multinode dynamic model; Dhupia et al. [5] applies the frequency-domain joint part model with weak nonlinear characteristic to the basic machine tool modeling and predicts the processing performance. When it comes to identifying the parameter of the joint part, there are several main ways such as frequency response function identification method [6], response coupling method [7], and contact interface parameter optimization method which is based on finite element modeling [8]. At present, the research of joint part is in the contending stage, and there are numerous problems of the mechanism and modeling to be solved.

There are certain difficulties to apply the mechanism model to engineering, because although the fundamental performance model and parameter statistic obtained have a versatility, it should be based on a great amount of experiments.

On the basis of the researches before [9–11] and starting from the material and structure aspects, this paper puts forward a combined “analysis-experiment” dynamic modeling method for linear guideway and studies the method of identifying the parameters of guideway contact interface. Furthermore, this paper builds a four-in-one joint part

research method of “static stiffness model-vibration model-experimental parameter-parameter identification” and applies this method to the guideway contact part and overall dynamic modeling of the gantry machining center. The research results provide an effective and feasible solution channel and practical method to predict the machine performance in the pattern design phase.

2. Modeling of the Rolling Guideway Contact Part

2.1. Modeling Method. From the perspective of material parameters and structural distortion and based on the elastic mechanics and vibration mechanics theory, the guideway contact part dynamic modeling method is established from these four aspects, “static stiffness model, vibration model, experimental parameter, and parameter identification.” Figure 1 clearly shows the relationship of each part of the method and the parameter that should be provided for next part. First of all, obtain the relationship between the stress and deformation through analyzing the static dynamics while the rolling linear guideway is under the general loading, solve the static stiffness via building the guideway static stiffness model, and provide the input parameter for the guideway vibration model. Secondly, employ the Lagrange method to analyze the vibration property of the linear rolling guideway and solve the basic modal vibration frequency of the vibration model through using analytic method, to provide the input requirement for the dynamic model parameter identification. At last, with the help of the hammer-hitting experiment, the multipoint frequency response function of the guideway is acquired. Combined with the guideway vibration basic modal frequency, the single degree modal fitting method is used to obtain the stiffness and damping value of the guideway equivalent model.

2.2. Static Stiffness Model. In the condition that the linear rolling guideway is under the load in vertical direction F_V , the load in horizontal direction F_H , and vertical moment M_γ (rotating around the y axis) (shown in Figure 2), the relation [12] between the stress and deformation is obtained via the mechanical analysis while the linear guideway is under general loading. The equation system is presented as

$$\begin{aligned} F_V &= f_1(\delta_V, \delta_H, \gamma), \\ F_H &= f_2(\delta_V, \delta_H, \gamma), \\ M_\gamma &= f_3(\delta_V, \delta_H, \gamma). \end{aligned} \quad (1)$$

This nonlinear equation system can be solved by numerical computation method.

Substitute a series of different external applied load F_V , F_H , and M_γ into the algorithm above and then a series of relative displacement δ_V , δ_H , and γ under different loading of the guideway can be obtained. Using the computer calculating method, the vertical direction stiffness, horizontal direction

TABLE 1: MRB35 rolling guideway model parameters.

Parameter Items	Values
The length of the slider L_G	103 mm
The diameter of the roller D_a	4.5 mm
The length of the roller L	5.1 mm
The mass of the slider M	2.2 kg
The moment of inertia around x -axis of the slider J_x	$3.927626 \times 10^{-3} \text{ kg m}^2$
The moment of inertia around y -axis of the slider J_y	$5.761594 \times 10^{-3} \text{ kg m}^2$
The moment of inertia around z -axis of the slider J_z	$2.467506 \times 10^{-3} \text{ kg m}^2$
The number of the rollers Z_L	21
The horizontal distance between each roller e	32 mm
The vertical distance between loaded point and upper rollers h_1	23.5 mm
The vertical distance between each roller h_0	10.7 mm
Tight deformation amount Δ	0.00072 mm
Contact angle β	45°

stiffness, and the rotation angular rigidity can be, respectively, solved:

$$\begin{aligned} K_V &= \left. \frac{\partial F_V}{\partial \delta_V} \right|_{F_V}, \\ K_H &= \left. \frac{\partial F_H}{\partial \delta_H} \right|_{F_H}, \\ K_\gamma &= \left. \frac{\partial M_\gamma}{\partial \gamma} \right|_{M_\gamma}. \end{aligned} \quad (2)$$

On the basis of the computing method above and MATLAB software, taking the Schneeberger MRB35 linear rolling guideway as an example, the parameters of the guideway measure are shown in Table 1.

Figure 3 shows the relation between the vertical bearing load and vertical direction deformation while the load is within the bearing capacity. In Figure 3, the line marked with red five-pointed stars represents computing data by theoretical model, and the line marked with blue points represents the experimental data provided by the guideway manufacturer. As is shown in the figure, the relation between the loading and deformation is moderate linear and it can be approximately treated as a linear relation. In the simplified model, it can be replaced by a linear model.

2.3. Dynamic Model. Ignoring the mass of the roller, linear rolling guideway vibration model only takes the normal direction stiffness of the roller (the normal direction means the direction that is vertical to the contact interface) and simplifies the roller between the two contact interfaces as a spring that is vertical to the contact interface, whose stiffness is K_n . It is shown in Figure 4.

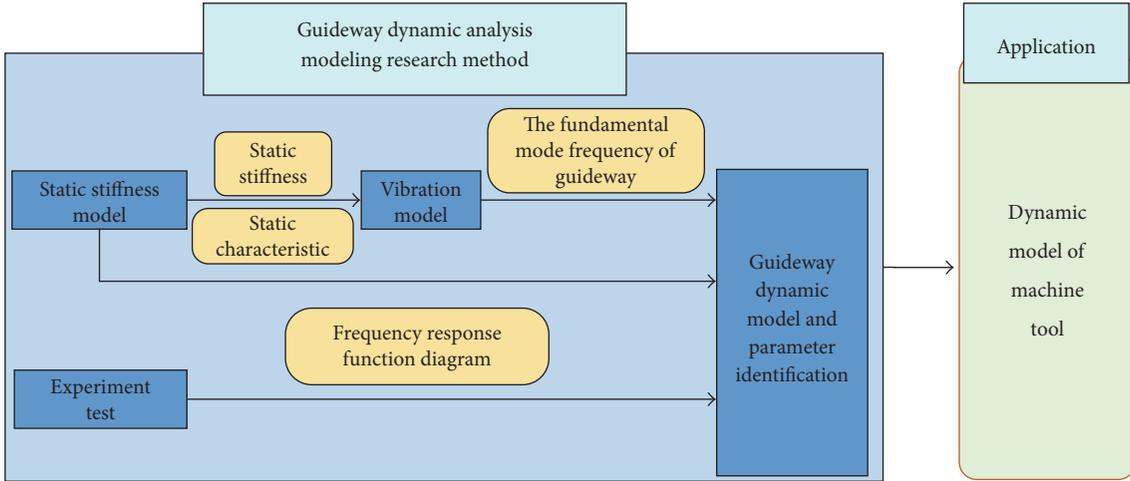


FIGURE 1: Mind map of the guideway dynamic model research method.

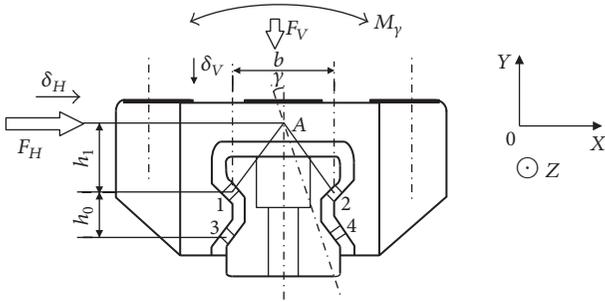


FIGURE 2: Bearing loading and guideway deformation graph.

The solution for the stiffness K_n [9, 10] is as follows.

The simplified spring normal stiffness of the single sphere roller is

$$K_n = \frac{dF}{d\alpha} = \frac{3}{2} k_h \alpha^{1/2}. \quad (3)$$

The simplified spring normal stiffness of the single cylinder roller is

$$K_n = \frac{dQ}{d\delta} = \frac{10}{9} k_l L^{8/9} \delta^{1/9} \quad (4)$$

or $K_n = \frac{K_V}{4Z \sin \beta}$,

where K_n means the spring normal stiffness which is simplified from the single roller; k_h means the roller contact rigidity of the spherical guideway; α means the elastic deformation of the roll ball; k_l means the roller contact stiffness of the roller guideway; L means the length of the roller guideway; δ means the elastic deformation of the roller guideway; K_V means the vertical stiffness of the linear guideway; Z means the number of the single row roller; β means the contact angle of guideway and groove.

Using the Lagrange equation, the dynamic equation of the linear guideway can be detected. Lagrange equation uses the

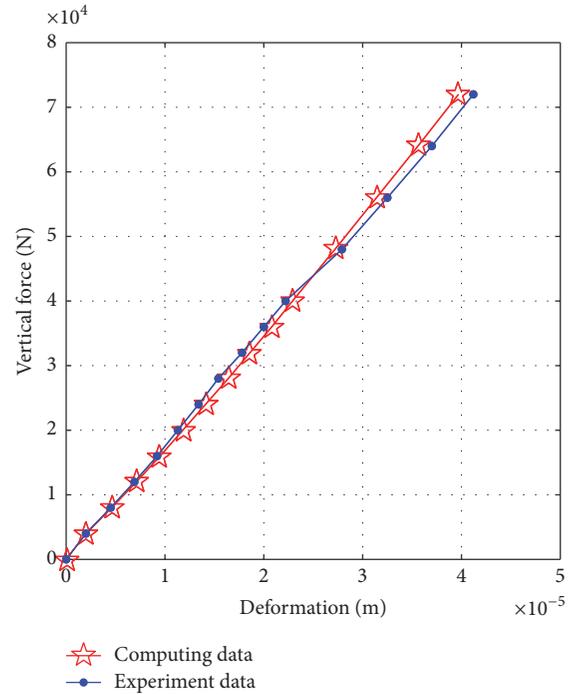


FIGURE 3: The relationship of vertical stress and deformation of MRB35 guideway.

variation of the generalized coordinates to present the virtual displacement of the particle in the particle system. This kind of system is more useful than Newton's laws of motion while solving some problems (e.g., the small oscillation theory and rigid body dynamics).

The Lagrange equation of linear rolling guideway is given in the following:

$$\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_j} \right) - \left(\frac{\partial T}{\partial q_j} \right) = Q_j \quad (j = 1, 2, \dots, k), \quad (5)$$

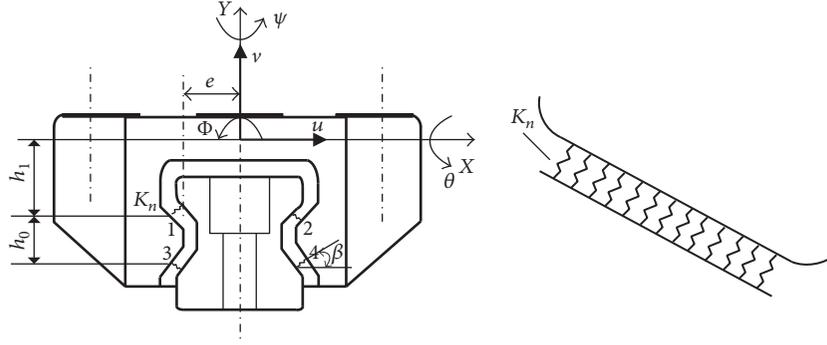


FIGURE 4: Simplified linear rolling guideway model.

where T means the kinetic energy of the system; q_j means the generalized coordinate of the system; \dot{q}_j means the generalized velocity of the system; Q_j means the generalized force which corresponds to the generalized coordinate of the system; k means the number of the generalized coordinates of the system, meaning the free degree of the system.

If all the force applied on the particle is potential force, using V to represent the potential function of the system; the generalized force corresponding to the generalized coordinate q_j is

$$Q_j = -\frac{\partial V}{\partial q_j}. \quad (6)$$

Defining $L = T - V$, L is named the Lagrange function or dynamic potential. Then the Lagrange function in potential field can be written as

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_j} \right) - \left(\frac{\partial L}{\partial q_j} \right) = 0 \quad (j = 1, 2, \dots, k). \quad (7)$$

The Lagrange function of the linear rolling guideway system shown in Figure 4 can be established directly.

The overall dynamic energy E_k can be expressed as follows.

$$E_k = \frac{1}{2} M \dot{u}^2 + \frac{1}{2} M \dot{v}^2 + \frac{1}{2} J_x \dot{\theta}^2 + \frac{1}{2} J_y \dot{\psi}^2 + \frac{1}{2} J_z \dot{\phi}^2. \quad (8)$$

The potential function E_p can be expressed as

$$E_p = E_{p1} = \frac{1}{2} K_n \sum_{i=1}^{Z_L} (\delta_{1i}^2 + \delta_{2i}^2 + \delta_{3i}^2 + \delta_{4i}^2),$$

$$E_{p1} = \frac{1}{2} K_n \sum_{i=1}^{Z_L} \left[\{(u + h_1 \phi + l_{1i} \psi) \cos \beta + (v - e \phi - l_{1i} \theta) \sin \beta\}^2 + \{(-u - h_1 \phi - l_{2i} \psi) \cos \beta + (v + e \phi - l_{2i} \theta) \sin \beta\}^2 + \{(u + h_3 \phi + l_{3i} \psi) \cos \beta + (-v + e \phi + l_{3i} \theta) \sin \beta\}^2 + \{(-u - h_3 \phi - l_{4i} \psi) \cos \beta + (-v - e \phi + l_{4i} \theta) \sin \beta\}^2 \right],$$

$$Z_{L1} = Z_{L2} = Z_{L3} = Z_{L4} = Z_L,$$

$$l_{1i} = l_{2i} = l_{3i} = l_{4i} = \left(\frac{Z_L - 2i + 1}{2(Z_L - 1)} \right) l_L \quad (i = 1, 2, \dots, Z_L),$$

$$h_3 = h_1 + h_0,$$

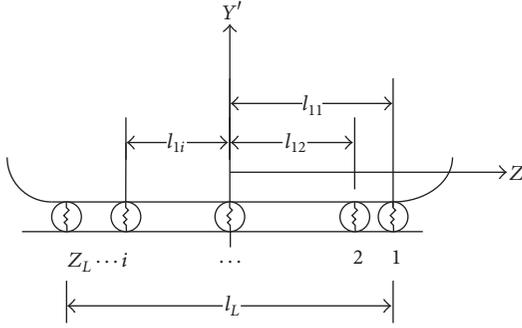
(9)

where u means the displacement in X direction; v means the displacement in Y direction; θ means the angular displacement around x -axis (pitching motion); ψ means the angular displacement around y -axis (yawing motion); ϕ means angular displacement around z -axis (rolling motion); M means the mass of slider; J_x means the moment of inertia around x -axis of the slider; J_y means the moment of inertia around y -axis of the slider; J_z means the moment of inertia around z -axis of the slider; e means the moment of inertia around z -axis of the slider; h_1 means Y direction distance from the center of roller 1 to the origin of coordinate; h_0 means Y direction distance from the center of roller 1 to the center of roller 3; β means the angle between the normal direction of the contact interface and the horizontal X direction; E_{p1} means the elastic potential energy of the normal direction spring; Z_{Lj} ($j = 1, 2, 3, 4$) means the number of the rollers in the i th raceway; l_{ji} ($j = 1, 2, 3, 4$) means the z -axis coordinate of the i th roller in j th raceway as is shown in Figure 5; the angle between Y' axis and y -axis is $(\pi/2 - \beta)$; l_L means the length of single raceway roller.

Establishing the linear guideway system dynamic function,

$$M \ddot{u} + 4Z_L K_n \cos^2 \beta u + 2Z_L K_n (h_1 + h_3) \cos^2 \beta \phi = 0, \quad (10)$$

$$M \ddot{v} + 4Z_L K_n \sin^2 \beta v = 0, \quad (11)$$

FIGURE 5: The position of the roller in $Y'Z$ plane.

$$J_x \ddot{\theta} + K_n \sum_{i=1}^{Z_L} (l_{1i}^2 + l_{2i}^2 + l_{3i}^2 + l_{4i}^2) \sin^2 \beta \theta = 0, \quad (12)$$

$$J_y \ddot{\psi} + K_n \sum_{i=1}^{Z_L} (l_{1i}^2 + l_{2i}^2 + l_{3i}^2 + l_{4i}^2) \cos^2 \beta \psi = 0, \quad (13)$$

$$J_z \ddot{\phi} + 2Z_L K_n (h_1 + h_3) \cos^2 \beta u + \{2Z_L K_n (h_1^2 + h_3^2) \cos^2 \beta + 4Z_L K_n e^2 \sin^2 \beta + 4Z_L K_n h_0 e \sin \beta \cos \beta\} \phi = 0. \quad (14)$$

Detected from (11) the vibration frequency of the direction that vertical to y -axis of the guideway can be obtained:

$$f_V = \frac{1}{\pi} \sqrt{\frac{Z_L K_n \sin^2 \beta}{M}}. \quad (15)$$

Detected from (12), the vibration frequency of the pitching motion around x -axis of the guideway can be obtained:

$$f_{XP} = \frac{1}{2\pi} \sqrt{\frac{K_n \sum_{i=1}^{Z_L} (l_{1i}^2 + l_{2i}^2 + l_{3i}^2 + l_{4i}^2) \sin^2 \beta}{J_x}}. \quad (16)$$

Detected from (4), the vibration frequency of the yawing motion around y -axis of the guideway can be obtained:

$$f_{YY} = \frac{1}{2\pi} \sqrt{\frac{K_n \sum_{i=1}^{Z_L} (l_{1i}^2 + l_{2i}^2 + l_{3i}^2 + l_{4i}^2) \cos^2 \beta}{J_y}}. \quad (17)$$

The displacement along X direction u and the angular displacement around z -axis couple together so (10) and (14) are simultaneously solved to get the vibration frequency.

$$M \ddot{u} + 4Z_L K_n \cos^2 \beta u + 2Z_L K_n (h_1 + h_3) \cos^2 \beta \phi = 0, \\ J_z \ddot{\phi} + 2Z_L K_n (h_1 + h_3) \cos^2 \beta u + \{2Z_L K_n (h_1^2 + h_3^2) \cos^2 \beta + 4Z_L K_n e^2 \sin^2 \beta + 4Z_L K_n h_0 e \sin \beta \cos \beta\} \phi = 0. \quad (18)$$

Suppose $u = Ue^{j\omega t}$ and $\phi = \Phi e^{j\omega t}$, while ω represents the frequency and $j = \sqrt{-1}$.

Then the simultaneous equation above can be written as follows.

$$\begin{bmatrix} c_1 - M\omega^2 & c_2 \\ c_2 & c_3 - J_z \omega^2 \end{bmatrix} \begin{bmatrix} U \\ \Phi \end{bmatrix} = 0. \quad (19)$$

In this expression,

$$c_1 = 4Z_L K_n \cos^2 \beta, \\ c_2 = 2Z_L K_n (h_1 + h_3) \cos^2 \beta, \\ c_3 = 2Z_L K_n (h_1^2 + h_3^2) \cos^2 \beta + 4Z_L K_n e^2 \sin^2 \beta + 4Z_L K_n h_0 e \sin \beta \cos \beta. \quad (20)$$

Solve (19):

$$M J_z \omega^4 - (c_3 M + c_1 J_z) \omega^2 + c_1 c_3 - c_2^2 = 0. \quad (21)$$

Therefore,

$$\omega_{1,2}^2 = \frac{c_3 M + c_1 J_z \mp \sqrt{(c_3 M + c_1 J_z)^2 - 4M J_z (c_1 c_3 - c_2^2)}}{2M J_z}. \quad (22)$$

The displacement along X direction u and the angular displacement around z -axis couple together so we suppose that the frequency obtained is low order rolling frequency f_{RL} and high order rolling frequency f_{RH} :

$$f_{RL} = \frac{\omega_1}{2\pi}, \\ f_{RH} = \frac{\omega_2}{2\pi}. \quad (23)$$

The solution of the vibration model above is the vibration frequency of the linear rolling guideway.

The parameters of the experimental guideway Schneeberger MRB35-V2 are displayed in Table 1. The result of the vertical moving frequency, which is computed by the analytical method from the vibration model, is $f_V = 3529$ Hz.

3. Dynamic Parameter Identification of Rolling Guideway

Because the contact part of the guideway has multiple modes, which can be known from the hammer-hitting experiment, numbers of groups vertical and horizontal "spring-damper" can be used as simplified equivalent model of the guideway contact part. We make a use of the results of the vibration model above and combine them with experiment to identify the dynamic parameters of the equivalent model.

The experimental facility of the hammer-hitting test is shown in Figure 6. The main apparatuses are listed in the following:

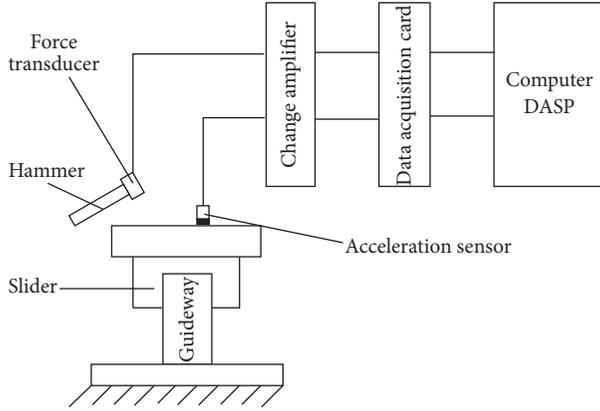


FIGURE 6: The schematic diagram of hammer-hitting experiment.

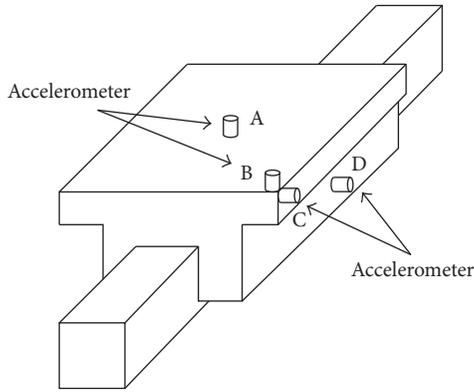


FIGURE 7: Measuring points map.

- (i) MSC-1 Impact Hammer, 500 kg Force Transducer
- (ii) YD 67D Miniature Accelerometer
- (iii) DLF-3 Four-in-one Two-Channel Charge Amplifier
- (iv) Four-Channel Data Collector with USB Interface INV306U, whose highest sampling frequency is 200 KHZ

The measure in the four points A, B, C, and D is shown in Figure 7.

The vibration spectrums of the four points A, B, C, and D are displayed in Figure 8 (including the transfer function amplitude and correlation coefficient). The measured value of the vertical vibration frequency is $f_V = 3450$ Hz, whose error is less than 3% compared with the theoretical calculation.

As is shown in Figure 8, the slider-roller-guideway is a system with multiple modes. Find out the necessary parameters by combining them with the analytical results of the theoretical vibration model, which is the vertical vibration frequency f_V and horizontal vibration frequency f_H . And then identify the dynamic stiffness in both vertical and horizontal direction. The damping ratio and damping value are calculated by using the half-power method. The detail of the calculation is displayed in (24)–(29).

The stiffness in vertical direction is

$$K_V = 4M\pi^2 f_V^2. \quad (24)$$

The stiffness in horizontal direction is

$$K_H = 4M\pi^2 f_H^2. \quad (25)$$

The damping ratio in vertical direction is

$$\xi_V = \frac{f_{V2} - f_{V1}}{2f_V}. \quad (26)$$

The damping value in vertical direction is

$$c_V = 4\pi M f_V \xi_V. \quad (27)$$

The damping ratio in horizontal direction is

$$\xi_H = \frac{f_{H2} - f_{H1}}{2f_H}. \quad (28)$$

The damping value in horizontal direction is

$$c_H = 4\pi M f_H \xi_H. \quad (29)$$

In these equations, M represents the mass; f_V represents the vertical direction vibration frequency; f_H represents the horizontal direction vibration frequency; f_{V1} , f_{V2} , f_{H1} , and f_{H2} , respectively, represent the homologous half-power frequencies in vertical and horizontal directions.

The measured value of vibration frequency in vertical direction, which is detected from the experiment, is $f_V = 3450$ Hz and the horizontal vibration frequency is $f_H = 1075$ Hz. The results can be computed by using (24)–(29).

Thus, the stiffness in vertical direction is

$$K_V = 1.03 \times 10^9 \text{ N/m}. \quad (30)$$

The damping value in vertical direction is

$$c_V = 2750 \text{ N}\cdot\text{s/m}. \quad (31)$$

The stiffness in horizontal direction is

$$K_H = 1.00 \times 10^8 \text{ N/m}. \quad (32)$$

The damping value in horizontal direction is

$$c_H = 3109 \text{ N}\cdot\text{s/m}. \quad (33)$$

4. The Dynamic Analysis and Experimental Verification of Machine Tool Considering Contact Interface of Guideway

4.1. The FEM Model of Machine Tool. We conduct the research on the gantry machining center driven by a linear electric motor (as is shown in Figure 9) and make use of the FEM software to analyze the dynamic property of the machining center. The guideway in the machine tool system, which is shown in Figure 10, is equivalent to the four-vertical-direction spring-damper system and eight-horizontal-direction spring-damper system. The dynamic parameters can be obtained by using the method mentioned above.

After adding the material property parameters, constraints, and loadings, we apply the modal analysis to the machining center and compute the first-six-order frequencies and modes of the overall machine tool (shown in Table 2).

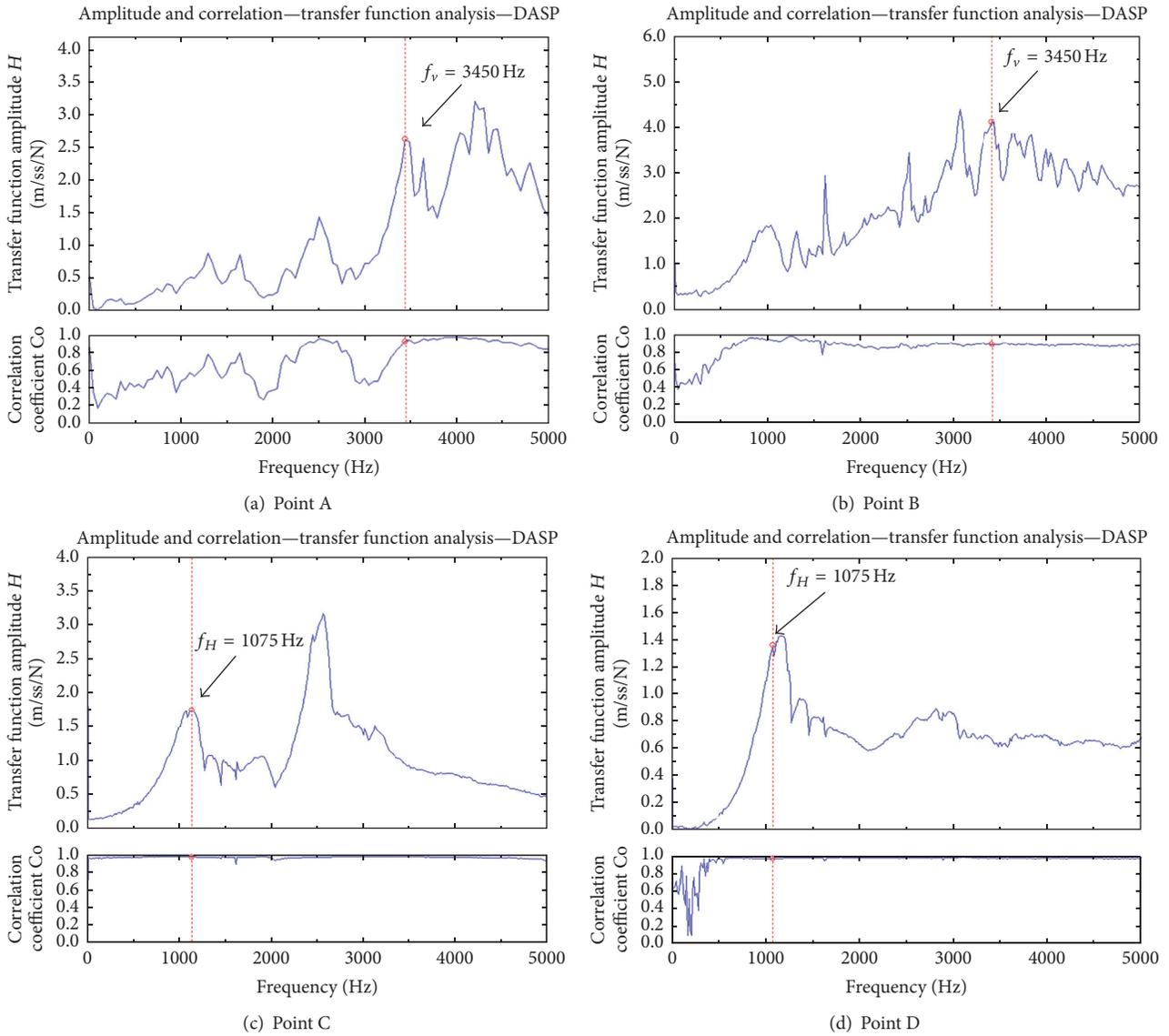


FIGURE 8: The frequency response functional graph of acceleration-force of linear guideway.

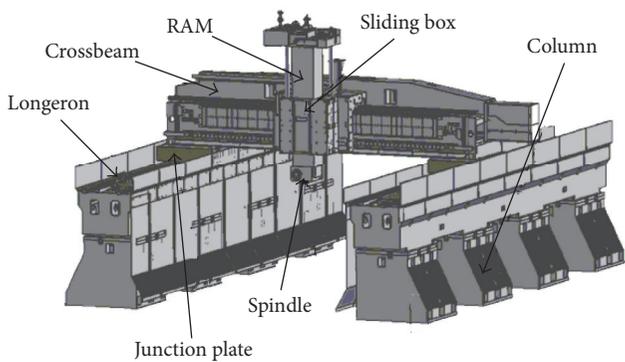


FIGURE 9: The model of gantry machining center. x -axis: the direction along the length of longeron, y -axis: the direction along the length of crossbeam, and z -axis: vertical direction.

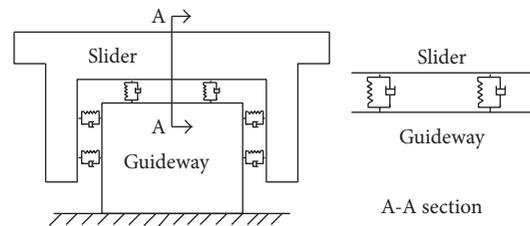


FIGURE 10: The dynamically equivalent model of the guideway.

4.2. *The Overall Test.* Using the hammer-hitting method, knock and pick up the signals, respectively, from these three directions X , Y , and Z . On account of the reason that the main consideration is the influence on the machine parts during

TABLE 2: The first-six-order frequencies and modes of machine tool computed by FEM method.

Order	Frequency of FEM model (Hz)	Modes	Reasons
1	27.5	Crossbeam bends around the horizontal Y axis	Junction plate and longeron's guideway pitching vibration
2	53.7	Crossbeam vibrating up and down. Sliding box bends around the horizontal Y axis	Crossbeam vibrates, sliding box, and crossbeam guideway's rolling vibration
3	55.1	Longeron and crossbeam wagging	Longeron's flexural vibration
4	78.7	Junction Plate's small amplitudes pitching vibration. Sliding box's guideway pitching vibration	Junction plate guideway and sliding box guideway's pitching vibration
5	83.0	Crossbeam bends around Z axis. Sliding box rolls around Y axis	Crossbeam's flexural vibration
6	115.7	Crossbeam's pitching vibration around Y axis	Junction plate guideway's pitching vibration

TABLE 3: The first-six-order natural frequencies of overall machine tool (unit: Hz).

Order	FEM model frequency	Experimental measured frequency	Error
1	27.5	25.8	6.6%
2	53.7	50.3	6.8%
3	55.1	55.9	1.4%
4	78.7	68.0	14.7%
5	83.0	85.6	3.0%
6	115.7	107.9	7.2%

the process exerted by the dynamic property of the machining center, the acceleration sensor is attached to the bottom of the principal axis and the vibration signal is collected. Respectively, knock on eight main components, which are the column, longeron, sliding plate, crossbeam, sliding box, ram, spindle box, and the bottom of the principal axis. Test the natural frequency of the overall machine and compare the results with frequency obtained from FEM analysis. The results are shown in Table 3. The method to calculate the error is shown as follows.

$$\text{Relative Error} = \frac{\text{Computed Value} - \text{Measured Value}}{\text{Measured Value}} \times 100\% \quad (34)$$

The result of the experiment indicates that the solution of the FEM analysis has a relative small error compared to the

experimental result, and the simulation of FEM model well matches the practical situation.

5. Conclusion

Based on the research of static property of the linear rolling guideway, this paper makes a research on the dynamic performance of the rolling guideway and the simplified model of the guideway and the identification of dynamic parameters. At the same time, the research achievement is applied to a gantry machining center driven by a linear electric motor, which solves the problem of dynamic modeling of contact interface in NC machine tool. On account of the research mentioned above, the conclusions are obtained as follows.

- (1) Based on the Hertz contact theory, the relation between the bearing load and deformation of the linear rolling guideway is detected to be approximate linear function. In the simplified model it can be replaced by the linear model.
- (2) Based on the Lagrange method, we conduct a research on the vibration property of the linear guideway and deduce the dynamic equations. It is explained through model analysis that the 'Slider-Roller-Guideway' system has multiple vibration modes called pitching vibration, yawing vibration, vertical vibration, low order rolling vibration, and high order rolling vibration, which means that it is a complex multiple modes system. Therefore, we use the multigroups of vertical and horizontal "spring-damper" system as the simplified equivalent model.
- (3) The paper proposed a type of guideway contact interface modeling method combining four stages which were static stiffness model, vibration model, experimental parameter, and parameter identification. This method is effective while solving the stiffness and damping value of the guideway equivalent model. Moreover, it solves the key problem in machine tool modeling—the problem of contact interface modeling.
- (4) Combined with the research and development of the practical machine tool, for the object of a type of gantry machining center, we establish an overall FEM model of the gantry machining center driven by linear motors which includes the guideway contact interface model; moreover, the results of natural frequencies and modes of the machine tool are obtained. By comparing the computed solution with the experimental result, the efficiency and accuracy of the guideway contact interface model is verified.
- (5) The spring model is used to replace the roller in the guideway system to improve the comprehension of stiffness and damping in the joint. However, when the interaction in the joint is needed to consider, the spring model is not suitable for simplified modeling of the roller, and it will be discussed in the future research.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This study was supported by the National Science Foundation of China (Grant no. 61403010).

References

- [1] S. Wen, X. Zhang, and M. Wu, "Fractal model and simulation of normal contact stiffness of joint interfaces and its simulation," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 40, no. 11, pp. 197–202, 2009.
- [2] S. Wen, X. Zhang, X. Wen, P. Wang, and M. Wu, "Fractal model of tangential contact stiffness of joint interfaces and its simulation," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 40, no. 12, pp. 223–227, 2009.
- [3] G. Zhang, W. Shi, and Y. Huang, "Analysis method of dynamic behavior of guideway joint and its application in machine tools design," *Jixie Gongcheng Xuebao/Chinese Journal of Mechanical Engineering*, vol. 38, no. 10, pp. 114–117, 2002.
- [4] K. Mao, B. Li, and B. Xie, "Rolling linear guide pair movable joint dynamic modeling," *Journal of Huazhong University of Science and Technology (Natural Sciences)*, vol. 36, no. 8, pp. 85–88, 2008.
- [5] J. S. Dhupia, B. Powalka, A. Galip Ulsoy, and R. Katz, "Effect of a nonlinear joint on the dynamic performance of a machine tool," *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, vol. 129, no. 5, pp. 943–950, 2007.
- [6] Y. Zhang and B. Liao, "Effective parameter identification method of joint part of machine tool," *Journal of Kunming University of Science and Technology*, vol. 23, no. 2, pp. 36–41, 1998.
- [7] S. S. Park, Y. Altintas, and M. Movahhedy, "Receptance coupling for end mills," *International Journal of Machine Tools and Manufacture*, vol. 43, no. 9, pp. 889–896, 2003.
- [8] X. Zhang, Y. Tang, and Y. Xu, "Identification method of the stiffness of contact interface by using experimental modality and FEM method," *Modular Machine Tool & Automatic Manufacturing Technique*, vol. 11, pp. 56–60, 2005.
- [9] H. Ohta and E. Hayashi, "Vibration of linear guideway type recirculating linear ball bearings," *Journal of Sound and Vibration*, vol. 235, no. 5, pp. 847–861, 2000.
- [10] J. S. Wu, J. C. Chang, and J. P. Hung, "The effect of contact interface on dynamic characteristics of composite structures," *Mathematics and Computers in Simulation*, vol. 7, no. 3, pp. 1–14, 2006.
- [11] J. Zheng, J. Zhang, and Q. Sun, "Parameter identification based on modal theory of slider contact interface of spherical guide way of machine tool," *Precise Manufacturing & Automation*, vol. 3, pp. 19–20, 2005.
- [12] D. Xu, Q. Liu, and S. Yuan, "Research on static stiffness of linear rolling guide way," *Machine Tool & Hydraulics*, vol. 36, no. 4, pp. 8–10, 2008.

Research Article

Control Performance and Robustness of Pounding Tuned Mass Damper for Vibration Reduction in SDOF Structure

Qichao Xue,^{1,2} Jingcai Zhang,¹ Jian He,¹ and Chunwei Zhang^{2,3}

¹College of Aerospace and Civil Engineering, Harbin Engineering University, Harbin, China

²School of Civil Engineering, Qingdao University of Technology, Qingdao, China

³Institute for Infrastructure Engineering, Western Sydney University, Sydney, NSW, Australia

Correspondence should be addressed to Chunwei Zhang; chunwei.zhang@westernsydney.edu.au

Received 31 August 2016; Accepted 27 October 2016

Academic Editor: Ganging Song

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

This paper investigates the control performance of pounding tuned mass damper (PTMD) in reducing the dynamic responses of SDOF (Single Degree of Freedom) structure. Taking an offshore jacket-type platform as an example, the optimal damping ratio and the gap between mass block and viscoelastic material are presented depending on a parametric study. Control efficiency influenced by material properties and contact geometries for PTMD is analyzed here, as well as robustness of the device. The results of numerical simulations indicated that satisfactory vibration mitigation and robustness can be achieved by an optimally designed PTMD. Comparisons between PTMD and traditional TMD demonstrate the advantages of PTMD, not only in vibration suppression and costs but also in effective frequency bandwidth.

1. Introduction

An offshore platform is essential equipment for ocean oil development with high construction cost, and it may lead to serious damage when subjected to seismic effects [1, 2]. Therefore, an effective vibration reduction solution is necessary for offshore platforms. Tuned mass damper (TMD) is a traditional control device installed on the offshore platform to reduce the seismic vibration responses. Kawano et al. [3] first investigated seismic responses of the offshore platform with TMD and indicated that TMD can reduce seismic vibration effectively. Since then, many researches have focused on TMD applications on platforms. For instance, Wu et al.'s study [4] on high response performance of TMD indicates that TMD can significantly reduce the vibration of the platform and a remarkable decreased power spectral density can be achieved.

However, TMDs have several disadvantages in utilizing structure vibration control. First of all, most of TMD devices need to be tuned to the first natural frequency of the primary structure to obtain better control performance. Once the excitation's frequency deviates from the first natural frequency, TMD's vibration reduction will decrease sharply

and even cause terrible adverse effects. At this time, although an extra damper attaching to TMD can suppress the adverse effects, it is costly and will lead to waste of space due to the large motion of the mass block.

In order to avoid the foregoing disadvantages and achieve satisfactory vibration control performance, many new analysis technologies have been applied to improve TMD. Almazan et al. [5] proposed a bidirectional and homogeneous tuned mass damper (BH-TMD) for passive control of vibrations. Both experimental and numerical simulation results demonstrate the superiority of the modified TMD system over the traditional one. A conceptual system for a semiactive tuned mass damper (STMD) with variable damping coefficients and stiffness values was presented by Chey et al. [6]. Sun and Nagarajaiah [7] presented a new algorithm for STMD to adjust the damping coefficient for different excitation. Jafarabad et al. [8] installed a hybrid damping system to control the seismic vibration which contains a friction damper device (FDD) and a tuned mass damper (TMD). Mohebbi et al. [9] analyzed optimum parameters of multiple tuned mass dampers (MTMDs) for different design criteria under seismic excitation.

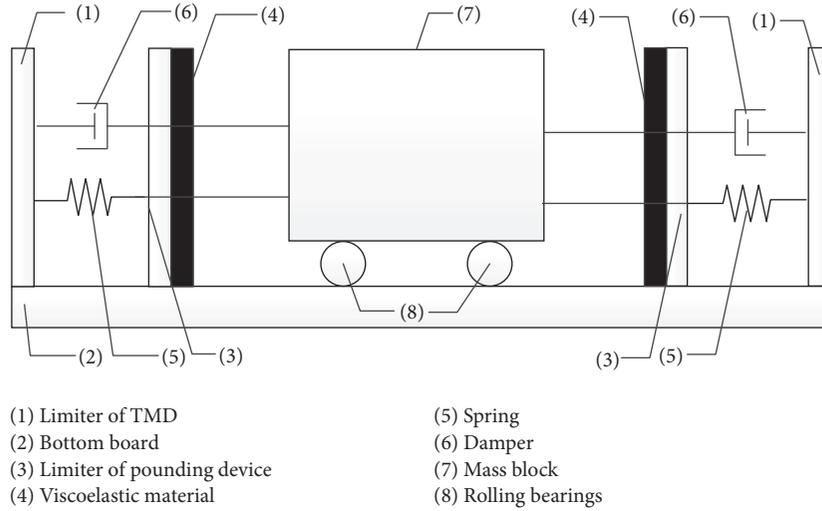


FIGURE 1: Schematic graph of PTMD with viscoelastic material layers.

Despite the former new analysis technologies, some novel control devices were proposed on platforms too. Huo and Li [10] developed a circular tuned liquid column damper (CTLCD) to control the platform's torsional responses under seismic loads. Li et al.'s [11] research shows that satisfactory vibration mitigation can be achieved by a new shape memory alloy (SMA) damper. Ou et al. [12] proposed a damping isolation system which was composed of rubber bearings and viscous dampers for controlling vibration. Komachi et al. [13] applied a friction damper device to a steel jacket platform located in an active zone and investigated the performance when subjected to seismic excitation. Mousavi et al. [14, 15] employed a tuned liquid column-gas damper (TLCGD) to suppress the seismic vibration and obtained optimum geometric parameters for a platform. Sarrafan et al. [16] presented an intelligent nonlinear neurofuzzy control strategy for magnetorheological (MR) damper system in order to adjust controlling force for different excitations. Lotfollahi-Yaghin et al. [17] verified the seismic control performance of tuned liquid damper (TLD) on typical offshore jacket-type platforms excited by El Centro, Kobe, and Tabas earthquakes.

Pounding tuned mass damper (PTMD) is proposed as a combination of traditional TMD and pounding effects by Zhang et al. [18]. Zhang et al. [18–22] investigated the vibration control performance of PTMD on power transmission tower and subsea jumpers, as well as traffic signal poles. There is rarely a result showing the performance of PTMD on offshore platform.

In this paper, taking an offshore jacket platform as an example, control performance and robustness of PTMD for vibration reduction in SDOF structure are analyzed to investigate the seismic control performance. Influences of different parameters are discussed, including damping ratio, gap between mass block and viscoelastic layer, contact geometries, robustness, and viscoelastic properties. Furthermore, a comparison between PTMD and traditional TMD is presented to verify the advantages of PTMD.

TABLE 1: Parameters of the offshore jacket platform.

	Test 1	Test 2		
Real platform	f/Hz	$\xi_1/\%$	m_1/t	
	0.90	0.85	4.0–5.5	3127
Simplified model	f/Hz	$k_1/\text{kN}\cdot\text{s}^{-1}$	$c_1/\text{kN}\cdot(\text{m}\cdot\text{s}^{-1})$	m_1/kg
	0.87	93436	1367	3127000

2. Schematic of PTMD with Viscoelastic layers

As an improvement of TMD, PTMD consists of two parts: a traditional TMD and a limiter layered by viscoelastic materials. Gaps between mass block and viscoelastic material are reserved as shown in Figure 1. When the relative motion occurs between mass block and viscoelastic layer without pounding effect, PTMD acts as a TMD. Otherwise, the mass block will impact the viscoelastic layer, which will produce pounding force as an extra controlling effect, and energy will be dissipated during the pounding process. One of the most remarkable advantages of this device is that controlling forces come from not only resonant force but also pounding force during the impact.

3. Modeling

3.1. Modeling of the System. As a widely used platform type in the Bohai Sea of China, JZ20-2MUQ offshore jacket platform plays an important role in the field of petroleum recovery. The platform consists of three parts: 3-story frame for living, 2-story frame for working, and the jackets. Ou et al.'s research [23] shows that the platform can be simplified to a Single Degree of Freedom (SDOF) structure since the mass of the upper quarter is dominant to the whole structure. Figure 2 shows the JZ20-2MUQ offshore jacket platform and the simplified model. The basic parameters are given in Table 1 based on Ou et al.'s research.

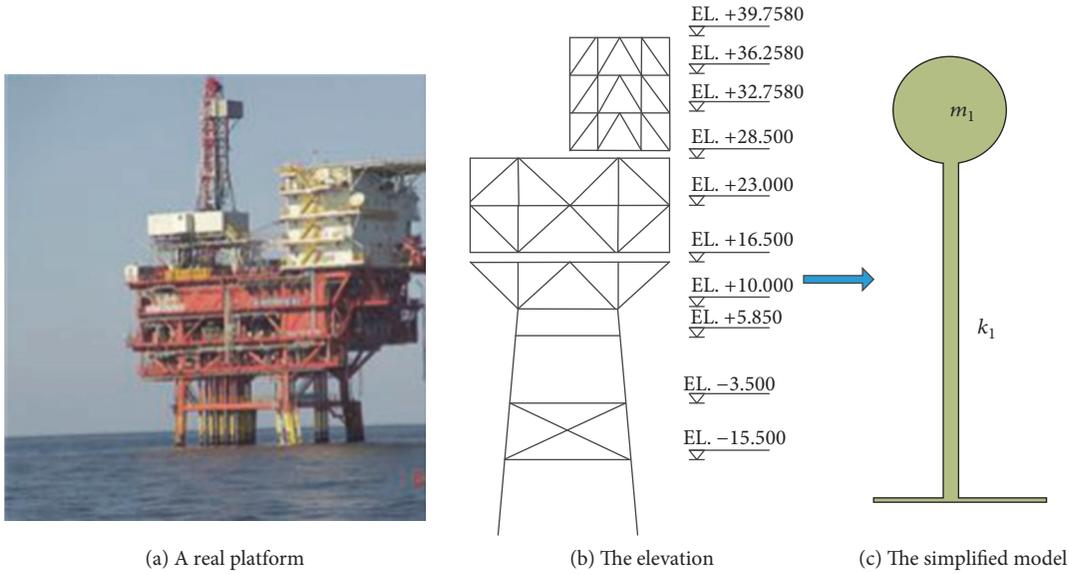


FIGURE 2: JZ20-2MUQ offshore jacket platform.

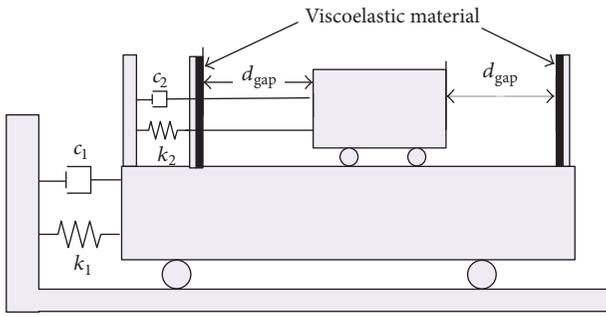


FIGURE 3: Schematic of a SDOF structure controlled by PTMD.

3.2. *Dynamic Equations of the System.* The schematic of a SDOF structure controlled by PTMD is illustrated in Figure 3, and the motion equation of the system can be expressed as

$$\begin{aligned} & \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \cdot \begin{Bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{Bmatrix} + \begin{bmatrix} c_1 + c_2 & -c_2 \\ -c_2 & c_2 \end{bmatrix} \cdot \begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{Bmatrix} \\ & + \begin{bmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_2 \end{bmatrix} \cdot \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} \\ & = \begin{Bmatrix} P \sin \omega t + H(d) F \\ -H(d) F \end{Bmatrix}, \end{aligned} \quad (1)$$

where m_1 and m_2 are mass of primary structure and PTMD, respectively. \ddot{x}_1 , \dot{x}_1 , and x_1 are acceleration, velocity, and displacement of the primary structure. \ddot{x}_2 , \dot{x}_2 , and x_2 are acceleration, velocity, and displacement of the PTMD. c_1 and c_2 are stiffness of primary structure and PTMD, respectively. k_1 and k_2 are stiffness of primary structure and PTMD,

respectively. F is the pounding force induced by collisions between mass block and viscoelastic layer. $H(d)$ is Havisd function

$$H(d) = \begin{cases} 1 & |d| - d_{\text{gap}} > 0 \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where d represents the relative displacement between the PTMD and the primary structure and is calculated by (3). d_{gap} is the gap between the mass block and the viscoelastic layer. Hence,

$$d = x_2 - x_1, \quad (3)$$

where P is the amplitude of the external force; in order to simplify the calculation process, $P = 1$.

3.3. Modeling of Pounding Force

3.3.1. *Pounding Force Expressions for Two Kinds of Contact Geometries.* Figure 4(a) is a sketch for a mass block with sphere head pounding to a plane surface, which will be utilized later in this paper to simulate vibration performance of PTMD. Figure 4(b) shows that an end plane of the column is pounding to plane, which will be calculated as a comparison to discuss influences of contact geometries. The pounding force can be calculated by (4) according to contact mechanics [24]:

$$F_c = \frac{4}{3} R_1^{1/2} E^* \delta_1^{3/2} \quad \text{for sphere to plane,} \quad (4)$$

$$F_c = 2R_1 E^* \delta_1 \quad \text{for column to plane,}$$

where R_1 is the radius of the hemisphere or ending round plane for column. $R = 0.5$ meters in this paper. δ_1 is

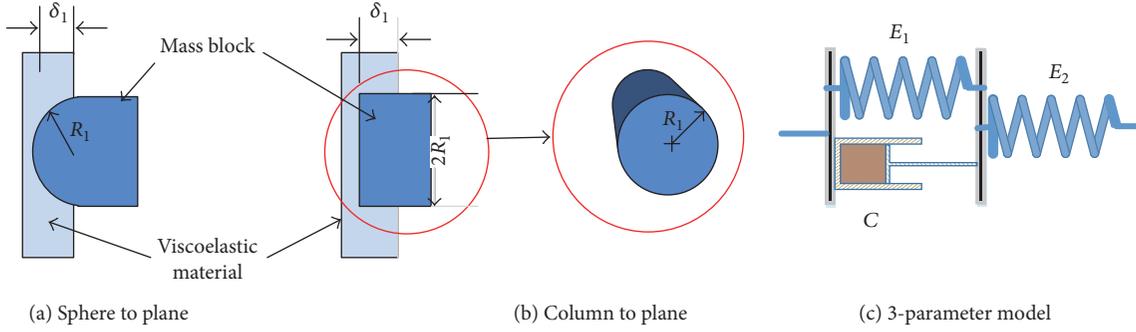


FIGURE 4: Schematic of contact geometries and 3-parameter viscoelastic model.

the penetration displacement. E^* is the equivalent elasticity modulus and can be derived by the following expression:

$$\frac{1}{E^*} = \frac{(1 - \vartheta_s^2)}{E_s} + \frac{(1 - \vartheta_p^2)}{E_p}, \quad (5)$$

where ν_p , ν_s , E_p , and E_s are Poisson's ratio and modulus of elasticity for polymer and steel, respectively. In case of polymer contact with steel, the modulus of elasticity for steel is far larger than polymer, and the expression of E^* can be simplified as follows:

$$E^* = \frac{E_p}{1 - \vartheta_p^2}. \quad (6)$$

According to the Correspondence Principle of viscoelasticity, pounding force expressions can be deduced by inverse Laplace transformation from (4) and (6):

$$F_{(t)} = \frac{4}{3} \frac{R_1^{1/2}}{1 - \vartheta_p^2} e^{-1} \left(\frac{\bar{Q}(s)}{\bar{P}(s)} \right) \delta_1^{3/2}(t) \quad \text{for sphere to plane,} \quad (7)$$

$$F_{(t)} = 2 \frac{R_1}{1 - \vartheta_p^2} e^{-1} \left(\frac{\bar{Q}(s)}{\bar{P}(s)} \right) \delta_1(t) \quad \text{for column to plane.}$$

For linear viscoelastic materials, $\bar{Q}(s)$ and $\bar{P}(s)$ are polynomial of parameter s and they can be calculated automatically as a transfer function with different initial values by MATLAB/Simulink.

4. Parametric Study

To obtain optimal parametric settings of PTMD, damping ratio of the PTMD, gap between mass block and viscoelastic layer, contact geometries, and viscoelastic properties are discussed in this section. Unless otherwise indicated, constitutive parameters of viscoelastic material in this paper are shown in Table 2. And E_1 , E_2 , and C_1 are the spring parameters and dash-pot parameter for 3-parameter model of viscoelastic material in Figure 4(c).

TABLE 2: Constitutive parameters of the viscoelastic material.

E_1/Pa	E_2/Pa	$C/\text{kN}\cdot(\text{m}\cdot\text{s}^{-1})$	E^*/Pa
2.041E5	2.041E5	3.9E4	1.08E5

For comparison, performance of an optimal TMD is presented based on Ioi and Ikeda's research [25], in which optimal frequency and damping ratio can be expressed as

$$f_{\text{opt}} = \frac{1}{1 + \mu} - (0.241 + 1.7\mu - 2.6\mu^2) \xi_s^2, \quad (8)$$

$$\xi_{\text{opt}} = \sqrt{\frac{3\mu}{8(1 + \mu)} + (0.13 + 0.12\mu + 0.4\mu^2) \xi_2} - (0.01 + 0.09\mu + 3\mu^2) \xi_2, \quad (9)$$

where f_{opt} and ξ_{opt} are the optimal frequency and damping ratio for the TMD, μ is mass ratio, and ξ_s is damping ratio of primary structure.

4.1. Damping Ratio. In order to illustrate the parameters' influence more accurately and efficiently, the dimensionless method is applied to the data analysis:

$$\delta = \frac{d_{\text{gap}}}{\delta_s},$$

$$\delta_s = \frac{P}{k_1}, \quad (10)$$

$$f_e = \frac{\omega}{\omega_1},$$

where d_{gap} is the gap between mass block and viscoelastic material. δ is the dimensionless gap, and δ_s is the static displacement of the primary structure. P is the amplitude of the sinusoidal wave which is input into the SDOF system, and k_1 is the structural stiffness of the primary structure. The frequency ratio which is defined as f_e is an index of degree for sinusoidal wave deviating from the natural frequency, and ω and ω_1 represent the sinusoidal wave frequency and natural frequency of the primary structure, respectively.

Dynamic amplification factor, defined as R in (11), is introduced to evaluate the performance of controlling devices.

In addition, the mass ratio is assumed to be 2.0% which is the most common selection in engineering practice. Hence,

$$R = \frac{x_{1,\text{peak}}}{\delta_s}, \quad (11)$$

where $x_{1,\text{peak}}$ is the peak value of the displacement in steady state subjected to sinusoidal wave excitation. The dimensionless displacement and acceleration of the primary structure, named X and A , respectively, can be calculated as follows:

$$\begin{aligned} X &= \frac{x}{\delta_s}, \\ A &= \ddot{X}, \end{aligned} \quad (12)$$

where x is the displacement of the primary structure during the excitation and A is the second derivative in time of X .

Damping ratio of PTMD, defined as ξ_2 , can be calculated as

$$\xi_2 = \frac{c_2}{2\mu m_1 \omega_2}, \quad (13)$$

where c_2 and ω_2 are damping coefficient and circular frequency of PTMD, m_1 is the mass of the platform, and μ is the mass ratio. Thus, we can conclude that only in case of $\xi_2 = 0$ is PTMD undamped.

Figure 5 illustrates the relationship between R and f_e with different damping ratios in the relative frequency domain. It can be observed from Figure 5(a) that an undamped PTMD has obvious magnification in structure responses. For instance, maximum value of R is 28.3 for PTMD when δ is equal to 2, and it is almost triple those controlled by TMD. In this situation, PTMD is not suitable as a vibration control device.

A damper in vibration control device can effectively eliminate magnification effect ($\xi_2 \neq 0$). Figure 5(c) displays the performance of PTMD with different gaps when an optimal damper parameter is applied. Although vibration control performance for PTMD is much better than TMD when f_e varies from 0.9 to 1.3, there is a significantly inverse effect on primary structure response for PTMD in case f_e is less than 0.9. That means the optimal damping ratio which is based on TMD is no longer applicable to PTMD due to pounding effects.

From Figures 5(b)–5(e), R diminishes with the increasing of damping ratio from 0 to $3.0\xi_{\text{opt}}$ when the gap value is a constant. As ξ_2 is $3.0\xi_{\text{opt}}$, PTMD shows a more remarkable performance compared to TMD. Moreover, R rises with the increase of gap in this situation which indicates that impact played a positive role in vibration reduction.

However, Figure 5(f) shows that the structures vibrate more severely in case of $\xi_2 = 4.0\xi_{\text{opt}}$ than those with the same gap and $\xi_2 = 3.0\xi_{\text{opt}}$. Further, PTMD shows a poorer performance than TMD when δ is equal to 10. In this situation, although an additional damper may enhance the efficiency of PTMD, an excessive damping ratio may lead to an opposite effect.

From Figures 5(a)–5(f), we can obtain an optimal damping ratio for the best performance of PTMD in this case: $\xi_{2,\text{opt}} = 3.0\xi_{\text{opt}}$.

4.2. Influences of Gap. From an energy viewpoint, energy can be dissipated from the damper in TMD or PTMD, as well as the impact process between mass block and viscoelastic material. A small gap in PTMD means that the mass will impact the viscoelastic material more severely and frequently. In an extreme case where the gap is large enough, the impact will not happen and PTMD turns into a traditional TMD.

Figure 5 clearly revealed that a sufficient gap is indispensable to an undamped PTMD to avoid the adverse effect caused by the impact. To quantify the energy dissipation of PTMD, the root mean square (RMS) of displacement and the acceleration of the primary structure are defined as in (14). A smaller RMS means more energy will be dissipated or transformed to dampers. Hence,

$$\begin{aligned} \text{RMS}_{X_1} &= \sqrt{\frac{\sum_1^n (X_{1i})^2}{n}}, \\ X_i &= \frac{Px_i}{k_1}. \end{aligned} \quad (14)$$

Figure 6 plots the changing trend of RMS_{A_1} and RMS_{X_1} along with dimensionless gap. In this case, $\xi_2 = 3\xi_t$ with mass ratio 2%, and n is the number of values in the time history record.

It can be observed from Figure 6 that both of RMS_A and RMS_X decrease with the increasing of δ and finally tend to be a constant under a low frequency of sinusoidal wave excitation (where f_e is 0.7 and 0.8). Considering that PTMD will degenerate to TMD when RMS turns to a constant, impact has a negative effect on PTMD's performance in this situation. That means a large gap is required for low frequency excitation.

However, for relatively high frequency excitation with f_e higher than 0.9, both RMS_A and RMS_X increase with the rising of the dimensionless gap. An optimal gap value will avoid the negative effect and maximize the positive effect. This is always needed for PTMD due to the randomness and variety of earthquakes. From Figure 6, the optional dimensionless gap δ is about 2 for offshore jacket platform in this paper.

4.3. Material Properties. Properties of the viscoelastic material layer in PTMD will give different vibration reduction results. It is clear that the material with a higher E^* can induce a larger pounding force under the same situation, which may enhance the controlling force and get more momentum exchanged during an impact. Two additional materials, named type A and B (shown in Table 3), together with former materials in Table 2, are utilized to investigate the influence of material properties on PTMD. The results are illustrated in Figure 7.

Figure 7(b) shows time history of the dimensionless displacement of the platform controlled by PTMD with different materials in case of $f_e = 1.1$. It can be seen in this case that a higher equivalent elastic modulus can improve the controlling performance. However, in Figure 7(a) which corresponds to the case $f_e = 0.85$, it is indicated that the equivalent elasticity modulus has an opposite effect which

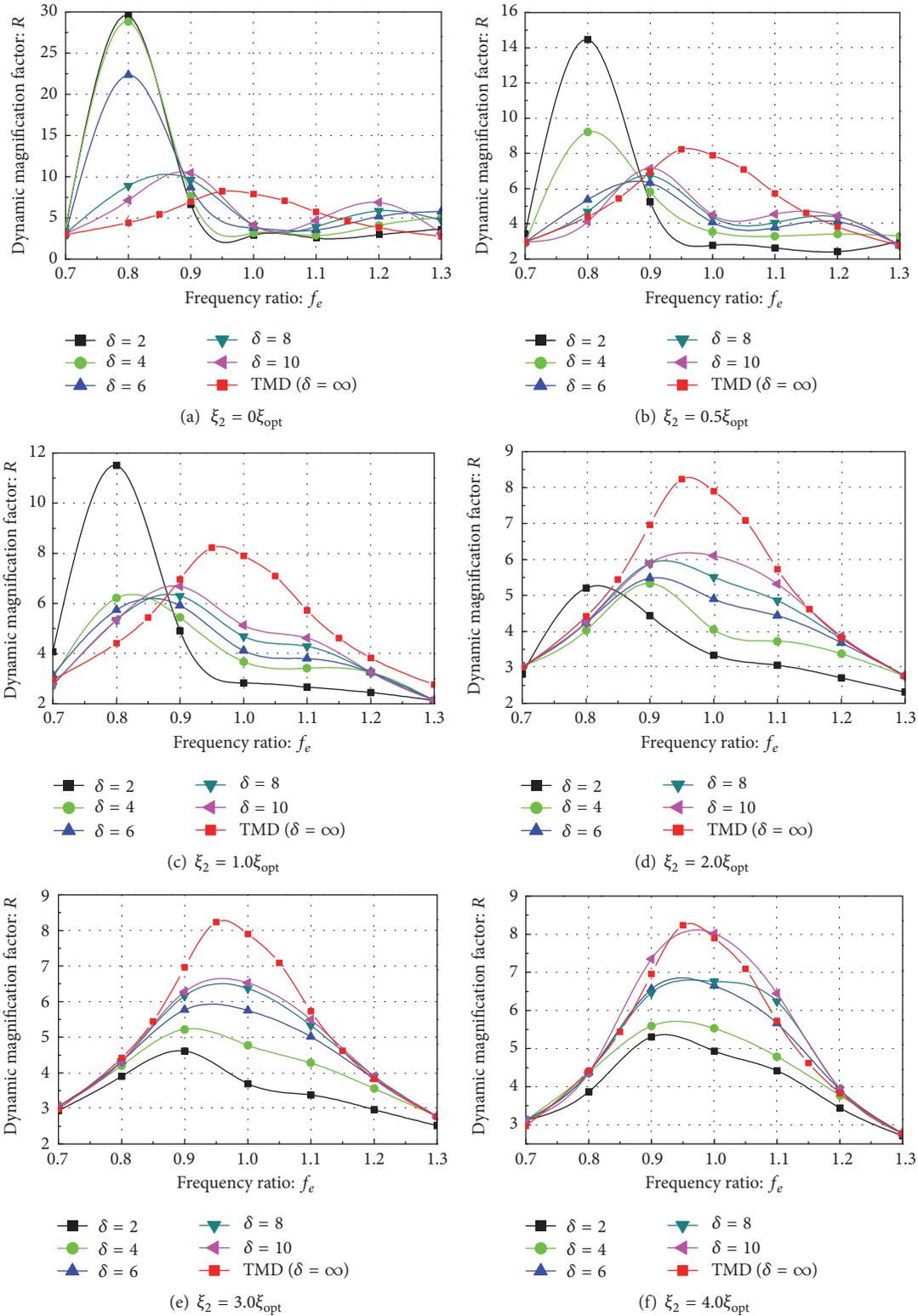


FIGURE 5: Responses of the controlled structure with different damping ratios.

leads to a poorer control result. Comparing Figures 7(a) and 7(b), we can conclude that a material with high equivalent elasticity modulus can enhance the efficiency when the

excitation's frequency is higher than the natural frequency of primary structure. Otherwise, it will weaken the vibration reduction.

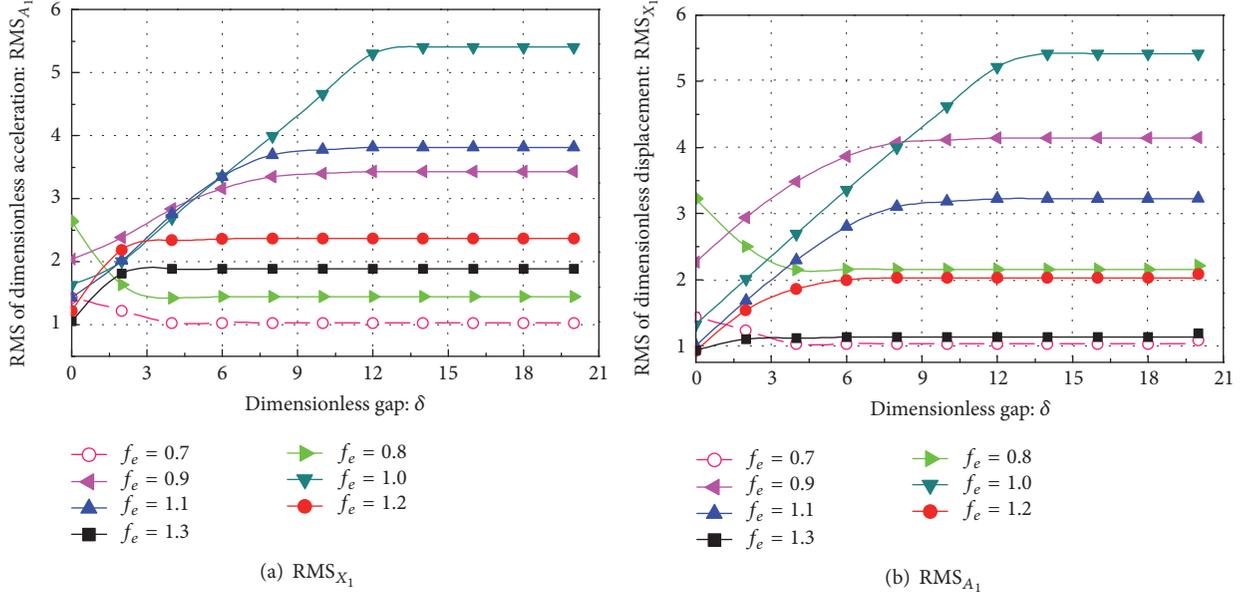


FIGURE 6: Primary structure RMS curve along with the change of δ ($\xi_2 = 3\xi_{opt}$, $\mu = 2\%$).

TABLE 3: Parameters of the additional materials.

Type ID	E_1/Pa	E_2/Pa	$C/\text{kN}\cdot(\text{m}\cdot\text{s}^{-1})$	E^*/Pa
A	2.041E6	2.041E6	3.9E5	1.08E6
B	2.041E7	2.041E7	3.9E6	1.08E7

4.4. Contact Geometries. As an indispensable part of controlling force for PTMD, pounding force is determined not only by material properties but also by the geometry of colliding bodies. Figure 8 illustrates the responses of the platform controlled by PTMD with two types of contact geometries, which were shown in Figures 4(a) and 4(b). For materials calculated in this paper, a better performance can be achieved by a sphere head than by a plane one. But differences of reduction between the two types of contact geometries are no more than 7% for both peak value and RMS. For type A material which is presented in Table 3, there is almost no difference between the two kinds of geometries. It can be concluded that geometric properties of the contact have a little effect on the performance of PTMD.

5. Robustness of PTMD

Tuning to the natural frequency of primary structure is the most important purpose for traditional TMD design. However, it is difficult to measure the natural frequency of the structure accurately. The primary structure natural frequency for the platform is not a constant due to the operating environment loads, such as depth of water and equipment replacement, which will lead to robustness of PTMD. Influences of tuned sinusoidal vibration, detuning vibration, and free vibration will be investigated here to study the robustness of PTMD.

To quantify the degree of deviation between an optimal frequency and tuned frequency in TMD and PTMD, the detuning ratio (DTR) is defined as follows:

$$\text{DTR} = \frac{f_{\text{damper}} - f_{\text{opt}}}{f_{\text{opt}}} \cdot 100\%, \quad (15)$$

where f_{damper} is the frequency of PTMD or TMD and f_{opt} is the optimal frequency which was expressed by (8).

5.1. Tuned Excitation Vibration. The value of frequency 0.74 Hz corresponds to -15% of the detuning ratio and 1.00 Hz corresponds to 15% . Dimensionless displacement X_1 in steady state can reach as large as 32.8 for the structure without control. However, for an installed PTMD, the maximum X_1 can be limited below 2.5 and 3.4 which correspond to 0.74 Hz and 1.00 Hz. In other words, X_1 can be sharply reduced by 89.6% and 74.9%, respectively.

The simulation results in Figure 9(a) illustrate that PTMD is effective even without tuning, which proves its good performance with robustness. Responses with the same detuning ratio TMD are presented in Figure 9(b), and the maximum value of X_1 can be reduced to 15.4 and 12.7, which correspond to reductions of 53.0% and 61.3% separately. This verifies that PTMD has superior performance over the traditional TMD in this case.

5.2. Free Vibration. Figure 10 shows free vibration responses of the primary structure when dimensionless initial displacement X_1 is equal to 40. It shows that displacement responses are damped rapidly for both frequencies 0.74 Hz and 1.00 Hz with PTMD control in Figure 10(a). The peak value of X_1 reduces by 80% in 22.4 s if a 1.00 Hz PTMD is installed. And when it comes to 0.74 Hz, the duration time will drop to 15 s.

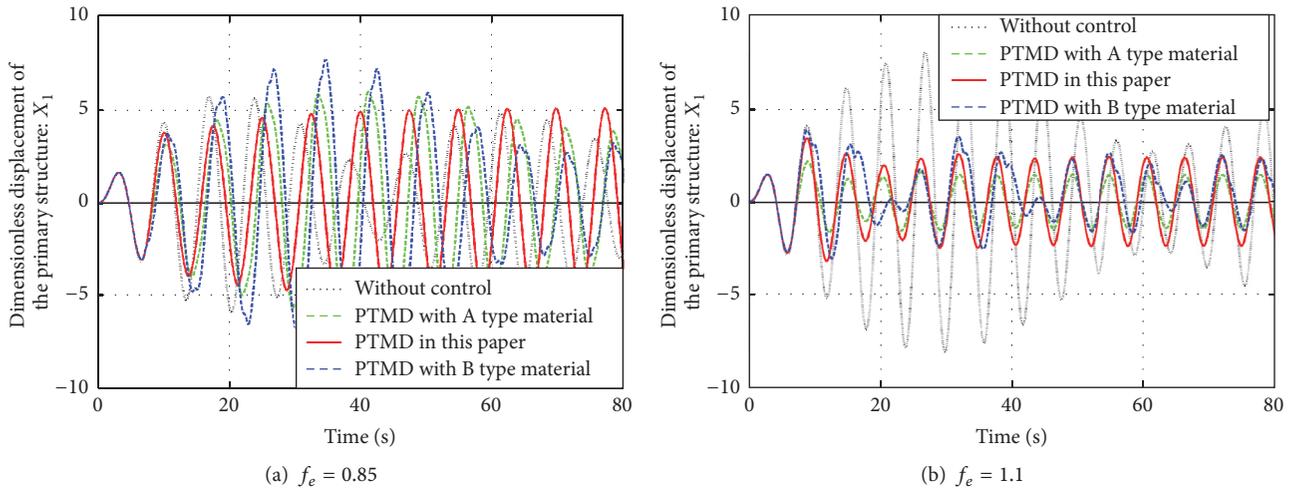


FIGURE 7: Control performance of PTMD with different types of materials.

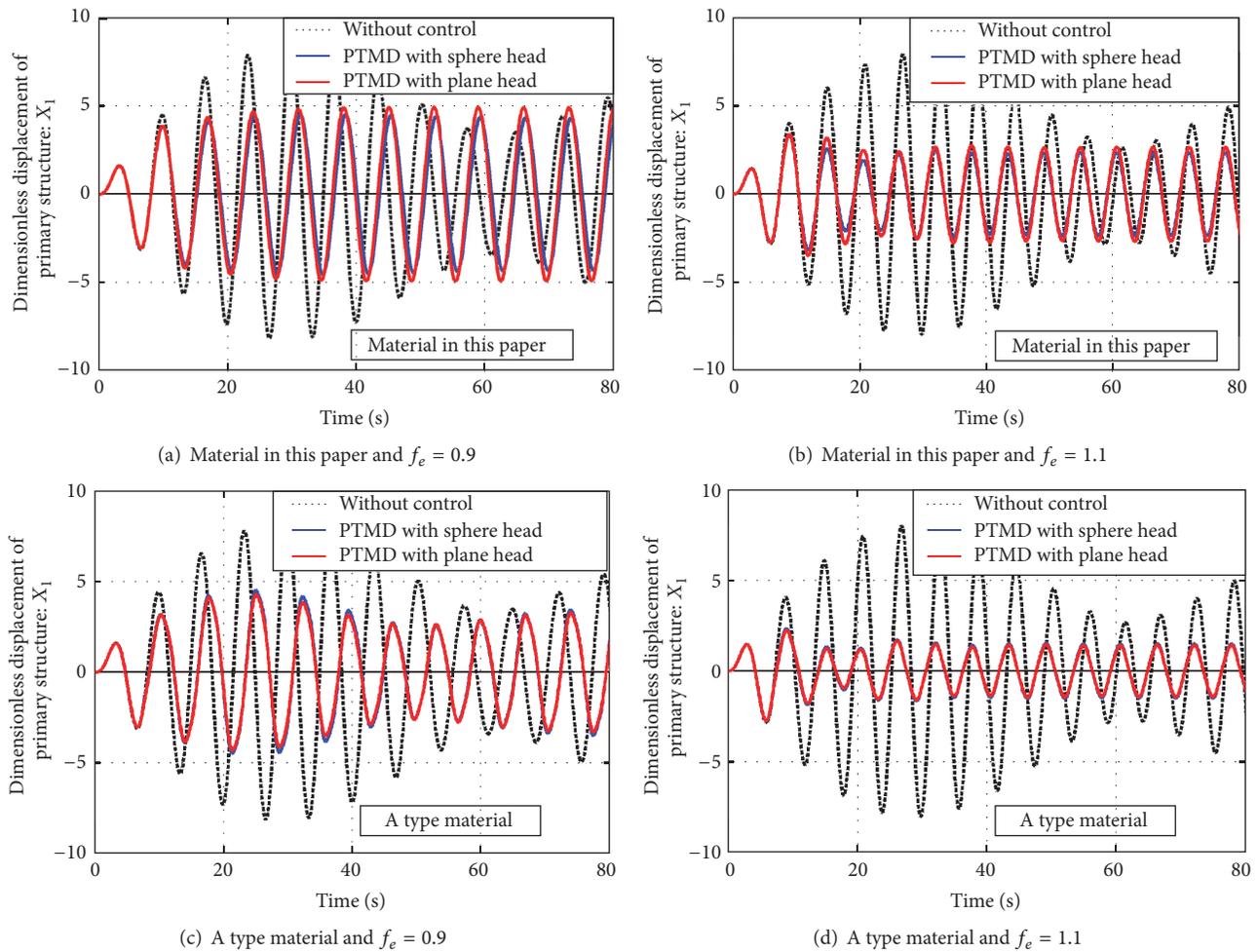


FIGURE 8: Control performance of PTMD with different contact geometries.

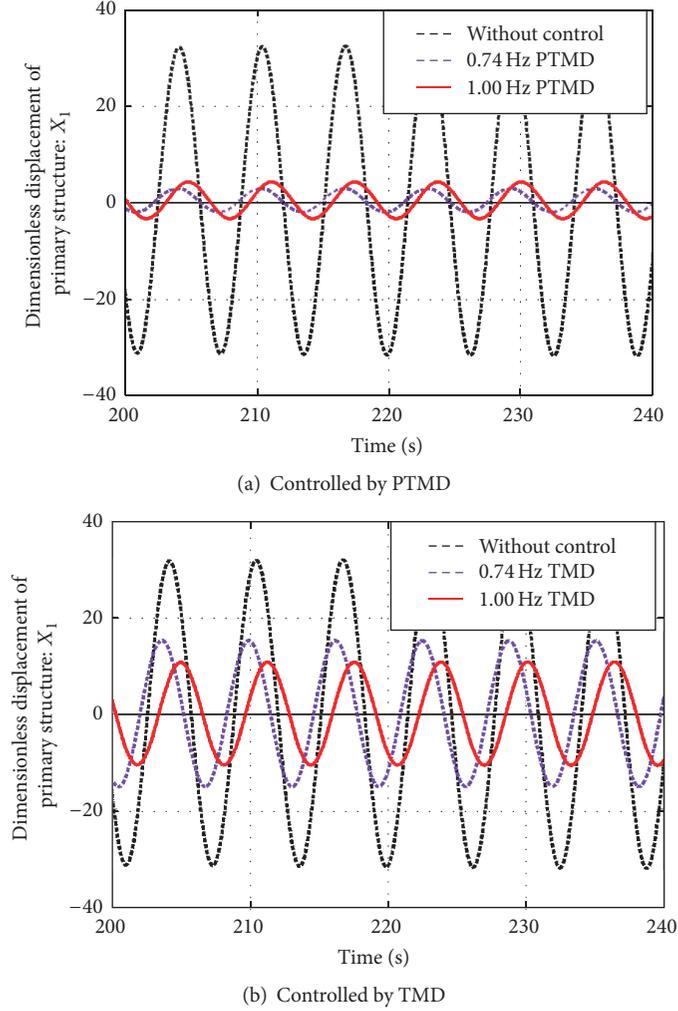


FIGURE 9: Control performance of TMD and PTMD in tuned excitation vibration.

As a comparison, the control performance of the TMD with the same detuning ratio is plotted in Figure 10(b). It will take 63.7 s and 61.3 s to reduce the maximum displacement to 8 mm for 0.74 Hz and 1.00 Hz, respectively. This is nearly three times as much as PTMD under 0.74 Hz and four times of a 1.00 Hz TMD. The simulation result demonstrates that PTMD can suppress the vibration and dissipate energy more effectively and show excellent adaptability and strong robustness.

5.3. Detuning Excitation Vibration. To evaluate the control performance of PTMD under free vibration excitation, four vibration reduction ratios named $\eta_{A1,peak,PTMD}$, $\eta_{RMSA1,PTMD}$, $\eta_{X1,peak,PTMD}$, and $\eta_{RMSX1,PTMD}$ are defined as (16). For TMD, $\eta_{A1,peak,TMD}$, $\eta_{RMSA1,TMD}$, $\eta_{X1,peak,TMD}$, and $\eta_{RMSX1,TMD}$ have similar definitions:

$$\eta_{A1,peak,PTMD} = \frac{A_{1,peak,NO} - A_{1,peak,PTMD}}{A_{1,peak,NO}} \times 100\%$$

$$\eta_{X1,peak,PTMD} = \frac{X_{1,peak,NO} - X_{1,peak,PTMD}}{X_{1,peak,NO}} \times 100\%$$

$$\eta_{RMSA1,PTMD} = \frac{RMS_{A1,NO} - RMS_{A1,NO}}{RMS_{A1,NO}} \times 100\%$$

$$\eta_{RMSX1,PTMD} = \frac{RMS_{X1,NO} - RMS_{X1,NO}}{RMS_{X1,NO}} \times 100\%$$

(16)

Figure 11 illustrates the relationship between reduction of peak value and DTR when the primary structure is applied by a series of detuning sinusoidal waves. From Figures 11(a) to 11(d), we can see that the performance will stay in a desired level and is barely affected by the DTR in case of being subjected to uptuned sinusoidal wave excitation ($f_e \geq 1$).

In Figure 11(a), PTMD shows an excellent performance with reduction of about 82.5% in case of $f_e = 1$ and only has tiny deterioration with DTR increasing. For $f_e = 1.1$ and $f_e = 1.2$, the reductions are almost kept as constants with reduction values of 50.5% and 33.8%, respectively. For downtuned cases ($f_e < 1$), the values of DTR have significant influences on reductions. In case of $f_e = 0.9$, reduction decreases sharply from 58.7% to 2.3% with DTR changing

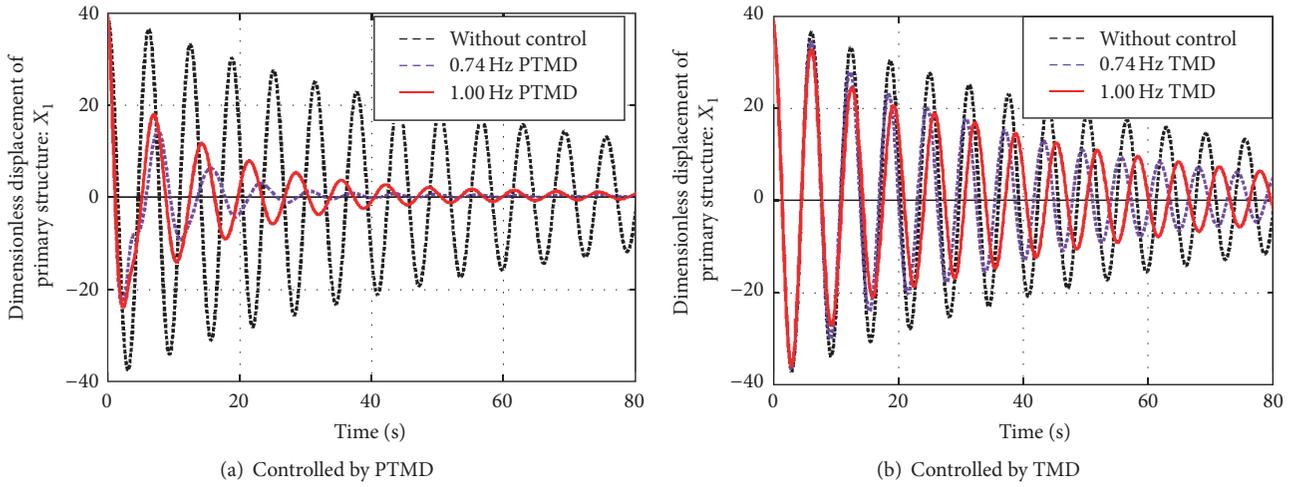


FIGURE 10: Control performance of PTMD and TMD in free vibration excitation.

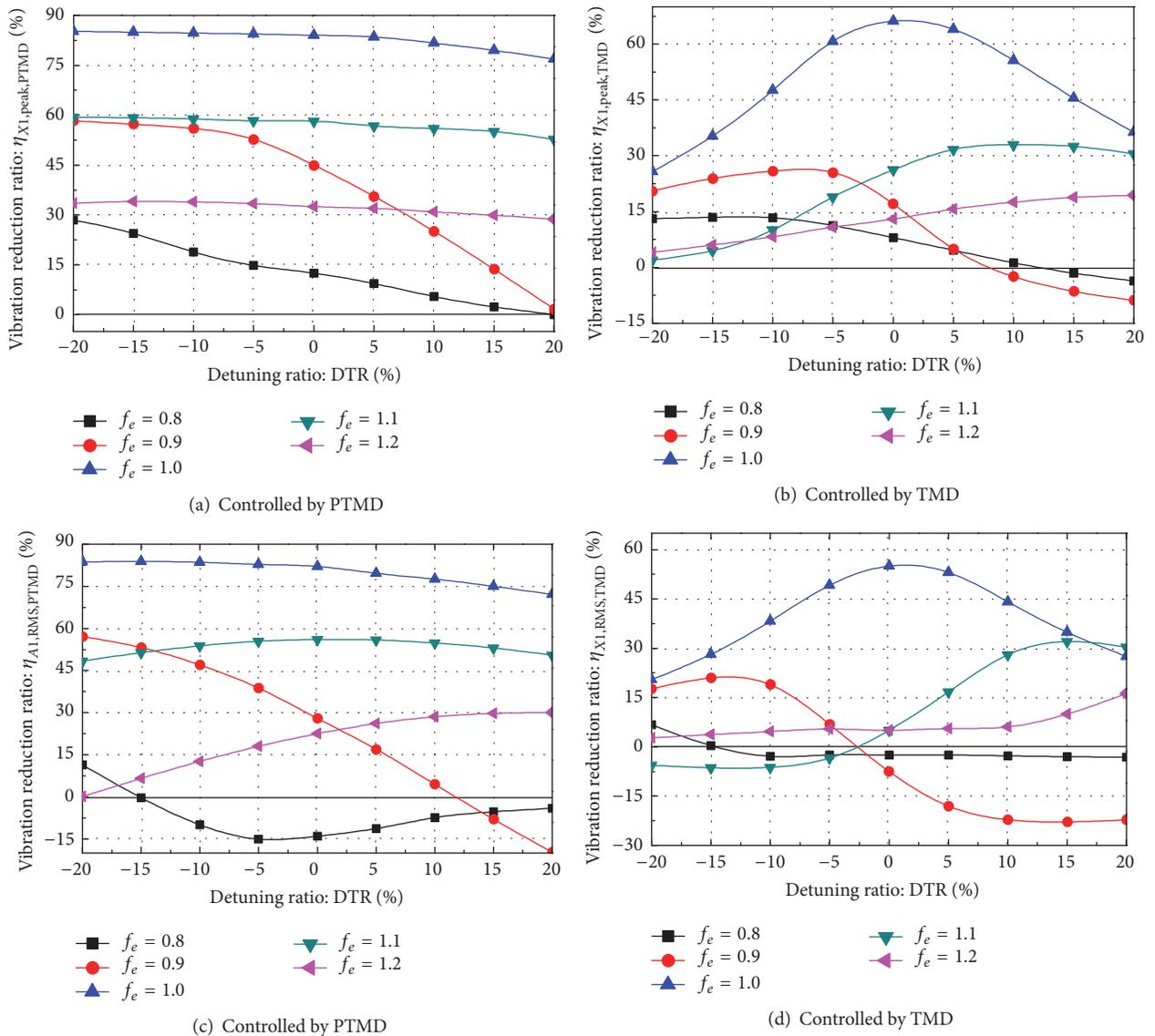


FIGURE 11: Control performance of TMD and PTMD in detuning excitation vibration.

from -20% to 20% . In case of $f_e = 0.8$, it will decline from 28.9% to 0 nearly in a linear relationship.

From Figure 11(c), vibration reduction of PTMD on RMS of displacement changes from 72.8% to 82.0% in case of $f_e = 1.0$ and varies between 48.4% and 56.1% when f_e is equal to 1.1 . This is almost the same as the maximum reduction value in Figure 11(a). However, PTMD shows a relatively poor performance with the value of f_e equal to 1.2 . Although the reduction is 30.2% , PTMD rarely works when DTR is -20% in that situation. In addition, reduction decreases sharply from 57.2% to -19.7% when f_e is equal to 0.9 , which has a similar tendency in Figure 11(a). In case of $f_e = 0.8$, reduction is minus as DTR varies from -15% to 20% . This means that PTMD has an adverse effect on vibration control performance.

Figures 11(b) and 11(d) display control performance of TMD. It can be seen that TMD is effective when the structure is subjected to tuned excitation ($f_e = 1.0$) especially when the TMD is tuned (DTR = 0). However, TMD's performance is not as stable as PTMD subjected to tuned excitation. For instance, the value of $\eta_{X1,peak,TMD}$ is 66.2% for tuned TMD, but only 25.8% for a downtuned TMD (corresponding to DTR = -20%).

As an improvement of TMD, PTMD's seismic vibration reduction mechanism mainly includes two parts: on the one hand, controlling force comes from the relative motion between mass block and primary structure, which is the most important reduction mechanism of a traditional TMD. On the other hand, pounding force induced by impact can be regarded as an extra controlling force just as an impact damper, in which energy can be dissipated during the impact between mass block and viscoelastic material. When a tuned wave ($f_e = 1$) is input into the system and a sufficient relative motion can be achieved, both PTMD and TMD show satisfactory reduction results.

However, reduction of TMD decreases due to dramatic falling in relative motion when the controlled platform is under detuning excitation ($f_e \neq 1$). As evidence shown in Figure 11(c), $\eta_{X1,peak,TMD}$ decreases from 66.22% to 26.2% as f_e changes from 1 to 1.1 . Unlike TMD, PTMD shows a superior performance even when the structure is subjected to a detuning wave. This is because controlling force from impact is considerable and high energy dissipation happens during the impact process even when the relative motion is very small. As a comparison result in Figure 11(a), the value of $\eta_{X1,peak,PTMD}$ for a tuned PTMD declines from 84.3% to 58.3% when f_e changes from 1 to 1.1 . For a larger detuning ratio of seismic effect ($f_e = 1.2$), $\eta_{X1,peak,PTMD}$ turns to 32.5% compared to 13.4% for TMD.

6. Comparisons between PTMD and TMD with Optimal Parameters

6.1. Vibration Reduction. The variations of reduction responses for PTMD and TMD which are subjected to different frequency sinusoidal waves are plotted in Figure 12.

In terms of acceleration (Figures 12(a) and 12(b)), the traditional TMD can significantly reduce both of the peak

value and the RMS when the seismic frequency is tuned to the natural frequency ($f_e = 1$), which can be up to 71.7% and 63.1% , respectively. However, higher reduction ratios of 85.8% and 80.6% can be achieved by the PTMD, which shows a better control performance than TMD. When the seismic effect deviates from the natural frequency of primary structure ($f_e \neq 1$), more excellent effects can also be obtained for PTMD by an average of 14% higher than TMD on peak acceleration reduction (Figure 12(a)).

For RMS of acceleration, vibration reduction of TMD ratios is minus when f_e varies from 0.78 to 0.91 and from 1.09 to 1.16 (Figure 12(b)). This means that the TMD magnifies the vibration responses of the offshore jacket platform, and, at this time, TMD is not recommended in the primary structure. In contrast, PTMD maintains an excellent performance within all the range of frequency ratios.

In terms of displacement (Figures 12(c) and 12(d)), more superior reductions are also achieved by the PTMD. The peak value and RMS of dimensionless displacement can be reduced by 86.6% and 84.5% when $f_e = 1$. This reduction is higher than TMD, whose corresponding reductions are 71.5% and 64.9% . Moreover, more effective performance has been observed with the PTMD in most cases except for f_e that ranges from 0.7 to 0.8 for RMS. The analysis results illustrate the fact that, in almost all cases, PTMD is far more effective than TMD on vibration control performance.

6.2. Effective Frequency Bandwidth. To achieve the best vibration control effect, the frequency of sinusoidal wave must be limited to a certain scope when the structure is equipped with TMD. In general, vibration control technique often fails when the frequency of applied load deviates from the structure frequency by more than 10% . Given the randomness of earthquake wave, this technique has its own limitations unavoidably. To quantify the effective bandwidth of TMD and PTMD, l_e is defined as in the following equation:

$$l_e = |f_{e1}(\eta_0) - f_{e2}(\eta_0)|, \quad (17)$$

where η_0 is the required control vibration reduction ratio and $f_{e1}(\eta_0)$ and $f_{e2}(\eta_0)$ are two responding frequencies. l_e measures the range of frequencies in which TMD or PTMD can meet the given requirements. And the larger the value of l_e is, the wider the effective frequency bandwidth is. In other words, a control device with a wider bandwidth is much more adaptive in seismic control.

Figures 13 and 14 illustrate the comparison of effective bandwidths between PTMD and TMD under a given requirement. It is clear that PTMD shows a wider bandwidth on acceleration (Figure 13) and displacement (Figure 14) than TMD for the same required control vibration reduction ratio. Furthermore, it can be observed from Figure 14 that effective bandwidth of PTMD expands approximately twice as much as TMD on RMS displacement.

For peak value of displacement shown in Figure 14(a), PTMD still has distinct advantages. When the reduction requirement is 10% , peak values of displacement are 0.67 for PTMD and 0.47 for TMD. This difference is the closest gap

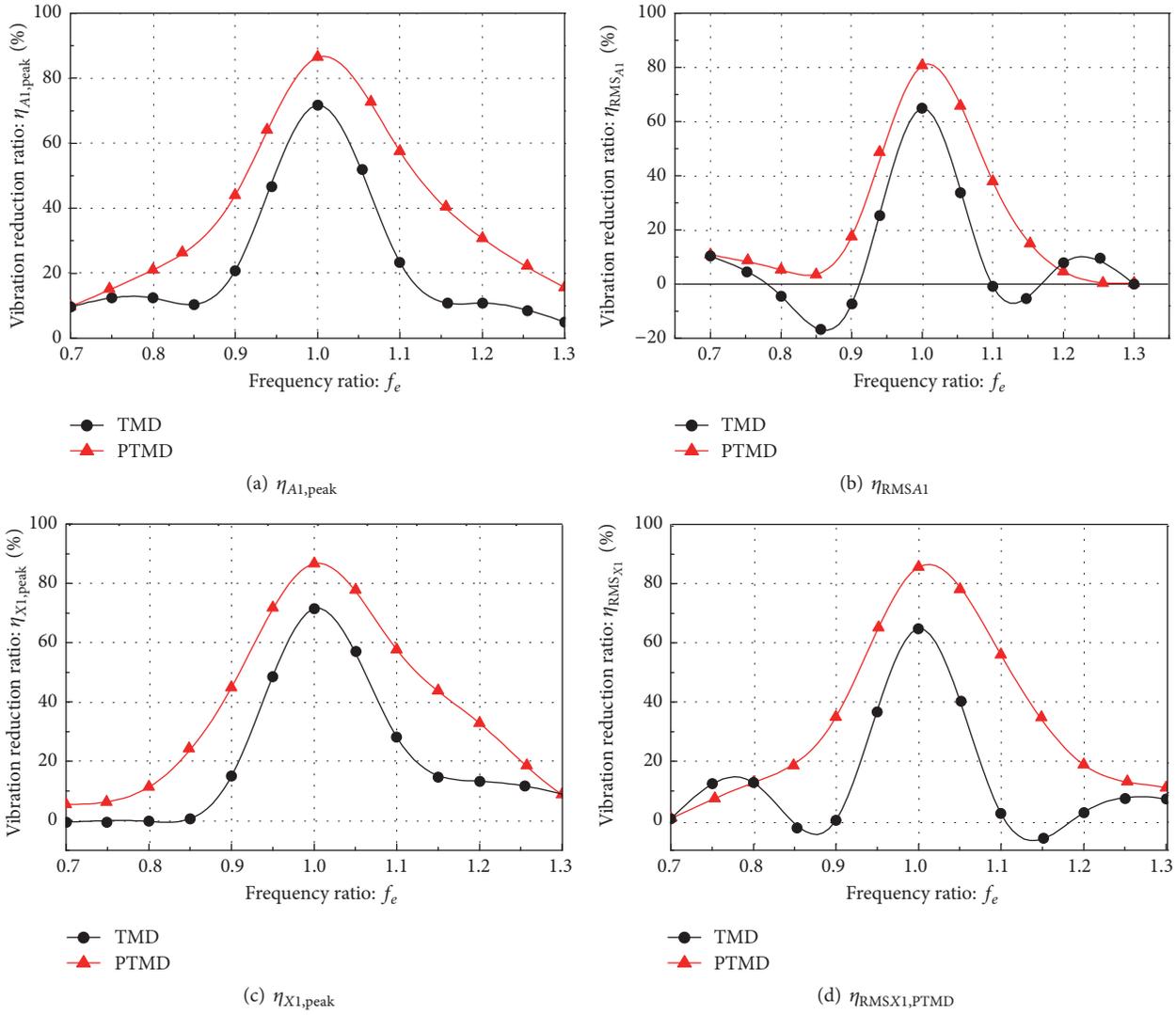


FIGURE 12: The comparison of η between PTMD and TMD with optimal parameter (PTMD: $\xi_2 = 3\xi_{opt}$, $\delta = 2$, and $\mu = 2\%$; TMD: $\xi_2 = 3\xi_{opt}$, $\mu = 2\%$).

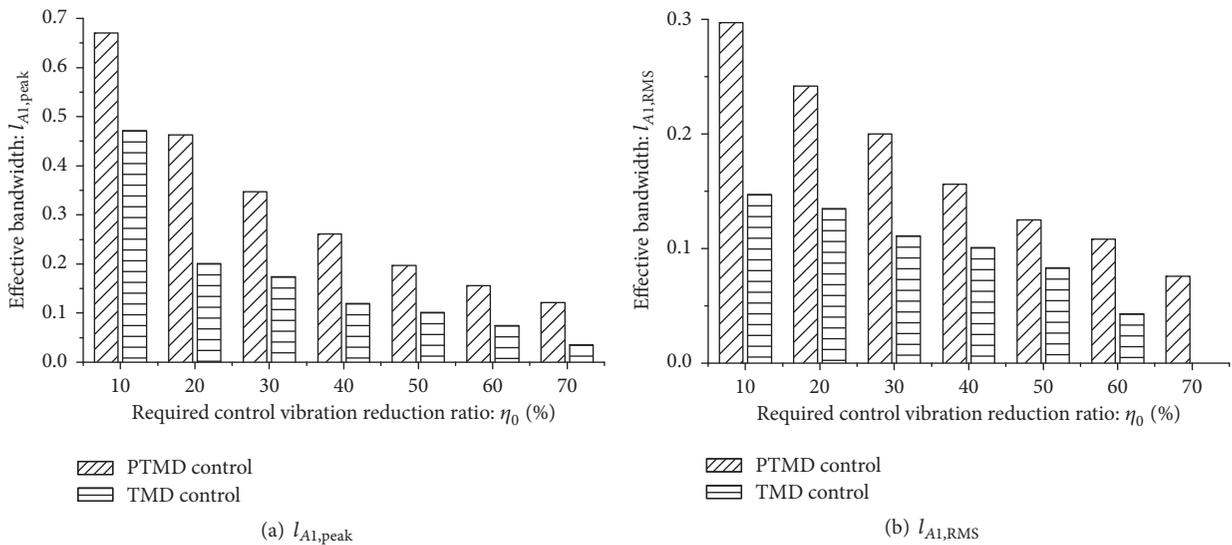


FIGURE 13: Comparison of effective bandwidth on dimensionless acceleration.

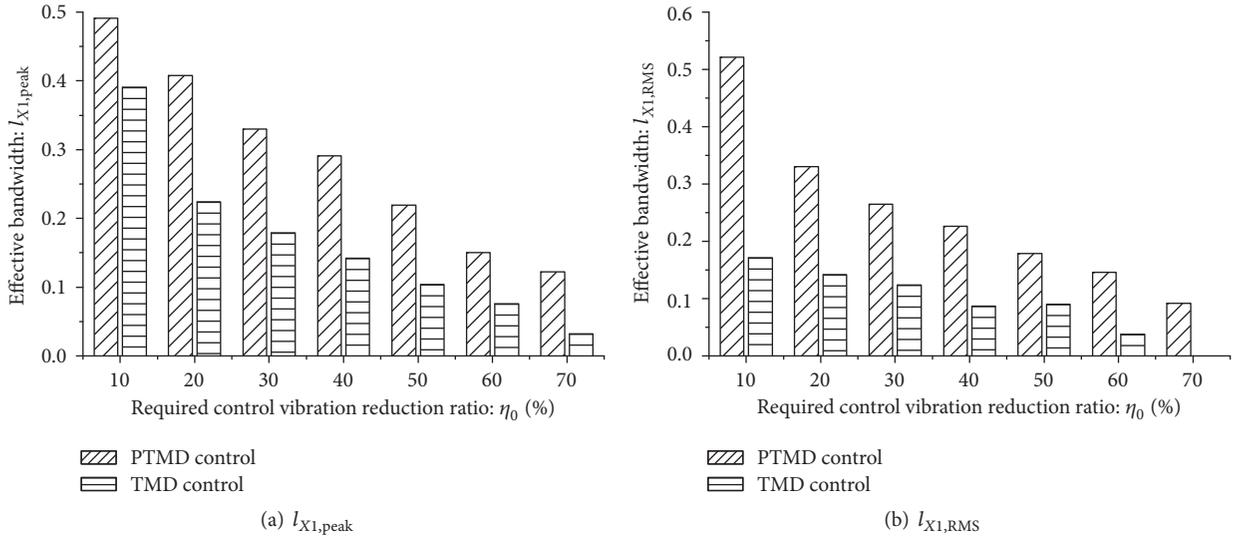
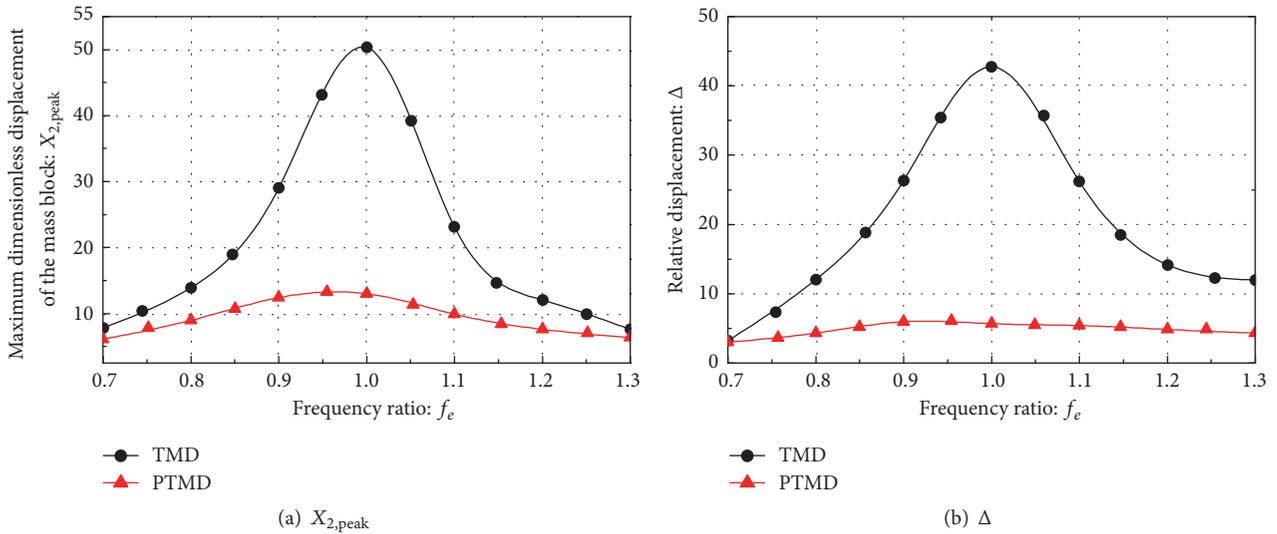


FIGURE 14: Comparison of effective bandwidth on dimensionless displacement.

FIGURE 15: Response of the mass block and relative displacement (PTMD: $\xi_2 = 3\xi_{opt}$, $\delta = 2$, and $\mu = 2\%$; TMD: $\xi_2 = 3\xi_{opt}$, $\mu = 2\%$).

between the devices among others. Otherwise, PTMD makes nearly double the reduction that TMD does.

In terms of acceleration, PTMD has superior performance over the traditional TMD both in peak value (Figure 14(a)) and in the RMS (Figure 14(b)). When the goal for peak value of displacement is 70%, the bandwidth is only 0.032, which means TMD only allows the seismic effect to deviate by 1.6% from the tuned sinusoidal wave since the reduction is almost center symmetry by $f_e = 1$ (Figure 12). In other words, the device is almost useless since it is practically impossible for the structure subjected to a tuned seismic effect. But for PTMD, the bandwidth is 0.122. Even for a lower requirement, like 30%, values of $l_{A1,peak}$ are 0.347 and 0.174 for PTMD and TMD, respectively. In terms of $l_{X1,peak}$, they are 0.330 and 0.179 for PTMD and TMD, respectively. The bandwidth results indicate that PTMD can be applied to different earthquake patterns.

6.3. Costs and Installing Space. A superpower continuous working damper needs to be equipped for TMD due to the large mass of block and continuous movement. Except for high cost, a huge space for installing is a primary need to achieve satisfying control effects.

Figure 15(a) presents the peak value of the dimensionless displacement of the mass block under different excitations. The maximum of the value for TMD can be as large as 50.3 while it is only 12.4 for PTMD. The value of TMD decreases rapidly, but it is smaller in the frequency domain for PTMD. Therefore, the movement of the mass block is significantly reduced due to the limiter, and this means that the cost of additional dampers can be drastically discarded.

The relative displacement is defined as Δ in the following equation:

$$\Delta = |X_1 - X_2|_{\max}, \quad (18)$$

where X_1 and X_2 are dimensionless displacement values for primary structure and mass block, respectively. Δ represents the minimum space requirement for installing. Figure 15(b) illustrates the relationship between Δ of different controlling techniques. It can be observed that at least the minimum value of dimensionless space is 42.8 which must be provided to TMD for motion. But for PTMD, the value is 5.9, which is only 13.7% of TMD.

7. Conclusion

A SDOF structure simplified from JZ20-2MUQ type offshore jacket platform is taken as an example to investigate the seismic control performance and robustness of PTMD. The following conclusions can be obtained:

- (1) Damping ratio is extremely important for PTMD. Although an additional damper will enhance the efficiency of PTMD, an exceeding damping ratio may have the opposite effect. And the optimal damping ratio which is based on TMD is no longer applicable for PTMD due to the impact effect. The optimal damping ratio value for PTMD is about three times that of the TMD in this paper.
- (2) The gap between the mass block and the viscoelastic layer is another key factor for vibration reduction. Since impact may have adverse influences on the control performance when the platform is subjected to low frequency excitation, an appropriate gap should be reserved.
- (3) Control performance of PTMD is sensitive to properties of material. A material with higher equivalent elasticity modulus can enhance the efficiency when the excitation frequency is higher than the natural frequency of primary structure. Otherwise, it will weaken the vibration reduction. But contact geometries have little effect on PTMD's vibration reduction.
- (4) Tuned excitation vibration, detuning vibration, and free vibration are present to investigate the robustness of PTMD. The simulation results show more excellent vibration mitigation in all cases compared to TMD. Comparison between PTMD and traditional TMD verifies the advantages of PTMD not only in vibration suppression and costs, but also in effective frequency bandwidth.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

The research is supported by the National Natural Science Foundation of China (Projects nos. 51678322 and 51409056), the Natural Science Foundation of Heilongjiang Province (E2015047, E2015045), the Fundamental Research Funds for the Central Universities (HEUCF160202), and the Taishan

Scholar Priority Discipline Talent Group program funded by Shandong Province.

References

- [1] S. Chandrasekaran, A. K. Jain, and N. R. Chandak, "Seismic analysis of offshore triangular tension leg platforms," *International Journal of Structural Stability and Dynamics*, vol. 6, no. 1, pp. 97–120, 2006.
- [2] A. Ajamy, M. R. Zolfaghari, B. Asgarian, and C. E. Ventura, "Probabilistic seismic analysis of offshore platforms incorporating uncertainty in soil-pile-structure interactions," *Journal of Constructional Steel Research*, vol. 101, pp. 265–279, 2014.
- [3] K. Kawano, K. Furukawa, and K. Venkataramana, "Seismic response of offshore platform with TMD," in *Proceedings of the 10th World Conference on Earthquake Engineering*, pp. 2241–2246, Balkema, Rotterdam, 1992.
- [4] Q. Wu, X. Zhao, R. Zheng, and K. Minagawa, "High response performance of a tuned-mass damper for vibration suppression of offshore platform under earthquake loads," *Shock and Vibration*, vol. 2016, Article ID 7383679, 11 pages, 2016.
- [5] J. L. Almazan, J. C. De la Llera, J. A. Inaudi, D. Lopez-Garcia, and L. E. Izquierdo, "A bidirectional and homogeneous tuned mass damper: a new device for passive control of vibrations," *Engineering Structures*, vol. 29, no. 7, pp. 1548–1560, 2007.
- [6] M.-H. Chey, J. G. Chase, J. B. Mander, and A. J. Carr, "Semi-active tuned mass damper building systems: design," *Earthquake Engineering and Structural Dynamics*, vol. 39, no. 2, pp. 119–139, 2010.
- [7] C. Sun and S. Nagarajaiah, "Study on semi-active tuned mass damper with variable damping and stiffness under seismic excitations," *Structural Control and Health Monitoring*, vol. 21, no. 6, pp. 890–906, 2014.
- [8] A. Jafarabad, M. Kashani, M. R. A. Parvar, and A. A. Golafshani, "Hybrid damping systems in offshore jacket platforms with float-over deck," *Journal of Constructional Steel Research*, vol. 98, pp. 178–187, 2014.
- [9] M. Mohebbi, H. Rasouli, and S. Moradpour, "Assessment of the design criteria effect on performance of multiple tuned mass dampers," *Advances in Structural Engineering*, vol. 18, no. 8, pp. 1141–1158, 2015.
- [10] L.-S. Huo and H.-N. Li, "Torsionally coupled response control of offshore platform structures using circular tuned liquid column dampers," *China Ocean Engineering*, vol. 18, no. 2, pp. 173–183, 2004.
- [11] H.-N. Li, X.-Y. He, and L.-S. Huo, "Seismic response control of offshore platform structures with shape memory alloy dampers," *China Ocean Engineering*, vol. 19, no. 2, pp. 185–194, 2005.
- [12] J. Ou, X. Long, Q. S. Li, and Y. Q. Xiao, "Vibration control of steel jacket offshore platform structures with damping isolation systems," *Engineering Structures*, vol. 29, no. 7, pp. 1525–1538, 2007.
- [13] Y. Komachi, M. R. Tabeshpour, A. A. Golafshani, and I. Mualla, "Retrofit of Resselat jacket platform (Persian Gulf) using friction damper device," *Journal of Zhejiang University: Science A*, vol. 12, no. 9, pp. 680–691, 2011.
- [14] S. A. Mousavi, S. M. Zahrai, and K. Bargi, "Optimum geometry of tuned liquid column-gas damper for control of offshore jacket platform vibrations under seismic excitation," *Earthquake Engineering and Engineering Vibration*, vol. 11, no. 4, pp. 579–592, 2012.

- [15] S. A. Mousavi, K. Bargi, and S. M. Zahrai, "Optimum parameters of tuned liquid column gas damper for mitigation of seismic-induced vibrations of offshore jacket platforms," *Structural Control and Health Monitoring*, vol. 20, no. 3, pp. 422–444, 2013.
- [16] A. Sarrafan, S. H. Zareh, A. A. A. Khayyat, and A. Zabihollah, "Neuro-fuzzy control strategy for an offshore steel jacket platform subjected to wave-induced forces using magnetorheological dampers," *Journal of Mechanical Science and Technology*, vol. 26, no. 4, pp. 1179–1196, 2012.
- [17] M. A. Lotfollahi-Yaghin, H. Ahmadi, and H. Tafakhor, "Seismic responses of an offshore jacket-type platform incorporated with tuned liquid dampers," *Advances in Structural Engineering*, vol. 19, no. 2, pp. 227–238, 2016.
- [18] P. Zhang, G. Song, H.-N. Li, and Y.-X. Lin, "Seismic control of power transmission tower using pounding TMD," *Journal of Engineering Mechanics*, vol. 139, no. 10, pp. 1395–1406, 2013.
- [19] H. Li, P. Zhang, G. Song, D. Patil, and Y. Mo, "Robustness study of the pounding tuned mass damper for vibration control of subsea jumpers," *Smart Materials and Structures*, vol. 24, no. 9, pp. 135–142, 2015.
- [20] L. Li, G. Song, M. Singla, and Y.-L. Mo, "Vibration control of a traffic signal pole using a pounding tuned mass damper with viscoelastic materials (II): experimental verification," *JVC/Journal of Vibration and Control*, vol. 21, no. 4, pp. 670–675, 2015.
- [21] G. Song, P. Zhang, L. Li et al., "Vibration control of a pipeline structure using pounding tuned mass damper," *Journal of Engineering Mechanics*, vol. 142, no. 6, Article ID 04016031, 2016.
- [22] P. Zhang, L. Li, D. Patil et al., "Parametric study of pounding tuned mass damper for subsea jumpers," *Smart Materials & Structures*, vol. 25, no. 1, Article ID 015028, 2016.
- [23] J. P. Ou, X. Long, Y. Q. Xiao, and B. Wu, "Damping isolation system and its vibration-suppressed effectiveness analysis for offshore platform jacket structures," *Earthquake Engineering and Engineering Vibration*, vol. 6, no. 1, pp. 115–122, 2002.
- [24] Q. Xue, C. Zhang, J. He, G. Zou, and J. Zhang, "An updated analytical structural pounding force model based on viscoelasticity of materials," *Shock and Vibration*, vol. 2016, Article ID 2596923, 15 pages, 2016.
- [25] T. Ioi and K. Ikeda, "On the houle damper for a damped vibration system," *Bulletin of the JSME*, vol. 23, no. 176, pp. 273–279, 1980.

Research Article

Reliability Analysis with Multiple Dependent Features from a Vibration-Based Accelerated Degradation Test

Fuqiang Sun,^{1,2} Jingcheng Liu,¹ Xiaoyang Li,¹ and Haitao Liao²

¹Science and Technology on Reliability and Environmental Engineering Laboratory, School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China

²Department of Industrial Engineering, University of Arkansas, Fayetteville, AR 72701, USA

Correspondence should be addressed to Haitao Liao; liao@uark.edu

Received 14 July 2016; Accepted 27 October 2016

Academic Editor: Minvydas Ragulskis

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

Accelerated degradation testing (ADT) has been widely used for reliability prediction of highly reliable products. In many applications, ADT data consists of multiple degradation-related features, and these features are usually dependent. When dealing with such ADT data, it is important to fully utilize the multiple degradation features and take into account their inherent dependency. This paper proposes a novel reliability-assessment method that combines Brownian motion and copulas to model ADT data obtained from vibration signals. In particular, degradation feature extraction is first carried out using the raw vibration signals, and a feature selection method quantifying feature properties, such as trendability, monotonicity, and robustness, is adopted to determine the most suitable degradation features. Then, a multivariate s-dependent ADT model is developed, where a Brownian motion is used to depict the degradation path of each degradation feature and a copula function is employed to describe the dependence among these degradation features. Finally, the proposed ADT model is demonstrated using the vibration-based ADT data for an electric motor.

1. Introduction

Nowadays, products are made more reliable due to the advances of design and manufacturing and to the improvement of materials technology. As a result, traditional reliability analysis methods based on lifetime data are becoming less capable of evaluating the reliability of such highly reliable products. In practice, the failure mechanism of a product often can be traced to the degradation of the product's key performance parameters, such as wear, fatigue crack, and material corrosion. The product is considered to be failed if any of the performance parameters exceeds its critical value. Therefore, an alternative way for evaluating the product's reliability is to analyze the degradation data obtained from a degradation test. To collect degradation data in a short time, accelerated degradation testing (ADT) technique has been proposed, where some units of the product are exposed to harsher-than-normal conditions in order to speed up product degradation. The resulting ADT data will be used to develop

an ADT model that enables the extrapolation of the product's reliability under the normal operating conditions.

In order to find the degradation trend of a product and predict its lifetime using ADT, the product's performance indicators must be defined first. Such indicators may include not only the product's functional or performance parameters, but also those special features extracted from raw sensor data [1], such as vibration, force and acoustic signals, temperature, and voltage. In particular, vibration monitoring has been widely used in fault diagnostics and remaining useful life (RUL) prediction for mechanical and electromechanical systems [2]. Nectoux et al. [3] developed a bearing experimental platform, called PRONOSTIA, to perform bearing ADT, where both vibration and temperature signals were monitored. Liu et al. [4] introduced the confidence value (CV) as the degradation index for a product and used a fuzzy regression model to analyze the bearing vibration data given in [3]. However, in the field of prognostic and health management (PHM), little effort has been focused on

the mathematical modeling of multiple degradation features obtained from vibration-based ADT.

Generally, raw vibration data may not be directly used for degradation modeling. As a solution, signal processing techniques are usually utilized to extract some features from the vibration data [5]. It is not uncommon that quite a few features can be extracted in the time, frequency, and time-frequency domains as well as via information entropy analysis [3]. However, since not all features are required in describing the product's degradation process, it is necessary to perform dimension reduction for the purpose of degradation modeling. There are two popular ways to do that. The first one is to perform principal component analysis, singular value decomposition, self-organizing map, or clustering [2]. The other one is feature selection that selects a subset of features by assessing the fitness of candidate features [6–10]. The purpose of performing feature selection is to identify those features that can best describe the fault progression process [8]. Coble and Hines [6] proposed three metrics for feature selection, that is, trendability, monotonicity, and prognosability. Javed et al. [9] performed feature selection by evaluating feature fitness using the monotonicity metric defined in [6] and a new trendability metric, and Liao [8] only used the monotonicity criterion. Camci et al. [10] divided a degradation feature into segments and the average separability of all segments was defined for feature selection. Zhang et al. [7] defined three other metrics, that is, correlation, monotonicity, and robustness, for selecting more relevant degradation features.

It is quite common that there is an s -dependent relationship among the stochastic processes describing different degradation features. Clearly, ignoring such dependency in ADT analysis will lead to inaccurate reliability estimates. For modeling multiple degradation processes, most work considers an independence assumption or multivariate normal distributions [11, 12]. However, these assumptions may not be appropriate in many engineering applications. Copulas provide a convenient way to model the dependence of random variables characterized by their marginal distributions [13]. In addition, no constraints on the univariate marginal distributions are required in implementing copula. Because of its advantages, copula has become a popular approach in multivariate statistical analysis. Sari et al. [14] used a copula function to describe the correlation between two performance characteristics of light emitting diode (LED) and combined it with a generalized linear regression model to model bivariate constant-stress ADT. Pan et al. [15] discussed the bivariate degradation modeling approaches based on Wiener processes and copulas under constant-stress accelerating scenarios as well. Similarly, Peng et al. [16], Liu et al. [17], Wang and Pham [18], and Li and Xue [19] also employed the copula method to model the dependency between bivariate degradation features. However, previous research is limited on modeling the dependence between bivariate degradation processes, and little effort has been focused on multiple dependent degradation processes.

Another key issue in ADT modeling is how to choose appropriate mathematic models to depict the degradation paths. In the literature, two kinds of degradation models that

have been widely used are stochastic process model (e.g., Brownian motion, gamma process, and inverse Gaussian process) and general path model [20]. Ye and Xie [20] conducted a comprehensive literature review on Brownian motion based degradation models. Liao and Elsayed [21] developed an ADT model based on Brownian motion to predict the field reliability of LED considering stress variations. Guan et al. [22] adopted Brownian motion to model the constant-stress ADT (CSADT) and estimated the model parameters using an objective Bayesian method.

Based on the related literature, one can see that some technical problems have not been resolved in the study of vibration-based ADT with multiple features, including fitness analysis of degradation indicators, development of multivariate s -dependent ADT model, and parameter estimation. To overcome these challenges, a novel reliability modeling method for vibration-based CSADT with multiple dependent features is proposed in this paper. To the best of our knowledge, this study is the first attempt in the related areas. The remainder of this paper is organized as follows. The framework of the proposed method is given in Section 2. Section 3 introduces the feature extraction and selection methods. Section 4 provides a multivariate ADT model based on copulas. In Section 5, a case study is used to illustrate the validity of the proposed model. Finally, conclusions are drawn in Section 6.

2. Framework of the Proposed Model

Figure 1 provides the proposed framework for reliability and life evaluation using vibration-based ADT data. The framework consists of two parts: fitness analysis of degradation indicators and modeling of multivariate ADT data.

The first part is to extract and select the degradation indicators from the raw vibration data. It is often necessary to first denoise the vibration data to reduce the negative influence of environmental noise on the subsequent analysis. Then, features are extracted in four domains, that is, time, frequency, time-frequency, and information domains. Since the dimension of the features is high, three goodness metrics are used for the selection of the most suitable degradation indicators from these candidate features.

The other part is to model the multiple accelerated degradation processes describing the selected features. At first, a Brownian motion model is adopted to depict the degradation path of each feature and to extrapolate the reliability related to each feature under the normal stress with an acceleration model and a specified failure threshold. Afterwards, a copula function is used to describe the dependence of different degradation features resulting in a multivariate s -dependent ADT model. A two-stage statistical inference procedure, called Inference Functions for Margins (IFM) method, is employed to estimate the unknown parameters of the proposed model. Finally, Akaike information criterion (AIC) is used to determine the best fitting copula function, and the system joint reliability under the normal operating condition can be obtained.

The following sections will elaborate on the analysis of vibration-based ADT with multiple features.

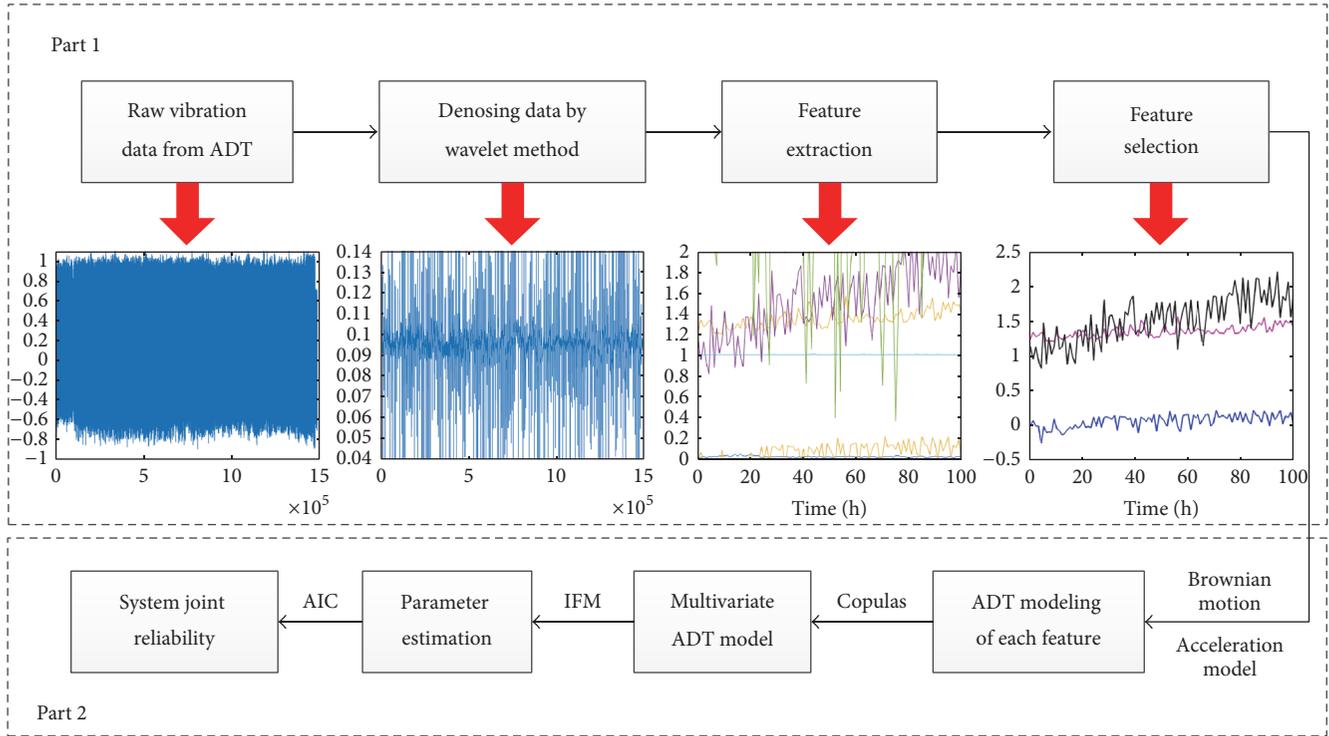


FIGURE 1: The framework of reliability and life evaluation for vibration-based ADT.

3. The Degradation Indicator Fitness Analysis

3.1. Signal Processing and Feature Extraction. Denoising is an important step in effective fault detection, diagnosis, and prognostics [23]. In this paper, wavelet automatic denoising technique [24] is used to remove the noise from vibration signals prior to feature extraction. This technique is capable of deconstructing complex signals into basis signals of finite bandwidth and then reconstructing them again with very little loss of information, which is best for nonstationary signal analysis.

Then, different signal processing methods are utilized to find the most valuable features relevant to the product's performance state. Indeed, feature extraction via signal processing is considered one of the major elements and the most challenging aspect in analyzing vibration-based ADT data [25]. The existing signal processing and feature extraction techniques dealing with vibration signals mainly include four categories [2].

(1) *Time Domain.* Time domain feature extraction is performed directly on the waveforms of vibration signals, which are suitable for fault detection. Usually, the extracted features include mean, variance, waveform factor, crest factor, clearance factor, and kurtosis factor.

(2) *Frequency Domain.* Frequency domain analysis can identify and isolate those important frequency components of the signals. The most popular method is the fast Fourier transform, with which features such as center frequency,

mean square frequency, and frequency variance can be obtained.

(3) *Time-Frequency Domain.* Time-frequency analysis aims at investigating the signals in both time and frequency domains [9]. In particular, Hilbert-Huang Transform (HHT), based on empirical mode decomposition, is the most powerful technique for the analysis of nonstationary signals [26]. HHT marginal energy is calculated as the feature in this domain.

(4) *Information Entropy.* Information entropy is a measure of uncertainty of signals [4]. A higher value indicates a higher uncertainty of the signal. Following this idea, the Hilbert marginal energy spectrum entropy is computed in feature extraction.

Let x_1, x_2, \dots, x_N be the recorded vibration signals, f be the frequency components of the signals, and $S(f)$ be the power spectrum. Table 1 gives various features obtained by processing the original vibration signals.

3.2. Feature Selection. The selection of a subset of appropriate features is essential for prognostic modeling. It is often performed on the basis of feature characteristics by comparing candidate features with respect to their fitness. In other words, the features with high attribute values, such as monotonicity, correlation, trendability, and robustness, should lead to more accurate prognostic results and thus be included for further degradation modeling.

Technically, after removing unwanted noise, a feature is decomposed into its trend and random part by applying a

TABLE 1: Candidate degradation features.

Feature	Formula	Feature	Formula
Mean	$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$	Clearance factor	$L = \frac{\max \{ x_i \}}{\left[(1/N) \sum_{i=1}^N \sqrt{ x_i } \right]^2}$
Variance	$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$	Frequency center	$FC = \frac{\int_0^{+\infty} f S(f) df}{\int_0^{+\infty} S(f) df}$
Crest factor	$C = \frac{\max \{ x_i \}}{\sqrt{(1/N) \sum_{i=1}^N x_i^2}}$	Frequency variance	$VF = \frac{\int_0^{+\infty} (f - FC)^2 S(f) df}{\int_0^{+\infty} S(f) df}$
Kurtosis factor	$\alpha = \frac{1}{N} \sum_{i=1}^N \frac{(x_i - \bar{x})^4}{\sigma^4}$	HHT marginal energy (HME)	$HME(\omega) = \int_0^T H^2(\omega, t) dt$
Waveform factor	$S = \frac{\sqrt{(1/N) \sum_{i=1}^N x_i^2}}{(1/N) \sum_{i=1}^N x_i }$	HHT marginal energy spectrum entropy (HMESE)	$H(\omega, t) = \text{Re} \sum_{j=1}^m c_j(t) e^{j 2\pi\omega_j t}$ $HMESE = \frac{-\sum_{j=1}^m p_j \log_2 p_j}{\log_2 m}$ $p_j = \frac{HME(j)}{\sum_{j=1}^m HME(j)}$ $\sum_{j=1}^m p_j = 1$

Note: m is the number of frequency components in Hilbert-Huang Transform (HHT).

robust locally weighted scatter plot smooth (RLOESS) with a span value of 0.5 (i.e., 50%), as shown in (1). Unlike locally weighted scatter plot smooth (LOESS) that is a popular smoothing method based on a locally weighted regression function and a quadratic polynomial, RLOESS uses a robust regression weight function to make the LOESS process insensitive to outliers

$$x = x_T + x_R, \quad (1)$$

where x is the actual feature value, x_T is its trend value, and x_R is the residual.

Then, a set of mathematical criteria in favor of the features with high attribute values are considered in feature selection.

- (1) Monotonicity characterizes the underlying increasing or decreasing trend of feature. It can be measured by the average absolute difference of derivatives for each feature [6] as follows:

$$M = \text{mean} \left(\left| \frac{\text{No. of } d/dx > 0}{n-1} - \frac{\text{No. of } d/dx < 0}{n-1} \right| \right), \quad (2)$$

where n is the number of observations, d/dx is the derivative, and mean represents averaging. The value of M is from 0 to 1, with $M = 1$ indicating highly monotonic features and $M = 0$ indicating nonmonotonic features.

- (2) Trendability is related to the linear relationship between the feature and the time. Clearly, it shows how the degrading state varies with time. The metric of trendability is given by [9]

$$T = \frac{|n(\sum x_T \cdot t) - (\sum x_T)(\sum t)|}{\sqrt{[n \sum x_T^2 - (\sum x_T)^2][n \sum t^2 - (\sum t)^2]}}, \quad (3)$$

where $T \in [0, 1]$ is the absolute correlation coefficient between feature trend x_T and time t . Obviously, a constant function of feature has zero correlation with time, while a linear function indicates a strong correlation between the feature and the time.

- (3) Robustness reflects the tolerance of the feature to outliers [7], which is given by

$$R = \frac{1}{n} \sum \exp \left(- \left| \frac{x_R}{x} \right| \right). \quad (4)$$

One can see that all the three metrics of features are confined in the range of $[0, 1]$. Due to the fact that one metric only partially measures the suitability of candidate feature, a weighted sum of the three metrics is used in this paper as the fitness function for feature selection

$$\begin{aligned} \max \quad & \text{fitness} = \omega_1 \cdot M + \omega_2 \cdot T + \omega_3 \cdot R \\ \text{s.t.} \quad & \omega_i > 0 \\ & \sum_i \omega_i = 1, \\ & i = 1, 2, 3, \end{aligned} \quad (5)$$

where constants ω_1 , ω_2 , and ω_3 control the importance of each metric in the optimization problem.

4. Modeling of CSADT with Multiple Features

4.1. *The Univariate Accelerated Degradation Model.* The premise of ADT-based reliability evaluation is to choose an appropriate model based on the degradation physics or the degradation data [20]. We first focus on each degradation feature.

4.1.1. Basic Assumptions. Among many stochastic process models, Brownian motion is the most widely used in degradation modeling and analysis. This paper employs this model to describe each degradation process. To evaluate the reliability characteristics of the product under the normal operating condition based on ADT data, the relationship between lifetime characteristics and accelerated stress levels should be modeled. To this end, the following assumptions are made.

- (A1) The test specimens are independent, and no catastrophic failures occur during the test.
- (A2) Under either the normal condition S_0 or accelerated conditions $S_1 < S_2 < \dots < S_K$, degradation process $Y(t)$ over time can be described by

$$Y(t) = \sigma B(t) + \mu \cdot t + y_0, \quad (6)$$

where t is time, $B(t)$ is a standard Brownian motion, y_0 is the initial degradation level at time zero, $\mu > 0$ is the drift parameter, and $\sigma > 0$ is the diffusion parameter describing the variability due to unit-to-unit variation and the variation of operating and environmental conditions.

The drift parameter μ , commonly known as the degradation rate, is regarded as a function of stress conditions. Usually, such acceleration relationships, such as Arrhenius model, Eyring model, or inverse power model [27], are obtained from physical mechanism analysis or empirical experiences. In particular, the following log-linear relationship is considered:

$$\ln \mu = \ln d(S) = a + b \cdot \varphi(S), \quad (7)$$

where a and b are unknown parameters, $d(S)$ is the degradation rate under stress level S , and $\varphi(S)$ is a function of stress level depending on the stress type. For example, if S is absolute temperature, $\varphi(S) = 1/S$; if S is voltage, $\varphi(S) = \ln S$.

- (A3) The diffusion parameter σ does not change with stress, that is, $\sigma_0 = \sigma_1 = \dots = \sigma_K = \sigma$ for the normal condition S_0 and accelerated conditions $S_1 < S_2 < \dots < S_K$.

4.1.2. Reliability Function for Each Degradation Process. Because of the independent increment property of Brownian motion, the degradation increment Δy over nonoverlapped interval Δt is normally distributed with mean $\mu \Delta t$ and variance $\sigma^2 \Delta t$. Its probability density function (PDF) is

$$f(\Delta y) = \frac{1}{\sqrt{2\pi\sigma^2\Delta t}} \exp \left\{ -\frac{[\Delta y - \mu \cdot \Delta t]^2}{2\sigma^2\Delta t} \right\}. \quad (8)$$

Generally, a product is supposed to be failed if its performance degradation exceeds a critical threshold for the first time. In other words, the first passage time (FPT) distribution is used as the product's lifetime distribution. For a given critical threshold D , the lifetime T_D of the product is

the instant when degradation process $Y(t)$ exceeds D for the first time; that is,

$$T_D = \inf \{t > 0 \mid Y(t) \geq D\}. \quad (9)$$

It is well-known that the first passage time follows the inverse Gaussian distribution with the following PDF [28]:

$$f(t; y_0, D) = \frac{D - y_0}{\sigma \sqrt{2\pi t^3}} \exp \left\{ -\frac{[(D - y_0) - \mu \cdot t]^2}{2\sigma^2 t} \right\}. \quad (10)$$

Combined with (7), the associated reliability function under normal operational condition S_0 can be expressed as

$$R(t) = \Phi \left[\frac{D - y_0 - d(S_0)t}{\sigma \sqrt{t}} \right] - \exp \left(\frac{2d(S_0)(D - y_0)}{\sigma^2} \right) \cdot \Phi \left[-\frac{D - y_0 + d(S_0)t}{\sigma \sqrt{t}} \right], \quad (11)$$

where $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution.

4.2. Use of Copulas for Multiple Features

4.2.1. Definition and Basic Properties. Sklar's theorem lays the theoretical foundation for copulas and their applications [13].

Theorem 1 (Sklar's theorem [29]). *Let $X = (X_1, \dots, X_N)$ be a random vector with marginal distributions $F_1(x_1), \dots, F_N(x_N)$ and F be their joint distribution function. Then, there exists a copula function C such that*

$$F(x_1, \dots, x_n, \dots, x_N) = C(F_1(x_1), \dots, F_n(x_n), \dots, F_N(x_N)). \quad (12)$$

If $F_1(x_1), \dots, F_N(x_N)$ are continuous, the copula function C is unique. Conversely, if $F_1(x_1), \dots, F_N(x_N)$ are univariate distributions, the function F defined by (12) is the joint distribution function associated with margins $F_1(x_1), \dots, F_N(x_N)$.

Based on the theorem, a multivariate copula function can be defined as follows.

Definition 2 (n -dimensional copula or n -copula [13]). An N -dimensional copula is a function C mapping from $\mathbf{I}^N = [0, 1]^N$ to \mathbf{I} with the following properties:

- (1) Grounded: for every $\mathbf{u} = (u_1, \dots, u_N)$ in \mathbf{I}^N , $C(\mathbf{u}) = 0$ if at least one coordinate of \mathbf{u} is 0.
- (2) Uniform marginals: if all coordinates of \mathbf{u} are ones except u_k , then

$$C(\mathbf{u}) = C(1, \dots, 1, u_k, 1, \dots, 1) = u_k. \quad (13)$$

TABLE 2: Typical multivariate copulas.

Copulas	$C(u_1, u_2, \dots, u_n)$	Parameters
Gaussian copula	$\Phi_\rho [\Phi^{-1}(u_1), \Phi^{-1}(u_2), \dots, \Phi^{-1}(u_n)]$	$\rho \in (-1, 1)$
Student's t -copula	$t_{\rho, \nu} [t_\nu^{-1}(u_1), t_\nu^{-1}(u_2), \dots, t_\nu^{-1}(u_n)]$	$\rho \in (-1, 1)$ $\nu > 2$
Clayton copula	$\left(\sum_{i=1}^n u_i^{-\delta} - n + 1 \right)^{-1/\delta}$	$\delta > 0$
Frank copula	$-\delta^{-1} \log \left\{ 1 + \frac{\prod_{i=1}^n [\exp(-\delta u_i) - 1]}{[\exp(-\delta) - 1]^{n-1}} \right\}$	$\delta \in (-\infty, \infty) \setminus \{0\}$ $\delta > 0$ for $n \geq 3$
Joe copula	$1 - \left\{ 1 - \prod_{i=1}^n [1 - (1 - u_i)^\delta] \right\}^{1/\delta}$	$\delta \in [1, \infty)$

(3) N -increasing: for each hyperrectangle $B = \prod_{i=1}^N [u_i, v_i] \subseteq [0, 1]^N$, the C -volume of B is nonnegative:

$$\int_B dC(\mathbf{u}, \mathbf{v}) = \sum_{\mathbf{z} \in \times_{i=1}^N \{u_i, v_i\}} (-1)^{N(\mathbf{z})} C(\mathbf{z}) \geq 0, \quad (14)$$

where $N(\mathbf{z}) = \#\{k : z_k = u_k\}$.

Furthermore, the PDF corresponding to joint distribution function $F(x_1, \dots, x_n, \dots, x_N)$ can be calculated as

$$f(x_1, \dots, x_n, \dots, x_N) = c(F_1(x_1), \dots, F_n(x_n), \dots, F_N(x_N)) \prod_{n=1}^N f_n(x_n), \quad (15)$$

where $f_n(\cdot)$ is the PDF of a marginal distribution $F_n(\cdot)$, and $c(\cdot)$ is the density of copula function C :

$$c(u_1, \dots, u_n, \dots, u_N) = \frac{\partial^N C(u_1, \dots, u_n, \dots, u_N)}{\partial u_1 \dots \partial u_n \dots \partial u_N} \quad (16)$$

$$\forall \mathbf{u} = (u_1, \dots, u_n, \dots, u_N) \in \mathbf{I}^N.$$

4.2.2. Selection of Copula Family. It is important to select an appropriate copula family for the given data, as the selection significantly influences the prediction performance. Table 2 presents five popular multivariate copula families, where u_1, \dots, u_n are n random variables.

To balance the statistical fitting and model complexity, the values of classic Akaike information criterion (AIC) are computed for the candidate copula families:

$$\text{AIC} = 2k - 2 \ln L, \quad (17)$$

where k is the number of estimated parameters in the model, and L is the value of the corresponding maximum likelihood. By comparing the AIC values, the family with the minimum value will be selected.

4.3. Multivariate ADT Model. To analyze the multivariate CSADT data, additional assumptions are considered.

(A4) For a product with multiple degradation indicators, the dependency among the features can be characterized by a copula function.

(A5) The degradation measurements of all specimens are collected at the same time.

(A6) The specimen is considered to be failed if one of the features reaches its corresponding failure threshold for the first time.

Suppose that a product has P degradation features and each of the accelerated degradation processes is modeled by a drift Brownian motion. $Y_p(t)$, $p = 1, 2, \dots, P$, denotes the degradation process of the p th characteristic index at operating condition S_0 , and its corresponding failure threshold is D_p . Let the failure time of the p th characteristic index be T_p , and the lifetime of the product be T , so $T = \min(T_1, \dots, T_P)$. Clearly, the system reliability under normal condition S_0 can be expressed as

$$R(t) = P(T > t) = P(T_1 > t, \dots, T_P > t) = P(Y_1(t) < D_1, \dots, Y_P(t) < D_P). \quad (18)$$

Apparently, if the P degradation features are independent, (18) can be rewritten as

$$R(t) = P(Y_1(t) < D_1) \times \dots \times P(Y_P(t) < D_P) = R_1(t) \times \dots \times R_P(t), \quad (19)$$

where $R_p(t)$, $p = 1, 2, \dots, P$ is the p th characteristic index's reliability function under normal operating condition S_0 , as defined in (11). However, this is often not true in many engineering applications. To address the dependence of P degradation features, a copula function can be utilized. In particular, the product's reliability at time t under normal operating condition S_0 can be performed as

$$R(t) = C(R_1(t), R_2(t), \dots, R_P(t); \boldsymbol{\theta}), \quad (20)$$

where $\boldsymbol{\theta}$ is the parameter set of the copula.

4.4. Parameter Estimation. Consider a copula-based multivariate distribution for the random vector $[Y_1, \dots, Y_P]$, with joint PDF

$$f(\mathbf{y}; \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_P, \boldsymbol{\theta}) = c(F_1(y_1; \boldsymbol{\alpha}_1), \dots, F_P(y_P; \boldsymbol{\alpha}_P); \boldsymbol{\theta}) \prod_{p=1}^P f_p(y_p; \boldsymbol{\alpha}_p), \quad (21)$$

where $F_p(\cdot; \boldsymbol{\alpha}_p)$ and $f_p(\cdot; \boldsymbol{\alpha}_p)$, $p = 1, \dots, P$, are the p th marginal CDF and PDF with parameter set $\boldsymbol{\alpha}_p$, and $c(\cdot; \boldsymbol{\theta})$ is the density of copula function $C(\cdot; \boldsymbol{\theta})$ with parameter $\boldsymbol{\theta}$.

For a sample of size n , the full log-likelihood function can be expressed as

$$\ln L(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_P, \boldsymbol{\theta})$$

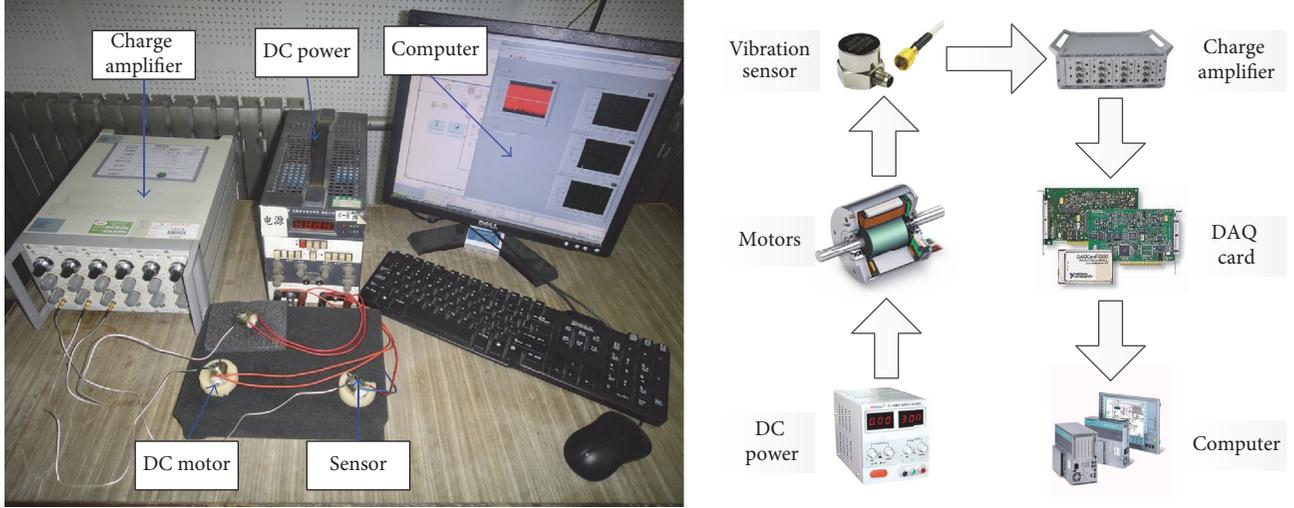


FIGURE 2: Overview of the motor vibration-based ADT platform.

$$\begin{aligned}
 &= \underbrace{\sum_{i=1}^n \ln c(F_1(y_{1i}; \alpha_1), \dots, F_P(y_{Pi}; \alpha_P); \theta)}_{\text{dependence structure } L_C} \\
 &+ \underbrace{\sum_{p=1}^P \sum_{i=1}^n f_p(y_{pi}; \alpha_p)}_{\text{marginals } \sum_{p=1}^P L_p}.
 \end{aligned} \tag{22}$$

The maximum likelihood estimates (MLE) of the parameters $(\alpha_1, \dots, \alpha_P, \theta)$ can be obtained by maximizing (22). However, as the dimension of parameters increases, it is difficult to get the optimal solution. Joe [30, 31] proposed the IFM method as a computationally attractive alternative to estimate the parameters in multivariate copula models. Technically, (22) is first decomposed into two parts. The first part, denoted by L_C , is the contribution from the dependence structure in data, and the other part is the contributions from each margin, L_p , $p = 1, \dots, P$. Then, The IFM method separately estimates the parameters of marginal distributions and the parameters of the copula in two stages. In the first stage, the estimates of the marginal distribution parameters are obtained from the log-likelihood L_p of each margin. In the second stage, the copula parameter estimates are computed by maximizing L_C based on the marginal distribution parameters estimated in the first stage. In this paper, we use this method to estimate the parameters of the proposed multivariate ADT model.

$$\ln L = -\frac{1}{2} \sum_{l=1}^K \sum_{i=1}^{n_l} \sum_{j=1}^{M_l-1} \left\{ \ln(2\pi\sigma_p^2\Delta t) + \frac{[\Delta y_{plij} - \exp(a_p + b_p\varphi(S_l)) \cdot \Delta t]^2}{\sigma_p^2\Delta t} \right\}, \quad p = 1, \dots, P \tag{24}$$

and the parameter set $\alpha_p = (a_p, b_p, \sigma_p)$ of each marginal distribution can be estimated by maximizing the log-likelihood function.

4.4.1. Marginal Distribution Parameters Estimation. Suppose that in a CSADT with K stress levels the total sample size is n . There are n_l specimens under stress level S_l . During the CSADT, all specimens are measured once in every Δt time interval and there are M_l inspections under S_l . Then, the observation of the p th characteristic index at time t_{lij} is $y_p(t_{lij})$, $p = 1, \dots, P$, $l = 1, \dots, K$, $i = 1, \dots, n_l$, $j = 1, \dots, M_l$, where t_{lij} is the time of the j th measurement of the i th unit under the l th stress level. According to (8), the likelihood function of the p th characteristic indices is given by

$$\begin{aligned}
 L_p &= \prod_{l=1}^K \prod_{i=1}^{n_l} \prod_{j=1}^{M_l-1} \frac{1}{\sqrt{2\pi\sigma_p^2\Delta t}} \\
 &\cdot \exp \left\{ -\frac{[(y_p(t_{li(j+1)})) - y_p(t_{lij})) - \mu_{pl} \cdot \Delta t]^2}{2\sigma_p^2\Delta t} \right\}, \tag{23} \\
 & \quad p = 1, \dots, P.
 \end{aligned}$$

Considering the acceleration model given by (7), μ_{pl} can be expressed as $\mu_{pl} = d_p(S_l) = \exp(a_p + b_p\varphi(S_l))$, where a_p and b_p are unknown parameters of the p th characteristic index. So the model parameter set of the characteristic index is $\alpha_p = (a_p, b_p, \sigma_p)$, $p = 1, \dots, P$. Let $\Delta y_{plij} = y_p(t_{li(j+1)}) - y_p(t_{lij})$; then (23) can be expressed as

4.4.2. Copula Parameter Estimation. By replacing the marginal parameters α_p , $p = 1, \dots, P$, with their first-stage estimate $\hat{\alpha}_p$, the copula parameters can be estimated by

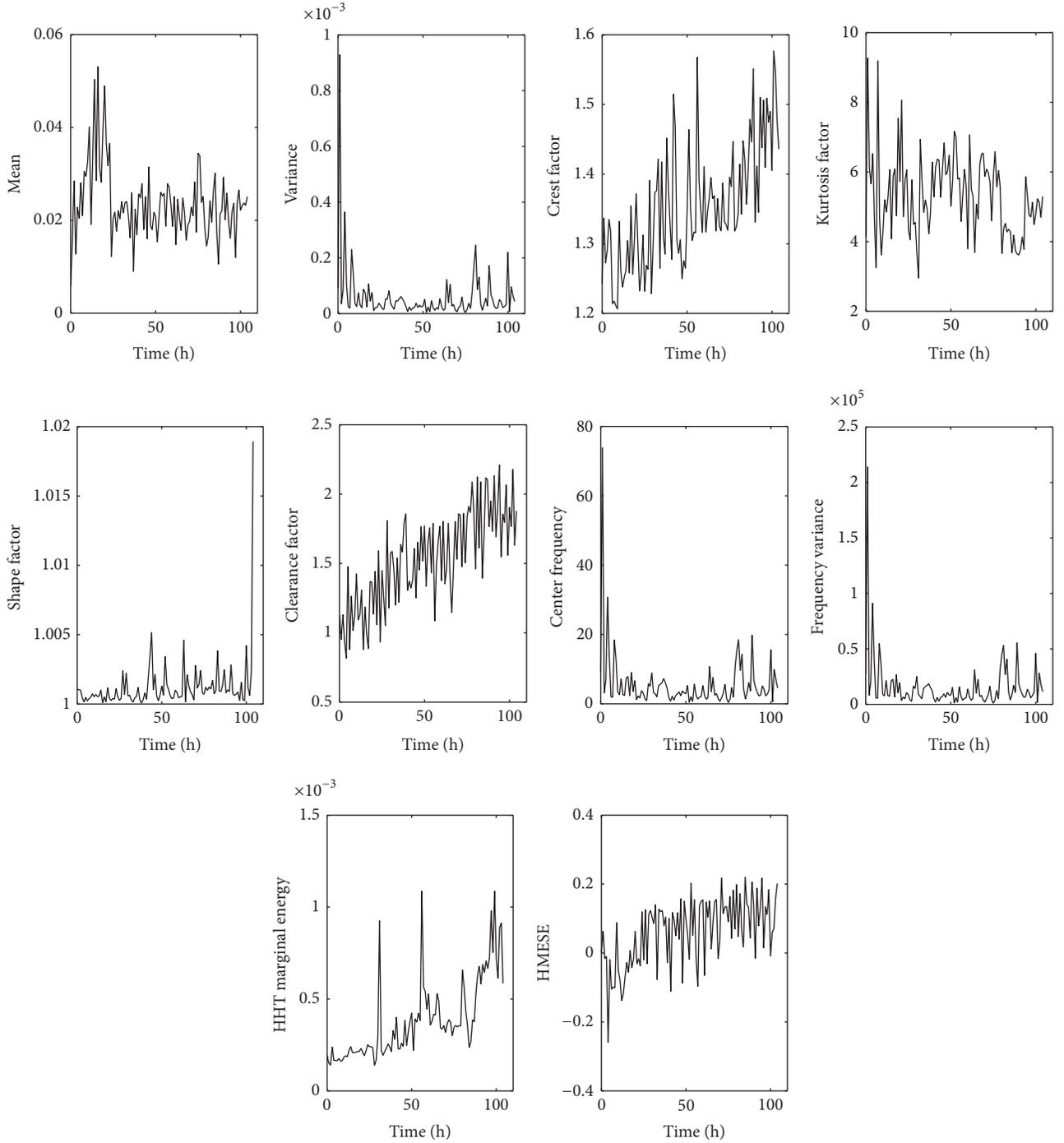


FIGURE 3: Feature set of a motor.

maximizing the copula likelihood contribution. Then, the copula parameter θ in (20) is given by

$$\hat{\theta} = \arg \max_{\theta} \sum_{p=1}^P \sum_{i=1}^n \ln c(R_1(y_{1i}; \hat{\alpha}_1), \dots, R_P(y_{pi}; \hat{\alpha}_P); \theta). \quad (25)$$

In addition, the AIC method mentioned in Section 4.2.2 is used to quantitatively select the best fitting model from candidate copula functions.

5. Case Study

A case study is provided in this section to demonstrate the use of the proposed framework in real-world industrial applications.

5.1. Description of the ADT Experiment. DC motor converts DC power into mechanical energy. It is a key part of robots, electric vehicle, numerical control machines and household appliances, and so forth. The failure of motors

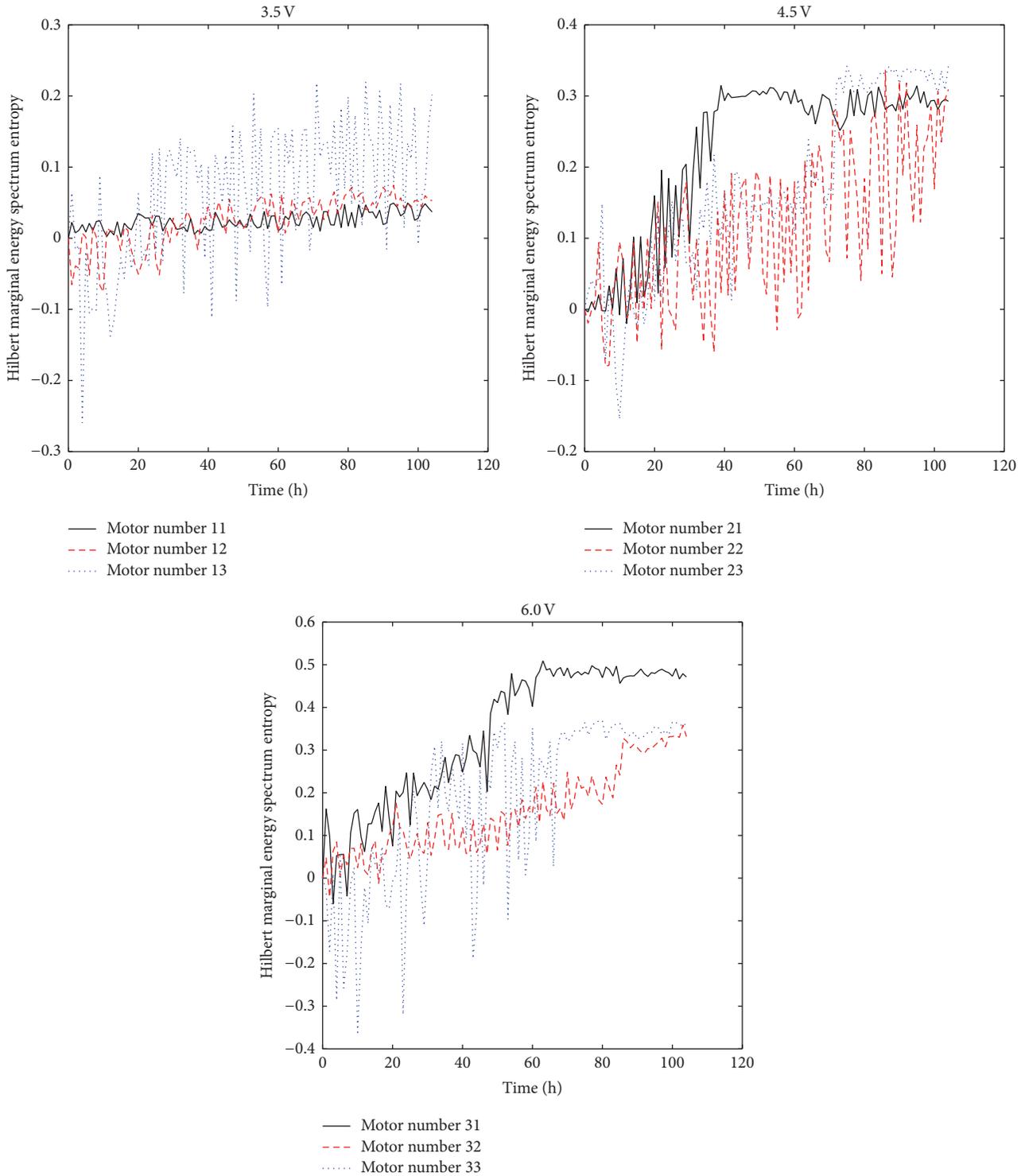


FIGURE 4: The CSADT data of Hilbert marginal energy spectrum entropy.

will significantly decrease the availability and safety of such machines and even lead to big economic losses.

To assess the life and reliability of PZB131 brushed DC motors, a dedicated CSADT platform was designed as shown in Figure 2. The platform is composed of DC power, electric motors, vibration sensors, charge amplifier, data acquisition

card, and computer. The key failure mechanism of this type of DC motors is the wear-out between brush and commutator, which is mainly related to the rotating speed and load. To speed up the degradation process, the input voltage was employed as the accelerated stress that controls the rotating speed of motors under a nonload condition.

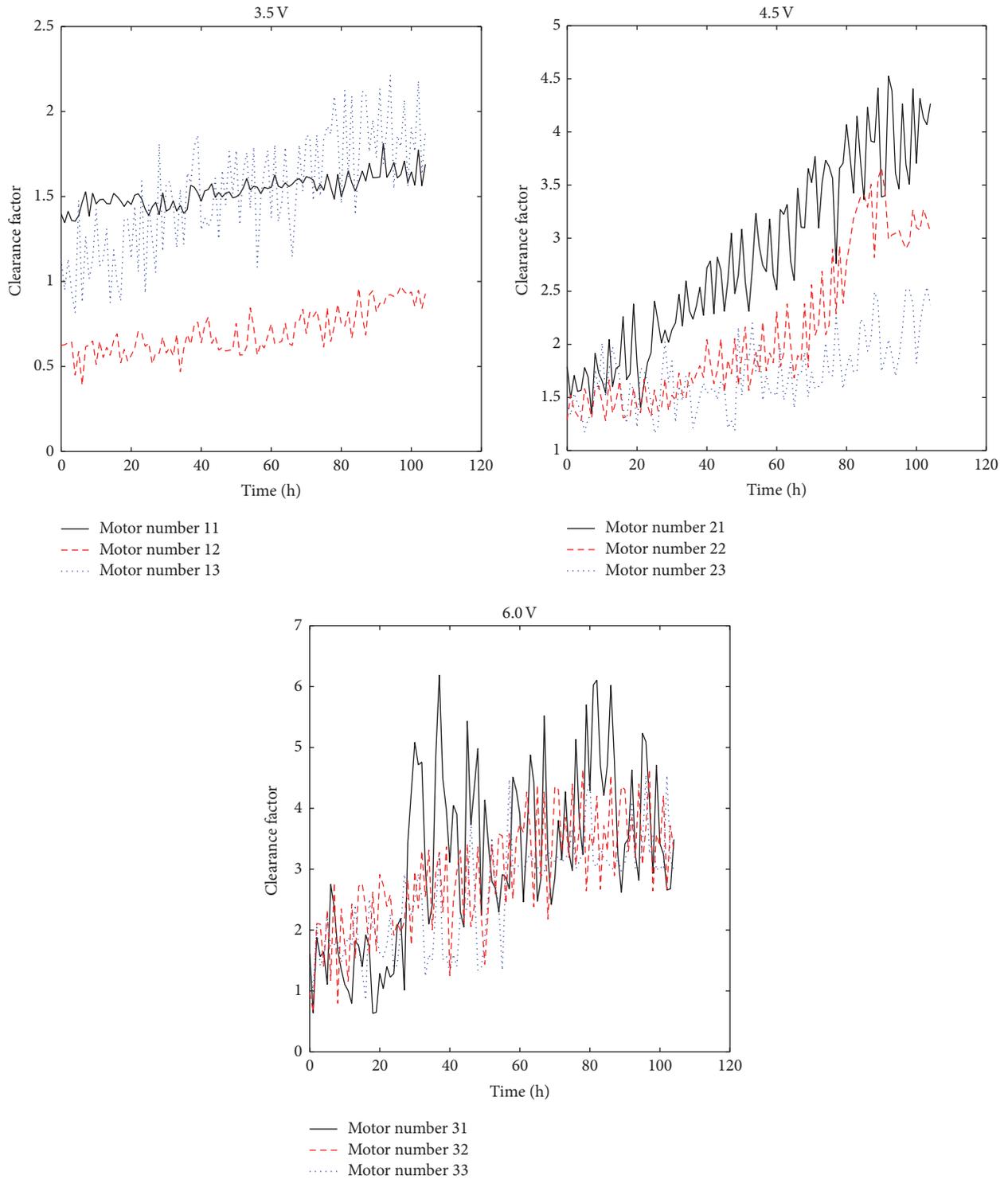


FIGURE 5: The CSADT data of clearance factor.

The rated voltage of the motor is 2.0 V, and the accelerated stress levels are 3.5 V, 4.5 V, and 6.0 V. The sample size under each stress level was 3. The vibration signals on the surface of motor were collected every hour. The sampling time is 2 seconds and the total number of sampling points is 3000.

5.2. Feature Extraction and Selection. First, denoising of raw vibration signals was performed using the Daubechies (db8) wavelet method based on the principle of Stein's unbiased risk with soft-thresholding [24]. Ten features of each motor were extracted using the signal processing methods addressed in

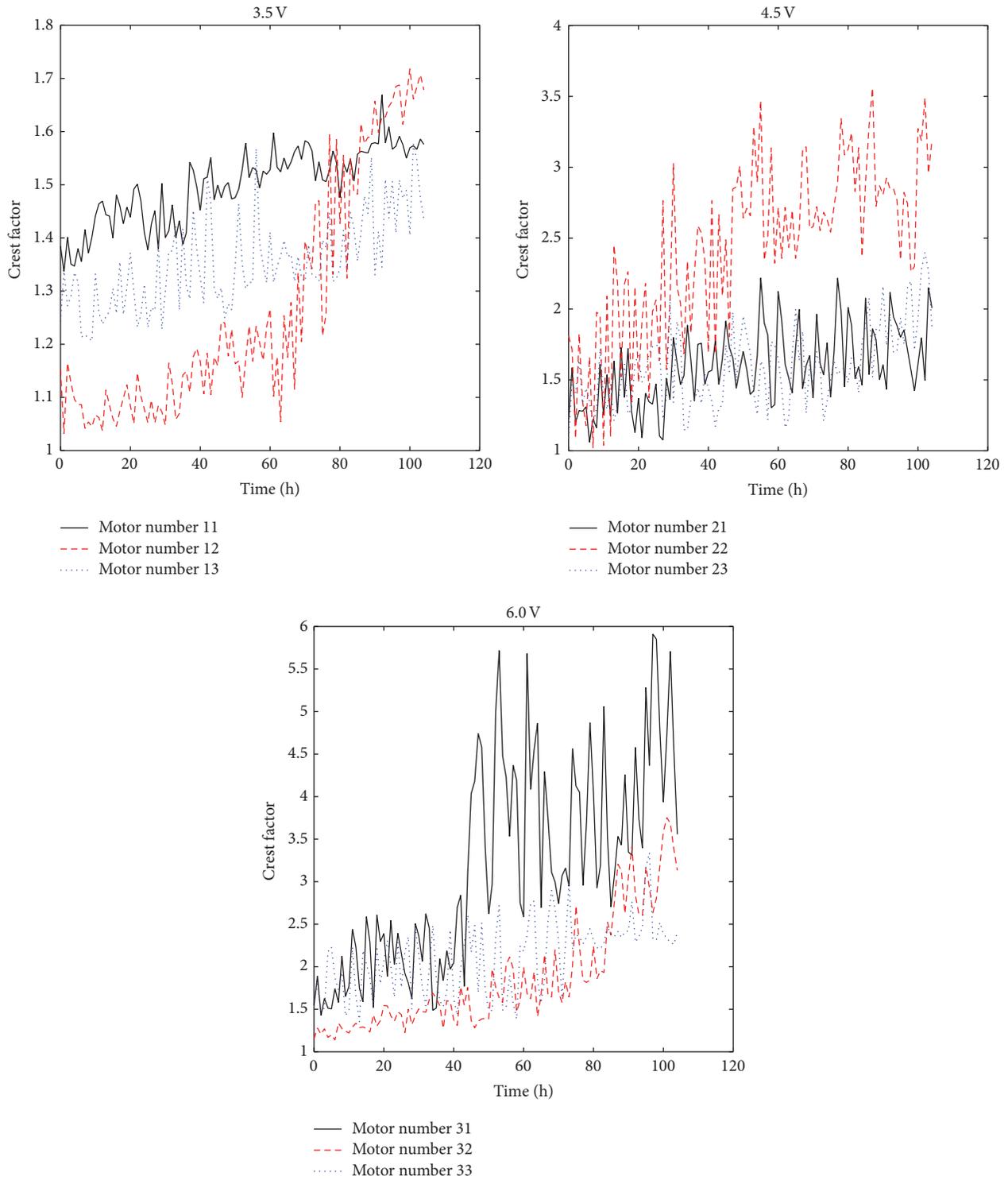


FIGURE 6: The CSADT data of crest factor.

Section 3.1. Figure 3 shows all the candidate features collected from one of the motors.

Afterwards, feature selection was implemented using the proposed feature selecting procedure given in Section 3.2, and the most suitable features were selected from the ten

candidate features. Because prognostic features showing significant monotonic trends are more desired [7], different weights, 0.3, 0.6, and 0.1, were given to trendability, monotonicity, and robustness, respectively. The feature selection results are presented in Table 3.

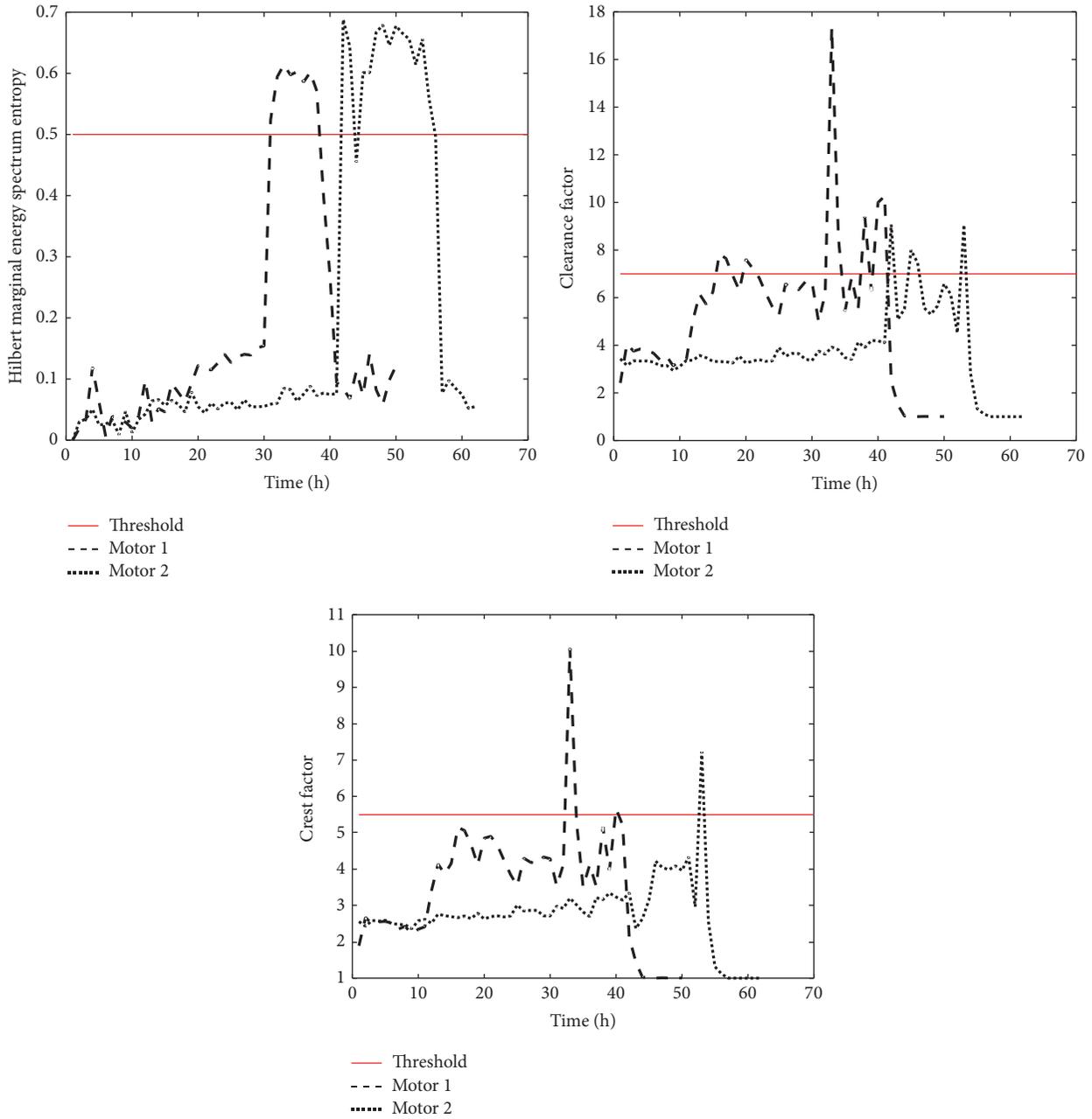


FIGURE 7: Determination of the failure threshold.

Because the clearance factor, crest factor, and HHT marginal energy spectrum entropy (HMESE) fitness were all greater than 0.7, they were selected as the degradation indicators for the motors. The degradation processes of three selected optimal features are illustrated in Figures 4–6.

5.3. Reliability Modeling. Defining the failure threshold for each feature is the first step for degradation-based reliability modeling. To this end, two motors were tested until failure, and the failure thresholds were determined by analyzing the clearance factor, Hilbert marginal energy spectrum entropy, and crest factor of the motors upon failure. As can be seen in Figure 7, the relative failure threshold $D\text{-}y_0$ of Hilbert

marginal energy spectrum entropy, clearance factor, and crest factor is 0.5, 3, and 3, respectively.

Then, the univariate accelerated degradation models for the three features were established, respectively, using the Brownian motion models addressed in Section 4.1. The model parameters were estimated using the method given in Section 4.4.1. Figure 8 shows the reliability functions of the three features calculated by (11). One can see that the reliability curve of crest factor tends to be higher than those of the other two features.

Next, the multivariate ADT model was developed using the copula function that describes the dependence among three features, and the IFM method was used to estimate

TABLE 3: Feature evaluation results.

No.	Feature	Stress	T	M	R	Fitness	Mean (f)	Rank
1	Mean	3.5 V	0.2491	0.5556	0.9500	0.5425	0.5102	8
		4.5 V	0.1616	0.3333	0.9839	0.3469		
		6.0 V	0.7032	0.5556	0.9682	0.6411		
2	Variance	3.5 V	0.6042	0.5556	0.8549	0.6001	0.5350	7
		4.5 V	0.4839	0.3333	0.7852	0.4237		
		6.0 V	0.5634	0.5556	0.7892	0.5813		
3	Crest factor	3.5 V	0.8614	0.5556	0.9808	0.6898	0.7509	3
		4.5 V	0.7359	1	0.9263	0.9134		
		6.0 V	0.7430	0.5556	0.9319	0.6494		
4	Kurtosis factor	3.5 V	0.3349	0.5556	0.9438	0.5282	0.4236	10
		4.5 V	0.3380	0.3333	0.7947	0.3809		
		6.0 V	0.3179	0.3333	0.6620	0.3616		
5	Waveform factor	3.5 V	0.5856	0.3333	0.9987	0.4756	0.6896	4
		4.5 V	0.7208	0.7778	0.9962	0.7825		
		6.0 V	0.8182	0.7778	0.9858	0.8107		
6	Clearance factor	3.5 V	0.8603	1	0.9436	0.9525	0.8961	1
		4.5 V	0.8570	1	0.9460	0.9517		
		6.0 V	0.7649	0.7778	0.8804	0.7842		
7	Frequency center	3.5 V	0.6258	0.5556	0.8631	0.6074	0.6260	5
		4.5 V	0.6474	0.5556	0.7387	0.6014		
		6.0 V	0.8213	0.5556	0.8940	0.6691		
8	Frequency variance	3.5 V	0.6124	0.5556	0.8679	0.6039	0.6194	6
		4.5 V	0.6907	0.5556	0.7590	0.6164		
		6.0 V	0.7157	0.5556	0.8978	0.6378		
9	HHT marginal energy (HME)	3.5 V	0.5675	0.5556	0.7430	0.5779	0.5057	9
		4.5 V	0.4637	0.5556	0.8415	0.5566		
		6.0 V	0.3360	0.3333	0.8192	0.3827		
10	HHT marginal energy spectrum entropy (HMESE)	3.5 V	0.7513	0.7778	0.9545	0.7875	0.8569	2
		4.5 V	0.8404	1	0.9643	0.9486		
		6.0 V	0.9094	0.7778	0.9518	0.8347		

the copula parameters. Figure 9 shows the system reliability curves of the motor using different copulas. To determine the best copula function, the AIC values of the five candidate copula families were obtained and compared. From Table 4, one can see that the AIC value of Frank copula function is the minimum, so it was adopted as the best choice among the candidates.

Figure 10 shows the marginal reliability and system reliability functions under both s -dependent and s -independent assumptions. One can see that the reliability function by assuming s -independent degradation features tends to be much lower than that based on the s -dependent assumption. Clearly, the dependence among the multiple features cannot be ignored.

6. Conclusions

This study is focused on the reliability modeling of vibration-based ADT data. The proposed framework for data analysis

TABLE 4: Goodness-of-fit for the candidate copulas.

Copula family	AIC value	Ranking
Gaussian copula	42961	5
t -copula	-2278	2
Clayton copula	-1179	3
Frank copula	-17136	1
Joe copula	-899	4

focuses on feature selection and multivariate accelerated degradation modeling. A set of feature selection criteria were used to select the most suitable features after extracting features from raw vibration signals. Then, a multivariate s -dependent ADT model was presented, where Brownian motion was applied to establish accelerated degradation model for each feature and a copula function was employed to characterize the dependence among multiple degradation

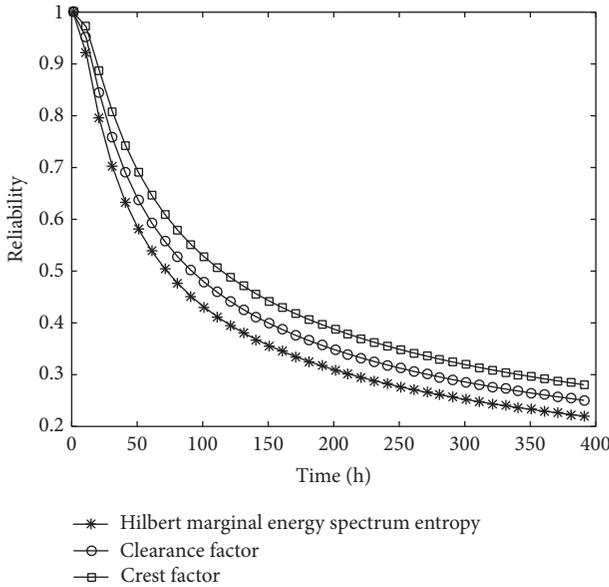


FIGURE 8: Reliability of three features under normal operating condition.

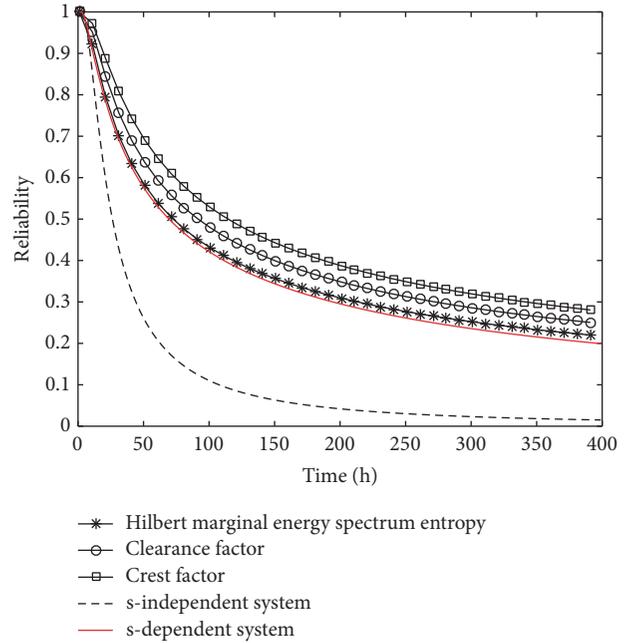


FIGURE 10: Comparison of the system reliability.

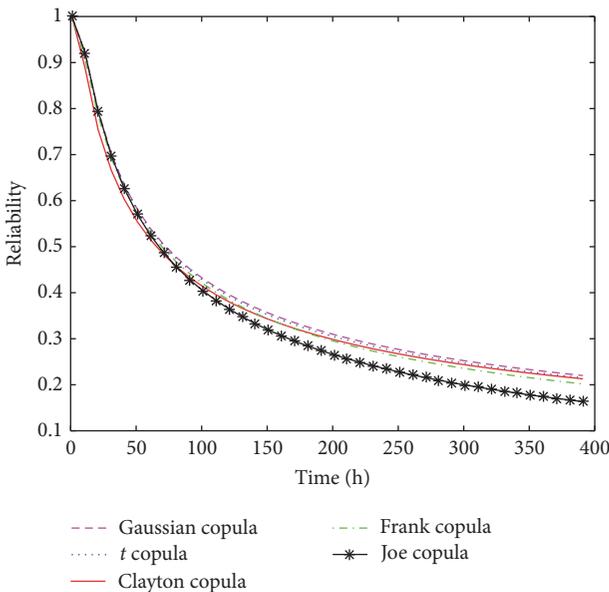


FIGURE 9: Comparison of different copula functions.

processes. A case study on motor vibration-based CSADT data was conducted to demonstrate the usefulness and validity of the proposed model and statistical method. The results show that the proposed framework can provide a powerful tool for modeling the reliability of products exhibiting multiple dependent degradation features.

Competing Interests

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

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China (Grant nos. 61603018 and 61104182) and the China Scholarship Council. The work of Dr. Haitao Liao was partially supported by the U.S. National Science Foundation under Grant #CMMI-1238304.

References

- [1] S. Kumar and M. Pecht, "Modeling approaches for prognostics and health management of electronics," *International Journal of Performability Engineering*, vol. 6, no. 5, pp. 467–476, 2010.
- [2] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems—reviews, methodology and applications," *Mechanical Systems and Signal Processing*, vol. 42, no. 1-2, pp. 314–334, 2014.
- [3] P. Nectoux, R. Gouriveau, K. Medjaher et al., "PRONOSTIA: an experimental platform for bearings accelerated degradation tests," in *Proceedings of the IEEE International Conference on Prognostics and Health Management (PHM '12)*, pp. 1–8, Denver, Colo, USA, 2012.
- [4] L. Liu, X.-Y. Li, W. Zhang, and T.-M. Jiang, "Fuzzy reliability prediction of rotating machinery product with accelerated testing data," *Journal of Vibroengineering*, vol. 17, no. 8, pp. 4193–4210, 2015.
- [5] J. Lee, J. Ni, D. Djurdjanovic, H. Qiu, and H. Liao, "Intelligent prognostics tools and e-maintenance," *Computers in Industry*, vol. 57, no. 6, pp. 476–489, 2006.
- [6] J. Coble and J. W. Hines, "Applying the general path model to estimation of remaining useful life," *International Journal of Prognostics and Health Management*, vol. 2, no. 1, pp. 71–82, 2011.
- [7] B. Zhang, L. Zhang, and J. Xu, "Degradation feature selection for remaining useful life prediction of rolling element bearings," *Quality and Reliability Engineering International*, vol. 32, no. 2, pp. 547–554, 2016.

- [8] L. Liao, "Discovering prognostic features using genetic programming in remaining useful life prediction," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2464–2472, 2014.
- [9] K. Javed, R. Gouriveau, N. Zerhouni, and P. Nectoux, "Enabling health monitoring approach based on vibration data for accurate prognostics," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 647–656, 2015.
- [10] F. Camci, K. Medjaher, N. Zerhouni, and P. Nectoux, "Feature evaluation for effective bearing prognostics," *Quality and Reliability Engineering International*, vol. 29, no. 4, pp. 477–486, 2013.
- [11] K. Yang and G. Yang, "Performance degradation analysis using principal component method," in *Proceedings of the Annual Reliability and Maintainability Symposium*, pp. 136–141, Philadelphia, Pa, USA, January 1997.
- [12] S. Lu, H. Lu, and W. J. Kolarik, "Multivariate performance reliability prediction in real-time," *Reliability Engineering & System Safety*, vol. 72, no. 1, pp. 39–45, 2001.
- [13] R. B. Nelsen, *An Introduction to Copulas*, Springer Science & Business Media, New York, NY, USA, 2nd edition, 2007.
- [14] J. K. Sari, M. J. Newby, A. C. Brombacher, and L. C. Tang, "Bivariate constant stress degradation model: LED lighting system reliability estimation with two-stage modelling," *Quality and Reliability Engineering International*, vol. 25, no. 8, pp. 1067–1084, 2009.
- [15] Z. Q. Pan, N. Balakrishnan, and Q. Sun, "Bivariate constant-stress accelerated degradation model and inference," *Communications in Statistics—Simulation and Computation*, vol. 40, no. 2, pp. 247–257, 2011.
- [16] W. Peng, Y.-F. Li, Y.-J. Yang, S.-P. Zhu, and H.-Z. Huang, "Bivariate analysis of incomplete degradation observations based on inverse Gaussian processes and copulas," *IEEE Transactions on Reliability*, vol. 65, no. 2, pp. 624–639, 2016.
- [17] Z. Liu, X. Ma, J. Yang, and Y. Zhao, "Reliability modeling for systems with multiple degradation processes using inverse Gaussian process and copulas," *Mathematical Problems in Engineering*, vol. 2014, Article ID 829597, 10 pages, 2014.
- [18] Y. Wang and H. Pham, "Modeling the dependent competing risks with multiple degradation processes and random shock using time-varying copulas," *IEEE Transactions on Reliability*, vol. 61, no. 1, pp. 13–22, 2012.
- [19] X. Li and P. Xue, "Multivariate storage degradation modeling based on copula function," *Advances in Mechanical Engineering*, vol. 2014, Article ID 503407, 7 pages, 2014.
- [20] Z.-S. Ye and M. Xie, "Stochastic modelling and analysis of degradation for highly reliable products," *Applied Stochastic Models in Business and Industry*, vol. 31, no. 1, pp. 16–32, 2015.
- [21] H. Liao and E. A. Elsayed, "Reliability inference for field conditions from accelerated degradation testing," *Naval Research Logistics*, vol. 53, no. 6, pp. 576–587, 2006.
- [22] Q. Guan, Y. C. Tang, and A. C. Xu, "Objective Bayesian analysis accelerated degradation test based on Wiener process models," *Applied Mathematical Modelling*, vol. 40, no. 4, pp. 2743–2755, 2016.
- [23] J. Lin, M. J. Zuo, and K. R. Fyfe, "Mechanical fault detection based on the wavelet de-noising technique," *Journal of Vibration and Acoustics*, vol. 126, no. 1, pp. 9–16, 2004.
- [24] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.
- [25] D. Goyal and B. S. Pabla, "Condition based maintenance of machine tools—a review," *CIRP Journal of Manufacturing Science and Technology*, vol. 10, pp. 24–35, 2015.
- [26] N. E. Huang and Z. Wu, "A review on Hilbert-Huang transform: method and its applications to geophysical studies," *Reviews of Geophysics*, vol. 46, no. 2, Article ID RG2006, 2008.
- [27] L. A. Escobar and W. Q. Meeker, "A review of accelerated test models," *Statistical Science*, vol. 21, no. 4, pp. 552–577, 2006.
- [28] G. K. Bhattacharyya and A. Fries, "Fatigue failure models—Birnbbaum-Saunders vs. Inverse Gaussian," *IEEE Transactions on Reliability*, vol. 31, no. 5, pp. 439–441, 1982.
- [29] A. Sklar, "Fonctions de répartition à n dimensions et leurs marges," *Publications de l'Institut de Statistique de l'Université de Paris*, no. 8, pp. 229–231, 1959.
- [30] H. Joe, *Multivariate Models and Multivariate Dependence Concepts*, CRC Press, Boca Raton, Fla, USA, 1st edition, 1997.
- [31] H. Joe, "Asymptotic efficiency of the two-stage estimation method for copula-based models," *Journal of Multivariate Analysis*, vol. 94, no. 2, pp. 401–419, 2005.

Research Article

Research of Fault Diagnosis Based on Sensitive Intrinsic Mode Function Selection of EEMD and Adaptive Stochastic Resonance

Zhixing Li and Boqiang Shi

School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China

Correspondence should be addressed to Boqiang Shi; shiboqiang@ustb.edu.cn

Received 10 July 2016; Accepted 12 October 2016

Academic Editor: A. El Sinawi

Copyright © 2016 Z. Li and B. Shi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A novel methodology for the fault diagnosis of rolling bearing in strong background noise, based on sensitive intrinsic mode functions (IMFs) selection of ensemble empirical mode decomposition (EEMD) and adaptive stochastic resonance, is proposed. The original vibration signal is decomposed into a group of IMFs and a residual trend item by EEMD. Constructing weighted kurtosis index difference spectrum (WKIDS) to adaptively select sensitive IMFs, this method can overcome the shortcomings of the existing methods such as subjective choice or need to determine a threshold using the correlation coefficient. To further reduce noise and enhance weak characteristics, the adaptive stochastic resonance is employed to amplify each sensitive IMF. Then, the ensemble average is used to eliminate the stochastic noise. The simulation and rolling element bearing experiment with an inner fault are performed to validate the proposed method. The results show that the proposed method not only overcomes the difficulty of choosing sensitive IMFs, but also, combined with adaptive stochastic resonance, can better enhance the weak fault characteristics. Moreover, the proposed method is better than EEMD and adaptive stochastic resonance of each sensitive IMF, demonstrating the feasibility of the proposed method in highly noisy environments.

1. Introduction

Rolling bearings are widely used in large or small mechanical equipment [1, 2]. In severe cases when rolling bearings early break down during serving the machines, the machines can be damaged. In fact, the unpredictable failures may cause serious damage to machinery and equipment, so the extraction of early defects is very important to ensure the reliable operation of the machinery [3]. It is difficult for the vibration signal of the rolling bearing which is nonstationary and strongly modulated to detect weak fault characteristics under strong background noise [4]. The traditional signal processing methods are mainly to reduce or eliminate the noise. Empirical mode decomposition method is an effective tool for processing the complex nonstationary signal, but it has a disadvantage which is the mode mixing [5]. Later, Wu and Huang [6] have proposed ensemble empirical mode decomposition (EEMD) by introducing the additional noise, and this method has overcome the shortcoming of the traditional empirical mode decomposition to some extent.

The complex nonstationary vibration signal can be decomposed into many intrinsic mode functions (IMFs), and these functions can reflect the real physical information in the vibration signals. However, some of these IMFs are useless or interfering components and the others are sensitive components to fault characteristics. Many researchers extracted the fault features by selecting the effective and sensitive IMFs to the mechanical faults, thereby eliminating the noise components or the unrelated-fault components. Hu et al. [7, 8] proposed a method of selecting the sensitive IMFs by using the largest kurtosis index; many kurtosis values may be greater than 3; the maximum kurtosis value corresponding to IMFs may lose part of the fault information. Tan et al. [9] proposed a method to select sensitive IMFs by using correlation coefficient, but the method needs to determine a threshold to select sensitive IMFs components. According to different signals, the threshold value required to be adjusted artificially. In order to overcome the above shortcomings, this paper proposed a method that constructs weighted kurtosis index difference spectrum for selecting effective

and sensitive IMFs to the faults. However, in the strong background noise, even if the effective IMFs are obtained, the fault feature is still very weak, which cannot identify the fault characteristics effectively. For the fault diagnosis of weak signal, stochastic resonance is widely used for its unique advantages. Stochastic resonance was first proposed in 1981 in the study of Earth's glaciers and it has attracted a lot of attention in the past thirty years, especially in denoising and signal enhancement [10]. When the useful signal and the noise are mixed, the traditional denoising methods will damage the useful signal and lose the vibration features. At the same time, the traditional methods cannot extract the fault features where the signal-to-noise ratio is less than -15 dB. Stochastic resonance is used to enhance the weak fault characteristics and reduce the noise, but the performance of stochastic resonance methods is mostly decided by system parameters of the potential function of stochastic resonance [11]. The existing stochastic resonance methods ignore the interactive effect between parameters through subjectively selecting parameters or optimizing only one parameter [12]. Zhang et al. [13] extended the frequency range of stochastic resonance by scaling transform and applied it to the fault characteristic detection of bearing outer race. In addition, multiple parameters of stochastic resonance system need to be optimized to ensure the effectiveness of stochastic resonance. This paper quoted the method using the ant colony algorithm to parallelly optimize multiple parameters of stochastic resonance [14–16]. Therefore, combining the advantages of EEMD reduced noise and stochastic resonance enhanced weak fault characteristics, this paper proposes a fault diagnosis method based on sensitive IMFs selection of EEMD and adaptive stochastic resonance. Firstly, the original vibration signal is decomposed by EEMD. The modulated vibration signal is demodulated effectively and the frequency features of vibration signal are decomposed into several IMFs components. These obtained IMFs components are selected by weighted kurtosis index difference spectrum (WKIDS). Finally, the selected sensitive IMFs are input into the stochastic resonance system. In order to further reduce noise, the enhanced sensitive IMFs via stochastic resonance are ensemble-averaged. Finally, the weak fault characteristics are realized in the strong background noise, and the validity of the sensitive IMFs selection of EEMD and adaptive stochastic resonance method is verified by simulation and experiment.

The other parts of this paper are as follows. Section 2 introduces the basic principles of EEMD and stochastic resonance. Section 3 describes the method for the sensitive IMFs selection of EEMD and adaptive stochastic resonance. Constructing WKIDS, the selected IMFs are input into adaptive stochastic resonance. To reduce noise further and enhance the weak signal and the adaptive stochastic resonance of each sensitive IMF, ensemble average is used to eliminate the stochastic noise. The effectiveness of the proposed method is verified by the periodic pulse signal simulation in Section 4. The inner ring fault of bearing is detected by the proposed method in Section 5. Finally, the conclusion is given in Section 6.

2. Principle Analyses

2.1. Ensemble Empirical Mode Decomposition. Huang et al. [17, 18] proposed empirical mode decomposition to decompose any complicated signal, which has been widely studied and applied in the fault diagnosis field. However, the empirical mode decomposition method still needs improvement in some areas. An important limitation is the mode mixing problem. Recently, Wu and Huang [6] proposed a new method (EEMD) to solve the mode mixing which can restore the nature of the original signal. It is known to be a great improvement of empirical mode decomposition and is widely used in the machinery field [19].

The principle of EEMD method used the Gaussian white noise with a uniform frequency distribution of the statistical characteristics; when the signal is added to the Gaussian white noise, the signal will be continued at different scales. In order to reduce the degree of mode mixing, the specific decomposition steps and principles are shown as follows.

Step 1. In the original signal, $x(t)$ is added to the Gaussian white noise $n_i(t)$ with the amplitude mean value of 0, and the standard deviation is constant. Hence,

$$x_i(t) = x(t) + n_i(t). \quad (1)$$

In the formula, $x_i(t)$ represents the Gaussian white noise signal that is added for i times, and the range size of the Gaussian white noise will directly affect the signal EEMD to avoid the decomposition effect of the mode mixing.

Step 2. $x_i(t)$ is decomposed by empirical mode decomposition and hence we get the IMF component $c_{ij}(t)$ and the remainder $r(t)$. $c_{ij}(t)$ indicates that, after the i times, Gaussian white noise is added and the j th component is obtained.

Step 3. Repeat Steps 1 and 2 N times. Using the principle of uncorrelated random sequence with the statistical mean value of 0, the IMF corresponding to the above will have total average operation and will eliminate the influence on IMF while the white Gaussian noise is added repeatedly. Finally, the IMFs are obtained after EEMD:

$$c_j(t) = \frac{1}{N} \sum_{i=1}^N c_{ij}. \quad (2)$$

In formula (2), $c_j(t)$ represents the notion that the j th IMF component is obtained from the EEMD of the original signal. The greater the value of N is, the more the IMFs of Gaussian white noise tend to be 0. So, the results of EEMD are as follows:

$$x(t) = \sum_j c_j(t) + r(t). \quad (3)$$

$r(t)$ is the final residual component and represents the average trend of the signal. The EEMD method can decompose any signal into the sum of a set of IMFs and a residual component. The intrinsic mode component $c_j(t)$ ($j = 1, 2, \dots$) represents a component of a signal from high to low frequency range.

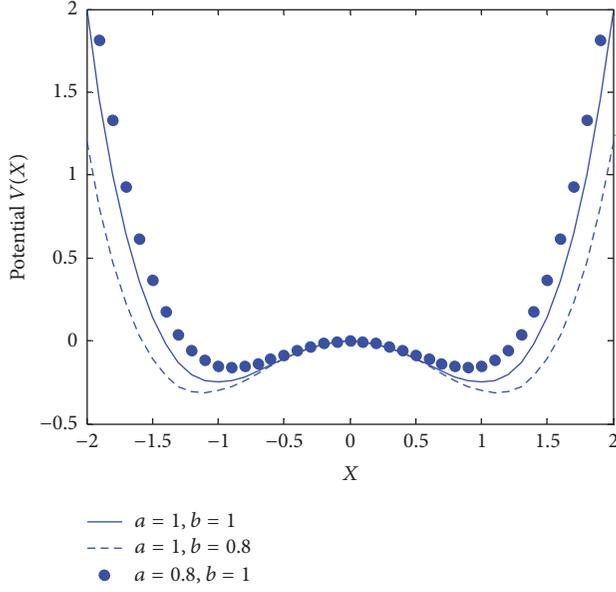


FIGURE 1: The potential function in different cases.

Each frequency band contains different frequency components and may vary with the change of the vibration signal $x(t)$.

2.2. Model of Stochastic Resonance. Nowadays, the bistable system is widely studied which is driven by the Gaussian white noise and external periodic signal. This bistable system can be defined as follows:

$$\begin{aligned} \frac{dx}{dt} &= \frac{-\partial V(x)}{\partial x} = S(t) + N(t), \\ \langle N(t) \rangle &= 0, \\ \langle N(t), N(0) \rangle &= 2D\delta(t), \end{aligned} \quad (4)$$

where $V(x)$ is potential function of nonlinear bistable system

$$V(x) = -\frac{a}{2}x^2 + \frac{b}{4}x^4, \quad a > 0, \quad b > 0, \quad (5)$$

where periodic external force $S(t) = A \cos(\omega t)$ and $N(t)$ is Gaussian white noise in which the intensity is D and the mean value is 0. Accordingly, we can obtain (6) by operating (4) and (5):

$$\frac{dx}{dt} = ax - bx^3 + S(t) + N(t), \quad (6)$$

where x is the output signal of stochastic resonance system and a and b are the parameters of stochastic resonance system; meanwhile, the system parameters satisfy $a > 0$, $b > 0$. This equation indicates that, in two potential wells, the overdamped motion of the particle is driven by periodic signal and noise. The potential function $V(x)$ has two stable fixed points at $x = \pm\sqrt{a/b}$, the height of potential barrier is $\Delta V(x) = a^2/4b$, and the barrier top is located at $x = 0$. To be specific, we choose $a = b = 1$; $a = 0.8, b = 1$; $a = 1, b = 0.8$. Figure 1 shows the different states of potential functions

$V(x)$ at different stochastic resonance system parameters a and b . It is clear that two potential wells are symmetric and separated by a barrier with the height changes with different system parameters a and b , so as to achieve the best stochastic resonance [20, 21]. Hence, we can achieve the aim of detecting weak signals successfully by adjusting the system parameters [22, 23].

3. The Proposed Diagnosis Method Based on Sensitive IMFs Selection and Stochastic Resonance

3.1. Selection of Sensitive IMFs. Kurtosis index is very sensitive to the fault characteristic in the strong background noise. Generally, kurtosis index is used for measuring the impulsive characteristics of the vibration signals. Kurtosis index formula is written as [24]

$$K = \frac{E(r^4)}{E^2(r^2)} = \frac{(1/T) \sum_{t=1}^T (r_t - \bar{r})^4}{\sigma^4}. \quad (7)$$

In the formula, $r = (r_1, r_2, \dots, r_t, \dots, r_T)$ is the signal sequence, T is the signal length, \bar{r} is the mean value of the signal, and σ is the standard deviation of the signal

$$\sigma = \sqrt{\frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^2}. \quad (8)$$

Equation (7) turns to

$$K = \frac{T^2 \sum_{t=1}^T (r_t - \bar{r})^4}{T \left(\sum_{t=1}^T (r_t - \bar{r})^2 \right)^2} = \frac{TM_1}{M_2}, \quad (9)$$

$$M_1 = \sum_{t=1}^T (r_t - \bar{r})^4, \quad (10)$$

$$M_2 = \left(\sum_{t=1}^T (r_t - \bar{r})^2 \right)^2, \quad (11)$$

$$M_2 - M_1 = 2 \sum_{t=1}^T (r_t - \bar{r})^2 \sum_{i=1, i \neq t}^T (r_i - \bar{r})^2 = P(r). \quad (12)$$

Because \bar{r} is the mean value of the signal, when the pulse signal increases or decreases, it will have little influence on \bar{r} . When the pulse signal increases, $P(r)$ increases and thus K decreases. Similarly, when the pulse signal decreases, K increases. When there are multiple pulse signals in the strong background noise, one or more of the pulse signals are hidden in the noise that may not be detected; K does not decrease but increases. The loss of the pulse signal is called “leakage peak phenomenon.” Therefore, kurtosis index curve cannot reflect all the information of IMFs.

C represents the correlation of the two signals. The more the similarity they have, the greater the value of C . The cross-correlation coefficient between signals $x(t)$ and $y(t)$, C , is defined as

$$C = \frac{\sum_{n=0}^T (x(t) - \bar{x})(y(t) - \bar{y})}{\left[\sum_{n=0}^T (x(t) - \bar{x}) \sum_{n=0}^T (y(t) - \bar{y})\right]^{1/2}}. \quad (13)$$

Because the fault vibration signal $x(t)$ is decomposed into IMFs components by EEMD, each IMF component x_{IMF} is similar to $x(t)$, and their cross-correlation coefficient is C_1 . x_{IMF} is also similar to normal working condition signal $x_{\text{nor}}(t)$, and each IMF component and normal working condition signal's cross-correlation coefficient is C_2 . C_1 and C_2 are defined as

$$\begin{aligned} C_1 &= \frac{\sum_{n=0}^T (x(t) - \bar{x})(x_{\text{IMF}}(t) - \bar{x}_{\text{IMF}})}{\left[\sum_{n=0}^T (x(t) - \bar{x}) \sum_{n=0}^T (x_{\text{IMF}}(t) - \bar{x}_{\text{IMF}})\right]^{1/2}}, \\ C_2 &= \frac{\sum_{n=0}^T (x_{\text{nor}}(t) - \bar{x}_{\text{nor}})(x_{\text{IMF}}(t) - \bar{x}_{\text{IMF}})}{\left[\sum_{n=0}^T (x_{\text{nor}}(t) - \bar{x}_{\text{nor}}) \sum_{n=0}^T (x_{\text{IMF}}(t) - \bar{x}_{\text{IMF}})\right]^{1/2}}. \end{aligned} \quad (14)$$

Combined with the two cross-correlation coefficients, a new cross-correlation coefficient $C_K = C_1 - C_2$ is obtained which is associated with fault. C_K is considered not only the similarity between each IMF and the normal signal, but also the similarity of the fault vibration signal. Because of the absolute value C_K which is a real number and is no more than 1, it has the potential to be the weight.

In the strong background noise and multiple pulse signals, one or more pulse signals are hidden by the noise, and K increases. However, the similarity of pulse signals associated with fault is reduced; that is, C_K decreases, and the decrease of C_K corrects the increase of K . Therefore, this paper uses C_K as the weight of K to construct weighted kurtosis index K_W which is defined as

$$K_W = C_K K. \quad (15)$$

The weighted kurtosis index can not only prevent the loss of vibration signal, but also ensure the similarity of vibration signal. Hence, the weighted kurtosis index curve reflects all the information of IMFs. Let the sequence of IMFs corresponding to the weighted kurtosis index defined as $S = (K_W^1, K_W^2, \dots, K_W^q)$ lead to the differential spectrum. The equation is $b^l = K_W^l - K_W^{l+1}$ ($l = 1, 2, \dots, q - 1$). Then, all b^l forming a sequence is

$$B = (b^1, b^2, \dots, b^{q-1}). \quad (16)$$

B is named weighted kurtosis index difference spectrum (WKIDS). Difference spectrum is used for describing the changes of the adjacent weighted kurtosis index. When the difference between adjacent weighted kurtosis indexes is large relatively, there is a maximum mutation point b_k in the whole difference spectrum corresponding to the weighted kurtosis

index sequence at k . b_k reflects the biggest difference between the weighted kurtosis index properties. In other words, signals generate a mutation between the useful IMFs signals and the noise signals at b_k . Therefore, a method of selecting effective IMFs signals can be proposed. In the weighted kurtosis index sequence, the points before the mutation point including it, k weighted kurtosis index corresponds to useful IMFs. After the mutation point, the weighted kurtosis index corresponds to useless IMFs. In this sense, the k sensitive IMFs are selected by WKIDS.

3.2. Adaptive Stochastic Resonance. Because the fault characteristics of mechanical equipment are hidden in strong background noise, a large amount of background noise greatly reduces the signal-to-noise ratio. Although k sensitive IMFs are selected by KWIDS method, it is difficult to extract the fault characteristics from them.

Therefore, sensitive IMFs of EEMD are further enhanced by adaptive stochastic resonance method. The traditional stochastic resonance is only applicable to small parameter signal (signal amplitude $A \leq 1$, noise intensity $D \leq 1$, and signal frequency $f \leq 1$). Therefore, the sensitive IMFs before inputting into stochastic resonance system are processed by frequency-shifted and scaling transform to satisfy the small parameter conditions. The principle of the method is to design a frequency compression scale, and the frequency of sensitive IMFs is compressed to determine the calculation step size. This way, the IMFs meet the requirements of small parameters. The system parameters a and b of stochastic resonance have a decisive effect on its performance [25]. In this paper, a good global optimization ability of ant colony algorithm is used to search the optimal parameters of stochastic resonance. The optimal combination of parameters a and b is brought into stochastic resonance and the signal-to-noise ratio of each sensitive IMF is calculated as the objection function of ant colony algorithm. Thus, the enhanced k IMFs are obtained. In order to reduce the stochastic interference in each enhanced IMF, each enhanced IMF was ensemble-averaged, eliminating the effects of noise in multiple IMFs. So, the final result is defined as

$$b(t) = \frac{1}{k} \sum_{i=1}^k b_i. \quad (17)$$

In formula (17), b_i represents the i th effective IMF component.

In the paper, the flowchart of the proposed method is shown in Figure 2. The procedure of the method is described as follows.

(1) The bearing vibration signals are decomposed by EEMD, obtaining N IMFs and a trend term.

(2) A method of WKIDS is proposed; k sensitive IMFs are selected by WKIDS, which are treated as the input of stochastic resonance system.

(3) Initiate the parameters of ant colony algorithm and select evaluation function of stochastic resonance. Here, the evaluation function is set as the signal-to-noise ratio of output signal of stochastic resonance system.

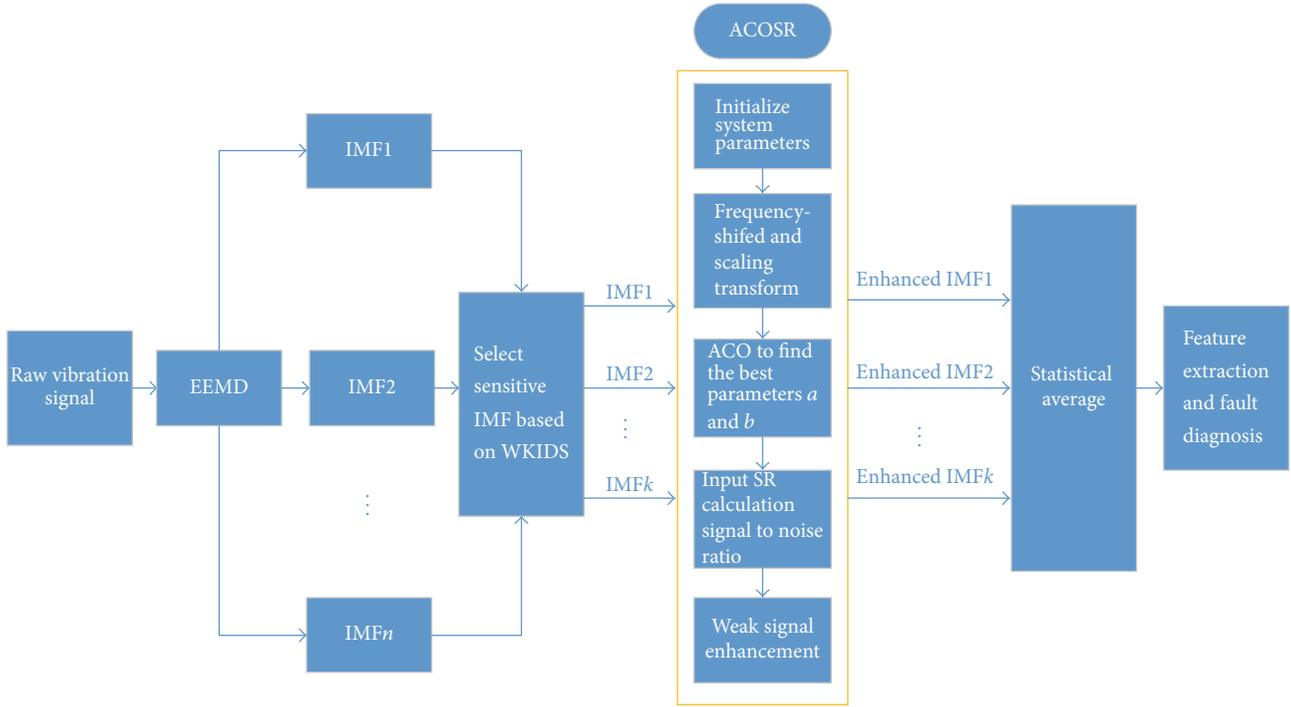


FIGURE 2: The flowchart of the proposed method.

(4) The k sensitive IMFs are initially processed by frequency-shifted and scaling transform to meet the requirements of small parameters.

(5) Employ ant colony algorithm to optimize the two parameters a and b of stochastic resonance and thereby obtain the optimal combination of a and b . The signal-to-noise ratio of k sensitive IMFs is calculated as an evaluation function and further k enhanced IMFs are obtained.

(6) The k enhanced IMFs are ensemble-averaged to get the final enhanced signal, and the fault characteristic frequency is extracted from it.

4. Simulated Verification

In order to validate the effectiveness of the method proposed, periodic pulse of the bearing fault simulation signal is employed in [26]. The sampling frequency of the periodic pulse signal is 10 kHz. The characteristic frequency is 50 Hz, and the sampling time is 1 s. Time-domain waveform and frequency spectrum of the periodic pulse signal are shown in Figure 3.

The noise intensity with Gaussian white noise of 0.5 is added to the periodic pulse signal in order to simulate the operating background of rotating machinery. It is shown in Figure 4(a) that the periodic pulse signal is submerged in strong background noise. Therefore, the simulated signal has certain practical significance.

Due to the fact that the periodic pulse signal is hidden in the strong noise, we cannot obtain the periodic pulse characteristics from the time-domain waveform in Figure 4(a) and the corresponding frequency spectrum in Figure 4(b).

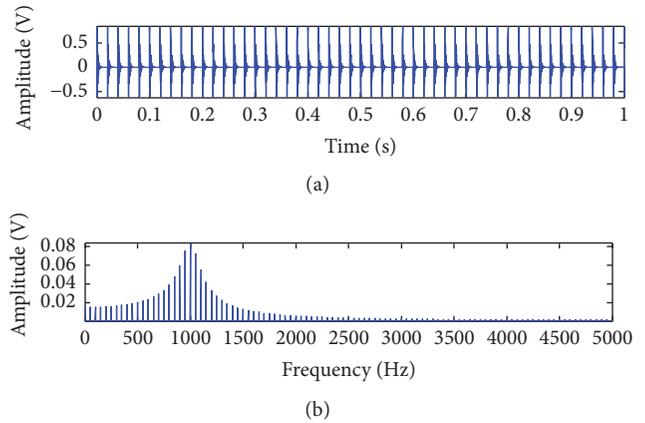


FIGURE 3: The periodic pulse signal: (a) time-domain waveform; (b) frequency spectrum.

We know that the envelope spectrum has a good effect on extracting the fault feature, but the fault characteristics are still hidden in Figure 4(c). Therefore, we conclude that the envelope spectrum cannot extract the characteristic frequency of the noisy signal in the strong background noise.

In order to obtain the characteristic frequency in strong background noise, the original signal is decomposed by EEMD, and then 13 IMFs and a residual trend item are obtained which are shown in Figure 5(a). IMF1~IMF6 are the high and low frequency components, IMF7~IMF9 are the low frequency components, and IMF10~IMF13 are the very low frequency components so that the characteristic

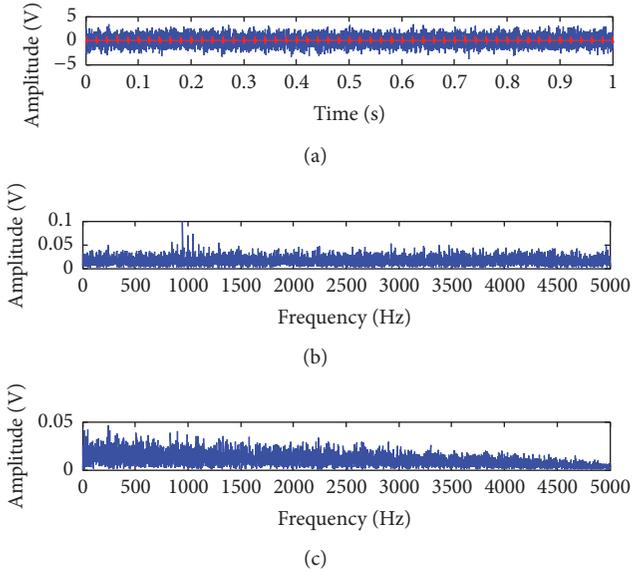


FIGURE 4: The noisy signal: (a) time-domain waveform; (b) frequency spectrum; (c) envelope spectrum.

frequency cannot be seen. The periodic signals cannot be clearly distinguished from the envelope spectrum in Figure 5(b). We use the method of KWIDS to select the sensitive IMFs. Firstly, according to (9), we can get the kurtosis index curves in Figure 6(a). From the graph, multiple kurtosis index values are greater than 3 which may cause the loss of some characteristics. Hence, the weighted kurtosis index is constructed by the cross-correlation coefficient to get all the fault information of the IMFs and the difference spectrum curve by formula (17) in Figure 6(b). Difference spectrum corresponding to IMFs is shown in Table 1, with IMF4 of 0.07206 and the maximum mutation point. Therefore, IMF1~IMF4 reflect the fault characteristics of the raw vibration signal and they are the sensitive IMFs input of adaptive stochastic resonance method.

According to frequency-shifted scaling transform method, the pass frequency and cutoff frequency of the high pass filter are 48 Hz and 40 Hz, respectively. The modulation frequency is 40 Hz and the variable scale compression rate is 400. The target frequency is compressed to $(50 - 40)/400 = 0.025 \leq 1$ Hz which meets the requirements of small parameters. IMF1~IMF4 input signals are input into stochastic resonance system, and two system parameters a and b of stochastic resonance are optimized by ant colony algorithm. Find the optimal parameter combination of adaptive stochastic resonance system in the range of $0 < a < 10$ and $0 < b < 10$, and the output signal-to-noise ratios of IMF1~IMF4 are 20.3957, 20.1430, 21.4371, and 18.5379, respectively, which is an evaluation function in Table 2.

The time-domain and the frequency spectrum of the IMF1~IMF4 components input into adaptive stochastic resonance are shown in Figures 7~10.

Because the adaptive stochastic resonance can enhance the weak fault characteristics buried in noise, the feature frequency of the component signal can be highlighted. The

TABLE 1: The difference spectrum of the simulation signal corresponding to IMFs.

IMF	Difference spectrum
IMF1	0.01753
IMF2	0.04468
IMF3	0.01653
IMF4	0.07206
IMF5	0.001666
IMF6	0.000916
IMF7	$9.036e - 5$
IMF8	0.002096
IMF9	0.000223
IMF10	0.000174
IMF11	0.000633
IMF12	$4.177e - 5$
IMF13	0.000682

adaptive stochastic resonance frequency spectrums of IMF1~IMF3 all have maximum value of 50 Hz in Figures 7(b), 8(b), and 9(b), which are consistent with the actual situation, and thus they have noise interference, where the noise interference of IMF2 is especially serious. In Figure 10, due to the strong noise of the interference, the maximum spectral peak is 51.67 Hz, instead of the target frequency. In order to reduce noise interference, overall average operation is performed on IMF1~IMF4, and the result is shown in Figure 11.

In Figure 11(b), the highest amplitude frequency is 50 Hz, which is the same as the characteristic frequency, and the period of vibration signal is very obvious. In addition, the noise is largely eliminated by using stochastic resonance and statistical average. Compared with EEMD in Figure 5(b) and stochastic resonance of each IMF component in Figures 7(b), 8(b), 9(b), and 10(b), it is obviously superior to all the above. In other words, the EEMD reduces noise but does not extract the features of the fault. Stochastic resonance of each IMF component is used to enhance the weak fault characteristics and this effect is not so perfect. In order to reduce noise further, the ensemble average is used for four enhanced IMFs. Finally, the fault characteristic frequency is extracted effectively and the peak is much higher than the surrounding noise. Therefore, we verified the validity of fault diagnosis method of sensitive IMFs selection of EEMD and adaptive stochastic resonance.

5. Experimental Demonstration

The equipment used in the experiment is a bearing fault test platform made by American spectra quest manufacturing company as shown in Figure 12. In the test, we used ZonicBook/618E type test system and the ER-10K cylindrical roller bearings as the failure bearing. The vibration signal is collected by the acceleration sensor which is mounted on the bearing seat. The cylindrical roller bearing geometry sizes are $n = 8$, $D = 33.5$ mm, $d = 7.9375$ mm, and $\alpha = 0^\circ$. In the test, sampling frequency is 5120 Hz and the speed of the shaft is 2100 r/min.

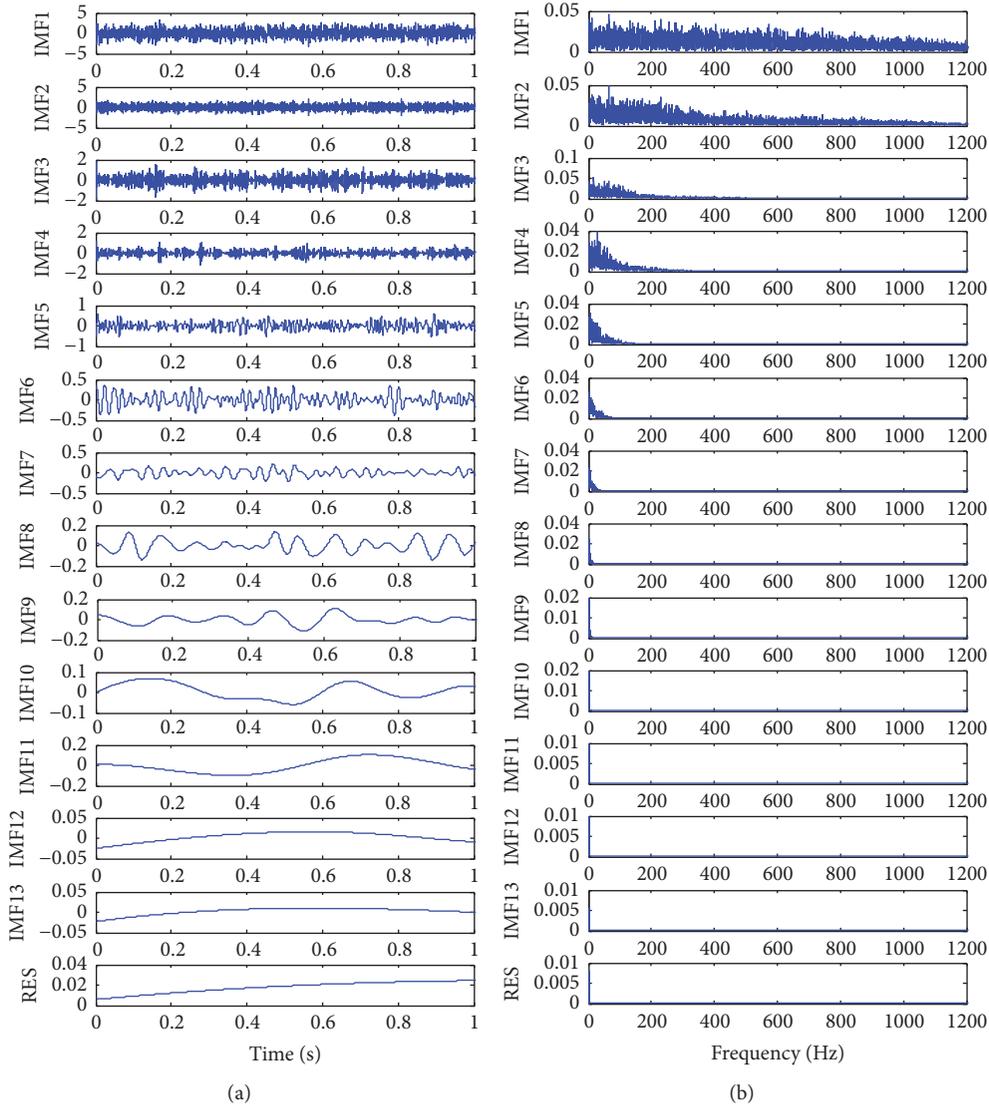


FIGURE 5: EEMD: (a) time-domain waveform; (b) envelope spectrum.

TABLE 2: The optimal parameter combination of adaptive stochastic resonance of IMF1~IMF4.

Parameter	IMF1	IMF2	IMF3	IMF4
a	0.2312	0.1187	0.1035	0.4112
b	9.7880	0.1378	0.543	10.0840
Signal-to-noise ratio	20.3957	20.1430	21.4371	18.5379

According to fault characteristic frequency theory about the rolling bearings, we can know that the inner race fault frequency is 173.25 Hz. Time-domain waveform of the bearing inner race is shown in Figure 13(a). Due to the fault signals being very weak in the strong background noise, it is not easy to find the fault characteristic frequency in the frequency spectrum and envelope spectrum that are shown in Figures 13(b) and 13(c). Therefore, we cannot determine whether the bearing inner race has a fault.

According to the method proposed in this paper, firstly, the raw vibration signal is decomposed by EEMD in which

the time-domain waveform of IMF1~IMF4 is shown in Figure 14(a). The characteristic component of the bearing inner race fault is completely hidden in the background noise with a low signal-to-noise ratio, which cannot detect obvious spectral peaks in the frequency spectrum of IMF1~IMF4 in Figure 14(b), and the fault features cannot be extracted. Therefore, the proposed method is used to select the sensitive IMFs of EEMD and adaptive stochastic resonance.

In order to obtain the sensitive IMFs, constructing KWDIS according to (9) and kurtosis index curve is shown in Figure 15(a). The kurtosis index curve illustrates multiple

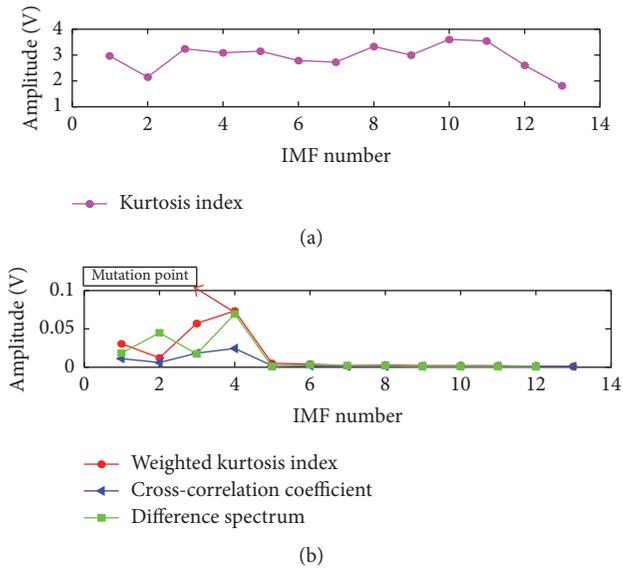


FIGURE 6: WKIDS of the inner race of the bearing: (a) kurtosis index; (b) weighted kurtosis index, cross-correlation coefficient, and difference spectrum.

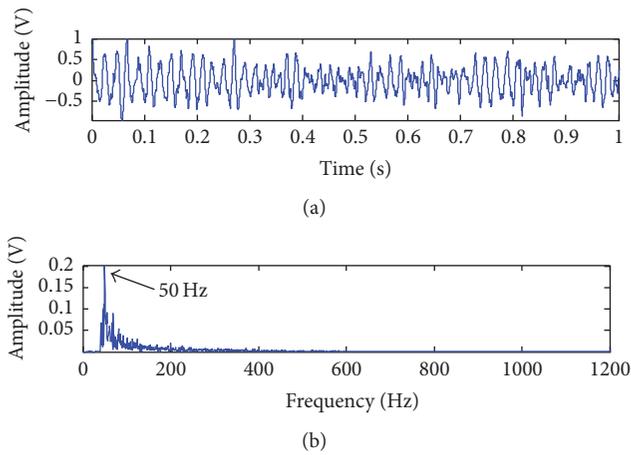


FIGURE 7: Optimal output signal of IMF1: (a) time-domain waveform; (b) frequency spectrum.

kurtosis index values greater than 3 which indicate that the extracted bearing inner race fault is distorted. Hence, the weighted kurtosis index is constructed by the cross-correlation coefficient to get all bearing inner race faults information and difference spectrum by (17) in Figure 15(b). Difference spectrums corresponding to IMFs are shown in Table 3, with IMF3 of 0.9087 and the maximum mutation point. Therefore, IMF1~IMF3 reflect the fault characteristics of the bearing inner race and they were the sensitive IMFs input as adaptive stochastic resonance signals.

The selected IMF1~IMF3 are input into the adaptive stochastic resonance. Due to the fact that the target frequency of the bearing inner race fault is 173.25 Hz, in the frequency-shifted and scaling transform pretreatment method, the pass frequency and cutoff frequency of the high pass filter are 170 Hz and 165 Hz, respectively. The modulation frequency

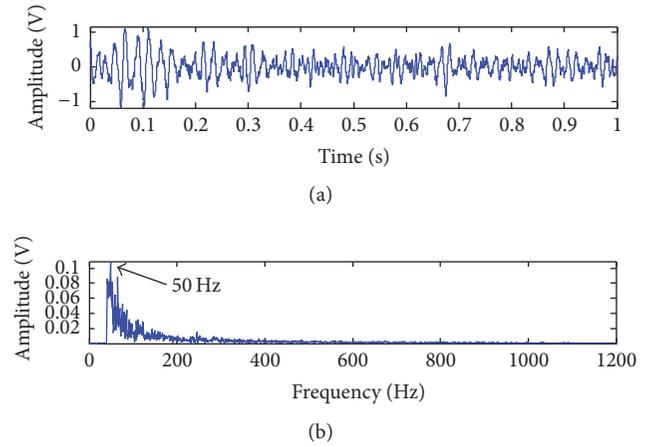


FIGURE 8: Optimal output signal of IMF2: (a) time-domain waveform; (b) frequency spectrum.

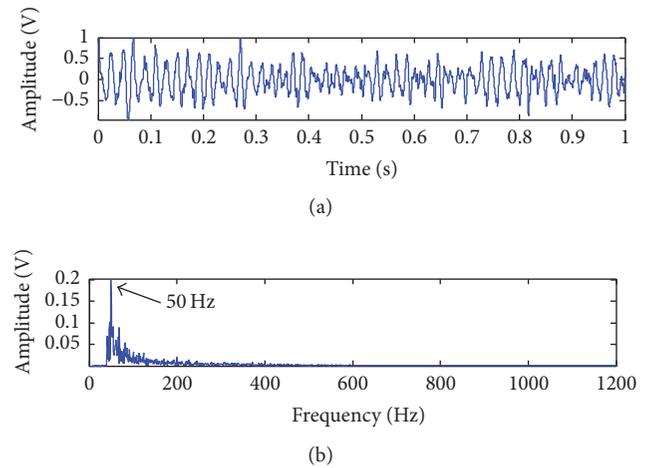


FIGURE 9: Optimal output signal of IMF3: (a) time-domain waveform; (b) frequency spectrum.

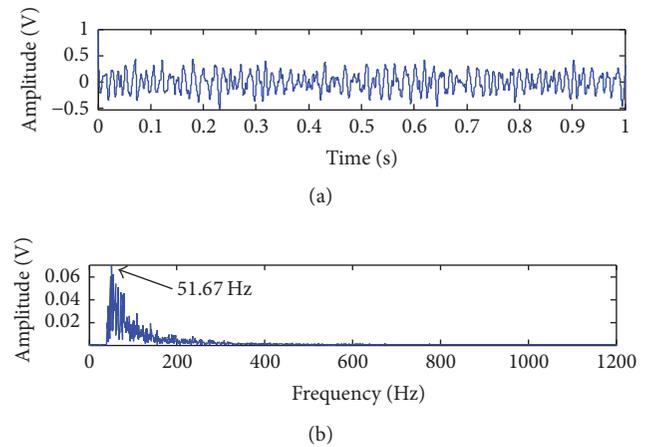


FIGURE 10: Optimal output signal of IMF4: (a) time-domain waveform; (b) frequency spectrum.

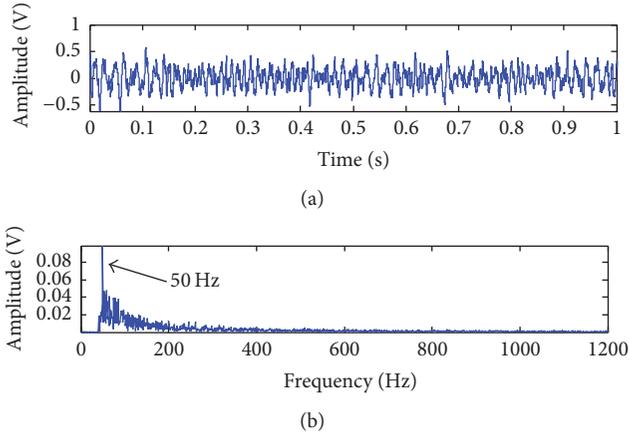


FIGURE 11: The statistical average of IMFs: (a) time-domain waveform; (b) frequency spectrum.



FIGURE 12: Rolling bearing fault test platform.

is 165 Hz and the variable scale compression rate is 400. The target frequency is compressed to $(173.25 - 165)/400 = 0.02065 \leq 1$ Hz which meets the requirements of small parameters. Because the actual signal is more complex than the simulation signal, the parameter optimization of stochastic resonance is based on ant colony algorithm. The optimal parameter combinations of adaptive stochastic resonance of IMF1~IMF3 and output signal-to-noise ratio of IMF1~IMF3 are 22.3860, 22.2677, and 22.3304, respectively, which are an evaluation function in Table 4.

Time-domain and the frequency spectrum of IMF1~IMF3 each in adaptive stochastic resonance are shown in Figures 16~18, respectively.

After adaptive stochastic resonance treatment, the fault frequencies of IMF1 and IMF2 are 170.7 Hz and 165.3 Hz which are greatly different from the theoretical value as shown in Figures 16(b) and 17(b). IMF3 is approximate to the characteristic frequency 173.25 Hz in Figure 18(b), but ambient noise has a large disturbance. From the above analysis, it can be known that even if we reduce the noise by EEMD and the stochastic resonance of each IMF component in the strong background noise, there is still a large deviation between the actual peak and the theoretical peak. The noise interference is still very serious.

In order to reduce noise further, statistical ensemble average on IMF1~IMF3 is performed and the final result is shown in Figure 19.

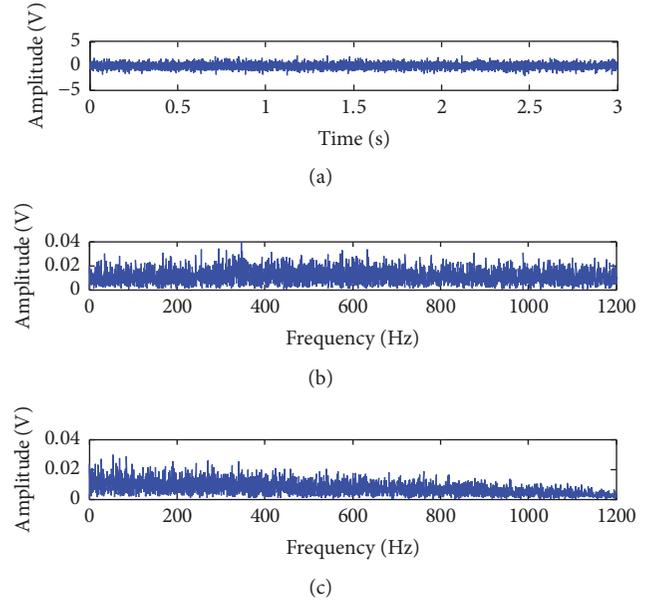


FIGURE 13: The fault signal of bearing inner race: (a) time-domain waveform; (b) frequency spectrum; (c) envelope spectrum.

TABLE 3: Difference spectrum of the bearing inner race corresponding to IMFs.

IMF	Difference spectrum
IMF1	0.4708
IMF2	0.4174
IMF3	0.9087
IMF4	0.07614
IMF5	0.01715
IMF6	0.02412
IMF7	0.0008135
IMF8	0.0009225
IMF9	$3.355e - 6$
IMF10	0.001409
IMF11	0.001771
IMF12	0.0004004
IMF13	0.0003416

TABLE 4: The optimal parameter combination of adaptive stochastic resonance of IMF1~IMF3.

Parameter	IMF1	IMF2	IMF3
a	0.001	0.0003	0.012
b	0.22420	0.1618	0.0443
Signal-to-noise ratio	22.3860	22.2677	22.3304

In Figure 19(b), the maximum spectral peak is 173.7 Hz (the theoretical value of the bearing inner race is 173.25 Hz) and is approximate to the fault characteristic frequency. Compared with Figures 14(b) and 19(b), we cannot see any fault characteristic frequency in Figure 14(b). Thus, the proposed method provides obvious advantages over the EEMD method in extracting weak characteristics from the signal with a

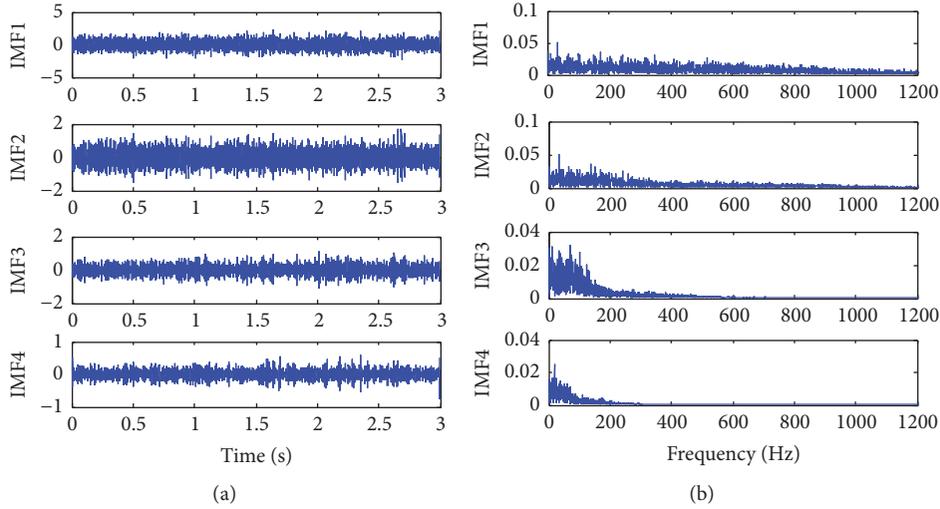


FIGURE 14: The IMF1~IMF4 of bearing inner race: (a) time-domain waveform; (b) frequency spectrum.

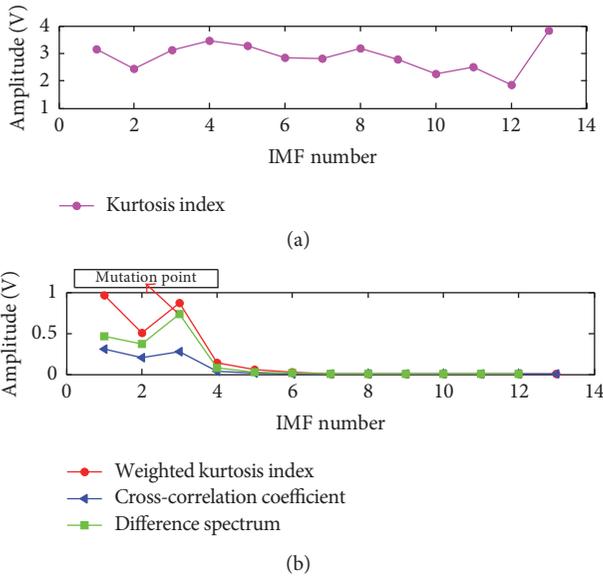


FIGURE 15: WKIDS of the inner race of the bearing: (a) kurtosis index; (b) weighted kurtosis index, cross-correlation coefficient, and difference spectrum.

low signal-to-noise ratio. Moreover, the frequency spectra of IMF1, IMF2, and IMF3 each in adaptive stochastic resonance are shown in Figures 16(b), 17(b), and 18(b), respectively; the characteristic frequencies are not the best results. We can assume that the weak fault characteristics cannot also be obtained only by adaptive stochastic resonance. In other words, the proposed method is not only better than EEMD method but also superior to adaptive stochastic resonance.

6. Conclusions

In this article, the fault diagnosis methods based on sensitive IMFs selection of EEMD and adaptive stochastic resonance

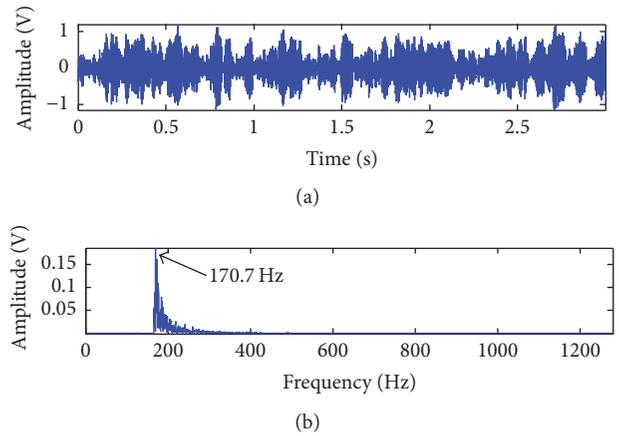


FIGURE 16: The enhanced result of IMF1 with an inner fault via adaptive stochastic resonance: (a) time-domain waveform; (b) frequency spectrum.

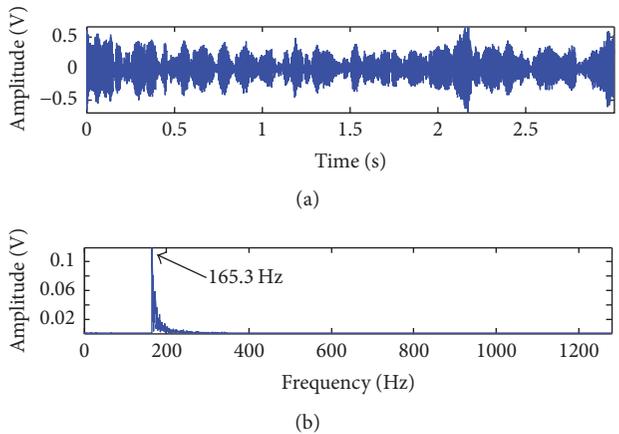


FIGURE 17: The enhanced result of IMF2 with an inner fault via adaptive stochastic resonance: (a) time-domain waveform; (b) frequency spectrum.

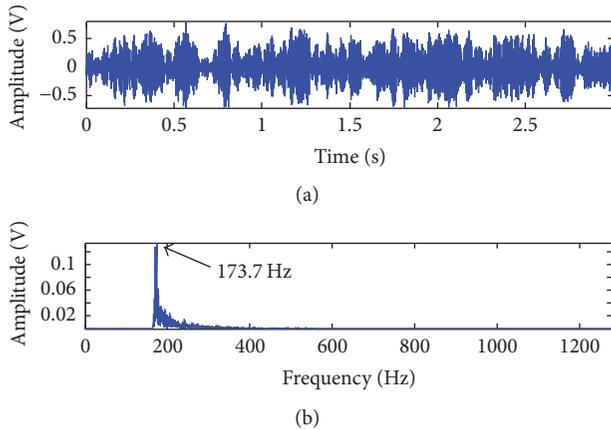


FIGURE 18: The enhanced result of IMF3 with an inner fault via adaptive stochastic resonance: (a) time-domain waveform; (b) frequency spectrum.

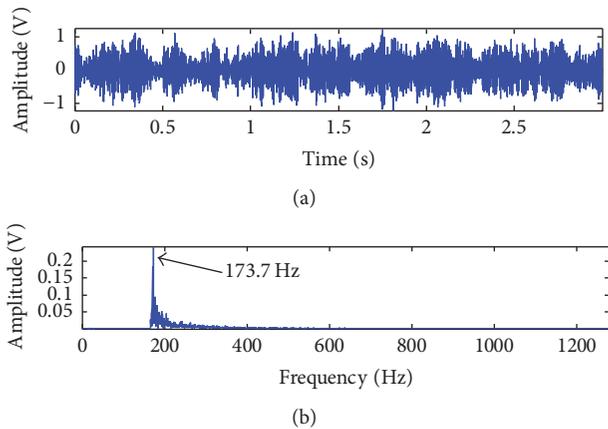


FIGURE 19: The statistical average of sensitive IMFs: (a) time-domain waveform; (b) frequency spectrum.

are studied, carrying out simulation and bearing experiment, mainly yielding the following conclusions.

(1) A new method of sensitive IMFs selection of EEMD is proposed. WKIDS is constructed by cross-correlation coefficient and kurtosis index. In WKIDS sequence, the maximum point reflects the mutation between the effective IMFs and the useless IMFs. This method can adaptively select the effective IMFs and eliminate the subjectivity of artificial selection IMFs.

(2) Combining effective IMFs selection of EEMD and adaptive stochastic resonance of each effective IMF and ensemble average, the noise of the original signal is greatly reduced and the signal-to-noise ratio is increased. The proposed method detects the weak fault feature accurately and the effect is very obvious.

(3) In the simulation and experiment, compared with EEMD and adaptive stochastic resonance of each sensitive IMF, it is found that EEMD cannot detect the weak fault feature in highly noisy environments. The selected effective IMFs are fed into adaptive stochastic resonance, but the

fault characteristic frequency in each enhanced IMF is still disturbed by the noise, and the fault frequency is not obvious. To reduce the noise further, all enhanced effective IMFs are ensemble-averaged. Finally, the weak fault characteristics hidden in the strong noise are extracted. After reducing the noise, adaptive stochastic resonance of the effective IMFs cannot effectively extract the fault, so we can conclude that stochastic resonance only can never detect weak fault characteristics. Hence, the proposed method is better than EEMD and adaptive stochastic resonance.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This research is supported by the National Science Foundation of China (51075029).

References

- [1] D. H. Pandya, S. H. Upadhyay, and S. P. Harsha, "Fault diagnosis of rolling element bearing by using multinomial logistic regression and wavelet packet transform," *Soft Computing*, vol. 18, no. 2, pp. 255–266, 2014.
- [2] Q. He, Y. Liu, Q. Long, and J. Wang, "Time-frequency manifold as a signature for machine health diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 5, pp. 1218–1230, 2012.
- [3] B. Hu and B. Li, "Blade crack detection of centrifugal fan using adaptive stochastic resonance," *Shock and Vibration*, vol. 2015, Article ID 954932, 12 pages, 2015.
- [4] Y. Lei, J. Lin, D. Han, and Z. He, "An enhanced stochastic resonance method for weak feature extraction from vibration signals in bearing fault detection," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 228, no. 5, pp. 815–827, 2014.
- [5] A. A. Tabrizi, L. Garibaldi, A. Fasana, and S. Marchesiello, "Performance improvement of ensemble empirical mode decomposition for roller bearings damage detection," *Shock and Vibration*, vol. 2015, Article ID 964805, 10 pages, 2015.
- [6] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.
- [7] A. Hu, W. Ma, and G. Tang, "Rolling bearing fault feature extraction method based on ensemble empirical mode decomposition and kurtosis criterion," *Proceedings of the Chinese Society of Electrical Engineering*, vol. 32, no. 11, pp. 106–111, 2012.
- [8] H. Wang, J. Chen, and G. Dong, "Fault diagnosis of rolling bearing's early weak fault based on minimum entropy deconvolution and fast Kurtogram algorithm," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 229, no. 16, pp. 2890–2907, 2015.
- [9] J. Tan, X. Chen, J. Wang et al., "Study of frequency-shifted and re-scaling stochastic resonance and its application to fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 23, no. 3, pp. 811–822, 2009.

- [10] Y. Xu, J. Li, J. Feng, H. Zhang, W. Xu, and J. Duan, "Lévy noise-induced stochastic resonance in a bistable system," *European Physical Journal B*, vol. 86, article 198, 2013.
- [11] Q. He, J. Wang, F. Hu, and F. Kong, "Wayside acoustic diagnosis of defective train bearings based on signal resampling and information enhancement," *Journal of Sound and Vibration*, vol. 332, no. 21, pp. 5635–5649, 2013.
- [12] R. Benzi, G. Parisi, A. Suter et al., "Stochastic resonance in climatic change," *Tellus*, vol. 34, pp. 10–16, 1982.
- [13] L. Zhang, F. Wang, and Z. Hui, "Detecting high frequency weak signal using parameter-adjusted frequency shifting based on Stochastic Resonance (SR)," *Physics Procedia*, vol. 24, pp. 382–389, 2012.
- [14] S. Lu, Q. He, H. Zhang, and F. Kong, "Rotating machine fault diagnosis through enhanced stochastic resonance by full-wave signal construction," *Mechanical Systems and Signal Processing*, vol. 85, pp. 82–97, 2017.
- [15] Y. Lei, D. Han, J. Lin, and Z. He, "Planetary gearbox fault diagnosis using an adaptive stochastic resonance method," *Mechanical Systems and Signal Processing*, vol. 38, no. 1, pp. 113–124, 2013.
- [16] J. Li, X. Chen, and Z. He, "Adaptive stochastic resonance method for impact signal detection based on sliding window," *Mechanical Systems and Signal Processing*, vol. 36, no. 2, pp. 240–255, 2013.
- [17] N. Huang, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London A*, vol. 454, no. 1971, pp. 903–995, 1998.
- [18] H. Dong, K. Qi, X. Chen, Y. Zi, Z. He, and B. Li, "Sifting process of EMD and its application in rolling element bearing fault diagnosis," *Journal of Mechanical Science and Technology*, vol. 23, no. 8, pp. 2000–2007, 2009.
- [19] F. Zhao, X. Mei, T. Tao, G. Jiang, and Y. Zhou, "Fault diagnosis of a machine tool rotary axis based on a motor current test and the ensemble empirical mode decomposition method," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 225, no. 5, pp. 1121–1129, 2011.
- [20] Z. Pan, Z. Qiao, and N. Zhang, "Weak signal detection of adaptive stochastic resonance based on shannon entropy of symbolic series," *Acta Metrologica Sinica*, vol. 36, pp. 496–500, 2015.
- [21] Z. Qiao, Y. Lei, J. Lin, and F. Jia, "An adaptive unsaturated bistable stochastic resonance method and its application in mechanical fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 84, pp. 731–746, 2017.
- [22] Q. He and J. Wang, "Effects of multiscale noise tuning on stochastic resonance for weak signal detection," *Digital Signal Processing*, vol. 22, no. 4, pp. 614–621, 2012.
- [23] S. Lu, Q. He, and F. Kong, "Stochastic resonance with Woods-Saxon potential for rolling element bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 45, no. 2, pp. 488–503, 2014.
- [24] Z. Qiao and Z. Pan, "SVD principle analysis and fault diagnosis for bearings based on the correlation coefficient," *Measurement Science and Technology*, vol. 26, no. 8, Article ID 085014, 2015.
- [25] Y. G. Leng, Y. S. Leng, T. Y. Wang, and Y. Guo, "Numerical analysis and engineering application of large parameter stochastic resonance," *Journal of Sound and Vibration*, vol. 292, no. 3–5, pp. 788–801, 2006.
- [26] R. Randall and J. Antoni, "Rolling element bearing diagnostics—a tutorial," *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 485–520, 2011.

Research Article

Prognostic and Remaining Life Prediction of Electronic Device under Vibration Condition Based on CPSD of MPI

Ying Chen, Ning Tang, and Zenghui Yuan

Science and Technology on Reliability and Environmental Engineering Laboratory, Reliability and System Engineering Department, Beihang University, Beijing 100191, China

Correspondence should be addressed to Ning Tang; 632057638@qq.com

Received 22 July 2016; Revised 29 September 2016; Accepted 20 October 2016

Academic Editor: Minvydas Ragulskis

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

Prognostic of electronic device under vibration condition can help to get information to assist in condition-based maintenance and reduce life-cycle cost. A prognostic and remaining life prediction method for electronic devices under random vibration condition is proposed. Vibration response is measured and monitored with acceleration sensor and OMA parameters, including vibration resonance frequency, especially first-order resonance frequency, and damping ratio is calculated with cross-power spectrum density (CPSD) method and modal parameter identification (MPI) algorithm. Steinberg vibration fatigue model which considers transmissibility factor is used to predict the remaining life of electronic component. Case study with a test board is carried out and remaining life is predicted. Results show that with this method the vibration response characteristic can be monitored and predicted.

1. Introduction

Prognostics and Health Monitoring (PHM) of electronic devices integrates sensor data with models that enable assessment of the degradation of a product from an expected normal operating condition and also assesses the future reliability of the product based on current and historic conditions [1]. The technique is intended for condition monitoring in high reliability applications where the knowledge of impending failure is critical and the risks in terms of loss-of-functionality are too high to bear. In avionics and aerospace electronics systems, electronic components, and dynamic loading are applied to electronic devices during normal operation, which includes high temperature and temperature cycling, vibration, and humidity. Data show that 20% [1] of electronic equipment failure is related to vibration and shock. PHM of electronic under vibration condition can help electronic devices to meet the safety, reliability, maintainability, and supportability requirements and reduce life-cycle cost.

PHM of electronic devices under vibration condition has been studied by some researchers, which is mainly based on two kinds of method, data-based method and model-based

method. Lall et al. developed a statistical method for prognostication of area-array electronics under shock and vibration loads during vibration testing, which is based on state space vectors from resistance spectroscopy measurements [2]. Kalman filter method is used to estimate the state space feature vector, extrapolate the future state, and predict the remaining useful life. Results will help to schedule downtime for maintenance and select the appropriate time to reorder a replacement part. Aiming at the phenomenon that the recent studies which predict impending failures of power electronic components' packaging are not explicit, they proposed monitor system for ceramic area-array-package with resistance spectroscopy and high speed cameras, predicted with system-transfer function and statistical pattern recognition techniques during shock test [3]. Results show that there is degradation in statistical confidence values, and the impending failure may have been detected before failure of the electronic assembly.

Compared with data-based prognostic method, model-based method provides the relation between failure and the characteristic of product, which helps to have the proper knowledge of failures and take measures to prevent them. Gu et al. proposed a model-based health monitoring and

prognostics method for assessing the reliability of a printed circuit board (PCB) under random vibration by using strain gauges and accelerometer to measure the vibration response [1]. The Physics-of-Failure fatigue model and Miner accumulation rule are used to assess the remaining life of a test board. Modal analysis with FEA software is conducted to transfer global sensor information to local sensor information. Miner rule is used to accumulate the damage of different loading condition. An uncertainty analysis of prognostics of electronics subject to random vibration, which includes measurement uncertainty, parameter uncertainty, failure criteria uncertainty, and future usage uncertainty, is also conducted [4]. A case study is presented whereby prognostics with uncertainty are applied to an electronic circuit board subject to random vibration. Gucik-Derigny et al. presented a model-based prognosis approach for assessing the remaining useful life of a PCB subject to low frequency thermal fatigue damage and high frequency of vibration damage [5]. The proposed prognosis method was also divided into fast dynamics part which includes the estimation of nonmeasurable state and slowly drifting PCB damping parameter and slow dynamic part which includes estimation of low cycle fatigue BGA solder joint interconnection damage state and parameter. Simulation is done to test the efficiency of their approach.

For model-based prognosis approach, modal parameters, such as frequency and damping, are associated with the prediction of remaining useful life, which will also change during the process of work. Recent research evaluates these parameters by simulation with FEA or CalcePWA software, which may add offline procedures and may need more experience of skilled person. Online measurement of these parameters and using them properly in the prediction model is critical. Operational Modal Analysis (OMA) technique can determine the inherent properties of a structure by measuring only the response of it without using an artificial excitation. Modal parameter identification is the technique to calculate the resonance frequencies, damping, and mode shapes of a structure. In OMA all modal parameters are to be determined without knowing the excitation forces. Therefore it is normally assumed that the excitation forces are Gaussian white noise [6]; recent research has extended the excitation to harmonic excitation [7], nonstationary excitation [8], and periodic excitations [9]. OMA has widely been used in performance assessment of structure whose properties would change with time [10–12], and recently the researchers have raised the attention on health diagnosis of structure and estimating the remaining service life in civil engineering [13] and oil industry [14].

The research on physics model of vibration fatigue failure can be traced back to the 1970s. After many years of practical experience, Steinberg proposed Steinberg model applied to lifetime estimation of electronic devices working under sinusoidal or random vibration conditions [15, 16]. Although Manson model and other models appeared in this field later, Steinberg model is still widely used in engineering for its obvious physical meaning [17]. Chesné and Deraemaeker studied the application of Steinberg model in tantalum capacitor [18]. Through experiment, they provided the S-N curve in different sinusoidal sweeping-frequency vibration conditions

and compared the experimental result with Finite Element Analysis (FEA) simulation, with the aim of determining parameter values of Steinberg model. Steenackers et al. proposed some principles of electronic systems to withstand high vibration and shock conditions [19]. Lim analyzed PCB's vibration with Steinberg model in CalcePWA [20]. Yang et al. studied the impact of transmissibility factor of Steinberg model on fatigue lifetime of electronic devices [21], which represent the coupling state of PCB and its case. Neglecting this factor will result in the inaccuracy of prediction.

From the above discussion, previous model based prognostic methods of electronic device under vibration condition are limited, for example, the methods lack online measurement and calculation of model parameters. In this paper, an online health prognostic method for electronic devices under random vibration condition is proposed. Vibration response is measured and monitored with acceleration sensor under white noise random excitation, and OMA parameters, including vibration resonance frequency, especially first-order resonance frequency, and damping ratio, are calculated with cross-power spectrum density (CPSD) method and modal parameter identification (MPI) algorithm. Steinberg vibration fatigue model which considers transmissibility factor is used to predict the remaining life of electronic component.

The remainder of the paper is organized as follows. Section 2 presents the PHM flowchart with detail steps. Section 3 introduces the cross-power spectrum density (CPSD) method and modal parameter identification (MPI) algorithm. Section 4 discusses the Steinberg model with transmissibility factor. Section 5 illuminates the real case with a test board. Section 6 gives conclusions as well as directions for the future work.

2. PHM Methodology for Electronics under Vibration Loading

The proposed health prognostic method for electronics device under random vibration is shown in Figure 1, in which either the vibration response is monitored and analyzed or the remaining life is predicted online.

The first step is to select the sensor location on the PCB and prepare the monitoring devices. Since some complicated PCBs have hundreds of components, the space is limited; it is essential to know where to place the sensors. Sensors will be placed at the site where components are most likely to fail. The vibration result of the PCB under given random vibration condition could be obtained. For example, CalcePWA software [1] can identify the natural frequency of the PCB and the locations of critical components at certain vibration loading levels. Piezoelectric sensor is used to measure the time-domain acceleration response of PCB. After that the cross spectrum of signal from different sensor is calculated, results will be used to generate the CPSD of reference point and measurement point. With frequency domain, modal extraction method calculates the first-order resonant frequency and damping of the PCB, which are the main parameters needed to calculate the remaining life time with PoF models. During the process, monitor the

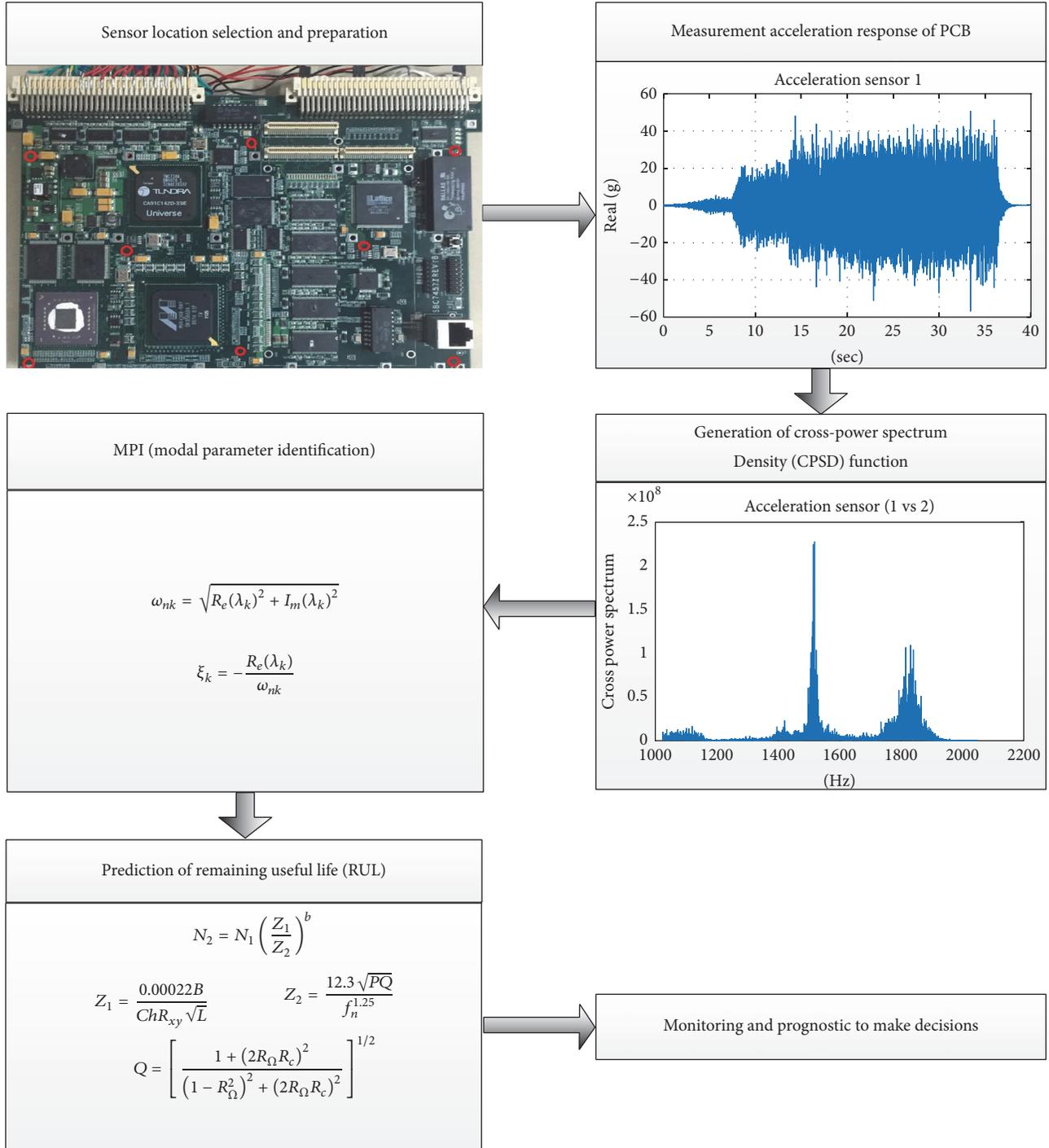


FIGURE 1: Proposed prognostic method of electronic devices under vibration condition.

vibration response signals and the modal parameters and make decisions when necessary.

3. CPSD Based Modal Parameter Identification

3.1. Cross-Power Spectrum Density Function. Piezoelectric acceleration sensor can measure the system response in time

domain and Discrete Fourier Transform (DFT) should be carried out to obtain the power spectral density matrices that will contain all the frequency information. The relationship between the input $x(t)$ and the output $y(t)$ in frequency domain can be written in the following form:

$$[G_{yy}(j\omega)] = [H(j\omega)]^* [G_{xx}(j\omega)] [H(j\omega)]^T, \quad (1)$$

where $G_{xx}(j\omega)$ is the input power spectral density matrix that is constant in the case of a stationary zero mean white noise input. This constant will be called C in the rest of the mathematical derivation. $G_{yy}(j\omega)$ is the output PSD matrix and $H(j\omega)$ is the frequency response function (FRF) matrix. As seen in (1), the output G_{yy} will be highly sensitive to the input constant C . The rest of the equation derivations and single degree of freedom identification will provide relevant results, only by assuming that the input is effectively represented by a constant value (mean Gaussian). It is therefore important to realize how this input assumption will be crucial to the technique.

The FRF matrix can be written in a typical partial fraction form (used in classical modal analysis), in terms of poles and residues where $[G_{xx}(\omega)]$ is the input power spectral density (PSD) matrix, $[G_{yy}(\omega)]$ is the output PSD matrix, and $[H(\omega)]$ is the frequency response function (FRF) matrix, and $*$ and superscript T denote complex conjugate and transpose, respectively. The FRF matrix can be written in a typical partial fraction form (used in classical Modal analysis), in terms of poles, λ , and residues, R , which is

$$[H(\omega)] = \sum_{k=1}^N \frac{[R_k]}{j\omega - \lambda_k} + \frac{[R_k]^*}{j\omega - \lambda_k^*}, \quad (2)$$

$$\lambda_k = -\xi_k + j\omega_{nk},$$

where N is the total number of modes of interest, λ_k the pole of the k th mode, ξ_k the modal damping (decay constant), and ω_{nk} the damped natural frequency of the k th mode.

$$\omega_{dk} = \omega_{0k} \sqrt{1 - \zeta_k^2} = \omega_{0k} \sqrt{1 - \frac{\sigma_k^2}{\omega_{0k}^2}}, \quad (3)$$

ζ_k being the critical damping and ω_{0k} the undamped natural frequency, both for mode k .

The transfer function matrix $[H]$ is symmetric and an element $H_{pq}(j\omega)$ of this matrix is then written in terms of the component $r_{kpq}(j\omega)$ of the residue matrix as follows:

$$H_{pq}(j\omega) = \sum_{k=1}^N \frac{r_k(p, q)}{j\omega - \lambda_k} + \frac{r_k(p, q)^*}{j\omega - \lambda_k^*}. \quad (4)$$

Using expression (1) for the matrix $[G_{yy}(\omega)]$ and the Heaviside partial fraction theorem for polynomial expansions, the following expression for the output PSD matrix is $[G_{yy}(\omega)]$ assuming the input is random in both time and space and has a zero mean white noise distribution; that is, $[G_{xx}(\omega)] = [C]$:

$$G_{yy}(j\omega) = \sum_{k=1}^N \left(\frac{[A_k]}{j\omega - \lambda_k} + \frac{[A_k]^*}{j\omega - \lambda_k^*} + \frac{[B_k]}{-j\omega - \lambda_k} + \frac{[B_k]^*}{-j\omega - \lambda_k^*} \right), \quad (5)$$

where $[A_k]$ is the k th residue matrix of the matrix $[G_{yy}]$. The matrix G_{xx} is assumed to be a constant value C , since the

excitation signals are assumed to be uncorrelated zero mean white noise in all the measured DOFs.

In (5), the form of $G_{yy}(\omega)$ is similar to frequency response function, the poles also including frequency and damping information. By replacing the traditional frequency response function with cross-power spectrum density function, it is possible to identify working modal parameters.

Cross-power spectrum is the product of Fourier transforms of test response signal and the reference response signal. Its amplitude is the product of the amplitude of test response signal and the reference. Its phase is the difference of the phase of test response signal and the reference. The relations between FRF and cross-power spectrum and working modal parameter identification algorithm based on cross-power spectrum will be demonstrated at the next section.

3.2. Modal Parameter Identification Algorithm. For a system with m measurement point and p reference point, the cross-power spectrum density matrix is

$$G(j\omega) = \begin{bmatrix} G_{11}(j\omega) & G_{12}(j\omega) & \cdots & G_{1p}(j\omega) \\ G_{21}(j\omega) & G_{22}(j\omega) & \cdots & G_{2p}(j\omega) \\ \vdots & \vdots & \ddots & \vdots \\ G_{M1}(j\omega) & G_{M2}(j\omega) & \cdots & G_{Mp}(j\omega) \end{bmatrix}_{M \times P} \quad (6)$$

and then (5) can be written as

$$G_{yy}(j\omega) = \sum_{k=1}^N \left(\frac{\{\psi_k\} \{W_k\}^T}{j\omega - \lambda_k} + \frac{\{\psi_k\}^* \{W_k\}^H}{j\omega - \lambda_k^*} + \frac{\{W_k\} \{\psi_k\}^T}{-j\omega - \lambda_k} + \frac{\{W_k\}^* \{\psi_k\}^H}{-j\omega - \lambda_k^*} \right) \quad (7)$$

in which $\{\psi_k\}$, $\{\psi_k\}^*$, $\{W_k\}$, and $\{W_k\}^*$ are complex modal matrix.

Let $s = j\omega$, $O(s) = (SI - Z)^{-1}W$, and (6) can be written as

$$G(s) = \phi(SI - Z)^{-1}W = \phi O(s). \quad (8)$$

The time derivative of $G(s)$ is

$$\begin{aligned} G(s) &= SG(s) - R(0) = SG(s) - \phi W \\ &= S\phi O(s) - \phi(SI - Z)O(s) = \phi ZO(s) \end{aligned} \quad (9)$$

which will get

$$\begin{aligned} \begin{bmatrix} G(s) \\ G(s) \end{bmatrix} &= \begin{bmatrix} \phi \\ \phi Z \end{bmatrix} O(s) \\ \text{or } \begin{bmatrix} G(s) \\ SG(s) \end{bmatrix} - \begin{bmatrix} 0 \\ \phi W \end{bmatrix} &= \begin{bmatrix} \phi \\ \phi Z \end{bmatrix} O(s). \end{aligned} \quad (10)$$

For Eigenvalue matrix and feature vector matrix, there must be a matrix, which will fulfill the Eigenvalue function of

$$A\phi + \phi Z = 0. \quad (11)$$

Another form of (9) is

$$\begin{bmatrix} A \\ I \end{bmatrix} \begin{bmatrix} \phi \\ \phi Z \end{bmatrix} = 0. \quad (12)$$

Multiply $O(s)$ to (10) and get

$$\begin{bmatrix} A \\ I \end{bmatrix} \begin{bmatrix} \phi \\ \phi Z \end{bmatrix} O(s) = \begin{bmatrix} A \\ I \end{bmatrix} \begin{bmatrix} G(s) \\ SG(s) - \phi W \end{bmatrix} = 0, \quad (13)$$

$$AG(s) + SG(s) - \phi W = 0 \quad (14)$$

Consider $S = j\omega$; then (12) can be given by

$$AG(j\omega_k) + j\omega_k G(j\omega_k) = \phi W, \quad k = 1, 2, \dots, K. \quad (15)$$

Let

$$D = [G(j\omega_1), G(j\omega_2), \dots, G(j\omega_{1K})], \quad (16)$$

$$\Omega = -\text{diag}[j\omega_1 I, j\omega_2 I, \dots, j\omega_k I]$$

and then

$$\begin{bmatrix} A \\ -\phi W \end{bmatrix} \begin{bmatrix} D \\ [I_1, I_2, \dots, I_k] \end{bmatrix} = D\Omega. \quad (17)$$

D and Ω are all functions of measured CPSD and ω , which are known; the matrix A and ϕW can be solved by (14):

$$\begin{aligned} \text{Modal frequency: } \omega_{nk} &= \sqrt{R_e(\lambda_k)^2 + I_m(\lambda_k)^2} \\ \text{Damping } \xi_k &= -\frac{R_e(\lambda_k)}{\omega_{nk}}, \quad n = 1, \dots, N. \end{aligned} \quad (18)$$

4. Remaining Useful Life Prediction

Steinberg model is used for remaining useful life prediction, which is derived from component fatigue characteristic fitting from fatigue experiments data [22], where

$$N_1 S_1^b = N_2 S_2^b \quad (19)$$

in which N_1 and N_2 are the stress cycle before fatigue. S_1 and S_2 are the stress when failure happens, and b is the fatigue index number related to the linear gradient of fatigue curve. For linear system, stress S is proportionate to displacement Z , and (19) can be written as

$$\begin{aligned} N_1 Z_1^b &= N_2 Z_2^b, \\ N_2 &= N_1 \left(\frac{Z_1}{Z_2} \right)^b. \end{aligned} \quad (20)$$

Under vibration condition, the displacement of simply supported PCB with four sides can expressed as follows, where the fatigue lifetime of these component can reach 1×10^7 stress cycles [22]:

$$Z_1 = \frac{0.00022B}{ChR_{xy}\sqrt{L}}, \quad (21)$$

where B is the length of the circuit board edge parallel to the component, L is the length of the electronic component, h is the circuit board thickness, C is the constant for different types of electronic components ($0.75 < C < 2.25$, and refer to [23]), and R_{xy} is the relative position factor for the component mounted on the board.

$$R_{xy} = \sin \frac{\pi X}{a} \sin \frac{\pi Y}{b}, \quad (22)$$

where X and Y are horizontal and longitudinal position of the component and a and b are length and width of PCB.

Based on the vibration fatigue curve, incorporating the system dynamic response property analysis, Steinberg proposed a vibration fatigue model for electronic device. In his model, PCB can be approximated as a single degree of freedom system, when it vibrates under the fundamental resonance.

When the input power spectral density (PSD) of random vibration is flat spectrum in resonance region, the mean square root acceleration response of a system is given by

$$G_{\text{out}} = \sqrt{\frac{\pi}{2}} P f_n Q (\text{RMS}). \quad (23)$$

The actual dynamic single amplitude displacement of PCB's center is given by [15]

$$Z_2 = \frac{9.8 G_{\text{out}}}{f_n^2} = \frac{12.3 \sqrt{PQ}}{f_n^{1.25}}, \quad (24)$$

where P is the input PSD at resonant frequency, f_n is resonant frequency, and Q is the transmissibility at resonant frequency, which is given by [15]

$$Q = \left[\frac{1 + (2R_\Omega R_c)^2}{(1 - R_\Omega^2)^2 + (2R_\Omega R_c)^2} \right]^{1/2}, \quad (25)$$

$$R_\Omega = \frac{\omega}{\omega_n}, \quad (26)$$

$$R_c = \frac{\xi_1}{\xi_c},$$

where $\omega_n = \sqrt{K/m}$, being the first-order angular resonance frequency, K is rigidity, m is mass, ξ_1 is the damping at first-order frequency, and

$$\xi_c = 2\sqrt{Km}. \quad (27)$$

The f_n and damping ratio is measured and calculated using (18).

5. Case Study

5.1. *The Test Board.* A certain type of test board (shown in Figure 2) is mainly used for ground automatic control systems. There are CPU microprocessor, memory and cache, sensor chips, counters, and various interfaces on the single board, and its function is like a microcomputer. There are

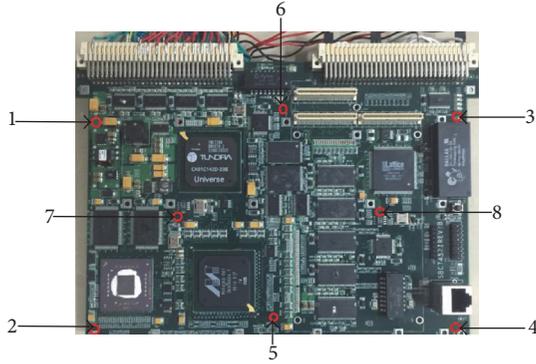


FIGURE 2: Test board and acceleration sensor layout.

many types of components on the test board including integrated circuits, chip capacitors, tantalum capacitors, metal-film resistors, inductors, connectors, crystals, switches, and PCB (including the solder joint, pad, PTH, vias, and metal interconnects). The material of PCB is FR-4; material of metal wires and plated through holes (PTH) is copper, and solder is 63Sn37Pb. The main types of package of the device are BGA, SOP, QFP, DIP, TSSOP, and so forth. And packaging materials are mainly plastic or ceramic. PCB is mounted through the holes on its four edges side, and it is considered that the installation site will not fail.

5.2. The Monitoring Devices. The monitoring system is shown in Figure 3, which mainly includes the vibration table (LDS), acceleration sensor, vibration measurement instrument, and data collection and analysis software. The working frequency of LDS vibration table is 5–4000 Hz, and vibration table can vibrate horizontally and vertically. The payload weight of the vibration table is about 200 Kg. Vibration Signal Control System is the brain of LDS vibration table, which has the function of random and sinusoidal vibration, harmony frequency lingering, and shock wave controlling. The vibration signal should be amplified by a power amplifier before sent to the vibration table. Endevco 2222c piezoelectric sensor whose weight is only 2 grams is used to measure the acceleration response of PCB. Vibration measurement instrument is DP730 24-channel device which could analyze polytype of vibration and noise in real time with software. The monitoring system is composed of data collection module, modal parameter identification module, and RUL calculation module.

The test board was mounted on vibration table, and its two edges were clamped, as shown in Figure 3. Random vibration loading condition with a frequency from 10 to 2000 Hz is applied to the vibration table, which is shown in Figure 4. Vibration spectrum is given by LDS vibration control system and the signal is exported into a power amplifier to drive the vibration table. The vibration table would generate vibration in the vertical direction.

There are mainly four ways to connect sensors and the test object, by bolt, adhesives, wax, or permanent magnet. The

Endevco 7703A-50 acceleration sensor in this test is with no magnetic base and requires high upper limited frequency. So we fix them via hard adhesives. Theoretically, more acceleration sensors can improve the measurement and prediction accuracy. But restricted to the space of PCB, the number of sensors is limited. This may be better if wireless and built-in sensors are used. In our case study, the layout of the test board is very compact and most of the components are on the front of the PCB. Modal analysis and random vibration analysis are carried out with CalcePWA. Results show the displacement response of this test board is well-distributed. Based on these pieces of information, we arranged eight acceleration sensors on the well-distributed eight locations where there is no component to avoid interaction between them.

The broadband random vibration spectrum is set close to white noise and loaded on the vibration table along the normal direction of the PCB. Via the fixture, excitation signal will be transmitted from the vibration table to the electronic equipment.

5.3. Prognostic Data and Remaining Life Prediction. The acceleration data were collected by the DP730 vibration instrument. Meanwhile, the acceleration data was sent to the data collection module which sample and record the data in real time. The modal parameter identification module will identify working modal parameters including the natural frequency and damping of the test board in accordance with the cross-power spectrum. Figure 4 is the exact vibration spectrum, which is a white noise spectrum. Figure 5 shows the time-domain data of the measurement points 1 to 8, which express the dynamic response of the PCB. Choose point 1 as the reference point. Figure 6 is the cross-power spectrum density curve of points 2–8 with reference point 1.

The times series data could be processed using the Fourier transform method. Multiplying the reference point's Fourier transform data by the conjugation of other measuring points, the cross-power spectrum could be obtained, which illustrate the correlation of the two-channel vibrate data. The cross-power spectrum density curves in Figure 6 indicate the correlation between the monitoring point and reference point. PCB is mounted through the holes on its four edges side. Monitoring points 1–6 are located around the PCB, but points 7 and 8 are located in the center of the PCB. So the curves' shape between monitoring points 2–6 and point 1 is consistent, and similarly, the curves' shape between monitoring points 2–6 and point 1 is consistent. Due to the transmission delay of the vibration signal on the board, the first curves and the last two curves are different. The transmission delay of the vibration signal has an effect on the cross-power spectrum density curves. We can find that the vibration magnitude of output signal is more dependent on the input signal when the frequency component of input signal approach 1550 Hz or 1850 Hz. The resonance frequency is calculated based on the modal parameter identification algorithm according to the first six curves. The peaks in Figure 6 indicate that the vibration data of other measuring point has significant correlation with the reference point.

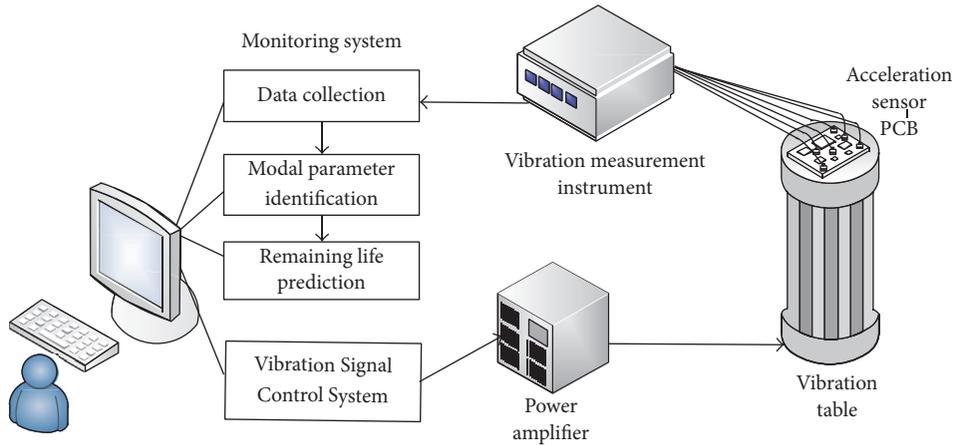


FIGURE 3: Monitoring devices of electronic devices under vibration.

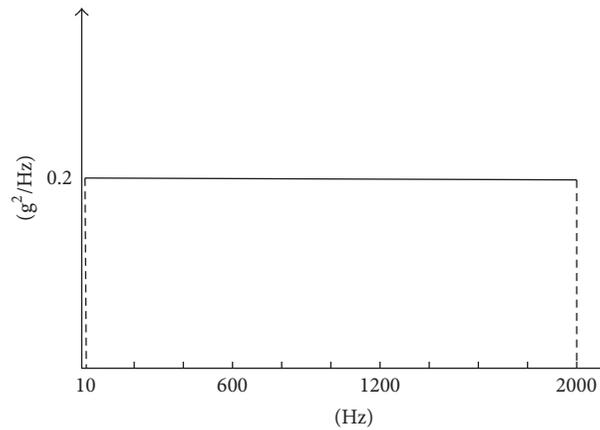


FIGURE 4: Exaction vibration spectrum, a white noise spectrum.

Through the identification of modal parameters, the first three work mode shapes of the PCB board would be obtained. Modal analysis and random vibration analysis are carried out with CalcePWA. The modal analysis results of cross-power spectrum estimation and the modal analysis results of the CalcePWA are consistent.

The remaining useful life of some important component predicted with (20)–(25) is presented in Table 1, in addition of the estimated first-order modal frequency, damping ratio at the beginning of the monitor ($t = 0$), and when $t = 28$ days, $t = 63$ days, and $t = 121$ days the data are also given. It can be seen from the results that, without consideration of measurement and computation deviation, the first-order modal frequency of the test board is not significantly changed; the damping will increase with the time.

Table 1 shows that the prediction for four components. All but BGA64 of the components lasted more than one year. BGA64 and QFP100, which have large dimension and are

located at the center of the test board, where the vibration response is strong than other components, have less lifetime.

6. Discussion and Conclusion

This paper presents an online health prognostic and monitoring method for assessing the reliability of electronic devices under random vibration condition, using acceleration sensor's information measured on the place where there are some critical components and the acceleration sensor could be placed perfectly. Based on cross-power spectrum density (CPSD) method and modal parameter identification (MPI) algorithm, the relationship between modal parameter and the remaining life is established. In the case study, we can assess the reliability of electronic devices without modal analysis and other simulation using software. As a result, this paper proved that an accelerometer is suitable for PHM and makes the field application of PHM much easier, since its

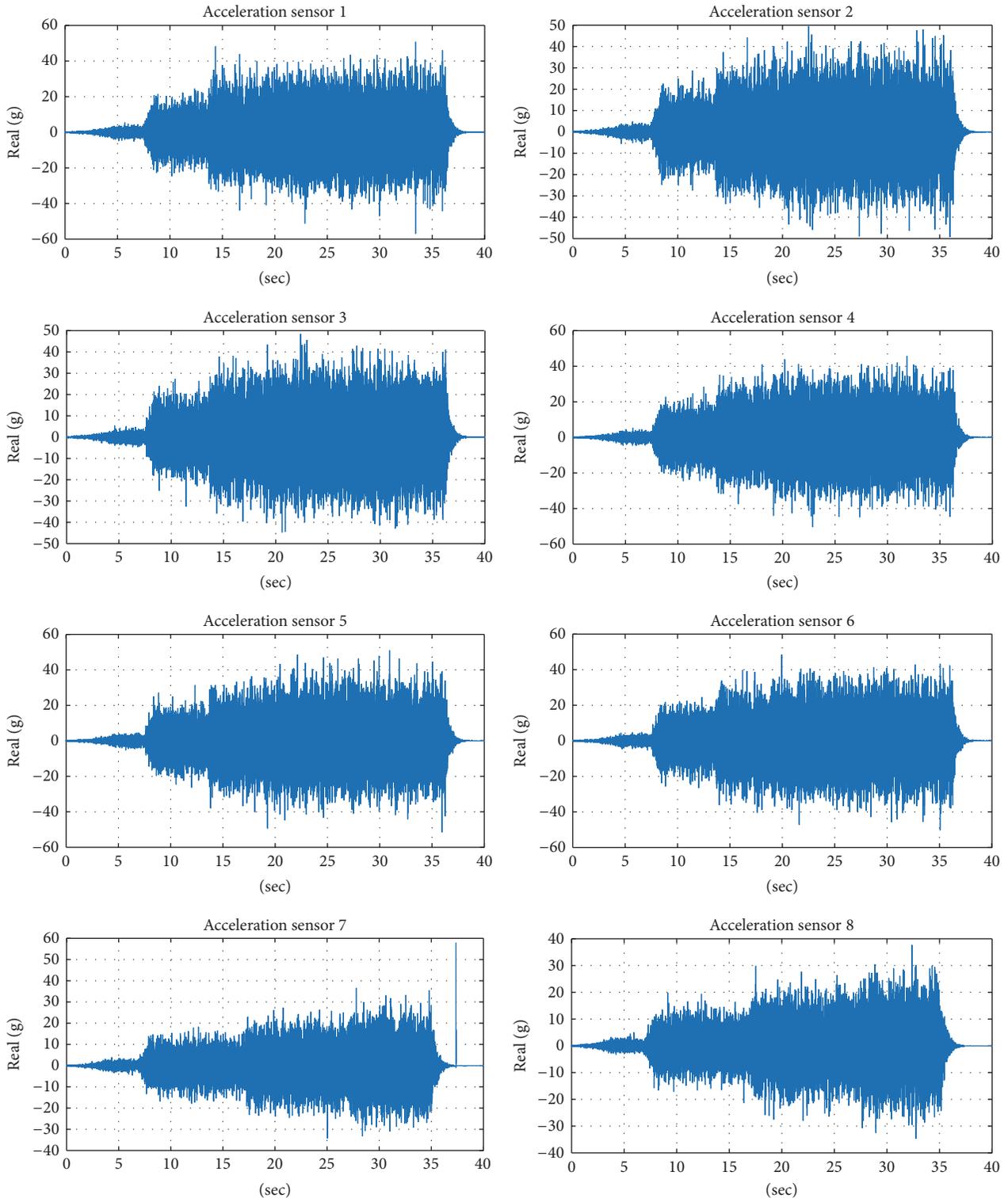


FIGURE 5: Time-domain dynamic response of measurement points 1 to 8.

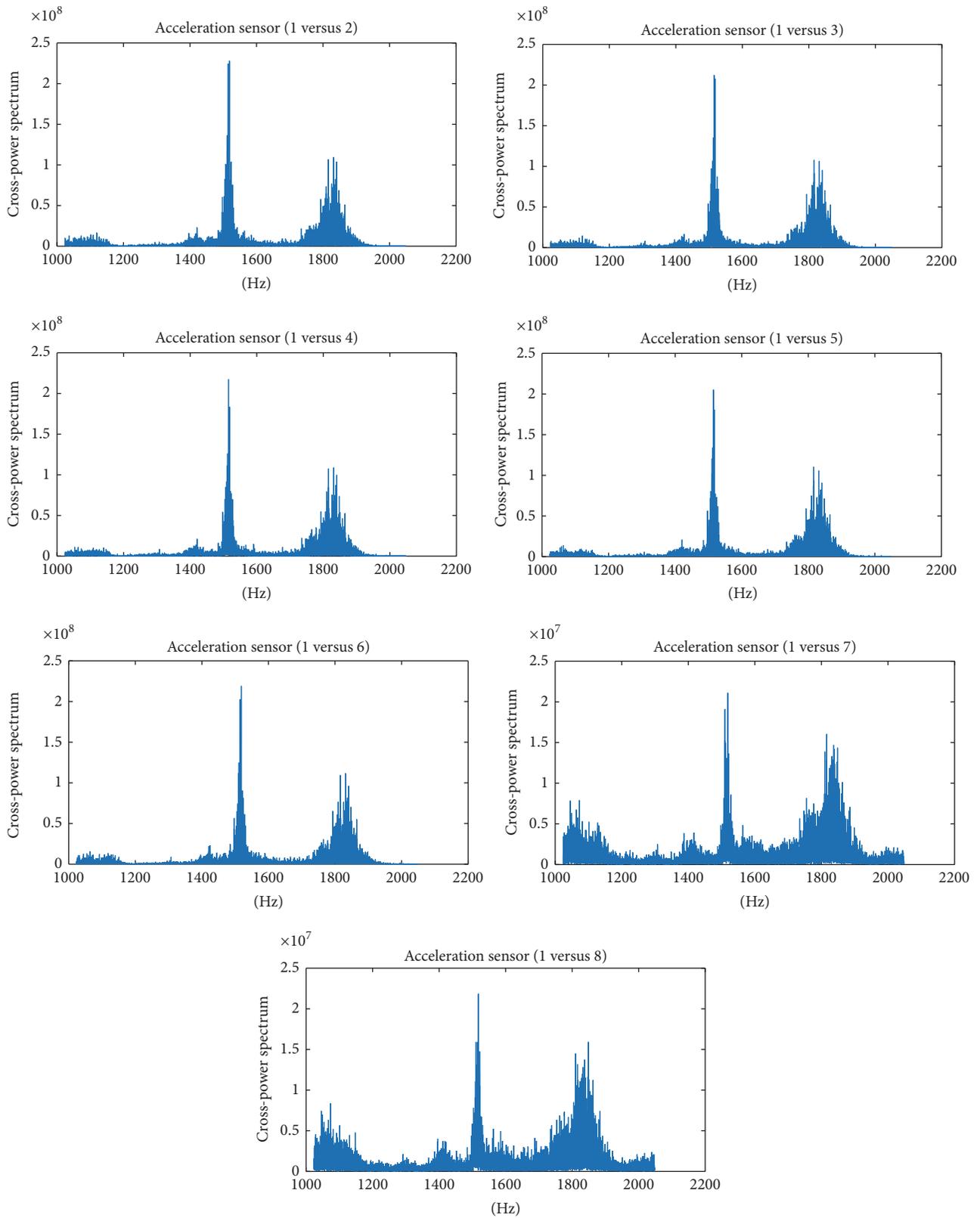


FIGURE 6: Cross-power spectrum density curve.

TABLE 1: Remaining life prediction results.

Component	Time (days)	First-order modal Frequency	Damping	Remaining life prediction (days)
BGA64	$t = 0$	231.97	3.69	332.15
	$t = 28$	227.41	3.73	298.73
	$t = 63$	222.76	3.76	220.67
	$t = 121$	220.32	3.78	183.32
QFP100	$t = 0$	231.97	3.69	384.29
	$t = 28$	227.41	3.73	349.50
	$t = 63$	222.76	3.76	281.25
	$t = 121$	220.32	3.78	249.66
QFP48	$t = 0$	231.97	3.69	435.77
	$t = 28$	227.41	3.73	405.41
	$t = 63$	222.76	3.76	330.92
	$t = 121$	220.32	3.78	387.45
SOP32	$t = 0$	231.97	3.69	592.86
	$t = 28$	227.41	3.73	563.43
	$t = 63$	222.76	3.76	478.32
	$t = 121$	220.32	3.78	449.66

installation and associated signal conditioning circuitry are simpler, cheaper, and more reliable.

Future work should focus on increasing PHM accuracy, by using wireless sensor systems to enhance prediction capability and accuracy by developing or application of algorithms for other vibration excitation, for example, nonwhite noise spectrum.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The authors acknowledge the support for the work presented in this paper from National Natural Science Foundation of China (Grant nos. 61503014 and 61573043).

References

- [1] J. Gu, D. Barker, and M. Pecht, "Health monitoring and prognostics of electronics subject to vibration load conditions," *Sensors Journal*, vol. 9, no. 11, pp. 1479–1485, 2009.
- [2] P. Lall, R. Lowe, and K. Goebel, "Prognostics using Kalman-Filter models and metrics for risk assessment in BGAs under shock and vibration loads," in *Proceedings of the 60th Electronic Components and Technology Conference (ECTC '10)*, pp. 889–901, June 2010.
- [3] P. Lall, R. Lowe, and K. Goebel, "Use of prognostics in risk-based decision making for BGAs under shock and vibration loads," in *Proceedings of the 12th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm '10)*, pp. 1–11, IEEE, June 2010.
- [4] J. Gu, D. Barker, and M. Pecht, "Uncertainty assessment of prognostics of electronics subject to random vibration," in *Proceedings of the AAAI Fall Symposium on Artificial Intelligence for Prognostics*, pp. 50–57, November 2007.
- [5] D. Gucik-Derigny, A. Zolghadri, E. Suhir, and L. Bechou, "A model-based prognosis strategy for prediction of Remaining Useful Life of Ball-Grid-Array Interconnections," *IFAC Proceedings*, vol. 47, no. 3, pp. 7354–7360, 2014.
- [6] R. Brincker, "Some elements of operational modal analysis," *Shock and Vibration*, vol. 2014, Article ID 325839, 11 pages, 2014.
- [7] P. Mohanty and D. J. Rixen, "Identifying mode shapes and modal frequencies by operational modal analysis in the presence of harmonic excitation," *Experimental Mechanics*, vol. 45, no. 3, pp. 213–220, 2005.
- [8] V. H. Vu, M. Thomas, A. A. Lakis, and L. Marcouiller, "Short-time autoregressive (STAR) modeling for operational modal analysis of non-stationary vibration," in *Vibration and Structural Acoustics Analysis*, pp. 59–77, Springer, Berlin, Germany, 2011.
- [9] D. Ning and Y. Gong, "Shocking fault component of abnormal sound signal in the fault engine extract method based on linear superposition method and cross-correlation analysis," *Advances in Mechanical Engineering*, vol. 7, no. 8, pp. 1–9, 2015.
- [10] L. Soria, B. Peeters, J. Anthonis, and H. V. der Auweraer, "Operational Modal Analysis and the performance assessment of vehicle suspension systems," *Shock and Vibration*, vol. 19, no. 5, pp. 1099–1113, 2012.
- [11] S.-D. Zhou, L. Liu, W. Yang, and Z.-S. Ma, "Operational modal identification of time-varying structures via a vector multistage recursive approach in hybrid time and frequency domain," *Shock and Vibration*, vol. 2015, Article ID 397364, 13 pages, 2015.
- [12] B. Peeters, H. Van der Auweraer, F. Vanhollebeke, and P. Guillaume, "Operational modal analysis for estimating the

- dynamic properties of a stadium structure during a football game,” *Shock and Vibration*, vol. 14, no. 4, pp. 283–303, 2007.
- [13] M. J. Whelan, M. V. Gangone, K. D. Janoyan, and R. Jha, “Real-time wireless vibration monitoring for operational modal analysis of an integral abutment highway bridge,” *Engineering Structures*, vol. 31, no. 10, pp. 2224–2235, 2009.
- [14] A. Mironov, P. Doronkin, A. Priklonsky, and I. Kabashkin, “Condition monitoring of operating pipelines with operational modal analysis application,” *Transport and Telecommunication Journal*, vol. 16, no. 4, pp. 305–319, 2015.
- [15] D. Steinberg, *Vibration Analysis for Electronic Equipment*, A Wiley Inter-Science Publication, 3rd edition, 2000.
- [16] Y. Chen, L. Yang, B. Liu, and D. Xue, “Applicability study of Steinberg vibration fatigue model in electronic products,” in *Proceedings of the Prognostics and System Health Management Conference (PHM-Hunan '14)*, pp. 31–35, Zhangjiajie City, China, August 2014.
- [17] F. Liu and G. Meng, “Random vibration reliability of BGA lead-free solder joint,” *Microelectronics Reliability*, vol. 54, no. 1, pp. 226–232, 2014.
- [18] S. Chesné and A. Deraemaeker, “Damage localization using transmissibility functions: a critical review,” *Mechanical Systems and Signal Processing*, vol. 38, no. 2, pp. 569–584, 2013.
- [19] G. Steenackers, C. Devriendt, and P. Guillaume, “On the use of transmissibility measurements for finite element model updating,” *Journal of Sound and Vibration*, vol. 303, no. 3–5, pp. 707–722, 2007.
- [20] J. H. Lim, “A correlation study of satellite finite element model for coupled load analysis using transmissibility with modified correlation measures,” *Aerospace Science and Technology*, vol. 33, no. 1, pp. 82–91, 2014.
- [21] L. Yang, Y. Chen, Z. Yuan, and L. Chen, “Effect of vibration transmissibility on fatigue lifetime of electronic devices,” in *Industrial Engineering, Management Science and Applications*, pp. 1031–1039, Springer, Berlin, Germany, 2015.
- [22] M. Vet, “Vibration analysis of thin rectangular plates,” *Machine Design*, vol. 39, no. 9, pp. 193–202, 1967.
- [23] T. Irvine, “Extending Steinberg’s Fatigue Analysis of Electronics Equipment Methodology to a Full Relative Displacement vs. Cycles Curve, Revision C, Vibrationdata,” 2013.

Research Article

Distance and Density Similarity Based Enhanced k -NN Classifier for Improving Fault Diagnosis Performance of Bearings

Sharif Uddin,¹ Md. Rashedul Islam,² Sheraz Ali Khan,¹ Jaeyoung Kim,¹
Jong-Myon Kim,¹ Seok-Man Sohn,³ and Byeong-Keun Choi⁴

¹School of Electrical, Electronics and Computer Engineering, University of Ulsan, Ulsan, Republic of Korea

²Department of Computer Science and Engineering, University of Asia Pacific, Dhaka, Bangladesh

³Power Generation Laboratory, KEPCO Research Institute, Jeollanam-do, Republic of Korea

⁴Department of Energy Mechanical Engineering, Gyeongsang National University, Gyeongsangnam-do, Republic of Korea

Correspondence should be addressed to Jong-Myon Kim; jongmyon.kim@gmail.com

Received 1 September 2016; Accepted 17 October 2016

Academic Editor: Lu Chen

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

An enhanced k -nearest neighbor (k -NN) classification algorithm is presented, which uses a density based similarity measure in addition to a distance based similarity measure to improve the diagnostic performance in bearing fault diagnosis. Due to its use of distance based similarity measure alone, the classification accuracy of traditional k -NN deteriorates in case of overlapping samples and outliers and is highly susceptible to the neighborhood size, k . This study addresses these limitations by proposing the use of both distance and density based measures of similarity between training and test samples. The proposed k -NN classifier is used to enhance the diagnostic performance of a bearing fault diagnosis scheme, which classifies different fault conditions based upon hybrid feature vectors extracted from acoustic emission (AE) signals. Experimental results demonstrate that the proposed scheme, which uses the enhanced k -NN classifier, yields better diagnostic performance and is more robust to variations in the neighborhood size, k .

1. Introduction

Rotary machines, in both industry and common households, use bearings to reduce friction and ensure steady and energy efficient operation. Bearings reduce the noise and vibration levels associated with a machine, which is essential for the long term health of both the machine and its operators. Although bearings are very sturdy components and have very long useful lives; nevertheless, material fatigue due to variations in operating load, currents due to electric discharge, thermal stresses due to variations in operating temperature, corrosion, and contaminants in the operating environment can cause them to fail abruptly. A bearing failure can result in the abrupt shutdown of a machine, which leads to tremendous financial losses. Bearings account for more than 50% of failures in induction motors alone [1], which makes their condition monitoring essential to preventing any abrupt failures. Thus, early and reliable detection of bearing defects is very important as these defects lead to bearing failure.

Many data driven techniques have been proposed for diagnosing faults in bearings. These techniques largely use time-frequency analysis of the fault signals for the extraction of meaningful information about underlying faults [2, 3]. Fault signals, such as stator current, vibration acceleration, and acoustic emissions, are inherently nonstationary and hence they are processed in the time-frequency domain, using the short-time Fourier transform (STFT) [4], wavelet transforms [5–10], empirical mode decomposition (EMD) [11–15], and the Hilbert-Huang transform [16–18], to extract characteristic information about different bearing defects. Acoustic emissions are characterized by their low energies and very high bandwidths. They are captured using wide-band acoustic sensors and are very effective in diagnosing nascent faults [19–21]. This paper presents a data driven approach for fault diagnosis in bearings, which extracts hybrid features from the acoustic emission (AE) signals and then employs the proposed enhanced k -NN classifier to diagnose different bearing defects.

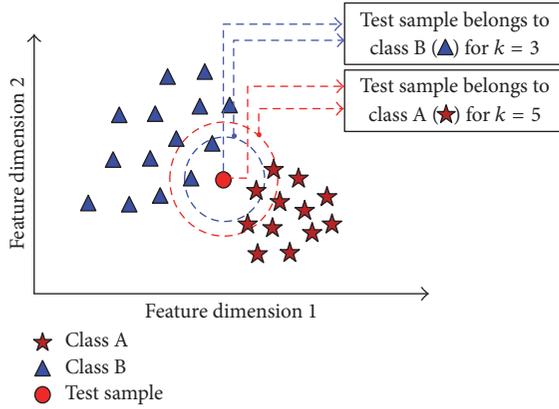


FIGURE 1: Limitation of classical k -NN using only the distance based similarity measure.

The hybrid feature vectors are constructed by calculating different statistical measures of the time and frequency domain AE signal and its envelope power spectrum. This rather extensive set of features is constructed to uniquely identify each fault condition; nevertheless, all features are not of equal utility in classifying a given fault correctly. Moreover, a high dimensional feature vector is bound to make the classification process computationally more expensive. Furthermore, if the feature vector contains too many redundant or irrelevant features, it may also degrade the classifier's accuracy. Hence, the dimensionality of the feature vector is reduced using feature selection methods, which prune the high dimensional feature vector by eliminating the suboptimal features and selecting only those, which would result in the highest classification accuracy. These optimal features are used to create a model of the data by training a classifier, which is then employed to classify the unknown fault signals.

Due to its simplicity and effectiveness, k -NN is usually the first choice in solving any classification problem. However, two factors can degrade its performance. First, k -NN determines the similarity between two samples using only a distance measure of similarity; the widely used distance measures are the Euclidean and Manhattan distance. Second, the classification decision and hence accuracy are sensitive to the neighborhood size, k . These problems have been highlighted in Figure 1, where the classification decision for the unknown test sample (shown as a red circle) changes with change in the neighborhood size. The test sample is labeled as "B" if $k = 3$, whereas it is labeled as "A" if $k = 5$. The limitations of traditional k -NN, due to its use of distance based similarity measure, can be overcome using the local outlier factor (LOF) [22, 23] and local correlation integral (LOCI) [24], which are measures of similarity, based on the density of data samples. Hence, in this study, hybrid similarity measures (i.e., both distance and density based) are proposed to improve the diagnostic performance of the classical k -NN and make it more resilient to the choice of neighborhood size, k .

The main contribution of this study is that an enhanced k -NN classifier is proposed, which uses hybrid measures of similarity between data samples to make it more resilient

to the choice of neighborhood size, k , and to increase its diagnostic performance relative to classical k -NN. The density based similarity measure (i.e., LOF) is used to boost the decision of classical k -NN, which classifies an unknown sample based only upon its Euclidean distance from its " k " nearest neighbors using the majority rule. In the proposed k -NN, when the k nearest neighbors of an unknown sample do not belong to the same class, then the LOF is used to decide the class membership of the unknown simple.

The organization of the rest of the paper is as follows. In Section 2, the fault simulator and data acquisition setup are presented. In Section 3, the fault diagnosis scheme and the proposed enhanced k -NN classifier are discussed in detail. In Section 4, a discussion of the achieved results is provided, whereas, in Section 5, conclusions of this work are provided.

2. Fault Simulator and Data Acquisition System

The acoustic emission (AE) signals are acquired using a machinery fault simulator, which is used to simulate different fault conditions. The fault simulator uses cylindrical roller element bearings (FAG NJ206-E-TVP2), which are ingrained with cracks on its different parts. AE signals are collected for bearings at the nondrive end of the simulator using a wide-band acoustic sensor and a PCI-2 based data acquisition system, which samples the AE signals at a rate of 250 KHz [25]. The acoustic sensor is connected to the top of the bearing housing and is at an approximate distance of 21.48 mm from the bearing, as shown in Figure 2. The nondrive end shaft is connected to the drive end through a gearbox with a reduction ratio of 1.52 : 1.

The bearings are seeded with cracks of two different sizes (e.g., 3 mm and 12 mm), and these cracks are introduced on either one or two components of the bearing to study both single and compound bearing defects. The AE signals recorded for bearings with 3 mm cracks and for bearings with 12 mm cracks are grouped into separate datasets. Moreover, for each crack size, the AE signals are recorded at two different shaft speeds (e.g., 300 RPM and 350 RPM). Thus, a total of four datasets are considered, each with AE signals recorded at a different shaft speed along with different crack sizes. The types of single and compound bearing defects are shown in Figure 3; they include cracks on the roller (BFR), inner raceway (BFI), outer raceway (BFO), inner and outer raceways (BFIO), inner raceway and roller (BFIR), outer raceway and roller (BFOR), and both inner and outer raceways and the roller (BFIOR). For each shaft speed, AE signal for a healthy bearing (FFB) is also recorded.

As mentioned earlier, the AE signals are divided into 4 datasets based upon the crack size and shaft speed, as given in Table 1. For every bearing defect, 90 AE signals are recorded; each signal is of 5-second duration. Similarly, 90 AE signals are recorded for the healthy bearing. Thus, every dataset contains a total of 720 AE signals.

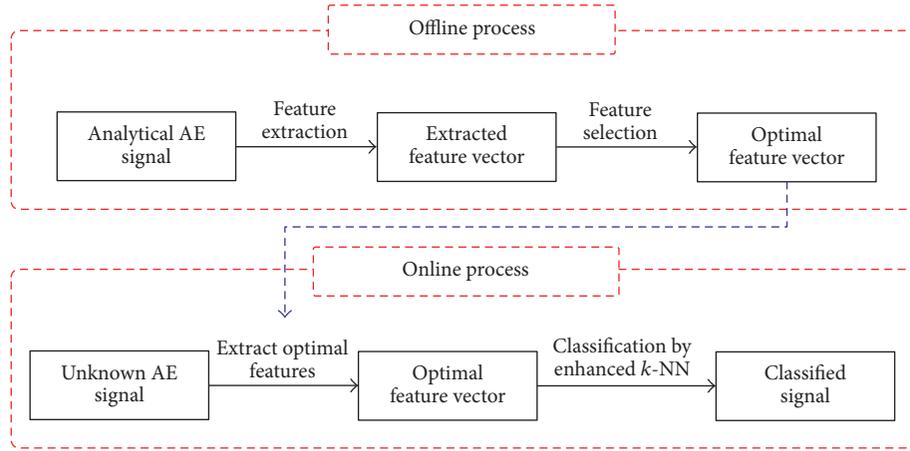


FIGURE 4: The proposed methodology for bearing fault diagnosis.

3. The Proposed Methodology for Bearing Fault Diagnosis

The proposed methodology for bearing fault diagnosis works in two phases, as illustrated in Figure 4. The first phase comprises an *offline process* that involves feature extraction and feature selection, which are discussed in detail in Sections 3.1 and 3.2, respectively. The offline process is used to determine the set of optimal features that would yield the highest classification accuracy. In the second phase, an *online process* is used to classify the unknown AE signals using the proposed enhanced k -NN classifier. The online process calculates only the optimal set of features for each AE signal and, using only those features, it labels the unknown AE signals.

3.1. Features Extraction. In order to accurately identify each bearing defect, a high dimensional hybrid feature vector is constructed using 22 different features of the AE signal. These features are useful in extracting maximum information about each fault [26] and include ten statistical measures of the time-domain AE signal and three statistical measures of the frequency domain AE signal. These features are listed in Table 2 along with the mathematical relationships for their calculation. Moreover, nine statistical measures, calculated over the envelope power spectrum of the AE signal, are also included in the hybrid feature vector. The features from the envelope power spectrum include the root mean square (RMS) values for each of the three defect frequencies and its first two harmonics. The defect frequencies include the ball pass frequency over inner race (BPFI), the ball pass frequency over the outer race (BPFO), and the ball spin frequency (BSF). The range of values for these defect frequencies and their harmonics is shown in Figure 5.

The range of values for the defect frequencies and their first two harmonics is calculated using (1), (2), and (3), respectively.

$$r_{\text{inner}} = 2 \times \{n_{\text{sidebands}} \times (f_s + f_s \times e_{\text{rate}}) + e_{\text{rate}} \times f_i\}, \quad (1)$$

$$r_{\text{outer}} = 2 \times e_{\text{rate}} \times f_o, \quad (2)$$

$$r_{\text{roller}} = 2 \times \{n_{\text{sidebands}} \times (f_c + f_c \times e_{\text{rate}}) + e_{\text{rate}} \times f_r\}, \quad (3)$$

TABLE 2: Statistical measures calculated over the time and frequency domain AE signal.

Parameter	Definition
Root mean square (RMS)	$\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$
Kurtosis value (KV)	$\frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma} \right)^4$
Peak-to-peak value (PPV)	$\text{PPV} = \max(x_i) - \min(x_i)$
Crest factor (CF)	$\frac{\max(x_i)}{\sqrt{(1/N) \sum_{i=1}^N x_i^2}}$
Shape factor (SF)	$\frac{\sqrt{(1/N) \sum_{i=1}^N x_i^2}}{(1/N) \sum_{i=1}^N x_i }$
Frequency center (FC)	$\frac{1}{N} \sum_{i=1}^N f_i$
RMS frequency (RMSF)	$\sqrt{\frac{1}{N} \sum_{i=1}^N f_i^2}$
Square root of amplitude (SRA)	$\left(\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i } \right)^2$
Skewness value (SV)	$\frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma} \right)^3$
Impulse factor (IF)	$\frac{\max(x)}{(1/N) \sum_{i=1}^N x_i }$
Margin factor (MF)	$\frac{\max(x_i)}{\left((1/N) \sum_{i=1}^N \sqrt{ x_i } \right)^2}$
Kurtosis factor (KF)	$\text{KF} = \frac{(1/N) \sum_{i=1}^N ((x_i - \bar{x})/\sigma)^4}{\left((1/N) \sum_{i=1}^N x_i^2 \right)^2}$
Root variance frequency (RVF)	$\sqrt{\frac{1}{N} \sum_{i=1}^N (f_i - f_c)^2}$

where $n_{\text{sidebands}}$ is the number of sidebands, f_s is the operating frequency, e_{rate} is the error rate, f_i is the inner defect

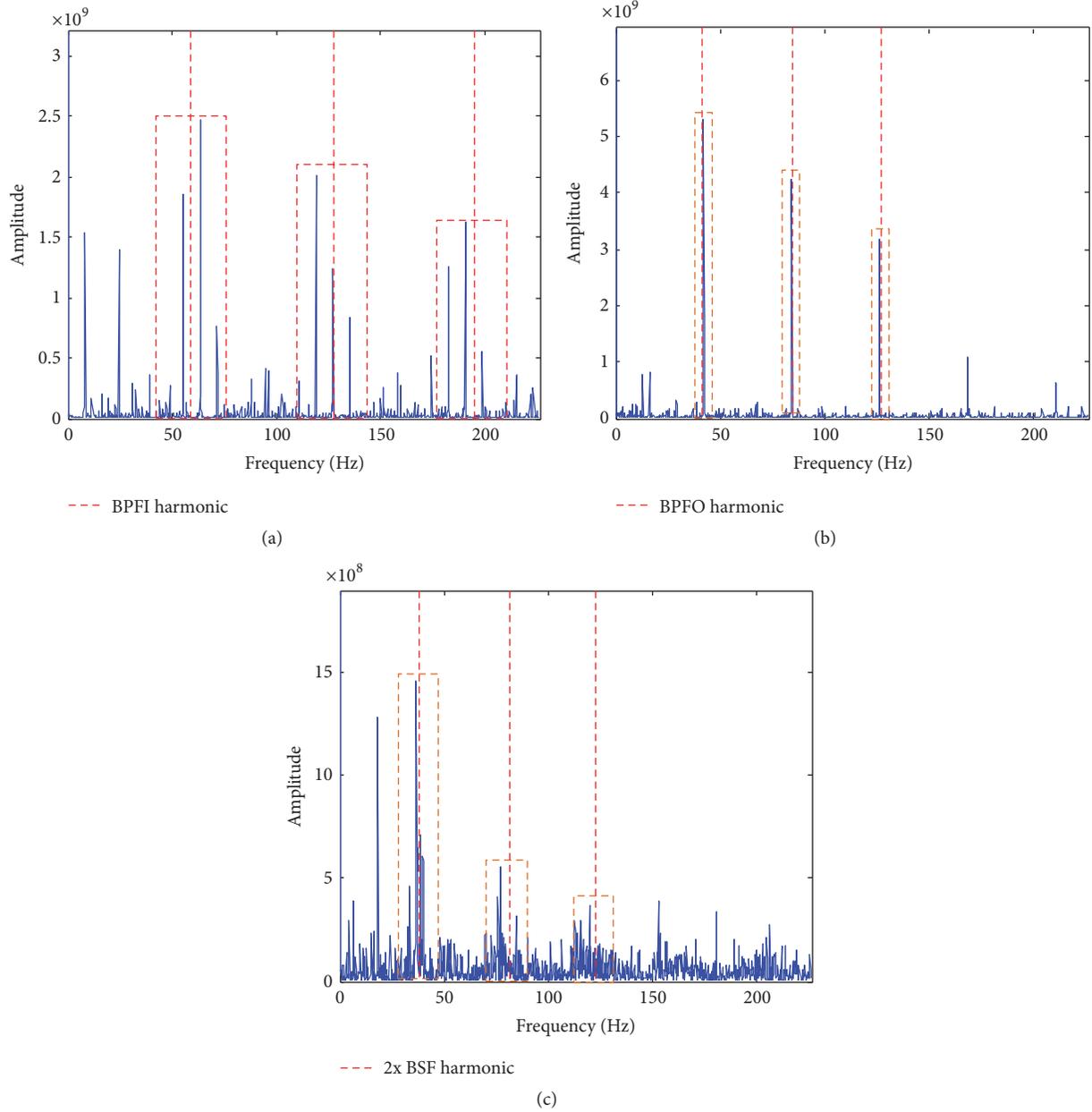


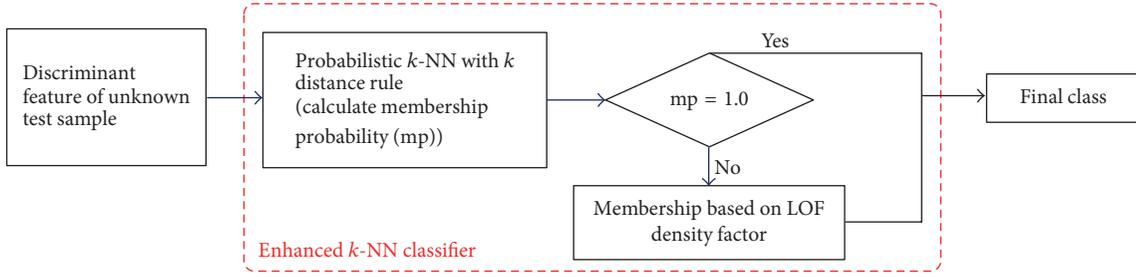
FIGURE 5: Fault frequency regions up to three harmonics at (a) inner, (b) outer, and (c) roller fault frequency.

frequency, f_o is the outer defect frequency, f_c is the cage frequency, and f_r is the roller defect frequency.

3.2. Feature Selection. Although a high dimensional hybrid feature vector is highly desirable to capture the characteristics of different types of defects, the diagnostic performance of the proposed method can be degraded by potentially irrelevant and redundant features. Moreover, a high dimensional feature vector entails an increased computational cost during feature extraction and classification, which involves the calculation of distances and densities between different samples [25–27]. Hence, the original feature vector is evaluated to determine the set of optimal features that would yield the best diagnostic

performance and reduce the computational cost of the proposed method.

In this study, sequential forward selection (SFS) is used for feature selection, which is a simple and fast greedy search algorithm. It starts with an initially empty set, $S = 0$, and then iteratively selects the most significant feature from the original set with respect to the set, S . This is done by first selecting a feature from the original set and then adding it to the set, S , only if the newly selected feature maximizes the value of the objective function for the set, S . The feature is discarded and the process moves to the next feature, if the selected feature decreases the value of the objective function for the set, S . The objective function for SFS is given by (4), which is basically

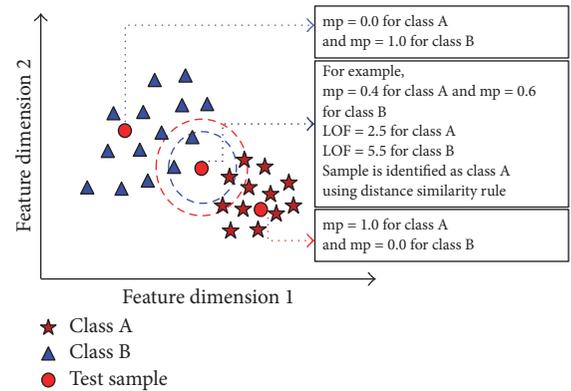
FIGURE 6: The proposed k -NN classifier.

the ratio of interclass separability to intraclass compactness [25]. The interclass separability is given by the interclass distance $d_{\text{inter_class}}$, whereas $d_{\text{intra_class}}$ is the intraclass compactness. Although SFS is simple, efficient, and reasonably accurate, it has its own disadvantages. It suffers from the nesting problem; that is, a feature retained once cannot be discarded, which can result in suboptimal feature selection [28–30].

$$f_{\text{objective}} = \frac{d_{\text{inter_class}}}{d_{\text{intra_class}}}. \quad (4)$$

3.3. Enhanced k -NN Classification Algorithm. The traditional k -NN classifier labels an unknown test sample according to the majority of its nearest neighbors in the training set. The nearest neighbors are determined using a distance measure, which is mostly the Euclidean distance between two samples. In multiclass classification problems, where the density of each class is different, the use of a distance based measure of similarity between the test and training samples can result in misclassification and render the classification result sensitive to the choice of neighborhood size, k , as illustrated in Figure 1. This happens because traditional k -NN does not take into account variation in densities across different classes. Therefore, an enhanced k -NN classifier is proposed, which uses both distance and density based similarity measures to improve its classification accuracy. For a given test sample, first its membership probabilities for different classes are calculated. This is done through voting by its k nearest neighbors, which in turn are determined using the Euclidean distance of the test sample from all the training samples. If the membership probability for the test sample is one (i.e., all its nearest neighbors belong to a single class), then the proposed k -NN classifier admits this result and labels the test sample according to its nearest neighbors. However, if the membership probability of the test sample is less than one, (i.e., all the nearest neighbors do not belong to a single class), then the LOF based density measure is used to determine the label of the test sample. The use of LOF in conjunction with Euclidean distance makes the classification performance, of the enhanced k -NN, insensitive to the neighborhood size, k .

As shown in Figure 6, the proposed k -NN first calculates the membership probabilities for the unknown test samples using probabilistic k -NN, which uses Euclidean distance as a

FIGURE 7: Classifying a test sample using the enhanced k -NN classifier.

measure of similarity. The probabilistic k -NN does not assign any class labels to the test samples; instead it only calculates their membership probabilities for all the classes.

If, for each class, the membership probability of a test sample is less than 1.0, then the output of the majority rule is ignored and the final membership of the test sample is determined using the LOF value, as shown in Figure 7.

3.4. Calculating the Local Outlier Factor (LOF). The local outlier factor (LOF) has been used for the detection of outliers or anomalous data points [22], which have relatively lower probabilities of being members of any class. An unknown sample is classified by comparing its density with that of its neighbors. Points with densities like their neighbors are classified accordingly; that is, points with lower densities are labeled according to their neighbors with lower densities, whereas points with higher densities are labeled according their neighbors with higher densities. The LOF can be calculated as follows:

- (i) *First*, the calculation of the distance of every data point “ q ” to its k th nearest neighbor (i.e., d_k^q is calculated), for $k = 3$, is illustrated in Figure 8(a).
- (ii) *Second*, for each data point “ q ”, its reachability distance with respect to the data point “ p ” (i.e., $d_r^{q:p}$ is calculated) is the true distance between points “ p ”

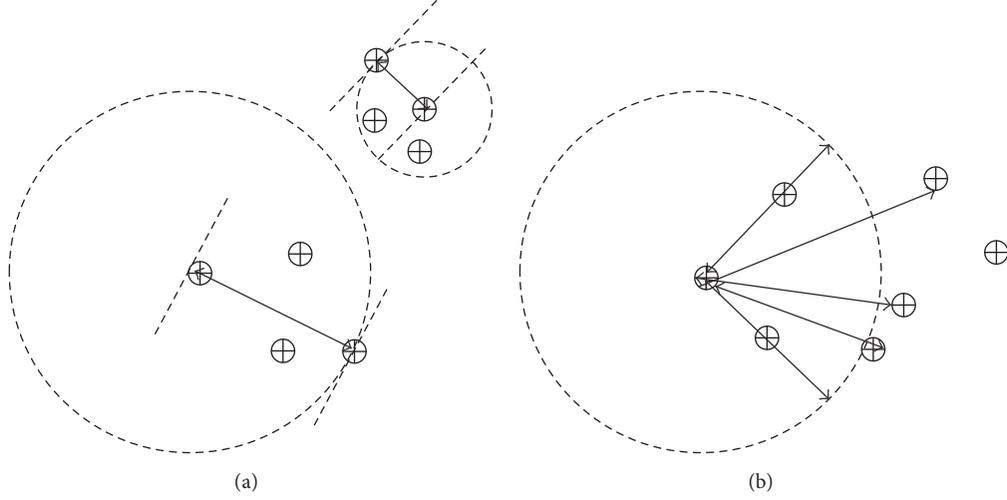


FIGURE 8: For $k = 3$, calculating (a) the k -distance and (b) the reachability distance.

and “ q ” with a minimum value of d_k^q , as illustrated in Figure 8(b). It can be calculated as follows:

$$d_r^{q,p} = \max_i \langle d_k^q, d_r^{q,p} \rangle. \quad (5)$$

- (iii) *Third*, for each data point “ q ”, its local reachability density (i.e., σ_r^q is calculated) is defined as the inverse of its average reachability distance from its “ M ” nearest neighbors, as given in (6). The value of “ M ” is set to 16, as given in Table 3:

$$\sigma_r^q = \frac{M}{\sum_M d_r^{q,p}}. \quad (6)$$

- (iv) *Finally*, for each data point “ q ”, its local outlier factor or LOF value is determined, by comparing its local reachability density to that of its “ M ” nearest neighbors using the following relation:

$$\text{LOF} = \frac{1}{M} \sum_{p=1}^M \frac{\sigma_r^p}{\sigma_r^q}. \quad (7)$$

The LOF values for all the training samples are computed using (7) during the training phase. The unknown test samples are classified based upon the similarity of their LOF values to that of their neighbors.

4. Results and Discussion

In this section, a discussion of the experimental results achieved by the proposed method for bearing fault diagnosis is provided. As mentioned earlier, four datasets are used to test the proposed method, details of which are given in Table 1. The method uses the enhanced k -NN classifier, which has been proposed to address the limitations of traditional k -NN. The enhanced k -NN classifier was used with the parameters given in Table 3.

TABLE 3: Values of various parameters for the enhanced k -NN classifier.

Property	Value
Neighborhood size for k -NN	3, 5, 7, and 9
Neighborhood size for local reachability density	16
Neighborhood size for LOF	12
Outlier threshold	$>2\sigma$

To demonstrate the effectiveness of the proposed k -NN classifier, the classification of inner race fault samples from dataset 1 is illustrated in Figure 9, using both the traditional and proposed k -NN classifiers with neighborhood sizes of 3 and 7 (i.e., $k = 3$ and $k = 7$). The samples shown inside the red ellipse are to be classified; their true label is “*inner_race_fault*” (i.e., these samples belong to the inner race fault class). However, the classification result of the traditional k -NN classifier varies with the value of k (i.e., for $k = 3$); it correctly classifies these samples as inner race fault samples, whereas, for $k = 7$, it classifies these as outer race fault samples, which is incorrect. It happens because traditional k -NN uses the majority rule to decide the class label for an unknown test sample. In this particular case, among the nearest three neighbors of these unknown test samples, two are inner race fault and one is outer race fault. Hence, for the case of $k = 3$, they are correctly classified as inner race fault samples. However, among the nearest seven neighbors of these unknown test samples, four are outer race fault and three are inner race fault. Hence, for the case of $k = 7$, they are incorrectly classified as outer race fault samples. In contrast, the proposed k -NN always classifies these samples as inner race fault samples, irrespective of the size of neighborhood (i.e., the value of k).

The proposed k -NN classifier correctly classifies these unknown test samples because it uses the LOF, which is a density based similarity measure. LOF is used only when the nearest neighbors of a given test sample do not belong to

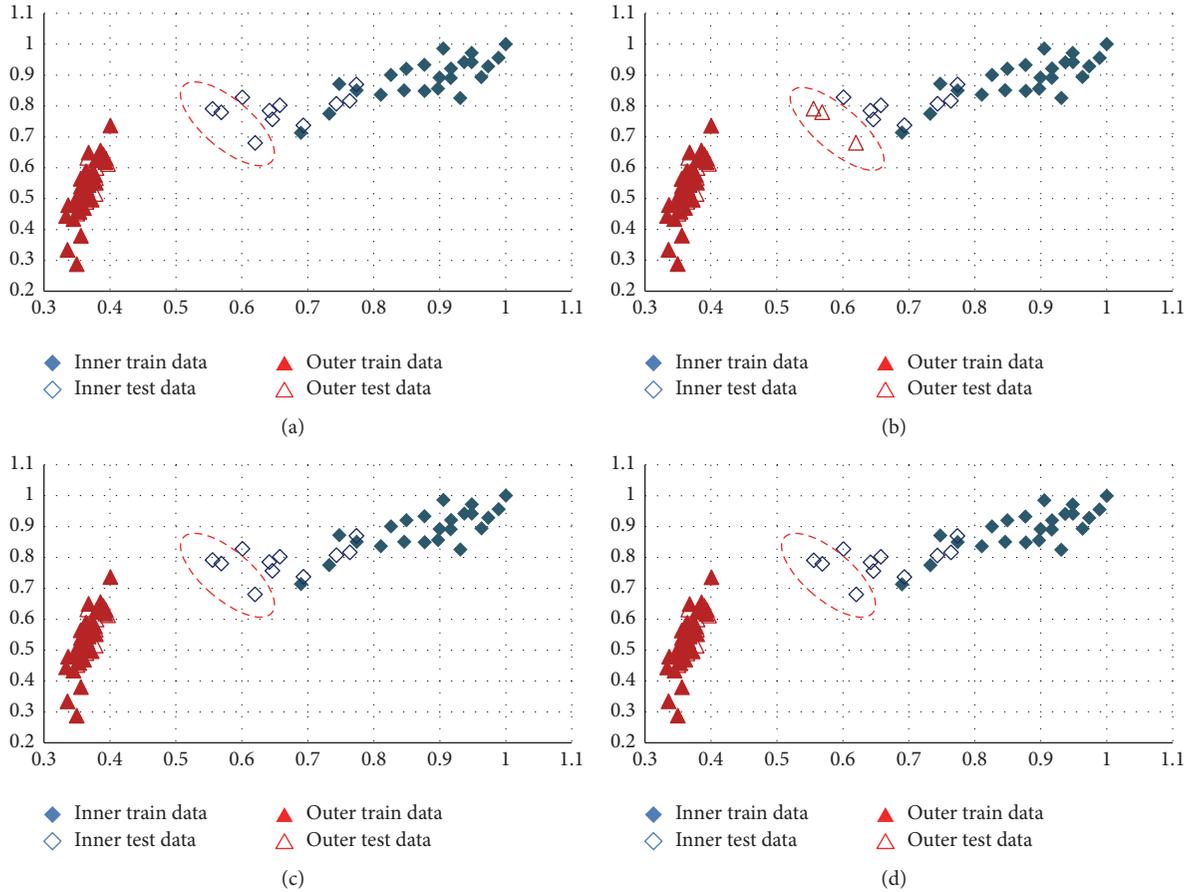


FIGURE 9: Classification of inner race fault samples from dataset 1 using the traditional k -NN classifier (a) with $k = 3$ and (b) with $k = 7$ and using the proposed k -NN classifier (c) with $k = 3$ and (d) with $k = 7$.

the same class (i.e., the vote is not unanimous). Therefore, the class membership probabilities for the unknown test samples are determined. In this particular case, for $k = 3$, the probability that a given test sample is a member of the inner race fault is 66.7%, and the probability that it belongs to the outer race fault is 33.33%. Since both class membership probabilities are less than one, the proposed k -NN classifier employs the LOF values of the unknown test samples and their neighbors to determine the final class labels. This is demonstrated in Figure 10, which shows the LOF values for the test samples and their nearest neighbors. The LOF values of the test samples for outer race fault class are 5.09, 5.069, and 4.979, whereas, for the inner race fault class, their LOF values are 3.33, 3.399, and 3.192, respectively. If the LOF values of these test samples for both the outer and inner race fault classes are compared to the LOF values of their nearest training samples, it can be observed that the LOF values of the test samples for inner race fault are similar to the LOF values of training samples from the inner race fault class. Hence, it can be argued that these test samples are outliers to the outer race fault class and inliers to or members of the inner race fault class.

Similarly, when $k = 7$, the probability that a given test sample is a member of the inner race fault is 42.86%, and

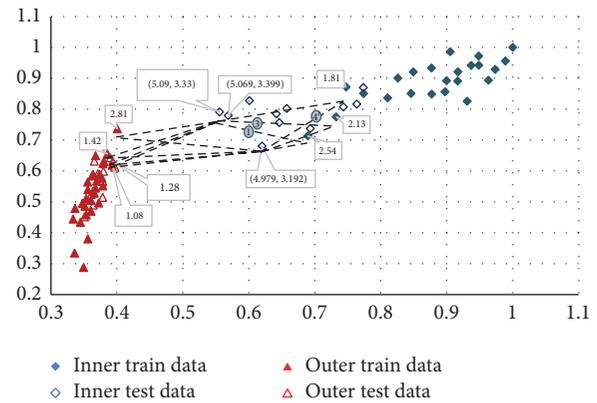


FIGURE 10: The classification of unknown test samples using LOF based density similarity measure.

the probability that it belongs to the outer race fault is 57.14%. Here again, the class membership probabilities are less than one, and, thus, the proposed k -NN classifier employs the LOF values of the unknown test samples and their neighbors to determine the final class labels. Using the LOF values of the test samples and their nearest training samples, the test samples are classified as members of the inner race fault class.

TABLE 4: Diagnostic performance of the two classifiers for different fault types and datasets.

$k = ?$	Model	BFI	BFO	BFR	BFIO	BFOR	BFIR	BFIOR	FFB	Avg. classification accuracy (%)
<i>Dataset 1: 300 rpm, 3 mm crack</i>										
$k = 3$	k -NN	94.65	79.50	96.42	98.53	95.67	97.08	91.67	100.00	94.19
	Enhanced k -NN	95.70	87.24	98.75	99.58	96.25	99.92	93.08	100.00	96.32
$k = 5$	k -NN	72.52	75.61	92.35	91.35	89.25	96.08	94.53	100.00	88.96
	Enhanced k -NN	96.25	92.35	99.45	98.25	99.50	97.92	96.25	100.00	97.50
$k = 7$	k -NN	98.76	78.50	97.75	96.25	90.75	94.92	87.42	99.83	93.02
	Enhanced k -NN	99.05	94.56	98.65	98.96	98.42	97.56	91.56	99.89	97.33
$k = 9$	k -NN	89.76	80.56	91.53	97.45	95.26	98.12	92.54	100.00	93.15
	Enhanced k -NN	98.68	94.56	98.16	98.56	99.92	99.92	93.89	100.00	97.96
<i>Dataset 2: 350 rpm, 3 mm crack</i>										
$k = 3$	k -NN	93.56	78.52	94.25	97.86	94.58	97.12	92.35	98.68	93.37
	Enhanced k -NN	95.59	88.45	99.02	98.95	96.12	99.43	93.56	99.26	96.30
$k = 5$	k -NN	73.48	76.53	90.89	90.75	87.62	96.89	95.42	100.00	88.95
	Enhanced k -NN	96.45	93.25	98.45	98.75	98.69	96.25	95.86	100.00	97.21
$k = 7$	k -NN	98.48	81.52	96.53	95.48	91.48	93.28	89.45	99.89	93.26
	Enhanced k -NN	99.75	95.48	98.15	98.75	97.98	98.06	92.56	100.00	97.59
$k = 9$	k -NN	94.58	84.83	94.25	96.53	94.56	97.85	92.56	100.00	94.40
	Enhanced k -NN	98.65	95.69	98.75	98.45	99.05	98.45	94.12	100.00	97.90
<i>Dataset 3: 300 rpm, 12 mm crack</i>										
$k = 3$	k -NN	97.05	86.52	98.45	98.53	96.98	97.08	93.58	100.00	96.02
	Enhanced k -NN	97.45	93.56	100.00	99.58	98.46	99.92	95.26	100.00	98.03
$k = 5$	k -NN	78.96	81.45	91.75	92.86	88.46	95.63	94.53	100.00	90.46
	Enhanced k -NN	98.96	95.86	99.26	99.43	99.50	98.76	96.25	100.00	98.50
$k = 7$	k -NN	98.76	86.46	97.75	96.25	95.36	96.46	94.85	99.83	95.72
	Enhanced k -NN	100.00	98.12	100.00	99.79	99.86	98.75	96.53	100.00	99.13
$k = 9$	k -NN	89.76	81.65	91.47	96.85	94.62	97.45	90.67	100.00	92.81
	Enhanced k -NN	100.00	98.45	100.00	100.00	99.45	99.92	93.89	100.00	98.96
<i>Dataset 4: 350 rpm, 12 mm crack</i>										
$k = 3$	k -NN	99.16	99.56	99.48	100.00	100.00	98.54	98.65	100.00	99.42
	Enhanced k -NN	99.86	100.00	100.00	100.00	100.00	99.75	99.46	100.00	99.88
$k = 5$	k -NN	89.45	80.54	93.25	92.45	90.74	95.86	98.56	100.00	92.61
	Enhanced k -NN	100.00	100.00	99.86	100.00	99.94	99.75	99.46	100.00	99.88
$k = 7$	k -NN	98.76	89.45	97.75	96.25	94.52	97.63	96.25	99.83	96.31
	Enhanced k -NN	99.80	100.00	100.00	100.00	100.00	99.75	99.70	100.00	99.91
$k = 9$	k -NN	97.84	84.25	96.57	94.56	95.26	98.12	95.68	100.00	95.29
	Enhanced k -NN	100.00	100.00	99.86	100.00	99.94	99.75	99.46	100.00	99.88

Likewise, for other datasets and fault types, this is how the proposed k -NN classifier improves the classification accuracy of traditional k -NN. It is clearly evident in Figure 11, which compares the performance of these two classifiers in terms of average classification accuracy, and Table 4, which lists the classification accuracies for each dataset and individual fault type. Moreover, it can also be observed that the accuracy of the proposed k -NN is not affected by the neighborhood size, k , whereas the accuracy of traditional k -NN varies with variations in the neighborhood size, k . It achieves a maximum accuracy for $k = 3$.

The size of the optimal neighborhood, which maximizes the classification accuracy of traditional k -NN, has to be determined on a case to case basis. There are no general rules that work equally well in all situations and for all classes, which can be challenging as it makes the whole process computationally expensive and inflexible. The robustness of the proposed k -NN to variations in the neighborhood size, k , makes it more flexible and efficient to use. It delivers better and steadier performance. Moreover, in multiclass problems like the one considered in this study, where the densities of different classes vary, traditional k -NN performs poorly as it

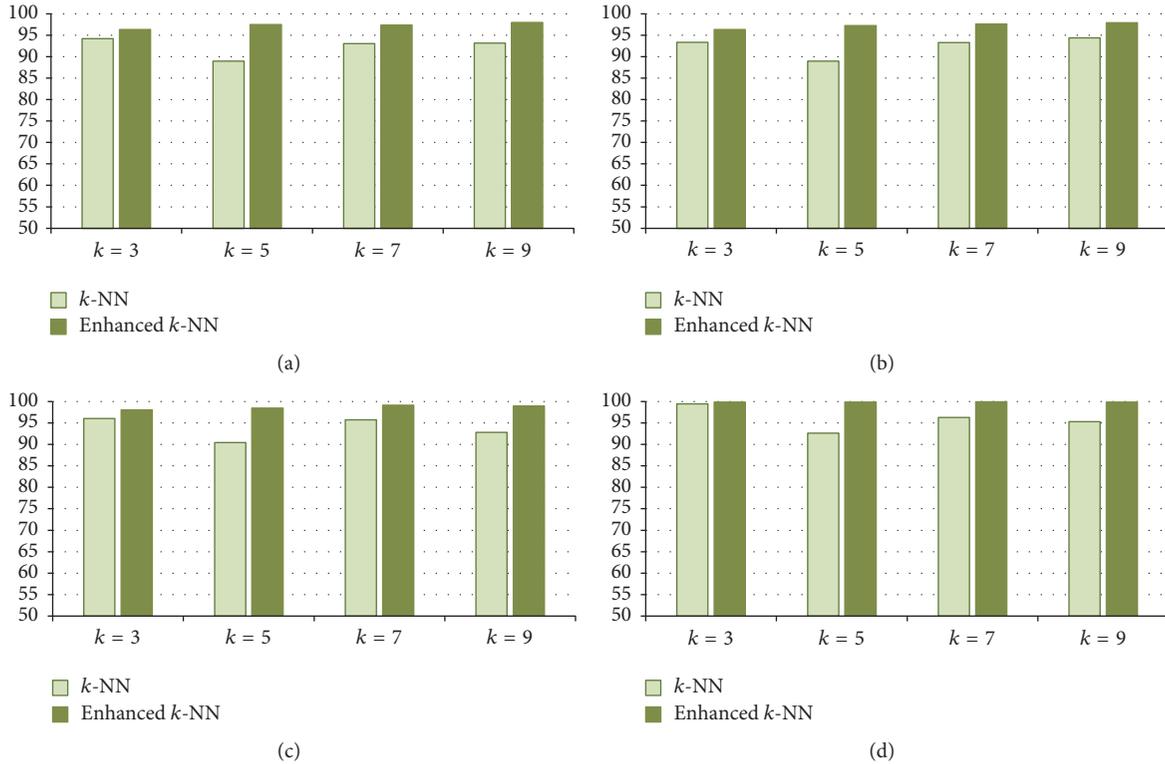


FIGURE 11: Performance comparison of traditional k -NN and the proposed enhanced k -NN in terms of average classification accuracy: (a) dataset 1, (b) dataset 2, (c) dataset 3, and (d) dataset 4.

does not consider variations in density. The proposed k -NN takes into account variations in density of different classes and uses the LOF to decide the class membership of test samples in such cases.

5. Conclusion

In this paper, an enhanced k -nearest neighbor (k -NN) classification algorithm was presented, which employs both density and distance based similarity measures to improve the diagnostic performance in bearing fault diagnosis. The density based similarity measure, LOF, was used to boost the classification performance of traditional k -NN, which deteriorates in case of overlapping samples, outliers, and multiple classes that show different feature distributions. Moreover, the distance based similarity measure makes the classification performance of traditional k -NN highly susceptible to the neighborhood size, k . These limitations were addressed through the use of both distance and density based similarity metrics, between the training and test samples. Using the enhanced k -NN classifier, the diagnostic performance of the proposed bearing fault diagnosis scheme was significantly improved, and the results were more robust to variations in the neighborhood size, k .

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by the development of a basic fusion technology in electric power industry (Ministry of Trade, Industry & Energy, 201301010170D); funded in part by The Leading Human Resource Training Program of Regional Neo Industry through the National Research Foundation of Korea (NRF), Ministry of Science, ICT, and Future Planning (NRF-2016H1D5A1910564); funded in part by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry, & Energy (MOTIE) of the Republic of Korea (no. 20162220100050); funded in part by the Business for Startup R&D funded by the Korea Small and Medium Business Administration in 2016 (Grants S2381631 and C0395147); and funded in part by the “Leaders INdustry-university Cooperation” Project supported by the Ministry of Education (MOE).

References

- [1] E. Cabal-Yepez, M. Valtierra-Rodriguez, R. J. Romero-Troncoso et al., “FPGA-based entropy neural processor for online detection of multiple combined faults on induction motors,” *Mechanical Systems and Signal Processing*, vol. 30, pp. 123–130, 2012.
- [2] H. Berriri, M. W. Naouar, and I. Slama-Belkhdja, “Easy and fast sensor fault detection and isolation algorithm for electrical drives,” *IEEE Transactions on Power Electronics*, vol. 27, no. 2, pp. 490–499, 2012.

- [3] E. Cabal-Yepez, A. G. Garcia-Ramirez, R. J. Romero-Troncoso, A. Garcia-Perez, and R. A. Osornio-Rios, "Reconfigurable monitoring system for time-frequency analysis on industrial equipment through STFT and DWT," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 2, pp. 760–771, 2013.
- [4] S. Nandi, T. C. Ilamparithi, S. B. Lee, and D. Hyun, "Detection of eccentricity faults in induction machines based on nameplate parameters," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 1673–1683, 2011.
- [5] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: a review with applications," *Signal Processing*, vol. 96, pp. 1–15, 2014.
- [6] J. Seshadrinath, B. Singh, and B. K. Panigrahi, "Investigation of vibration signatures for multiple fault diagnosis in variable frequency drives using complex wavelets," *IEEE Transactions on Power Electronics*, vol. 29, no. 2, pp. 936–945, 2014.
- [7] P. K. Kankar, S. C. Sharma, and S. P. Harsha, "Fault diagnosis of rolling element bearing using cyclic autocorrelation and wavelet transform," *Neurocomputing*, vol. 110, pp. 9–17, 2013.
- [8] P. Konar and P. Chattopadhyay, "Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs)," *Applied Soft Computing Journal*, vol. 11, no. 6, pp. 4203–4211, 2011.
- [9] J. Rafiee, M. A. Rafiee, and P. W. Tse, "Application of mother wavelet functions for automatic gear and bearing fault diagnosis," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4568–4579, 2010.
- [10] P. H. Nguyen and J.-M. Kim, "Multifault diagnosis of rolling element bearings using a wavelet kurtogram and vector median-based feature analysis," *Shock and Vibration*, vol. 2015, Article ID 320508, 14 pages, 2015.
- [11] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 108–126, 2013.
- [12] J. Zheng, J. Cheng, and Y. Yang, "Generalized empirical mode decomposition and its applications to rolling element bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 40, no. 1, pp. 136–153, 2013.
- [13] X. Zhang and J. Zhou, "Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines," *Mechanical Systems and Signal Processing*, vol. 41, no. 1-2, pp. 127–140, 2013.
- [14] G. F. Bin, J. J. Gao, X. J. Li, and B. S. Dhillon, "Early fault diagnosis of rotating machinery based on wavelet packets—empirical mode decomposition feature extraction and neural network," *Mechanical Systems and Signal Processing*, vol. 27, no. 1, pp. 696–711, 2012.
- [15] M. Amarnath and I. R. Krishna, "Empirical mode decomposition of acoustic signals for diagnosis of faults in gears and rolling element bearings," *IET Science, Measurement and Technology*, vol. 6, no. 4, pp. 279–287, 2012.
- [16] J. Yan and L. Lu, "Improved Hilbert-Huang transform based weak signal detection methodology and its application on incipient fault diagnosis and ECG signal analysis," *Signal Processing*, vol. 98, pp. 74–87, 2014.
- [17] G. Cheng, Y.-L. Cheng, L.-H. Shen, J.-B. Qiu, and S. Zhang, "Gear fault identification based on Hilbert-Huang transform and SOM neural network," *Measurement*, vol. 46, no. 3, pp. 1137–1146, 2013.
- [18] S. Osman and W. Wang, "An enhanced Hilbert-Huang transform technique for bearing condition monitoring," *Measurement Science and Technology*, vol. 24, no. 8, Article ID 085004, 2013.
- [19] A. Widodo, B.-S. Yang, E. Y. Kim, A. C. C. Tan, and J. Mathew, "Fault diagnosis of low speed bearing based on acoustic emission signal and multi-class relevance vector machine," *Nondestructive Testing and Evaluation*, vol. 24, no. 4, pp. 313–328, 2009.
- [20] D. H. Pandya, S. H. Upadhyay, and S. P. Harsha, "Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN," *Expert Systems with Applications*, vol. 40, no. 10, pp. 4137–4145, 2013.
- [21] S. A. Niknam, V. Songmene, and Y. H. J. Au, "The use of acoustic emission information to distinguish between dry and lubricated rolling element bearings in low-speed rotating machines," *International Journal of Advanced Manufacturing Technology*, vol. 69, no. 9–12, pp. 2679–2689, 2013.
- [22] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers," *ACM SIGMOD Record*, vol. 29, no. 2, pp. 93–104, 2000.
- [23] E. Schubert, A. Zimek, and H.-P. Kriegel, "Local outlier detection reconsidered: a generalized view on locality with applications to spatial, video, and network outlier detection," *Data Mining and Knowledge Discovery*, vol. 28, no. 1, pp. 190–237, 2014.
- [24] S. Papadimitriou, H. Kitagawa, P. B. Gibbons, and C. Faloutsos, "LOCI: fast outlier detection using the local correlation integral," in *Proceedings of the 19th International Conference on Data Engineering*, pp. 315–326, March 2003.
- [25] M. Kang, J. Kim, B.-K. Choi, and J.-M. Kim, "Envelope analysis with a genetic algorithm-based adaptive filter bank for bearing fault detection," *The Journal of the Acoustical Society of America*, vol. 138, no. 1, pp. EL65–EL70, 2015.
- [26] R. Islam, S. A. Khan, and J.-M. Kim, "Discriminant feature distribution analysis-based hybrid feature selection for online bearing fault diagnosis in induction motors," *Journal of Sensors*, vol. 2016, Article ID 7145715, 16 pages, 2016.
- [27] I.-K. Jeong, M. Kang, J. Kim, J.-M. Kim, J.-M. Ha, and B.-K. Choi, "Enhanced DET-based fault signature analysis for reliable diagnosis of single and multiple-combined bearing defects," *Shock and Vibration*, vol. 2015, Article ID 814650, 10 pages, 2015.
- [28] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," *Pattern Recognition*, vol. 33, no. 1, pp. 25–41, 2000.
- [29] D. P. Muni, N. R. Pal, and J. Das, "Genetic programming for simultaneous feature selection and classifier design," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 36, no. 1, pp. 106–117, 2006.
- [30] M. Last, A. Kandel, and O. Maimon, "Information-theoretic algorithm for feature selection," *Pattern Recognition Letters*, vol. 22, no. 6-7, pp. 799–811, 2001.

Research Article

Improvement of Roller Bearing Diagnosis with Unlabeled Data Using Cut Edge Weight Confidence Based Tritraining

Wei-Li Qin,¹ Wen-Jin Zhang,¹ and Zhen-Ya Wang^{1,2}

¹*School of Reliability and System Engineering, Beihang University, Beijing, China*

²*Science & Technology on Reliability & Environmental Engineering Laboratory, Beijing, China*

Correspondence should be addressed to Zhen-Ya Wang; wangzy@buaa.edu.cn

Received 9 May 2016; Revised 20 October 2016; Accepted 26 October 2016

Academic Editor: Minvydas Ragulskis

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

Roller bearings are one of the most commonly used components in rotational machines. The fault diagnosis of roller bearings thus plays an important role in ensuring the safe functioning of the mechanical systems. However, in most cases of bearing fault diagnosis, there are limited number of labeled data to achieve a proper fault diagnosis. Therefore, exploiting unlabeled data plus few labeled data, this paper proposed a roller bearing fault diagnosis method based on tritraining to improve roller bearing diagnosis performance. To overcome the noise brought by wrong labeling into the classifiers training process, the cut edge weight confidence is introduced into the diagnosis framework. Besides a small trick called suspect principle is adopted to avoid overfitting problem. The proposed method is validated in two independent roller bearing fault experiment vibrational signals that both include three types of faults: inner-ring fault, outer-ring fault, and rolling element fault. The results demonstrate the desirable diagnostic performance improvement by the proposed method in the extreme situation where there is only limited number of labeled data.

1. Introduction

Roller bearings are one of the most commonly used components in rotational machines and their faults may lead to huge economic losses, environment pollution, and human casualties. Hence, the fault diagnosis of the roller bearing is vital to guarantee the smooth and safe functioning of the mechanical systems.

There are a great deal of researches on vibration-based fault diagnosis of roller bearings and several powerful diagnostic methods are available [1]. Li et al. [2] presented an approach for motor roller bearing fault diagnosis using neural networks. Seryasat et al. [3] brought forward a ball bearing fault diagnosis method using fast Fourier transform (FFT) and wavelet energy entropy mean and root mean square (RMS). Peng and Chiang [4] used C4.5 decision tree and random forest algorithm to diagnose the fault of ball bearing of three-phase induction motor. Jin et al. [5] introduced a bearing fault diagnosis method using trace ratio linear discriminant analysis. And Liu et al. [6] proposed an extended wavelet spectrum analysis technique to achieve a more positive assessment of bearing health conditions. In fact, all

these methods yield a rather excellent performance for fault diagnosis of different bearings. However, the data used in those methods are all labeled data, the kind that are already marked according to the bearing states. In the case of bearing fault diagnosis, however, the labeled data are quite expensive to obtain since they require human effort while large amount of unlabeled data is readily available. For a better practical value, the use of unlabeled data ought to be considered. Therefore, semisupervised learning, a technique that exploits unlabeled data plus few labeled data to train a good classifier, might be promising candidates in the area of roller bearing diagnosis when there is limited number of labeled data.

Good reviews [7, 8] have given out a good review of semisupervised classification methods. Among them, generative models, self-training, and cotraining are three classic semisupervised learning methods. Generative models specify a joint probability distribution over observation and label sequences and thus are used for modeling data. Nigam et al. applied the expectation maximization (EM) algorithm, a classic generative model, on mixture of multinomial distribution for the task of text classification. And the result showed the classifiers performed better than those trained only from

labeled data [9]. However, the generative model must be carefully constructed to reflect reality; otherwise unlabeled data that are supposed to help may actually hurt accuracy [10]. Self-training is a technique where a classifier is first trained from the small amount of labeled data and then used to classify the unlabeled data that will be added to the training set for further retraining. Rosenberg et al. [11] applied self-training to object detection systems from images and show the semisupervised technique compares favorably with a state-of-the-art detector. But self-training suffers from wrong labeling; note that the classifier uses its own predictions to teach itself [12]. Cotraining, proposed by Blum and Mitchell [13], can be quite effective, where in the extreme case only one labeled point is needed to learn the classifier, which is utmost incredibly amazing [14]. However, cotraining makes more than strong assumptions that (1) features can be split into two sets; (2) each subfeature set is sufficient to train a good classifier; and (3) the two sets are conditionally independent given the class on the splitting of features, which generally cannot be met in real life. To deal with this problem, Zhou and Li [15] proposed a cotraining style semisupervised learning algorithm named tritraining. In tritraining process, three weak classifiers are generated from the original labeled example set and are then refined using unlabeled examples. Tritraining neither requires the instance space to be described with sufficient and redundant views nor puts any constraints on the supervised learning algorithm. In addition, it possesses the merits of good efficiency and generalization ability. Tritraining has been successfully applied in Chinese chunking [16], biomedical named entity recognition [17], and web spam detection [18]. With all these advantages and successful application in other areas, tritraining is supposed to be a promising method in bearing fault diagnosis too. However, the process of unlabeled data adopted in tritraining is the simplistic consistency principle. In detail, in each round of tritraining, an unlabeled example is labeled for the third classifier if the other two classifiers agree on the labeling, under certain conditions. This might undermine the performance stability of tritraining because the unlabeled data may often be wrongly labeled by both classifiers during the learning process [19]. In order to overcome this problem, the cut edge weight statistic (CEWS) [20] is utilized to give the confidence of each predicted label of the unlabeled data. Only when the confidence is high enough can the predicted label be added to training set. With this problem solved by cut edge weight confidence (CEWC) plus all its merits, there is no doubt that tritraining will be a promising semisupervised algorithm for improvement of bearing fault diagnosis.

Hence to fully appreciate the large amount of unlabeled data of roller bearing and thus improve the performance of bearing fault diagnosis, this paper presents a roller bearing fault diagnosis method based on the combination of tritraining and CEWC. And the remainder of the paper is organized as follows. In Section 2, a detailed description of the methodologies used in this paper is presented. In Section 3, the experiment setup and relative information of two independent roller bearing fault datasets are presented. In Section 4, the results are presented. In Section 5, the results

are discussed. And finally in Section 6, the conclusion of the research is given.

2. Methodology

2.1. Tritraining. Tritraining is semisupervised machine learning proposed by Zhou and Li [15]. The procedure of tritraining is as follows. First three diverse classifiers are initially trained from the bagging samples from the original labeled example set. The diversity of the classifiers is guaranteed by the manipulation of the original labeled example set through a popular ensemble learning algorithm, that is, Bagging [21]. Second, the three trained classifiers are used to predict the examples from the unlabeled set. Those who pass the consistency principle will be added to the labeled dataset. Third, the initial classifiers are updated and the process repeats.

Let L denote the labeled dataset with size $|L|$ and U denote the unlabeled dataset with size $|U|$. In standard tritraining algorithm, there are three diverse classifiers h_1 , h_2 , and h_3 initially trained from the original L . Then, for any classifier, an unlabeled example can be labeled for it as long as the other two classifiers agree on the labeling of this example. For example, if h_1 and h_2 agree on the labeling of an example x in U , then x can be labeled for h_3 . It is obvious that in such a scheme if the prediction of h_1 and h_2 on x is correct, then h_3 will get a solid new instance for further training. Otherwise, h_3 will get an example with noisy label. However, Zhou and Li [15] proved that, even in the worst case, the increase in the classification noise can be compensated if the amount of newly labeled examples is sufficient and the constraint condition (1) is met.

$$\begin{aligned} & |L \cup L^t| \left(1 - 2 \frac{\eta_L |L| + \hat{\epsilon}_1^t |L^t|}{|L \cup L^t|} \right)^2 \\ & > |L \cup L^{t-1}| \left(1 - 2 \frac{\eta_L |L| + \hat{\epsilon}_1^{t-1} |L^{t-1}|}{|L \cup L^{t-1}|} \right)^2, \end{aligned} \quad (1)$$

where L^t and L^{t-1} are the set of examples that are labeled for a classifier by other two classifiers in the t th round and the $(t-1)$ th round, respectively. $\hat{\epsilon}_1^t$ is the upper bound of the classification error rate of those other two classifiers in the t th round. And η_L is the classification noise rate of L ; that is, the number of examples in L that are mislabeled is $\eta_L |L|$.

It is noteworthy that if the labeled examples are not sufficient or the constraint condition is not met, it is rather doubtful whether the benefits outweigh the drawbacks in case that an unlabeled example is wrongly labeled. Therefore, it is still necessary to measure the confidence of the labeling of each classifier.

2.2. Cut Edge Weight Confidence. The CEWC is established by a two-step process. In the first step, by employing the k -nearest neighbor criterion, a neighborhood graph is constructed from the labeled examples $L = \{(l_p, y_p) \mid p = 1, 2, \dots, |L|\}$, where l_p is the attributes of p th example in set L and y_p the label. Concretely, each example $(l_p, y_p) \in L$

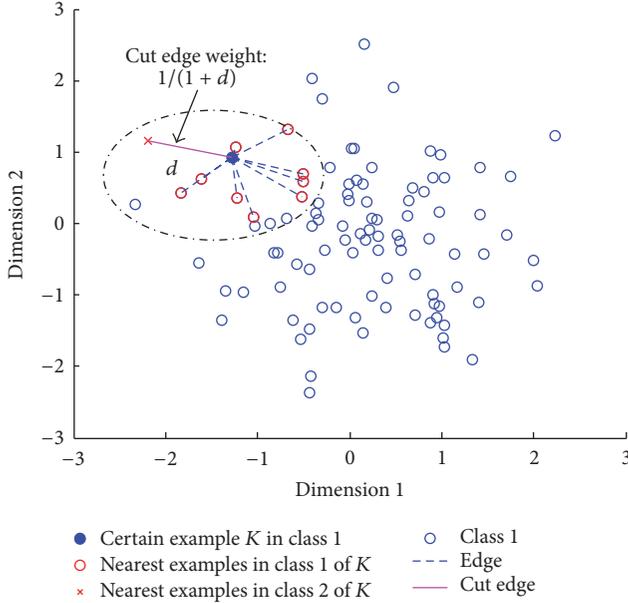


FIGURE 1: Cut edge illustration.

corresponds to a vertex in the graph G_L . There will be an edge \overline{pq} connecting the two vertices of l_p and l_q if either l_p is among the k -nearest neighbors of l_q or l_q is among the k -nearest neighbors of l_p . And a weight $w_{pq} \in [0, 1]$ is associated with the edge \overline{pq} computed as $(1 + d(l_p, l_q))^{-1}$, where $d(l_p, l_q)$ is the Euclidean distance between l_p and l_q .

In the second step, the confidence of whether the label y_p associated with l_p is correct is evaluated through exploring information encoded in G_L 's structure. As illustrated in Figure 1, an edge in G_L is called *cut edge* if the two vertices connected by it have different associated labels. The CEWS is as follows:

$$J_p = \sum_{l_q \in C_p} w_{pq} I_{pq}, \quad (2)$$

where C_p corresponds to the set of examples which are connected with l_p in G_L and I_{pq} corresponds to an *i.i.d.* Bernoulli random variable which takes value of 1 if y_p is different from y_q . When the size of C_p is sufficiently large, according to the central limit theorem, J_p can be approximately modeled by a normal distribution. Let J_p^s denote the standardized form of J_p . Then based on the left unilateral p value of J_p^s with respect to $N(0, 1)$, the labeling confidence of (l_p, y_p) is as follows:

$$CF_L(l_p, y_p) = 1 - \Phi(J_p^s), \quad (3)$$

where $CF_L(l_p, y_p)$ is the labeling confidence and $\Phi(J_p^s) = (1/\sqrt{2\pi}) \int_{-\infty}^{J_p^s} e^{-t^2/2} dt$ is the p value of J_p^s under standard normal distribution.

Note that $CF_L(l_p, y_p)$ represents only a heuristic way to estimate the labeling confidence of (l_p, y_p) and should by no means be deemed to represent the ground-truth probability

of y being the correct label of l . Though, experimental results in [22] validated the usefulness of this heuristic confidence estimation strategy in discriminating correctly labeled examples from incorrectly labeled ones.

2.3. Diagnosis Framework. The proposed approach combines the tritraining and CEWC to achieve bearing fault diagnosis and thus is called C-tritraining. The framework of it is illustrated in Figure 2. The data used for diagnosis are bearing vibration signals. First, the diagnostic features of the original vibration signals are extracted. Using ensemble empirical mode decomposition (EEMD) the original vibration signals can be broken down into intrinsic mode functions (IMFs) [23]. The information entropies of IMFs, which are surprisingly good features for bearing fault diagnosis [24], are used as the features, as the input of proposed method. Then, three bagging sample sets are drawn from the labeled feature set and each of them is used for the initial training of the weak classifier that we adopt BP neural network in this paper. Three weak classifiers will be obtained and used to predict certain proportion of unlabeled feature examples. In detail, the prediction of weak classifier 1 and weak classifier 2, if the CEWC of both are higher than the threshold, will be added to sample set 3 for updating of weak classifiers 3. The same goes for the updating of weak classifier 1 and classifier 2; that is, the training set is enlarged by the prediction of other two weak classifiers. Besides, the initial proportion of unlabeled features examples from the database is set to be 0.5. The proportion updates as follows:

$$\text{prop}_t = \text{prop}_{t-1} + (e_t - e_{t-1}), \quad (4)$$

where prop_t and prop_{t-1} are the proportion at t th and $(t-1)$ th iteration. e_t and e_{t-1} are the training error at t th and $(t-1)$ th iteration. The proportion updating process is rather intuitive. If the error decreases after the enlargement of training set with added unlabeled prediction, we naturally are confident that the weak classifiers are reliable and able to handle more unlabeled examples next time. However, if the error increases, we will have lower confidence of the weak classifiers and less unlabeled examples are trusted to them next time. The tritraining process keeps running until the termination condition is reached. The final output of the framework is the ensemble classifier that will be used to do the final bearing diagnosis using majority voting.

Trying to avoid the overfitting problem, a small trick named suspect principle is introduced into the classifiers updating process as the termination condition. The core of suspect principle lies in that when the three initial classifiers in tritraining have been updated to their best (the error rates stop decreasing) with the help of unlabeled examples, we remain doubtful whether they have reached their best or just fall into a local optimum. Therefore, the termination condition is set as the classifiers updating process keeps running after certain times that the error rates stop decreasing. It is worth discussing how we should set the suspect principle value. The experimental results in Section 4 show that four times is a good choice.

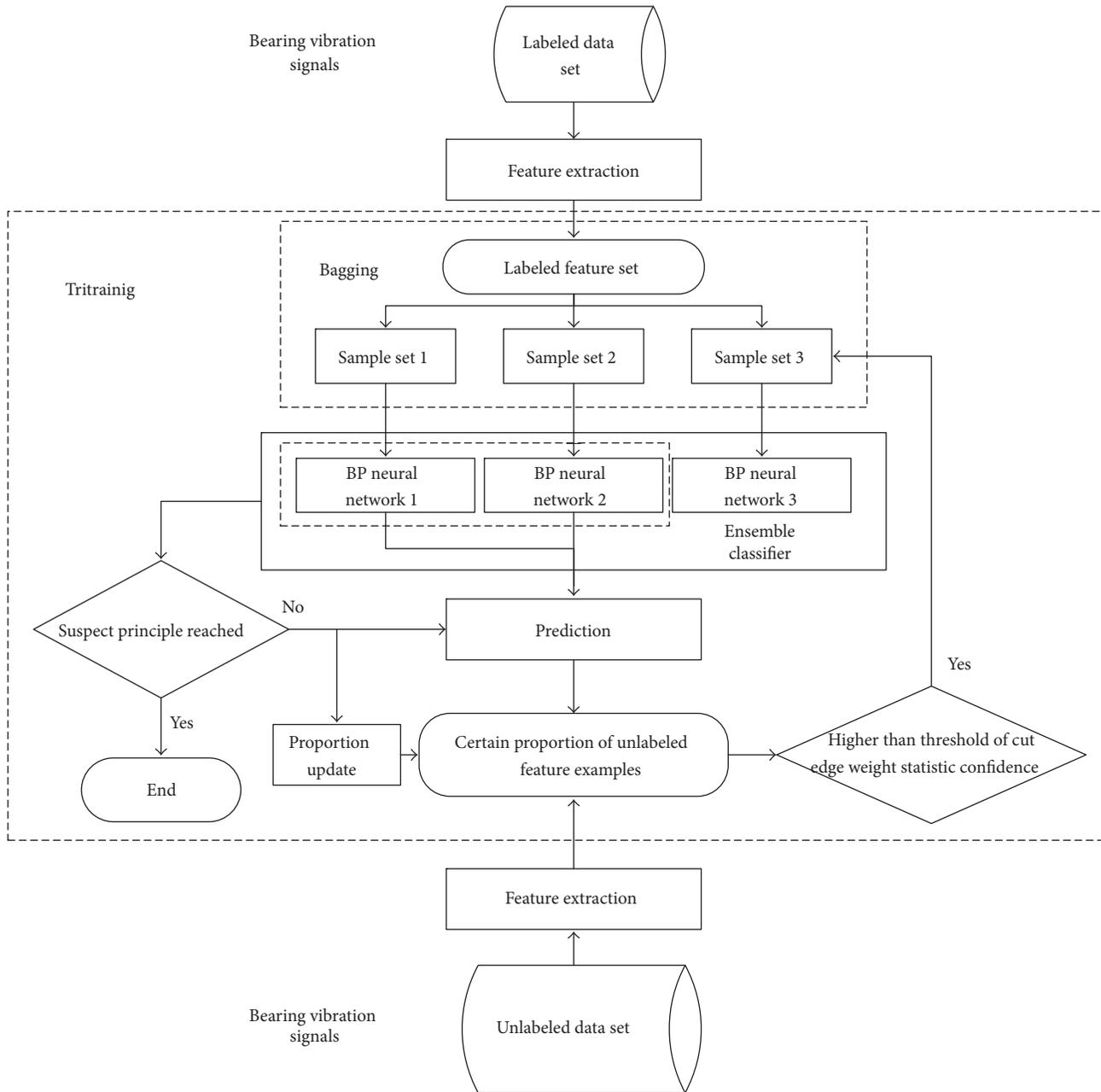


FIGURE 2: Framework of the proposed method.

3. Case Study Description

To verify the effectiveness and generalization ability of the proposed method, datasets from two individual bearing fault cases conducted by different groups were adopted.

Case 1. As shown in Figure 3, the first case was originally conducted on rotational machinery fault simulation test bed (QPZZ-II) by Prognostic and Health Management Laboratory of School of Reliability and Systems Engineering, BUAA.

The inner-ring fault, outer-ring fault, and roller element fault are introduced by wire-electrode cutting a crevice on the

surface of inner ring, outer ring, and one of the roller elements as marked in Figure 4. The vibrational signals are sampled at a frequency of 5120 samples per second and the rotation speed is 1500 revolutions per minute.

The test bearings used are cylindrical roller bearing (N205EM HRB CHINA), the detailed structure information of which is listed in Table 1.

Case 2. The second case was originally conducted by Institute of Intelligent instrument and Diagnosis, Xi'an Jiaotong University. The test rig shown in Figure 5 is completely designed and manufactured all by them. It consists mainly of a speed

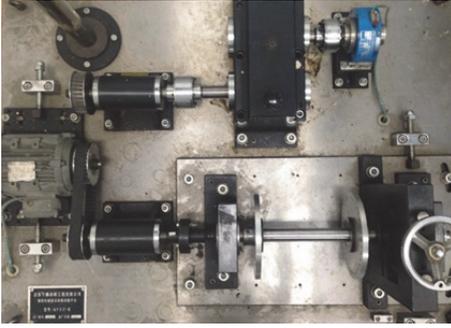


FIGURE 3: Test rig of Case 1.

TABLE 1: Bearing structure information.

Designation	Inside diameter	Outside diameter	Thickness	Element type
N205EM	25 mm	52 mm	15 mm	Cylindrical roller

TABLE 2: Bearing structure information.

Designation	Inside diameter	Outside diameter	Thickness	Element type
6308	40 mm	90 mm	23 mm	Deep groove ball

governor, driving motor, power supply box, horizontal and radial loading devices, and of course sensors.

Bearing faults in Case 2 include inner-ring fault, outer-ring fault, and roller element fault with an area of 3.8 mm^2 , 7 mm^2 , and 3 mm^2 circle-shaped spalling on the surface of inner ring, outer ring, and roller element, respectively. The test bearings used are deep groove ball bearing (6308), the detailed structure information of which is listed in Table 2. The sampling frequency is 20 K samples per second and the rotation speed is 1500 revolutions per minute.

4. Results

Through EEMD process, the original vibration signals collected from the two cases are transformed into two feature sets. According to [18], two parameters of EEMD, that is, the ratio of the standard deviation of the added noise and that of input, are set to be 0.15 and ensemble number is set to be 100. Information of the feature sets is tabulated in Table 3. For each feature set, 85 percent of the data are kept as training set while the rest are used as the test set to examine the trained classifiers. The training set, composed of labeled pool and unlabeled pool, that is, $L \cup U$, is partitioned under different *unlabeling rates* including 80 percent, 60 percent, 40 percent, and 20 percent. Take the data of Case 1 whose size is 400 examples; for example, the training set has 340 examples (85 percent) and test set 60 examples (15 percent). When the unlabeling rate is 80 percent, 68 examples out of

TABLE 3: Feature sets.

Feature set	Attribute	Size	Class	Nor./inner/outer/element
Case 1	24	400	4	25%/25%/25%/25%
Case 2	9	128	4	25%/25%/25%/25%

TABLE 4: Parameters of BPNN.

Number of layers	Input layer nodes	First hidden layer nodes	Second hidden layer nodes	Output layer nodes
4	24	8	4	4
<i>Train function</i>	<i>Function coefficient</i>	<i>Mean square error</i>	<i>Epoch</i>	<i>Validation checks</i>
Levenberg-Marquardt	$1.00e + 10$	$1.00e - 20$	1000	6

340 examples are then put into L and other 292 examples are put into U without their labels. To overcome the randomness of the results, 50 independent runs are performed and the averaged results are summarized as the final outcome.

Figure 6 shows the classification error rate of Cases 1 and 2 under different unlabeling rate and suspect principle value. When suspect principle value is set to four, the classification error rates are the lowest or the second lowest in most situations except only for the classification error rate in Case 2 under the unlabeling rate of 0.6. Therefore, it is naturally intuitive to determine that suspect principle value set to four is a practical optimal choice.

With suspect principle value set to be four, the averaged results are summarized in Table 4, which presents the classification error rate of the initial ensemble weak classifiers, that is, the combination of the three initial BP neural network classifiers only trained from L and the final ensemble classifiers generated by tritraining and the improvement of the latter over the former. The architecture and parameters of the BP neural network are shown in Table 4.

4.1. Comparative Experiments with Different Semisupervised Learning Models. In this paper, self-learning and tritraining models were conducted for comparison. The self-learning model is a traditional semisupervised learning method where the most confident unlabeled data samples, together with the predicted labels, are added to the initial training set, so that the neural network classifier can be retrained and the procedure repeats. The tritraining model is an elementary model whose parameters were the same as those of the C-tritraining except the CEWS optimization process. Detailed diagnosis is listed in Table 5–7 and Figure 7.

4.2. Comparative Experiments with Different Base Classifiers. For the purpose of investigating the diagnosis performance with different base classifiers, an additional experiment was conducted where the support vector machine (SVM) was built with a RBF kernel function whose kernel parameter is set to 0.08 and the penalty factor set to 128. The SVM model was trained using the one-versus-all criterion. Note that the SVM model could be regarded as a more stable classifier while



FIGURE 4: Faults of inner ring, outer ring, and roller element.

TABLE 5: Classification error rate of initial and final hypothesis and the corresponding improvements of C-tritraining under different unlabeled rate with suspect principle value set to 4.

Case	Unlabeling rate of 0.8			Unlabeling rate of 0.6			Unlabeling rate of 0.4			Unlabeling rate of 0.2		
	Initial	Final	Improvement									
Case 1	0.1670	0.1367	18.14%	0.0990	0.0857	13.43%	0.0613	0.06	2.12%	0.0500	0.0487	2.6%
Case 2	0.4589	0.3400	25.91%	0.3274	0.28	14.48%	0.2495	0.2211	11.38%	0.2453	0.2200	10.31%

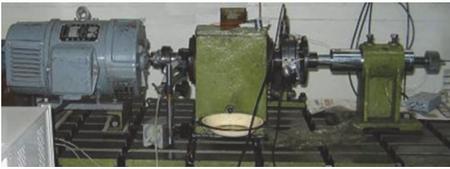


FIGURE 5: Test rig of Case 2.

neural network based classifiers are mostly unstable in terms of training mechanism. Taking Case 1 as an example, detailed diagnosis results are displayed in Table 8.

5. Discussion

- (1) Different from the supervised learning based diagnosis methods for fault detection and identification, this manuscript proposes a new incremental learning approach that takes advantage of unlabeled data to improve diagnosis performance of rolling bearings. Considering that fault samples are continuously attained over monitoring time, semisupervised ensemble learning is employed so as to avoid manual labeling error, as well as improving classification accuracy for health assessment utilizing prior learned knowledge and newly attained information in a real-time diagnosis mechanism. In this regard, tritraining, where three diverse classifiers are generated from the bagging samples and integrated for fault diagnosis, is conducted to improve classification performance of base classifiers. On this basis, CEWC is employed in this study to further mine salient characteristics of unlabeled data with a view to design a more intelligent diagnosis model. This method was applied

to two bearings with different proportions of unlabeled samples (20, 40, 60, and 80 percent, resp.). As shown in Table 5, the proposed method is able to effectively improve the performance of the initial ensemble classifiers under all unlabeled rates for both Cases 1 and 2. The improvement percentage ranges from 25.9% to 2.6%.

- (2) It is noteworthy in Figure 8 that the improvement percentage increases sharply as unlabeled rate increases in both Cases 1 and 2. This means, by utilizing unlabeled data, the proposed method really makes a difference where there is limited labeled data to train the classifiers. And when there is less labeled data to train the classifiers, the proposed method is able to more improve the classifiers' performance. However, the absolute value of improvement and diagnostic error rate in Case 2 are commonly higher than those in Case 1. The difference between two results is caused by their dataset size. The feature set of Case 1 has 400 examples while Case 2 has only 128 examples. For Case 2, when unlabeled rate is 0.8, then there are only $128 \times 85\% \times 20\% \approx 22$ labeled examples to train classifiers, which is apparently not enough to train good classifiers. No wonder the initial underfitting classifiers' classification error reaches 0.4589 when unlabeled rate is 0.8. The proposed method promotes 25.91% performance of the initial classifiers in this extreme situation. When there is enough labeled data, for example, Case 1 when unlabeled rate is 0.2, the classification error rate lowers to 0.0487 (95.13% diagnostic accuracy). It implies that roller bearing fault diagnosis based on tritraining is promising in either situation (a) where there is not enough labeled data to obtain good classifiers or situation (b) where

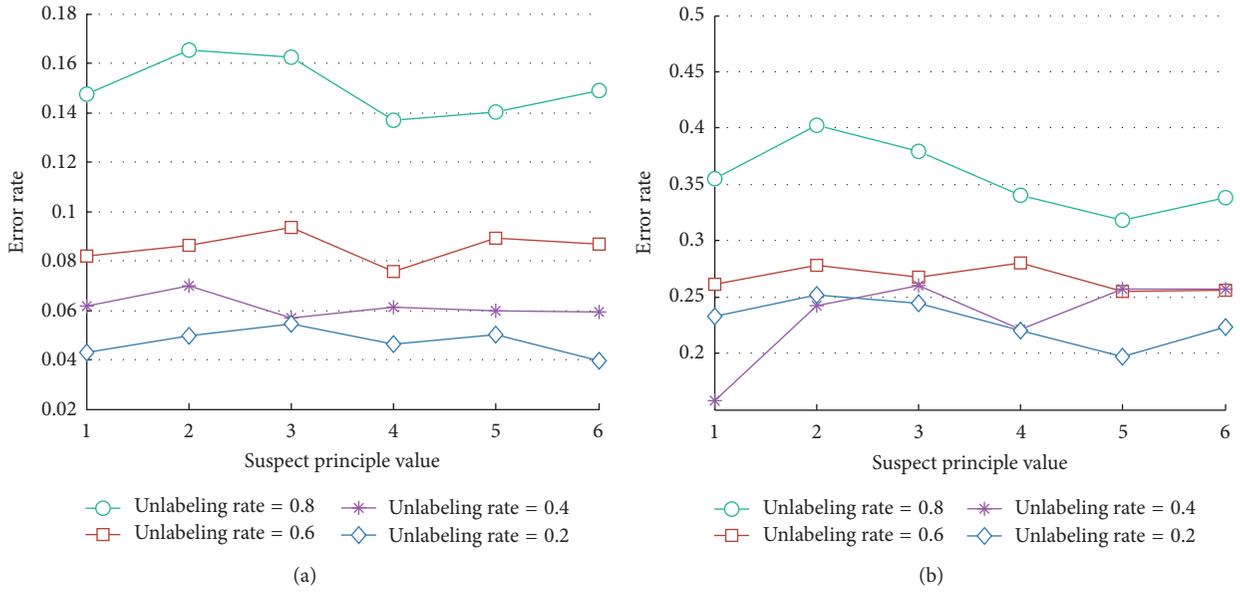


FIGURE 6: Classification error rate of Case 1 (a) and Case 2 (b) under different unlabeled rate and suspect principle value.

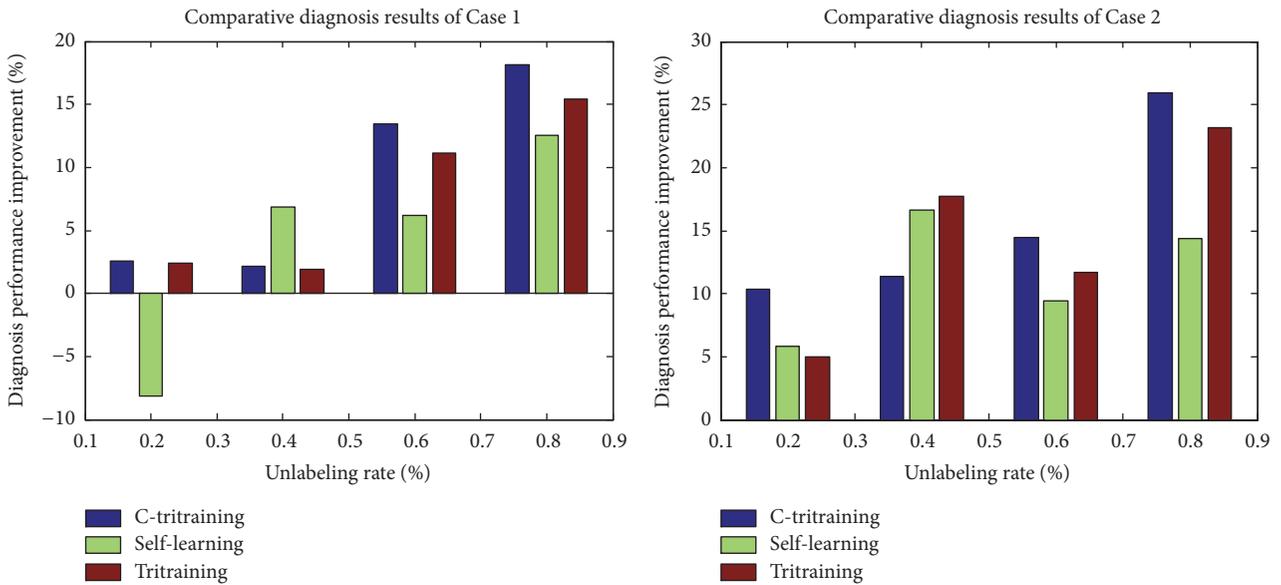


FIGURE 7: Performance improvement at different unlabeled rate in Cases 1 and 2.

there is enough labeled data. In situation (a) tritaining greatly improves classifiers' performance by utilizing unlabeled data that are easily available. And the performance will keep upgrading as long as there is more unlabeled data. In situation (b) tritaining can still be helpful though the initial classifiers are good enough for the bearing fault diagnosis.

- (3) In this study, the semisupervised learning methods including self-learning and traditional tritaining are conducted for comparison. Detailed diagnosis results are listed in Tables 3–7 and Figure 8. It is observed that although the classification rates of all methods were improved, the tritaining based methods appeared to

produce higher correct rates in most of the cases. Taking the diagnosis results with unlabeled rate of 0.8 in Case 2 as an example, the improvements of classification accuracies were 25.91%, 14.33%, and 23.13% for C-tritaining, self-learning, and tritaining, respectively. This is mainly because such ensemble process could effectively strengthen the learning ability through integrating multiple views of individual classifiers. In addition, compared to basic tritaining model, the proposed method attained better diagnosis results demonstrating the effectiveness of CEWS on capturing pivotal fault characteristics from unlabeled data in rolling bearing diagnosis issues. It is also noted

TABLE 6: Classification error rate of initial and final hypothesis and the corresponding improvements of self-training under different unlabeled rate.

Case	Unlabeling rate of 0.8			Unlabeling rate of 0.6			Unlabeling rate of 0.4			Unlabeling rate of 0.2		
	Initial	Final	Improvement									
Case 1	0.3353	0.2933	12.53%	0.2168	0.2033	6.23%	0.1002	0.0933	6.89%	0.0417	0.0454	-8.15%
Case 2	0.4737	0.4058	14.33%	0.4037	0.3658	9.39%	0.3158	0.2632	16.66%	0.2811	0.2648	5.80%

TABLE 7: Classification error rate of initial and final hypothesis and the corresponding improvements of tritraining under different unlabeled rate.

Case	Unlabeling rate of 0.8			Unlabeling rate of 0.6			Unlabeling rate of 0.4			Unlabeling rate of 0.2		
	Initial	Final	Improvement									
Case 1	0.2167	0.1833	15.41%	0.1500	0.1333	11.13%	0.0667	0.0654	1.95%	0.0533	0.0520	2.44%
Case 2	0.3684	0.2832	23.13%	0.3632	0.3205	11.73%	0.2559	0.2105	17.74%	0.2105	0.2000	4.99%

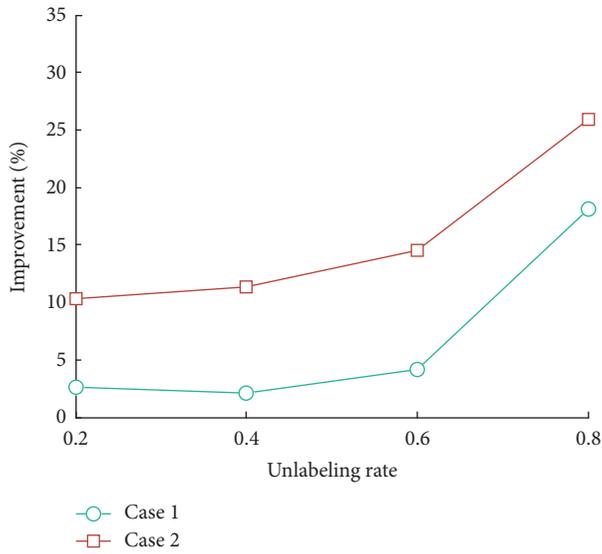


FIGURE 8: Comparative diagnosis results of rolling bearing diagnosis cases.

in Case 1 that the diagnosis result of self-learning decreased with unlabeled rate of 0.8, which may due to some negative effects of improper training such as overfitting problems.

- (4) From the diagnosis results of SVM based C-tri-training, it is noted that the fault classification performance was improved as well, demonstrating the effectiveness of the proposed semisupervised learning method in rolling bearing diagnosis; that is, such model can be appropriately applied using various base classifiers. However, the misclassification rates of the testing data were relatively high when compared to the BPNN based model, which may be due to the lesser difference between three SVM models. The ensemble process could only be more effective on condition that the base classifiers are of greater diversity. Therefore in this study, when determining the base classifier and architecture, we follow a simple

idea that the classifiers should be as different as possible in the bagging process so that more sufficient information could be learned from the unlabeled data.

6. Conclusion

In order to improve performance of bearing fault diagnosis when there is limited labeled data, this paper presents a roller bearing fault diagnosis method based on the combination of tritraining and CEWC. The method is validated in two roller bearing fault cases conducted by two independent groups. The results showed that, with the help of unlabeled examples, the method is able to effectively improve the fault diagnosis for both cylindrical roller bearing and deep groove ball bearing when there is limited labeled examples. The proposed method still helps even when there is enough labeled data and the diagnostic accuracy can reach up to 95%.

Although the proposed method is promising, there is something that could be improved in the future work. The feature extracted from the vibrational signal is information entropy of IMFs through EEMD, which is an iterative process and so is tritraining. That makes the proposed method time-consuming, which undermines its applicability in roller bearing online diagnosis. Hence, the efficiency improvement is among the priorities in future work.

Competing Interests

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

Authors' Contributions

Wei-Li Qin and Zheng-Ya Wang contributed equally to this work and should be considered joint first authors.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant nos. 51575021 and 51105019), the

TABLE 8: Comparative diagnosis results of SVM and BPNN in Case 1.

Classifier	Unlabeling rate of 0.8			Unlabeling rate of 0.6			Unlabeling rate of 0.4			Unlabeling rate of 0.2		
	Initial	Final	Improvement									
BPNN	0.1670	0.1367	18.14%	0.0790	0.0757	4.18%	0.0613	0.0600	2.12%	0.0500	0.0487	2.60%
SVM	0.3001	0.2617	12.80%	0.1697	0.1652	2.65%	0.1267	0.1250	1.34%	0.0833	0.0783	6.00%

Technology Foundation Program of National Defense (Grant no. Z132013B002), and the Fundamental Research Funds for the Central Universities (Grant no. YWF-16-BJ-J-18).

References

- [1] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—a tutorial," *Mechanical Systems & Signal Processing*, vol. 25, no. 2, pp. 485–520, 2011.
- [2] B. Li, M.-Y. Chow, Y. Tipsuwan, and J. C. Hung, "Neural-network-based motor rolling bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, pp. 1060–1069, 2000.
- [3] O. R. Seryasat, M. A. Shoorehdeli, F. Honarvar, and A. Rahmani, "Multi-fault diagnosis of ball bearing using FFT, wavelet energy entropy mean and root mean square (RMS)," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '10)*, pp. 4295–4299, Istanbul, Turkey, October 2010.
- [4] H. W. Peng and P. J. Chiang, "Control of mechatronics systems: ball bearing fault diagnosis using machine learning techniques," in *Proceedings of the 8th Asian Control Conference*, pp. 175–180, May 2011.
- [5] X. Jin, M. Zhao, T. W. S. Chow, and M. Pecht, "Motor bearing fault diagnosis using trace ratio linear discriminant analysis," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2441–2451, 2014.
- [6] J. Liu, W. Wang, and F. Golnaraghi, "An extended wavelet spectrum for bearing fault diagnostics," *IEEE Transactions on Instrumentation & Measurement*, vol. 57, no. 12, pp. 2801–2812, 2008.
- [7] X. Zhu, "Semi-supervised learning literature survey," *Computer Science*, vol. 37, pp. 63–77, 2008.
- [8] N. N. Pise and P. Kulkarni, "A survey of semi-supervised learning methods," in *Proceedings of the International Conference on Computational Intelligence and Security (CIS '08)*, vol. 2, pp. 30–34, Suzhou, China, December 2008.
- [9] K. Nigam, A. K. McCallum, S. Thrun, and T. Mitchell, "Text classification from labeled and unlabeled documents using EM," *Machine Learning*, vol. 39, no. 2, pp. 103–134, 2000.
- [10] F. G. Cozman, I. Cohen, and M. C. Cirelo, "Semi-supervised learning of mixture models," in *Proceedings of the International Conference on Machine Learning*, pp. 41–65, 2003.
- [11] C. Rosenberg, M. Hebert, and H. Schneiderman, "Semi-supervised self-training of object detection models," in *Proceedings of the 7th IEEE Workshop on Applications of Computer Vision/IEEE Workshop on Motion and Video Computing (WACV/MOTION '05)*, pp. 29–36, Breckenridge, Colo, USA, January 2005.
- [12] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning*, MIT Press, 2010.
- [13] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proceedings of the 11th Annual Conference on Computational Learning Theory*, pp. 92–100, 2000.
- [14] O. Chapelle, B. Schölkopf, and A. Zien, "An augmented PAC model for semi-supervised learning," *Semi-Supervised Learning*, pp. 397–419, 2006.
- [15] Z.-H. Zhou and M. Li, "Tri-training: exploiting unlabeled data using three classifiers," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 11, pp. 1529–1541, 2005.
- [16] W. Chen, Y. Zhang, and H. Isahara, "Chinese chunking with tri-training learning," in *Computer Processing of Oriental Languages. Beyond the Orient: The Research Challenges Ahead: 21st International Conference, ICCPOL 2006, Singapore, December 17–19, 2006. Proceedings*, vol. 4285 of *Lecture Notes in Computer Science*, pp. 466–473, Springer, Berlin, Germany, 2006.
- [17] Y. H. Cai and X. Cheng, "Biomedical named entity recognition with tri-training learning," in *Proceedings of the 2nd International Conference on Biomedical Engineering and Informatics (BMEI '09)*, pp. 1–5, Tianjin, China, October 2009.
- [18] H. Li, "Web spam detection based on improved tri-training," in *Proceedings of the 2nd IEEE International Conference on Progress in Informatics and Computing (PIC '14)*, pp. 61–65, IEEE, Shanghai, China, May 2014.
- [19] A. Blum and S. Chawla, "Learning from labeled and unlabeled data using graph mincuts," in *Proceedings of the 18th International Conference on Machine Learning*, pp. 19–26, 2001.
- [20] F. Muhlenbach, S. Lallich, and D. A. Zighed, "Identifying and handling mislabelled instances," *Journal of Intelligent Information Systems*, vol. 22, no. 1, pp. 89–109, 2004.
- [21] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [22] M.-L. Zhang and Z.-H. Zhou, "CoTrade: confident co-training with data editing," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 6, pp. 1612–1626, 2011.
- [23] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.
- [24] W.-L. Qin, W.-J. Zhang, and C. Lu, "Rolling bearing fault diagnosis based on ensemble empirical mode decomposition, information entropy and random forests," in *Proceedings of the International Conference on Vibroengineering*, pp. 211–216, September 2015.

Research Article

Vibration Based Diagnosis for Planetary Gearboxes Using an Analytical Model

Liu Hong,^{1,2} Yongzhi Qu,^{1,2} Yuegang Tan,^{1,2} Mingyao Liu,^{1,2} and Zude Zhou^{1,2}

¹School of Mechanical and Electronic Engineering, Wuhan University of Technology, 122 Luoshi Road, Wuhan 430070, China

²Hubei Digital Manufacturing Key Laboratory, Wuhan University of Technology, 122 Luoshi Road, Wuhan 430070, China

Correspondence should be addressed to Yongzhi Qu; yongzhiqu@hotmail.com

Received 21 July 2016; Revised 4 October 2016; Accepted 26 October 2016

Academic Editor: Ganging Song

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

The application of conventional vibration based diagnostic techniques to planetary gearboxes is a challenge because of the complexity of frequency components in the measured spectrum, which is the result of relative motions between the rotary planets and the fixed accelerometer. In practice, since the fault signatures are usually contaminated by noises and vibrations from other mechanical components of gearboxes, the diagnostic efficacy may further deteriorate. Thus, it is essential to develop a novel vibration based scheme to diagnose gear failures for planetary gearboxes. Following a brief literature review, the paper begins with the introduction of an analytical model of planetary gear-sets developed by the authors in previous works, which can predict the distinct behaviors of fault introduced sidebands. This analytical model is easy to implement because the only prerequisite information is the basic geometry of the planetary gear-set. Afterwards, an automated diagnostic scheme is proposed to cope with the challenges associated with the characteristic configuration of planetary gearboxes. The proposed vibration based scheme integrates the analytical model, a denoising algorithm, and frequency domain indicators into one synergistic system for the detection and identification of damaged gear teeth in planetary gearboxes. Its performance is validated with the dynamic simulations and the experimental data from a planetary gearbox test rig.

1. Introduction

With the advantages of coaxial shafting, compact structure, and high torque to weight ratio, planetary gearboxes have been widely employed in heavy-duty applications such as helicopters, offshore rigs, and wind turbines. Since unexpected failures of gearboxes in heavy-duty drive trains may lead to expensive downtime losses and catastrophic accidents, the condition monitoring of planetary gearbox is essential to reduce the maintenance costs and ensure the safety operations. In industrials, vibration based diagnosis using accelerometers mounted on the gear/bearing housing has prevailed in the recent decades because it does not interfere with the normal operation of rotating machines [1]. The gear failures in fixed-axis gearboxes can be directly detected by observing the presence of sidebands around the gear mesh harmonics in the measured vibration spectrum [2]. However, as planetary gear systems employ P number of identical planet pinions revolving around the sun gear with the

carrier, dominant sidebands around gear mesh harmonics are observed even when no gear failure or severe manufacturing error exists [3, 4]. Thus, the fault detection criteria for planetary gearboxes are not straightforward compared with the fixed-axis gearboxes due to the complexities of sidebands in the vibration spectrum measured by the fixed accelerometer. Existing literatures have also found that the sidebands are often asymmetric and the gear mesh harmonics may not even show up in the measured vibration spectrum for specific gear-set geometries, which are the result of manufacturing errors and the superposition of vibrations from P planets [5, 6]. Moreover, the noises and vibrations from other stages of planetary gearboxes may further mask the fault signatures of planetary gear-sets, which make the vibration spectrum more difficult to interpret.

To address the relatively complex behaviors of frequency components and the superposition of vibrations from P planets, a vibration separation (VS) methodology was developed for the fault detection of planetary gearboxes in helicopter

transmissions [7, 8]. The well-known denosing technique, time synchronous averaging (TSA), was also studied to isolate the vibration of planetary gear-sets from noisy sensory signals, which is equivalent to filter the raw vibration data through a comb filter [9]. However, the applicability of VS and TSA to planetary gearboxes is often limited in practice because they require extensive stationary vibration signals and additional knowledge of the time instant when each planet passes through the fixed accelerometer. Furthermore, in practical industrial applications, the simple detection of gear fault is often insufficient. The fault identification is also of interest as the replacement gears can be ordered before the actual disassembly of gearboxes, which in turn reduces the machine downtime [10]. Thus, to advance the state-of-art of diagnosis of planetary gearboxes, it is critical to develop a novel diagnostic system for industrials that is applicable to intelligently highlight and identify the fault signatures of planetary gear systems from the noisy and profuse measured raw data.

To this aim, a novel synergistic diagnostic scheme is proposed in this work that enables the automated detection and identification of damaged gear teeth for planetary gear-sets. The paper begins with an introduction of the analytical model of planetary gear systems developed by the authors [11], which can characterize the influence of different planetary gear faults on the sidebands of vibration spectrum measured by a fixed accelerometer mounted on the annulus gear. The presented analytical model has also been validated by extensive experimental studies in [12]. Then, to address the limitation of the requirement of extensive stationary data of TSA, a novel denoising algorithm for planetary gear-sets is proposed herein to highlight the fault introduced sidebands by employing a comb filter in the frequency domain. This fault signature enhanced filter is based on the summarized analytical model instead of the ensemble of the average of numerous revolutions. Afterwards, frequency domain indicators are proposed for the intelligent fault identification. The overall diagnostic scheme integrates the analytical vibration model, a proposed denoising algorithm, and frequency domain indicators into one synergistic system for the automated diagnosis of damaged gear teeth, which is capable of carrying out vibration based condition monitoring of planetary gearboxes. Finally, the effectiveness of the proposed diagnostic scheme is validated using both the simulated data from a dynamic model of planetary gearbox and experimental signals from a 4 kW planetary gearbox test rig.

2. Review of the Geometry Based Analytical Model of Planetary Gear Systems

Figure 1 illustrates a typical configuration of planetary gear-sets as an example, which includes P ($P = 4$) identical planet pinions. As a practical wind turbine planetary gearbox, the planet carrier c and the sun gear s are assumed to be connected with the input and output shaft, respectively. The annulus gear a is fixed with the gear housing. ψ_i is defined as the initial angular position of the i th planet, in

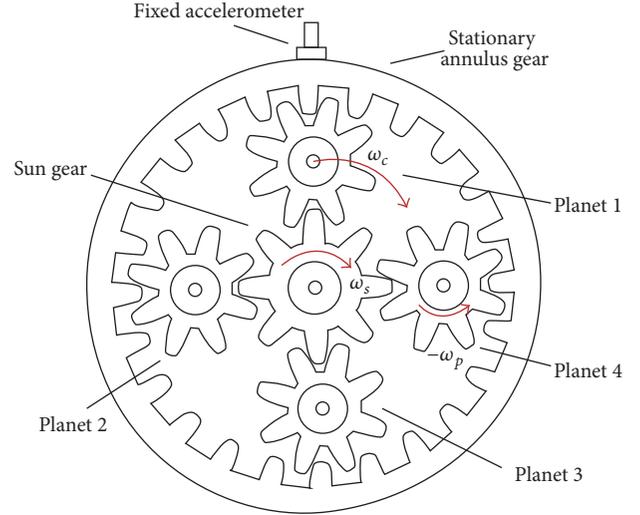


FIGURE 1: A typical configuration of planetary gear-set with 4 equally spaced planet pinions and an accelerometer mounted on the stationary annulus gear.

which $\psi_i = 2\pi(i - 1)/P$. If planet pinions rotate counter-clockwise with an angular rotational frequency ω_p as shown in Figure 1, the sun gear s and the carrier c would rotate clockwise with an angular rotational frequency ω_s and ω_c . The angular rotational frequency of the annulus gear is $\omega_a = 0$. The fundamental gear mesh frequency of the planetary gear system ω_m can be calculated as $\omega_m = N_a\omega_c = N_s(\omega_s - \omega_c) = N_p(\omega_p + \omega_c)$, in which N represents the number of teeth on the gear with respect to the subscript.

The vibration signal measured by the accelerometer mounted on the stationary annulus gear contains valuable diagnostic information of the planetary gear-set. This measured vibration signal $z(t)$ by the fixed accelerometer, which is the superposition of vibrations from P planets meshing with the annulus and sun gear, can be expressed as [11, 12]

$$z(t) = \sum_{i=1}^P w_i(t) [x_i(t) + y_i(t)], \quad (1)$$

where the window function $w_i(t)$ is used to describe the amplitude modulation phenomenon, which is caused by the relative motion between the fixed accelerometer and the revolving planets with the carrier [5, 6]. The details can be explained as follows: as a planet moves toward the accelerometer, its vibration dominates the sensory data. The level of its contribution reaches a peak when the planet is almost under the accelerometer and then decreases as the planet recedes from the accelerometer. Thus, the window function $w_i(t)$ is periodic with the carrier rotational frequency ω_c , the shape of which is usually modeled as either a Hann window or Turkey window [13].

The vibration signal $x_i(t)$ in (1) represents the vibration that resulted from the annulus gear and i th planet mesh, which is periodic at the fundamental gear mesh frequency ω_m . Note that the vibration waveforms resulting from P different annulus-planet gear meshes are similar but shifted

in time because of the variation in the initial phase during the gear meshes [6]. Therefore, the vibration signal $x_i(t)$ can be written as a Fourier series [6, 11]:

$$\begin{aligned} x_i(t) &= \sum_{j=1}^J X_{ij} \cos\left(jN_a\omega_c\left(t - \frac{\psi_i}{\omega_c}\right) + \theta_j\right) \\ &= \sum_{j=1}^J X_{ij} \cos\left(jN_a\omega_c t - jN_a\psi_i + \theta_j\right), \end{aligned} \quad (2)$$

where X_{ij} represents the Fourier coefficient of the j th gear mesh harmonic of the vibration signal from the annulus gear and i th planet mesh. $-\psi_i/\omega_c$ is the time shift between the annulus - i th planet and annulus - 1st planet mesh. θ_j is the initial phase of the j th gear mesh harmonic.

On the other hand, the vibration that resulted from the sun gear and i th planet mesh is presented as $y_i(t)$ in (1). Since the vibration signal $y_i(t)$ is periodic at the fundamental gear mesh frequency ω_m , it can be also written as a Fourier series [11]:

$$\begin{aligned} y_i(t) &= \sum_{j=1}^J Y_{ij} \\ &\cdot \cos\left(j\left(N_s(\omega_s - \omega_c)\left(t + \frac{\psi_i}{\omega_s - \omega_c}\right) + \gamma\right) + \delta_j\right) \quad (3) \\ &= \sum_{j=1}^J Y_{ij} \cos\left(jN_s(\omega_s - \omega_c)t + jN_s\psi_i + j\gamma + \delta_j\right), \end{aligned}$$

where Y_{ij} is the Fourier coefficient of the j th gear mesh harmonic of the vibration signal from the sun gear and i th planet mesh. As the sun gear is assumed to rotate clockwise in Figure 1, $N_s\psi_i$ is the phase difference caused by the time shift between the sun - i th planet and sun - 1st planet mesh. δ_j is the initial phase of the j th gear mesh harmonic. Additionally, the phase angle γ is the phase difference between the sun-planets meshes and annulus-planets meshes [14]. Under the ideal condition, $X_{ij} = X_{1j}$ and $Y_{ij} = Y_{1j}$ as the planet pinions are designed to share the equal load.

In case of a fixed-axis gear pair, the amplitude and phase modulation functions, $a_j(t)$ and $b_j(t)$, are employed to represent the variations of the vibration spectrum caused by a local gear fault. Similarly, when a local gear fault exists in the planetary gear-set, it would also result in amplitude and phase modulations in addition to the vibrations contributed from sun-planets meshes and annulus-planets meshes as described by (1)–(3). Thus, under the faulty condition, the vibration signal of the annulus - i th planet and sun - i th planet mesh can be further expanded, respectively, [11]:

$$\begin{aligned} x'_i(t) &= \sum_{j=1}^J X_j \left(1 + a_j\left(t - \frac{\psi_i}{\omega_c}\right)\right) \\ &\cdot \cos\left(jN_a\omega_c\left(t - \frac{\psi_i}{\omega_c}\right) + b_j\left(t - \frac{\psi_i}{\omega_c}\right)\right), \end{aligned}$$

$$\begin{aligned} y'_i(t) &= \sum_{j=1}^J Y_j \left(1 + a_j\left(t + \frac{\psi_i}{\omega_s - \omega_c}\right)\right) \\ &\cdot \cos\left(j\left(N_s(\omega_s - \omega_c)\left(t + \frac{\psi_i}{\omega_s - \omega_c}\right) + \gamma\right) + b_j\left(t + \frac{\psi_i}{\omega_s - \omega_c}\right)\right), \end{aligned} \quad (4)$$

where the amplitude and phase modulation functions, $a_j(t)$ and $b_j(t)$, can be presented:

$$\begin{aligned} a_j(t) &= \sum_{l=1}^L A_{jl} \cos(l \times \omega_{\text{characteristic}} t), \\ b_j(t) &= \sum_{l=1}^L B_{jl} \cos(l \times \omega_{\text{characteristic}} t). \end{aligned} \quad (5)$$

If there is a local fault on the annulus gear, its corresponding characteristic frequency is $\omega_{\text{characteristic}} = \omega_c$, which is the rotational frequency of the annulus relative to the carrier. If a damaged tooth exists on the sun gear, it would influence the gear mesh with a frequency that equals the rotational frequency of the sun gear relative to the carrier. Thus, the characteristic frequency caused by the sun gear fault is periodic with a fundamental frequency $\omega_s - \omega_c$. Analogically, the amplitude and phase modulation functions, $a_j(t)$ and $b_j(t)$, corresponding to a fault on a planet are periodic with a fundamental frequency that equals the rotational frequency of the planet relative to the carrier, $\omega_c + \omega_p$. Generally, the analytical model presented by (1)–(5) can predict the distinct behaviors of fault introduced sidebands picked by the fixed accelerometer and its prerequisite input is just the basic geometry of the planetary gear-set. Figure 2 plots the predicted results of this analytical model on the hypothesis of faults occurring at different components of a planetary gear-set.

3. Scheme of the Proposed Diagnostic Algorithm

A novel vibration based diagnostic scheme is proposed in this section to enable the automated enhancement and identification of fault signatures for planetary gear systems, which is shown in Figure 3. The details of each step are explained as follows.

Step 1. As discussed in the last sections, the vibration signal $v(t)$ contains rich diagnostic information, which is measured by an accelerometer mounted on the stationary annulus gear as shown in Figure 1. Although the behaviors of sidebands are relatively more complex than the cases of fixed-axis gear-sets, many of these sidebands in the frequency domain can still be used as fault signatures to detect and identify the local gear faults. However, in practical applications, the fluctuations of the operational speed can smear the frequency components of the spectrum and the rotary encoders/tachometers are

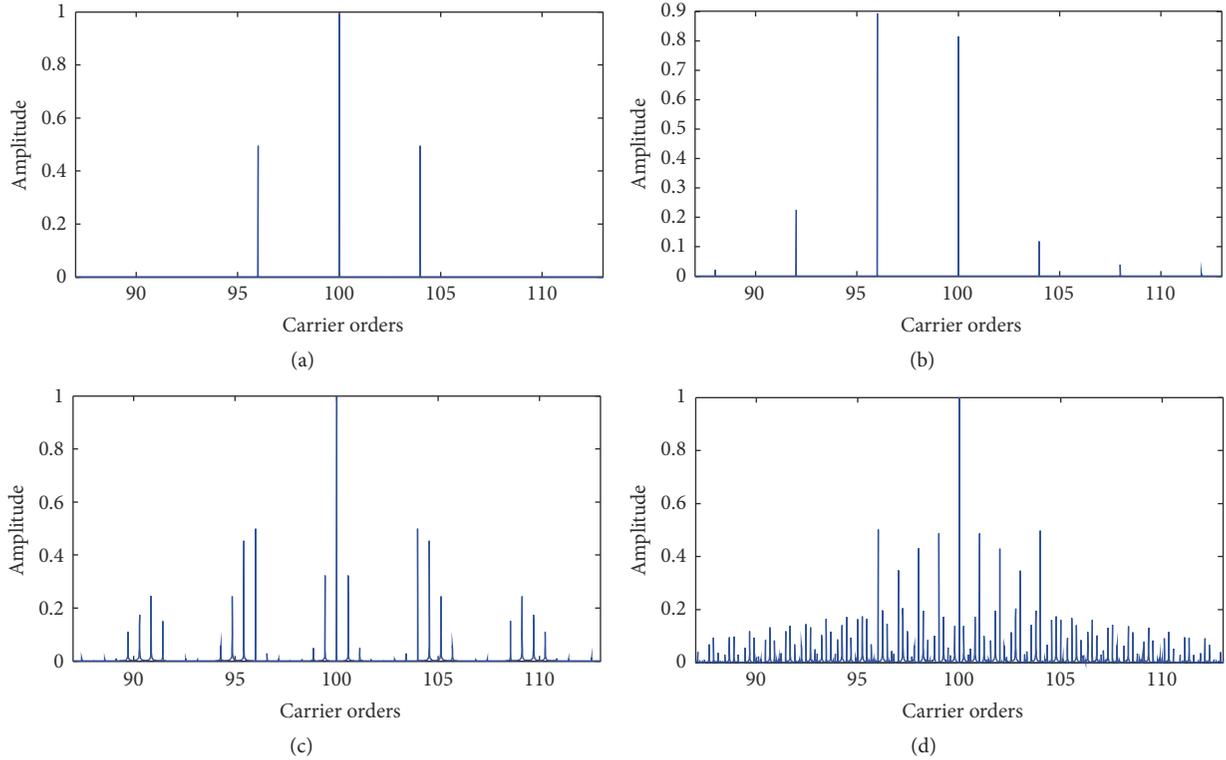


FIGURE 2: Predicted vibration spectrum around the fundamental gear mesh frequency picked by a fixed vibration sensor on the annulus for $N_a = 100$, $N_s = 28$, $N_p = 36$, and $P = 4$; (a) healthy case; (b), (c), and (d) a local gear fault on the annulus, sun, and one planet gear, respectively.

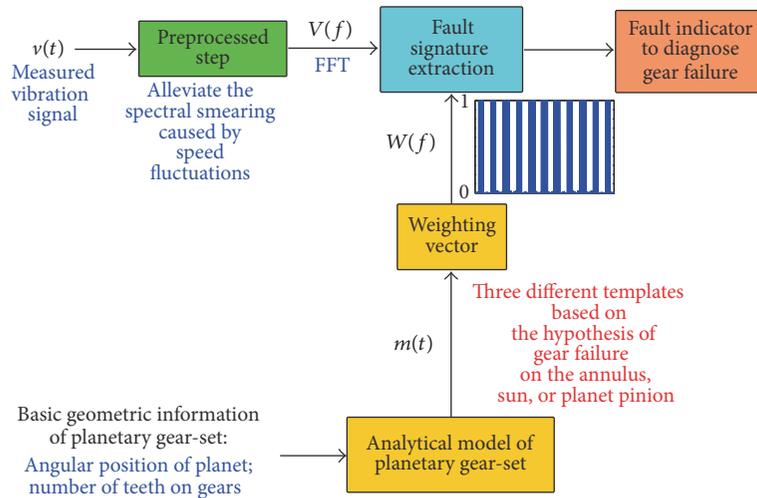


FIGURE 3: The architecture of the proposed vibration based diagnostic scheme.

usually absent in industrial drive trains. Fortunately, tachless order tracking algorithms have been developed to remove the smearing effect caused by the speed fluctuations, most of which are based on the estimation of the instantaneous frequency and sophisticated interpolations are required to resample the vibration data [15, 16]. Recently, an original pattern based tachless algorithm has been proposed by the author to remove the spectral smearing and squeeze the

smearing components as individual peaks [17]. After the preprocessing to alleviate the blurred spectrum caused by the speed fluctuations, the resulted vibration signal $\bar{v}(t)$ is then transformed into the frequency domain as the vibration spectrum $V(f)$.

Step 2. To extract the fault introduced information of the profuse sidebands masked in the noisy vibration spectrum

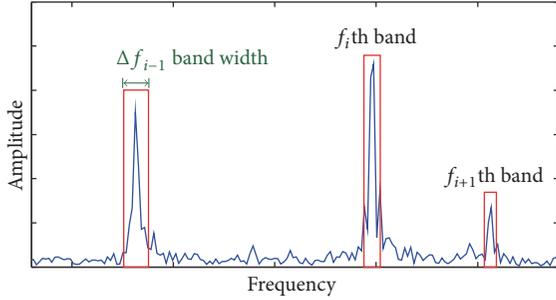


FIGURE 4: Illustration of the proposed comb filter.

$V(f)$, a comb filter is specifically designed herein for the planetary gear system based on the analytical model reviewed in the last section. The designed comb filter can be described with two variables vectors: f and Δf as presented in Figure 4, where the variable f_i represents the center frequency of the i th comb band and another variable Δf_i defines the corresponding bandwidth of f_i .

The center frequencies f of the comb filter are chosen as follows: the basic geometric information of the monitored planetary gearbox; that is, the number of teeth on the annulus, sun, and planet pinions, the number of planets, and the angular locations of planets serve as the input to the analytical model as given in (1)–(5). This model is used to generate vibration response templates $m(t)$ that summarize the sidebands characteristics of the healthy case and all the possible fault scenarios. Such model enables the capability of the identification of the gear fault. Then, the frequency components of the spectrum of the template $m(t)$, which have the nonzero magnitudes, are normalized to one to obtain a weighting vector $W(f)$. The normalization is necessary because the practical magnitudes of the frequency components in the measured vibration spectrum rely on the system specification of the gearbox, the operating conditions, and the severity of the faults, which cannot be estimated by the analytical model. These frequency components having the nonzero magnitudes correspond to the center frequencies f of the comb bands.

An important feature for the success of the proposed comb filter is that the bandwidth Δf should be selective enough so unwanted noisy components are correctly filtered out from the signal and all the energy of the fault introduced sidebands is kept in the filtered signal. Δf_i may be relatively small and it is independent of the center frequencies f of the comb bands. In order to automatically determine a proper bandwidth Δf_i for the i th comb band, statistical indicators are proposed to supervise the exhaustive search. The main principle of the proposed intelligent selective method is illustrated in Figure 5. The main peak plotted in blue color in Figure 5 mimics the i th sideband caused by the gear fault. In practice, the measured vibration signals are usually contaminated by the white noise, which is illustrated in Figure 5(b). The parasite peak highlighted in yellow color in Figure 5(c) is used to mimic the noise introduced by the other mechanical component of the gearbox. An arbitrary

chosen white noise series is given in Figure 5(d) for reference. It can be observed from Figure 5(c) that when the bandwidth Δf_i is smaller than the bandwidth of the main energy peak, the kurtosis of the spectrum in the selected frequency band decreases with the increase of the bandwidth Δf_i . However, when the bandwidth is larger than the desired main peak, the kurtosis value begins to increase with the increase of the bandwidth Δf_i . The same phenomenon is also applicable to the scenarios in Figures 5(a) and 5(b). Thus, the value of Δf_i can be selected according to this watershed as the proper bandwidth for the i th comb band. Generally speaking, a small RMS value of a comb band means that there is just background noise within this comb band as shown in Figure 5(d), the magnitudes of which can be set as zero to filter out the white noise.

Eventually, the obtained frequency lines f of $W(f)$ and their selected bandwidth Δf are used to form the comb filter to exclude the noise and extract the fault features. This comb filter allows for an automated highlight and identification of any additional frequency components caused by gear faults, which are located at frequencies depending on the geometry of the planetary gear-set and the location of the local fault.

Step 3. To intelligently diagnose the location of the local fault and the severity of the tooth damage, a frequency domain statistical metric, absolute-value logarithmic ratio (ALR), is designed as the fault indicator, which can be computed over the extracted signal after the denoising process:

$$\text{ALR} = \frac{1}{N} \sum_{n=1}^N \left| \log_e \frac{|\alpha(n)|_{\text{af}}}{|\alpha(n)|_{\text{bf}}} \right|, \quad (6)$$

where N is the number of the extracted nonzero frequency components and $\alpha(n)$ is the magnitude of the n th extracted frequency component. The subscripts “af” and “bf” denote after and before the local tooth damage affecting the vibration response, respectively. As manufacturing errors often exist in the practical gearboxes, previous publications found that the occurrence of an incipient damage at a single tooth can cause either the increase or decrease in magnitudes at certain frequency components with almost equal probability [18, 19]. Because of this, the value of the metric $\log_e(|\alpha(n)|_{\text{af}}/|\alpha(n)|_{\text{bf}}) = \log_e(|\alpha(n)|_{\text{af}}) - \log_e(|\alpha(n)|_{\text{bf}})$ can be either positive or negative. Thus, the absolute value is proposed in (6) to ensure that the accumulated fractional change $|\alpha(n)|_{\text{af}}/|\alpha(n)|_{\text{bf}}$ can always be regarded as incremental. As a consequence, the proposed fault indicator ALR assesses the changes, either increase or decrease, in the individual magnitudes of the extracted frequency components. The effect of using the natural logarithm in (6) is to substantially underweight the manufacturing errors so that they do not dominate the fault detection indicator in (6) during its real application [19]. Since the magnitudes of fault signatures increase with the fault degradation of the gear tooth, a larger value of the ALR implies a more severe fault [20]. Thus, the ALR is capable of assessing the severity of gear faults. The dynamic analysis of gearbox and experimental pretest can determine the appropriate thresholds of ALR for the classification of the severity of gear faults. In industrial

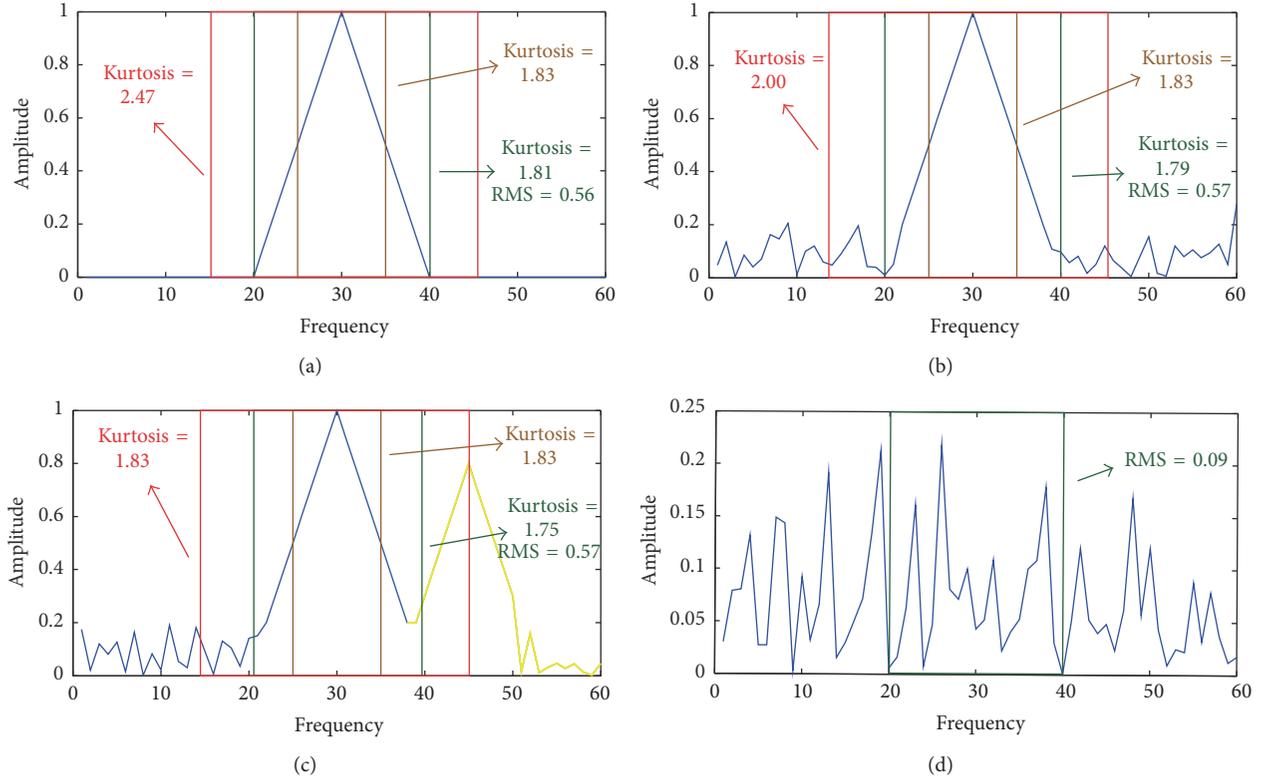


FIGURE 5: Illustration of the intelligent bandwidth selective principle: (a) single spectral peak; (b) single spectral peak with background noise; (c) single spectral peak with a parasite peak and background noise; (d) background noise.

environments, when the state and system parameters of a gearbox are unavailable, the historical data can be referred as the set of “bf” to track the degradation process of the gearbox. If the value of ALR equals 0, it implies no change in the individual sidebands and thus means free of fault in the monitored gearbox.

4. Simulation Results

4.1. Lumped Parameter Model of Planetary Gearbox. A two-stage gearbox is simulated to test the efficacy of the proposed diagnostic scheme. The schematic of this simulated planetary gearbox is illustrated in Figure 6. The first stage is an equally spaced planetary gear-set with 4 planets, in which $N_a = 100$, $N_s = 28$, and $N_p = 36$. The second stage is a fixed-axis gear pair, whose wheel gear has 48 teeth ($=N_{wg}$) and pinion gear has 30 teeth ($=N_{pg}$). The total gear ratio of the planetary gearbox equals $4.57 \times 1.6 = 7.31$. The rotational speed of the input shaft is set as 913 RPM. The fundamental gear mesh frequency of the planetary gear-set and the fixe-axis gear pair is 1521.8 Hz and 3338.1 Hz, respectively.

The dynamic model of planetary gearbox used in this work is illustrated in Figure 7 and described in detail in [10, 21]. The gears are simplified as rigid bodies connected to each other along the line of action through the corresponding gear mesh stiffness and viscous damping [22–24], the model

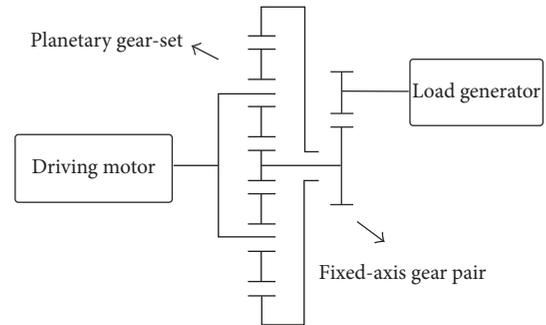


FIGURE 6: Schematic diagram for the simulated planetary gearbox.

of which has been widely applied to study the dynamics of industrial planetary gearbox. Since the gears are held by bearings, such rigid bodies are allowed to translate in x and y directions and to rotate freely around their centers in the x - y transverse plane of gears. Thus, the motion of the sun gear is defined with the translational displacement x_s and y_s and the angular coordinate θ_s . Similarly, the motion of the carrier is defined by x_c , y_c , and θ_c . α_i is the pressure angle and ψ_i is the initial angle location for planet i . The gear imperfections due to the manufacturing errors are simulated to use the error functions e with a $5 \mu\text{m}$ amplitude of saw-tooth [22] as shown in Figure 8(a). Thus, the gear mesh deformation along the line

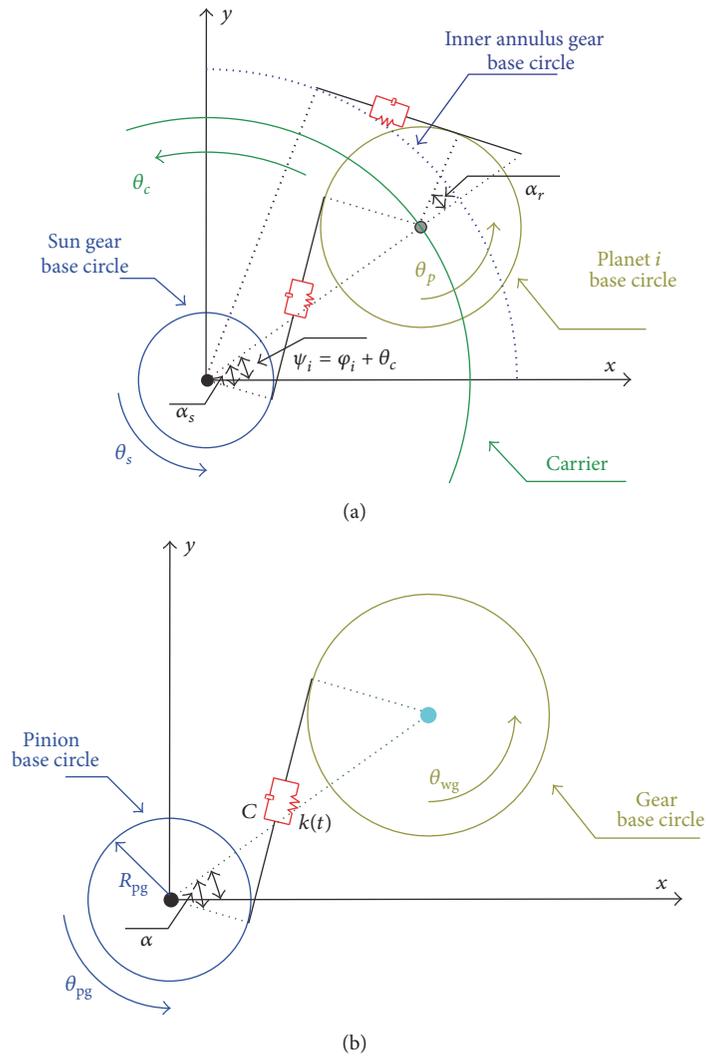


FIGURE 7: Two-dimensional lumped parameter model of a planetary gearbox: (a) the planetary gear-set and (b) the fixed-axis gear stage.

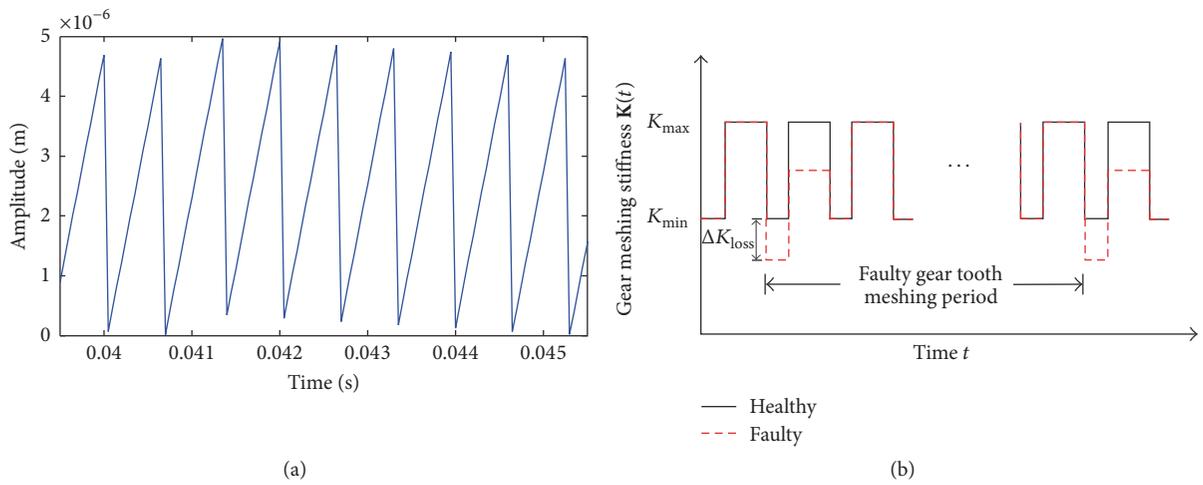


FIGURE 8: (a) Error function and (b) gear mesh stiffness under the healthy and faulty case.

of action between the sun gear and the i th planet gear can be defined as (R is the radius of the base circle):

$$\begin{aligned} \delta_{\text{spi}} = & (x_s - x_i + R_c \cos \psi_i) \sin(\alpha - \psi_i) \\ & + (y_s - y_i + R_c \sin \psi_i) \cos(\alpha - \psi_i) \\ & + (\theta_s - \theta_c) R_s - [-(\theta_p - \theta_c)] R_p + e_{\text{spi}}. \end{aligned} \quad (7)$$

The gear mesh deformation between the annulus gear and the i th planet gear can be written as

$$\begin{aligned} \delta_{\text{api}} = & (x_i - R_c \cos \psi_i) \sin(\alpha + \psi_i) \\ & - (y_i - R_c \sin \psi_i) \cos(\alpha + \psi_i) - (\theta_i - \theta_c) R_p \\ & - [-(0 - \theta_c)] R_r + e_{\text{api}}, \end{aligned} \quad (8)$$

where the i th planet pinion has rotation θ_i , x_i and y_i are the translational displacements of the i th planet.

The gear mesh deformation between the wheel gear and pinion gear along the line of action is expressed as

$$\begin{aligned} \delta = & (x_{\text{wg}} - x_{\text{pg}}) \sin \alpha + (y_{\text{wg}} - y_{\text{pg}}) \cos \alpha + \theta_{\text{wg}} r_{\text{wg}} \\ & - \theta_{\text{pg}} r_{\text{pg}} + e, \end{aligned} \quad (9)$$

where θ represents the rotational motion. x and y are the translational displacements of the gears. The subscripts wg and pg correspond to the wheel gear and pinion gear of the fixed-axis gear stage, respectively. R_{wg} and R_{pg} are the radius of the base circle of the wheel gear and pinion gear. α is the pressure angle of the fixed-axis gear pair.

Then, the dynamic gear mesh forces of the gear pair $F_m(t)$ can be further given as

$$F_m(t) = c\dot{\delta} + k(t)\delta, \quad (10)$$

where c is the damping coefficient and $k(t)$ is the time-varying gear mesh stiffness between a gear pair. The square waveform is often used to represent $k(t)$ for spur gear pairs, which is illustrated in Figure 8(b). The maximum and minimum values of gear mesh stiffness are estimated to be $K_{\text{max}} = 5 \times 10^8$ N/m and $K_{\text{min}} = 3 \times 10^8$ N/m for the simulated planetary gearbox. Furthermore, previous literatures have concluded that the local gear faults are always accompanied by a local reduction in the gear mesh stiffness. Thus, the common local gear faults are modeled by assuming a local drop from the general squared waveform of the gear mesh stiffness. In this work, the tooth damage-induced variation, 35% and 65% local stiffness loss ΔK_{loss} , is used to simulate the moderate and severe local sun gear faulted cases, respectively.

Finally, the global equation of motion for the gearbox can be expressed in the matrix form as

$$\mathbf{M}\ddot{\mathbf{Q}}(t) + [\mathbf{C} + \mathbf{C}_b]\dot{\mathbf{Q}}(t) + [\mathbf{K}(t) + \mathbf{K}_b]\mathbf{Q}(t) = \mathbf{F}(t), \quad (11)$$

where \mathbf{M} is the mass matrix, \mathbf{C} is the damping matrix, $\mathbf{K}(t)$ is the time-varying gear mesh stiffness matrix, \mathbf{C}_b and \mathbf{K}_b are the bearing damping and stiffness matrices, $\mathbf{F}(t)$ is the externally applied torques vector, and $\mathbf{Q}(t)$ is the degrees of freedom

vector that contains two coordinates for translational vibration and a coordinate for torsional motion for each gear in the plane containing the gear. The dynamic equations of the lumped parameter of gear-sets are numerically integrated in MATLAB/Simulink environment using ode45 solver and a 10 dB white noise is then superimposed to the results to simulate the background noise of the practical system.

4.2. Fault Detection. Figure 9 plots the simulated vibration spectrums around the second gear mesh harmonic of planetary gear-set (3043.5 Hz) under the healthy and severe local sun gear tooth faulted conditions. The raw spectrums in Figure 9 are not straightforward to interpret due to the complexities in the vibration response of the planetary gear-set itself. The vibration signals from the planetary gear-set are also superimposed with the background noise and the vibration contribution from the fixed-axis gear pair with run-out errors (the gear mesh frequency and its sidebands caused by the run-out error of the fixed-axis gear pair around 3338.1 Hz). From such raw spectrum of the multistage planetary gearbox, it is difficult to conclude whether there is any fault in the planetary gear-set or identify the location of the gear failure.

Thus, the developed diagnostic scheme shown in Figure 3 is applied to the simulated data to enhance the fault signatures. Figure 10 presents the processed spectrum with the proposed scheme, the filter template of which is based on the hypothesis of the sun gear fault scenario. The variations at the extracted frequency features in the spectrum caused by the local fault on the sun gear are now more easily visible in Figure 10.

Then, the proposed ALR is calculated using (6) with the spectrum from Figure 10. The calculated ALR of the moderate local sun gear fault case is 0.99 and the ALR of the severe local sun gear fault case equals 1.28, which indicates the capability of ALR to assess the severity of gear failure in the planetary gearbox. On the other hand, the calculated ALR nearly approaches 0 when applying the filter template on the hypothesis of the annulus and planet gear fault scenario. It implies that the examined planetary gear-set is free of the annulus and planet gear fault. Thus, this simulated case also verifies the effectiveness of the proposed scheme to diagnose the location of the gear failure.

In general, a gear tooth fault is always accompanied by a loss of gear mesh stiffness [1]. The reduction of gear mesh stiffness due to the incipient spalling or tooth breakage fault is within 10% of the original gear mesh stiffness [20]. The moderate spalling or tooth breakage fault can cause around 10% to 40% stiffness loss compared with the original gear mesh stiffness. When there is over 40% loss of the gear mesh stiffness, the tooth damage is labelled as the severe fault. Accordingly, extended simulations as presented in this section are performed to determine the appropriate thresholds of ALR for the assessment of fault severity, in which the simulated gearbox is assumed to have a 10% and 40% local stiffness loss, respectively. Based on the simulation results, the thresholds of ALR for the classification of fault severity in this work are suggested as follows: $0 < \text{ALR} \leq 0.34$ implies the existence of incipient gear fault, $0.34 < \text{ALR}$

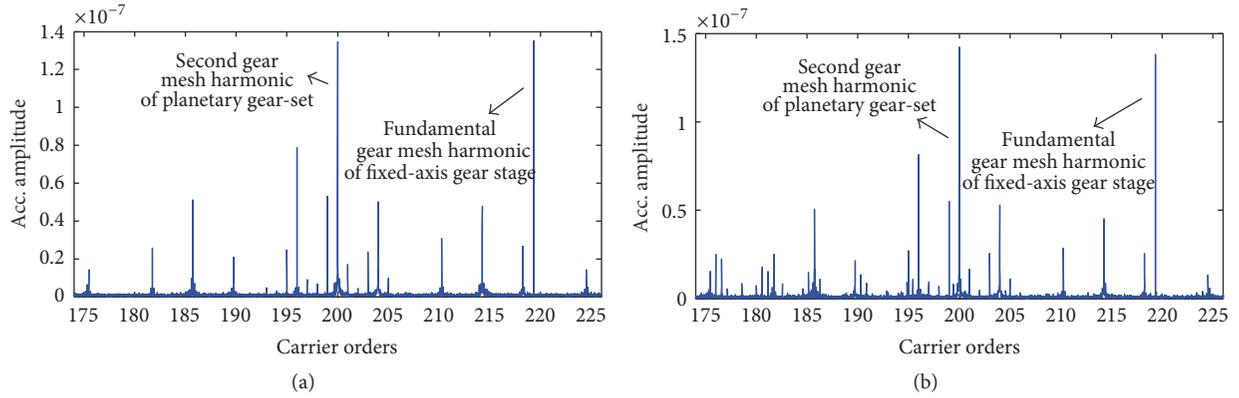


FIGURE 9: (a) The simulated raw vibration spectrum under the healthy case; (b) the simulated raw vibration spectrum under the severe sun gear faulted case.

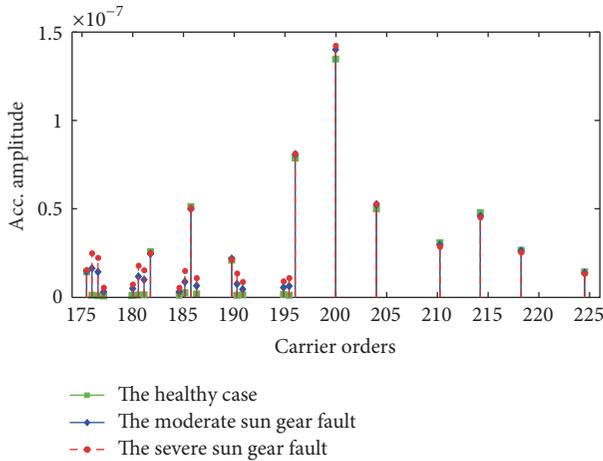


FIGURE 10: The processed spectrum after the proposed diagnostic scheme.

< 1.09 indicates the presence of moderate gear fault, and $ALR \geq 1.09$ means there is a severe gear fault. For a specific type of gearbox in practice, the exact threshold boundaries of ALR for the classification of fault severity should be adjusted against the actual system parameters of the gearbox using the proposed dynamic analysis discussed herein. Alternatively, the experiment based approach under typical operational conditions, which uses damaged gears under different stages of fault severity, can also give guideline to determine the thresholds of ALR for the classification of the severity of gear faults.

5. Experimental Validation

The proposed vibration based diagnostic scheme was applied to experimentally measured vibration signals from a planetary gearbox test rig shown in Figure 11. The input shaft of the gearbox was connected to a 3-phase AC motor (4 kW 4 pole) controlled using an ABB industrial drive and operated under the nominal speed of 1400 RPM. The load of the drive train was provided by a generator with a resistive load bank,

which was connected to the output shaft of the gearbox. The two tested gearboxes were designed as back-to-back planetary gear-sets, each having four equally spaced planet pinions with the number of annulus gear teeth $N_a = 84$, the number of sun gear teeth $N_s = 28$, and the number of planet pinion teeth $N_p = 28$. As the result of the back-to-back configuration, the overall gear ratio of the gearbox equaled 1 ($4 \times 1/4$). A moderate seeded spalled gear tooth fault was introduced to one of the annulus gear teeth (Figure 11(c)) using the electrodischarge machining (EDM). The vibration signals, plotted in Figure 12, were measured at a sampling rate of 20 KHz using an accelerometer mounted to the gearbox housing outside the annulus gear.

Since the raw vibration signals are often contaminated by the noises and profuse vibration features from other mechanical components of gearboxes, the manual visual investigation of the raw vibration spectrum as shown in Figure 12 is difficult and requires a deep skill and experience that is costly in industrials. Hence, the proposed automated diagnostic scheme for planetary gearbox described in Section 3 is employed to intelligently extract the fault signatures, which correspond to the gear failures in the planetary gear-set. The prerequisite input information of this automated diagnostic scheme is only the number of gear teeth of the monitored planetary gearbox and the approximated operational speed. The extracted fault signatures around the fundamental gear mesh frequency are shown in Figure 13, which is processed by the filter template on the hypothesis of the annulus gear fault scenario. The corresponding calculated ALR equals 0.84 that indicates the existence of a moderate annulus fault in the planetary gearbox. Thus, the effectiveness of proposed diagnostic scheme is validated, which is capable of the fault detection and identification from practical measured signals.

6. Conclusion

Failure diagnosis of gearboxes requires high quality fault signatures extracted from the raw measured vibration data to propagate through the fault diagnostic scheme. However, these fault signatures in the raw spectrum are often masked by noises and vibrations of other mechanical components as

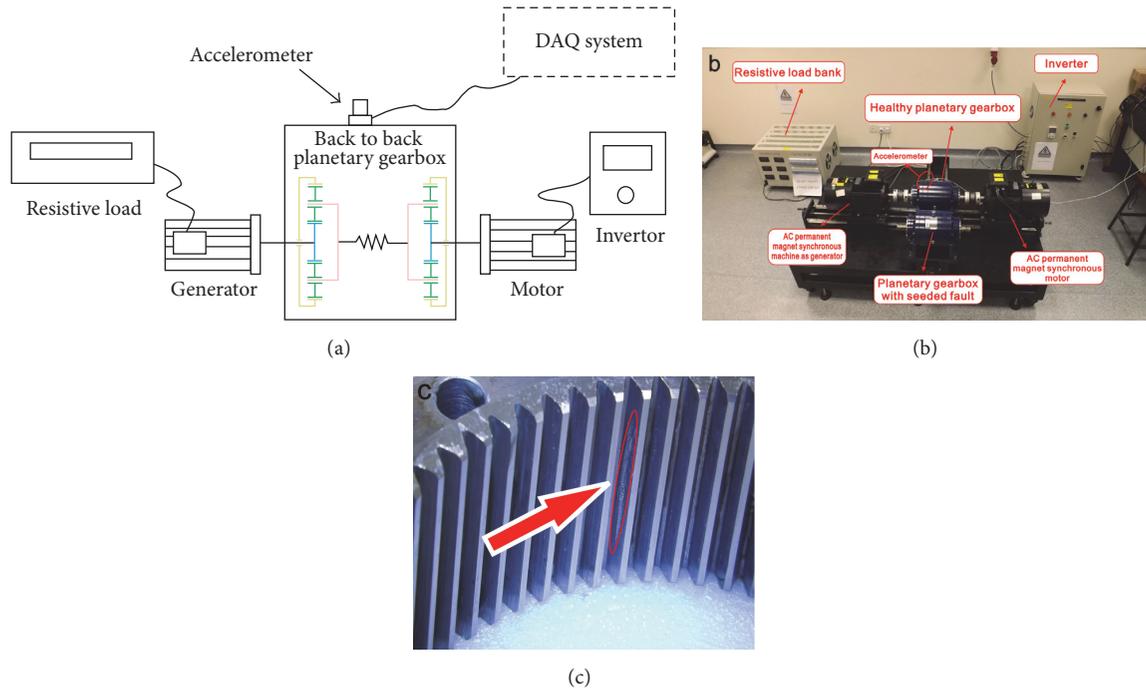


FIGURE 11: (a) Schematic of the test rig; (b) planetary gearbox test rig; and (c) seeded gear fault on the annulus gear.

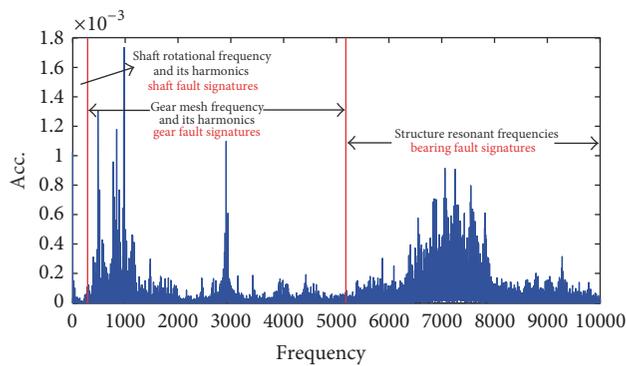


FIGURE 12: The vibration spectrum of the raw measured signal picked by the fixed accelerometer mounted on the annulus gear.

well as the white noise from the sensors. Due to the complex nature of sidebands in the measured vibration signal picked up by a fixed accelerometer, the application of conventional vibration based diagnostic techniques to planetary gearboxes is a challenge. Hence, this paper develops a novel vibration based diagnostic scheme for planetary gearboxes. The proposed diagnostic system synergistically integrates an analytical model, a denoising algorithm, and frequency domain indicators. This proposed diagnostic scheme can address the limitation of the requirement of extensive stationary data of TSA algorithm, which highlights the fault introduced sidebands by using a comb filter in the frequency domain based on the analytical model of planetary gear-sets. Its performance is firstly tested using simulated signals from a dynamic model of planetary gearbox. Then, the diagnostic

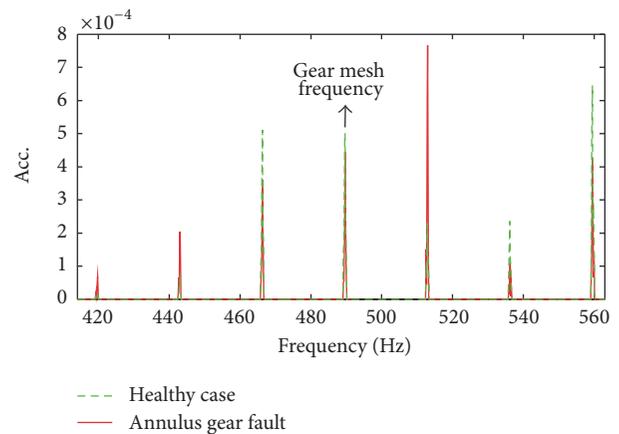


FIGURE 13: The processed spectrum after the diagnostic scheme around the fundamental gear mesh harmonic on the hypothesis of the annulus gear fault scenario.

scheme is also validated with experimental data from a 4 kW planetary gearbox test rig.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The authors are pleased to acknowledge the financial support of National Natural Science Foundation of China (Grant

no. 51505353) and the Fundamental Research Funds for the Central Universities (WUT: 163104002).

References

- [1] W. Q. Wang, F. Ismail, and M. Farid Golnaraghi, "Assessment of gear damage monitoring techniques using vibration measurements," *Mechanical Systems and Signal Processing*, vol. 15, no. 5, pp. 905–922, 2001.
- [2] J. Urbanek, J. Antoni, and T. Barszcz, "Detection of signal component modulations using modulation intensity distribution," *Mechanical Systems and Signal Processing*, vol. 28, pp. 399–413, 2012.
- [3] P. D. McFadden and J. D. Smith, "An explanation for the asymmetry of the modulation sidebands about the tooth meshing frequency in epicyclic gear vibration," *Proceedings of the Institution of Mechanical Engineers. Part C. Mechanical engineering science*, vol. 199, no. 1, pp. 65–70, 1985.
- [4] W. D. Mark and J. A. Hines, "Stationary transducer response to planetary-gear vibration excitation with non-uniform planet loading," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1366–1381, 2009.
- [5] J. McNames, "Fourier series analysis of epicyclic gearbox vibration," *ASME Journal of Vibration and Acoustics*, vol. 124, no. 1, pp. 150–152, 2002.
- [6] M. Inalpolat and A. Kahraman, "A dynamic model to predict modulation sidebands of a planetary gear set having manufacturing errors," *Journal of Sound and Vibration*, vol. 329, no. 4, pp. 371–393, 2010.
- [7] B. Zhang, T. Khawaja, R. Patrick, G. Vachtsevanos, M. E. Orchard, and A. Saxena, "Application of blind deconvolution denoising in failure prognosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 2, pp. 303–310, 2009.
- [8] D. G. Lewicki, K. E. LaBerge, R. T. Ehinger, and J. Fetty, "Planetary gearbox fault detection using vibration separation techniques," NASA Technical Report NASA/TM-2011-217127, 2011.
- [9] J. M. Ha, B. D. Youn, H. Oh, B. Han, Y. Jung, and J. Park, "Autocorrelation-based time synchronous averaging for condition monitoring of planetary gearboxes in wind turbines," *Mechanical Systems and Signal Processing*, vol. 70–71, pp. 161–175, 2016.
- [10] L. Hong and J. S. Dhupia, "A time domain approach to diagnose gearbox fault based on measured vibration signals," *Journal of Sound and Vibration*, vol. 333, no. 7, pp. 2164–2180, 2014.
- [11] L. Hong, J. S. Dhupia, and S. Sheng, "An explanation of frequency features enabling detection of faults in equally-spaced planetary gearbox," *Mechanism and Machine Theory*, vol. 73, pp. 169–183, 2014.
- [12] Y. Lei, Z. Liu, J. Lin, and F. Lu, "Phenomenological models of vibration signals for condition monitoring and fault diagnosis of epicyclic gearboxes," *Journal of Sound and Vibration*, vol. 369, pp. 266–281, 2016.
- [13] M. Inalpolat and A. Kahraman, "A theoretical and experimental investigation of modulation sidebands of planetary gear sets," *Journal of Sound and Vibration*, vol. 323, no. 3–5, pp. 677–696, 2009.
- [14] R. G. Parker and J. Lin, "Mesh phasing relationships in planetary and epicyclic gears," *Journal of Mechanical Design, Transactions of the ASME*, vol. 126, no. 2, pp. 365–370, 2004.
- [15] F. Bonnardot, M. El Badaoui, R. B. Randall, J. Danière, and F. Guillet, "Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation)," *Mechanical Systems and Signal Processing*, vol. 19, no. 4, pp. 766–785, 2005.
- [16] M. Zhao, J. Lin, X. Wang, Y. Lei, and J. Cao, "A tachometer-less order tracking technique for large speed variations," *Mechanical Systems and Signal Processing*, vol. 40, no. 1, pp. 76–90, 2013.
- [17] L. Hong, Y. Qu, J. S. Dhupia, and Y. Tan, "A novel fault diagnostic technique for gearboxes under speed fluctuations without angular speed measurement," in *Proceedings of the ASME/ISCIE International Symposium on Flexible Automation*, Cleveland, Ohio, USA, August 2016.
- [18] L. Hong, Y. Qu, J. S. Dhupia, and Y. Tan, "A novel synergistic diagnostic scheme for planetary gearboxes based on an analytical vibration model of planetary gear-sets," in *Proceedings of the Prognostics and System Health Management Conference*, Chengdu, China, October 2016.
- [19] W. D. Mark, H. Lee, R. Patrick, and J. D. Coker, "A simple frequency-domain algorithm for early detection of damaged gear teeth," *Mechanical Systems and Signal Processing*, vol. 24, no. 8, pp. 2807–2823, 2010.
- [20] F. Chaari, W. Baccar, M. S. Abbes, and M. Haddar, "Effect of spalling or tooth breakage on gearmesh stiffness and dynamic response of a one-stage spur gear transmission," *European Journal of Mechanics, A/Solids*, vol. 27, no. 4, pp. 691–705, 2008.
- [21] W. Kim, J. Y. Lee, and J. Chung, "Dynamic analysis for a planetary gear with time-varying pressure angles and contact ratios," *Journal of Sound and Vibration*, vol. 331, no. 4, pp. 883–901, 2012.
- [22] W. Bartelmus, "Mathematical modelling and computer simulations as an aid to gearbox diagnostics," *Mechanical Systems and Signal Processing*, vol. 15, no. 5, pp. 855–871, 2001.
- [23] V. K. Ambarisha and R. G. Parker, "Nonlinear dynamics of planetary gears using analytical and finite element models," *Journal of Sound and Vibration*, vol. 302, no. 3, pp. 577–595, 2007.
- [24] T. M. Ericson and R. G. Parker, "Planetary gear modal vibration experiments and correlation against lumped-parameter and finite element models," *Journal of Sound and Vibration*, vol. 332, no. 9, pp. 2350–2375, 2013.

Research Article

Adaptive Multiscale Noise Control Enhanced Stochastic Resonance Method Based on Modified EEMD with Its Application in Bearing Fault Diagnosis

Jimeng Li¹ and Jinfeng Zhang²

¹College of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China

²College of Liren, Yanshan University, Qinhuangdao 066004, China

Correspondence should be addressed to Jimeng Li; xjtuljm@163.com

Received 30 June 2016; Accepted 28 September 2016

Academic Editor: Ganging Song

Copyright © 2016 J. Li and J. Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The structure of mechanical equipment becomes increasingly complex, and tough environments under which it works often make bearings and gears subject to failure. However, effective extraction of useful feature information submerged in strong noise that is indicative of structural defects has remained a major challenge. Therefore, an adaptive multiscale noise control enhanced stochastic resonance (SR) method based on modified ensemble empirical mode decomposition (EEMD) for mechanical fault diagnosis is proposed in the paper. According to the oscillation characteristics of signal itself, the algorithm of modified EEMD can adaptively decompose the fault signals into different scales and it reduces the decomposition levels to improve calculation efficiency of the proposed method. Through filter processing with the constructed filters, the orthogonality of adjacent intrinsic mode functions (IMFs) can be improved, which is conducive to enhancing the extraction of weak features from strong noise. The constructed signal obtained by using IMFs is inputted into the SR system, and the noise control parameter of different scales is optimized and selected with the help of the genetic algorithm, thus achieving the enhancement extraction of weak features. Finally, simulation experiments and engineering application of bearing fault diagnosis demonstrate the effectiveness and feasibility of the proposed method.

1. Introduction

With the fast development of modern technology, the structure of mechanical equipment becomes increasingly complex, and the automation degree becomes more and more high, but tough environments under which it works often make bearings and gears subject to failure. The failure may deteriorate mechanical performance and even lead to fatal breakdowns. Therefore, how to ensure the safe and reliable operation of mechanical equipment and reduce economic losses is increasingly becoming a hot topic concerned by enterprises. The health monitoring and fault diagnosis technology provides an effective solution for the predictive maintenance of mechanical equipment operating state. In view of the inherent link between machine operation and vibration, vibration signals collected from mechanical equipment carry rich information on machine health conditions. Therefore, vibration signal analysis [1–3] has been extensively investigated

during the past decades. The emergence of periodic components is often associated with component failures in vibration analysis. However, fault signals collected from mechanical equipment are often corrupted by strong noise coming from other coupled machine components and the working environment, which increases the difficulty in fault identification. Thus, effective extraction of weak features submerged in strong noise that are indicative of structural defects has remained a major challenge. In order to obtain vital feature information from vibration signals, various signal processing techniques, like variational mode decomposition (VMD) [4, 5], synchrosqueezing transform (SST) [6, 7], wavelet transform (WT) [8–10], and so forth, have been extensively studied and used in machinery fault diagnosis. Traditionally, noise is always considered an undesirable disturbance, thus these signal processing techniques mainly focus on filtering and denoising the signals to extract fault features. In fact, noise is not just a source of signal contamination, but it also

represents a kind of signal energy. If the energy generated by noise can be used properly, it is conducive to extracting the feature information from the signals submerged under strong noise. Stochastic resonance (SR) [11] is a kind of typical noise-assisted data processing method. Unlike traditional denoising techniques, SR realizes the detection of weak signal by utilizing noise instead of eliminating noise, and the weak signal features are not weakened but enhanced.

The concept of SR was first introduced in the 1980s, which was applied to describe the periodicity associated with Earth's ice ages in climatology [12, 13]. SR, as a nonlinear physical phenomenon, emphasizes the synergistic effect between a nonlinear dynamic system, a small parameter signal, and noise. As a special "low-pass filter," SR can transfer noise energy to useful signal with the assistance of nonlinear system, thereby realizing the elimination of noise and enhancement of useful signal features. Consequently, by virtue of the unique advantage of using noise to enhance weak signals, SR has been widely studied in machinery fault diagnosis field [14, 15]. Qin et al. [16] proposed an adaptive and fast SR method based on dyadic wavelet transform and least square parameters solving algorithm to extract the fault feature of a rotor system, which can increase the noise utilization and does not need to set up the searching range and searching step size of system parameters. Aiming at the problem of detecting the multifrequency signals buried under strong noise, Han et al. [17] proposed a multistable SR method by using wavelet transform and parameter compensation band-pass algorithm, which enhanced the signal amplitude and achieved the effective extraction of bearing fault features. Wang et al. [18] proposed an adaptive multiscale noise tuning SR based on wavelet transform and weighted power kurtosis to diagnose the rolling bearing fault and realized the adaptive selection of control parameters through the artificial fish swarm algorithm. In order to realize the fault diagnosis of planetary gear transmission, a weak fault feature information extraction method based on ensemble empirical mode decomposition (EEMD) and adaptive SR was proposed [19]. Through EEMD, the effective IMFs containing main fault feature information are selected and reconstructed. Then the reconstructed signal is inputted into adaptive SR system, and the weak fault feature information can be extracted from the output signal of SR system. By analyzing the SR phenomenon of a two-dimensional duffing oscillator, Lai and Leng [20] proposed a generalized parameter-adjusted SR model, which can be used for the fault diagnosis of rolling bearing. In addition, many researches have also been conducted on the SR model and SR enhancement methods [21, 22].

The algorithm of multiscale noise tuning provides an effective control strategy for realizing the SR. However, there are some shortcomings in the method of multiscale noise tuning based on wavelet decomposition, such as the selection of optimal wavelet bases, which makes it have some limitations in processing diverse signals in engineering practice. Adaptive signal decomposition methods provide an effective solution for complicated signal analysis and feature extraction, which can adaptively decompose the signal into different scales according to the oscillation characteristics of signal itself, thus avoiding artificial intervention and optimizing

the data analysis flow. As a noise-assisted adaptive signal decomposition method, EEMD can eliminate the mode mixing phenomenon in empirical mode decomposition (EMD) through adding white noise to signal, and it decomposes the complicated signal into a set of simple components named intrinsic mode function (IMF) [23–25]. The IMFs represent the natural oscillatory mode embedded in the signal, which are determined by the signal itself, instead of predetermined kernels. Therefore, a new adaptive multiscale noise control enhanced SR method based on modified EEMD for bearing fault diagnosis is studied by using the adaptive decomposition capability of EEMD and the unique advantage of SR using noise to enhance weak signal features. EEMD is used for signal multiscale decomposition, and through improving the EEMD algorithm, the decomposition levels are reduced to increase the computation efficiency of the proposed method. The band-pass filters are constructed in terms of the spectrum distribution characteristics of IMFs, so as to eliminate some aliasing components and improve the orthogonality of adjacent IMFs through filter processing, which is conducive to enhancing and extracting of weak information features to some extent. And then, the signal, used as the input of SR system, is reconstructed with the IMFs obtained by the modified EEMD algorithm, and the noise control parameter of different scales is optimized and selected with the help of the genetic algorithm to achieve the extraction of weak features. Therefore, the proposed method can not only adaptively decompose the signal into different scales but also realize the adaptive selection of noise control parameter of different scales, which is well-suited for enhancement of rotating machine weak fault identification. Experiments and application demonstrate that the proposed method is validated to be effective in detecting the weak fault features in bearing fault diagnosis.

The rest of the paper is arranged as follows. A brief introduction to the theoretical background of SR is provided in Section 2. The method of adaptive multiscale noise control enhanced SR based on modified EEMD is described in detail in Section 3. Some experimental and practical data are applied to verify the effectiveness of the proposed method in Section 4. Finally, conclusions are drawn in Section 5.

2. SR Basic Theory

SR describes a phenomenon that the weak signal is enhanced and the noise is weakened through the interaction of the nonlinear system, small parameter signal, and noise, whose dynamic behavior can be represented by the Brownian motion equation of particles. And the overdamped SR equation with a nonlinear bistable model in the presence of periodic force and noise can be written as follows:

$$\frac{dx(t)}{dt} = ax(t) - bx(t)^3 + A \sin(2\pi ft + \varphi) + n(t), \quad (1)$$

where $x(t)$ denotes the system output, parameters a and b are positive real numbers, A and f are the amplitude and frequency of the periodic force, respectively, and $n(t)$ is a Gaussian white noise with zero mean and D variance. The potential function is $U(x, t) = -(a/2)x(t)^2 + (b/4)x(t)^4$ with

two stable points $x = \pm\sqrt{a/b}$ and one critical stable point $x_0 = 0$. The height of potential barrier is $\Delta U = a^2/4b$. From (1), the system output $x(t)$ is actually the Brownian particle trajectory in the potential function $U(x)$ under the combined action of periodic force and noise. And, the crucial process of using SR to detect weak signal is to adjust system parameters or noise intensity to make the transition rate of the particle caused by noise match the oscillation period of the input signal, thus amplifying the particle movements in single potential well to the transition motions between double potential wells to realize the feature enhancement of input signal.

It is generally known that the theory basis of SR is the adiabatic approximation theory, which requires the amplitude and frequency of periodic signal as well as noise intensity to be smaller than 1. However, the defective signals caused by components fault of rotating machine have difficulty in satisfying the above requirements in engineering application. Therefore, in order to solve the problem of using SR to detect large parameter signals, several parameter tuning SR methods have been researched based on a normalized scale transformation.

Mathematically, let $y = x\sqrt{b/a}$, $\tau = at$, and (1) be written as

$$\frac{dy}{d\tau} = y - y^3 + \sqrt{\frac{b}{a^3}} \left[A \sin\left(\frac{2\pi f\tau}{a} + \varphi\right) + n\left(\frac{\tau}{a}\right) \right]. \quad (2)$$

Equation (2) indicates that the frequency of the periodic signal is normalized to be $1/a$ times that of original signal through the normalized scale transform, and the normalized frequency of larger parameter signal can satisfy the small parameters requirements of SR for input signal by selecting a corresponding larger parameter a . And importantly, according to (1) and the derivation process of (2), the process of realizing SR by adjusting the system parameters is consistent with that of realizing SR by adjusting the input signal strength and noise intensity on the premise of the constant system parameters in nature. Consequently, we can adjust the input signal strength and noise intensity to realize the SR detection of weak signal.

3. Adaptive Multiscale Noise Control Enhanced SR Method Based on Modified EEMD

Recent studies demonstrate that the SR effect can be driven by different scales of noise but with different degree [26]. Therefore, according to the influence of noise at different scales on the SR, multiscale noise tuning SR methods based on wavelet transform have been studied in recent years [27, 28]. However, based on the inner product transform principle, the selection of wavelet bases has important effects on the decomposition results in wavelet transform. When the selected wavelet base is inappropriate, namely, it does not match up with the target signal, the useful signal features may be impaired. Additionally, the width of frequency bands obtained by wavelet decomposition is fixed, which cannot be adjusted adaptively according to the oscillation characteristics of signal itself. Therefore, the existing multiscale noise

tuning methods based on wavelet transform still have some shortcomings, which have difficulty in satisfying the demand of the diversity of signals in engineering practice. EMD, as an adaptive signal processing method, has been developed and widely applied in machinery fault diagnosis. Based on the local characteristic time scales of a signal, EMD can adaptively decompose the complicated signal into a set of IMF components. However, it has a major drawback, which is the mode mixing problem. Therefore, EEMD, as an improved method of EMD, is presented to alleviate the mode mixing problem in EMD. The principle of EEMD algorithm is the following: by using the statistical property that Gaussian white noise is uniformly distributed over the whole frequency range, Gaussian white noise is added to a signal, which makes the signal continuous in different scales to alleviate the mode mixing problem in EMD. Finally, the added white noise can be decreased or even completely canceled out through the ensemble mean of enough trials, and the ensemble mean is treated as the true answer. Therefore, according to the oscillation characteristics of signal itself, EEMD algorithm can realize the adaptive multiscale decomposition for input signal, which provides a new control strategy for realizing SR by adjusting multiscale noise. EEMD overcomes the mode mixing problem through enough iterations and trails which is at the cost of increasing the calculation amount. However, in consideration of the requirements of the multiscale noise control SR algorithm in this study, the algorithm of EEMD is modified to simplify the calculation, and a new adaptive multiscale noise control enhanced SR method based on modified EEMD is proposed. By using the oscillation characteristics of signal itself, the modified EEMD algorithm can adaptively decompose the signal into different scales, and by adjusting the noise intensity of different scales the proposed method can achieve the enhanced extraction of weak signal features. The important steps of the proposed method are described in detail below.

3.1. Modified EEMD Algorithm. As mentioned above, EEMD achieves the adaptive partition of frequency band for input signal, and the decomposition level n is not set by manual but depends on the local characteristic time scales of signal itself. Therefore, how to select the appropriate IMF components from the n IMFs obtained by EEMD, namely, that determines the reconstruction scale L to reconstruct the signal used as the input of SR system, is the key of realizing the multiscale noise control SR method based on EEMD.

According to the process of EMD algorithm extracting IMF components, the high-frequency components contained in the signal are extracted first, and the low-frequency components are extracted last. So, the obtained n IMFs are arranged in descending order by the center frequency of each IMF component. And the process of determining the reconstruction scale L is to seek out the L th IMF component containing the target signal from n IMFs, thereby obtaining the first L IMFs which are served as the reconstruction components of multiscale noise control algorithm, and the remainder $n - L$ IMFs below the target signal frequency are abandoned. Aiming at the requirements of the proposed method, EEMD is applied to decompose the signal into

- (1) Initialize: $r_0(t) \leftarrow x_i(t)$, $j \leftarrow 1$, and $Err_{j-1} \leftarrow f_s/2$
- (2) Extract the j th IMF $c_{ij}(t)$
 - (a) $h_{j(p-1)}(t) \leftarrow r_{j-1}(t)$, and $p \leftarrow 1$
 - (b) Extract the local maxima and minima of $h_{j(p-1)}(t)$, and obtain the mean $m_{j(p-1)}(t)$ of upper and lower envelopes of $h_{j(p-1)}(t)$ by cubic spline interpolation
 - (c) $h_{jp}(t) \leftarrow h_{j(p-1)}(t) - m_{j(p-1)}(t)$
 - (d) If $h_{jp}(t)$ is an IMF then set $c_{ij}(t) = h_{jp}(t)$, else go to step (a) with $p \leftarrow p + 1$
- (3) Calculate the difference value Err_j between the center frequency f_j of $c_{ij}(t)$ and the frequency f_0 of target signal
 - (a) Calculate the discrete Fourier transform $C_{ij}(k)$ of $c_{ij}(t)$
 - (b) Calculate the center frequency f_j of $c_{ij}(t)$

$$f_j \leftarrow \frac{\sum_{k=1}^{N/2} [k(f_s/N) |C_{ij}(k)|^2]}{\sum_{k=1}^{N/2} |C_{ij}(k)|^2}$$
 - (c) $Err_j \leftarrow |f_j - f_0|$
- (4) Determine the terminal condition
 If $Err_j > Err_{j-1}$, the decomposition process is finished, and the obtained IMFs are $c_i(t) = \{c_{i1}(t), c_{i2}(t), \dots, c_{ij}(t)\}$, $j = 1, 2, \dots, L + 1$, and the IMF c_{iL} corresponding to the decomposition scale L includes the target signal, else $r_j(t) = r_{j-1}(t) - c_{ij}(t)$. If $r_j(t)$ still has least 2 extremum then go to step (2) with $j = j + 1$.

ALGORITHM 1: The algorithm of modified EMD.

different scales, but the algorithm of multiscale noise control does not need all of the IMFs obtained by EEMD but just the first L IMFs. For this reason, the stopping criterion for iteration of EMD algorithm extracting IMF components is modified, and thus a modified EEMD algorithm is presented, which can effectively reduce the decomposition levels of EMD algorithm to improve the overall calculation efficiency of the proposed method in the paper.

The modified EEMD algorithm is described in detail below:

- (1) Initialize the number of ensemble Num, the standard deviation of added white noise σ .
- (2) Add a Gaussian white noise with the given standard deviation σ to the investigated signal $x(t)$ with the length of N and the sample frequency f_s ,

$$x_i(t) = x(t) + n_i(t) \quad i = 1, 2, \dots, \text{Num}, \quad (3)$$

where $n_i(t)$ denotes the i th added white noise and $x_i(t)$ indicates the noise-added signal of the i th trial.

- (3) Decompose the noise-added signal $x_i(t)$ with the modified EMD algorithm and determine the decomposition scale L by using the target signal frequency f_0 and the center frequency f_j of each IMF obtained in the decomposition process, thus obtaining $L + 1$ IMFs. The concrete decomposition process is described as shown in Algorithm 1.
- (4) Repeat steps (2) and (3) again and again, but with different white noise each time, and calculate the ensemble mean \bar{c}_j of the Num trials for each IMF:

$$\bar{c}_j = \frac{1}{\text{Num}} \sum_{i=1}^{\text{Num}} c_{ij}(t), \quad j = 1, 2, \dots, L + 1. \quad (4)$$

- (5) Conduct the mean \bar{c}_j of each of the $L + 1$ IMFs as the final IMFs.

It is clear that the $L + 1$ IMFs can be obtained through the modified EEMD algorithm, and the L th IMF \bar{c}_L of them includes the target signal. Accordingly, the first L IMFs are applied to reconstruct the signal used as the input of SR system. Obviously, the modified EEMD algorithm effectively reduces the computation load and improves the efficiency of the proposed method. And the IMF components below the target signal frequency are abandoned to avoid the influence of low-frequency components on the detection result of SR and improve the detection accuracy of target signal.

3.2. Construction of Filters Based on IMFs. According to the oscillation characteristics of signal itself, EEMD can partition the frequency band adaptively, and the bandwidth of each IMF component depends on the natural oscillation mode embedded in the signal, instead of being determined by human being. However, due to the influence of strong background noise, the obtained IMFs are not strictly orthogonal to each other, as shown in Figure 1. Therefore, based on the spectrum distribution characteristics of $L + 1$ IMFs obtained by modified EEMD, the $L + 1$ IMFs are employed to construct the corresponding filters to filter the L IMFs, so as to eliminate the spectrum energy leakage and improve the orthogonality of adjacent IMFs.

From Figure 1, it is observed that the frequency spectrum distribution of the first IMF component represents the high-pass characteristics and the rest of IMF components show the band-pass characteristics in frequency domain. Accordingly, the corresponding high-pass and band-pass filters can be constructed by using the frequency spectrum distribution characteristics of IMFs.

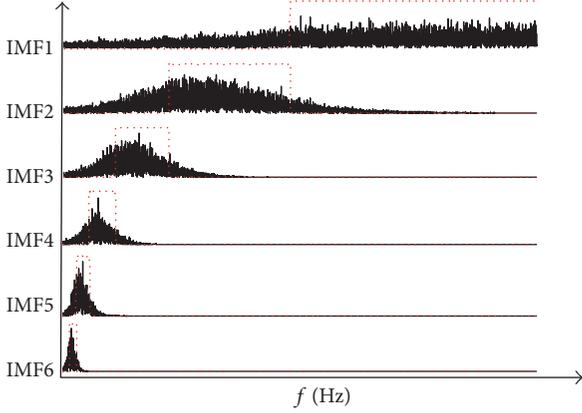


FIGURE 1: Construction sketch of filters based on IMFs.

First, according to the frequency spectrum characteristics of the first IMF component, the parameters of high-pass filter are designed as follows:

$$\begin{aligned} f_{1p} &= \frac{(f_1 + f_2)}{2}, \\ f_{1s} &= f_{1p} - 50, \end{aligned} \quad (5)$$

where f_{1p} is the cutoff frequency of passband and f_{1s} is the cutoff frequency of stopband. f_1 and f_2 are the center frequencies of the first and second IMF components, respectively.

Similarly, using the frequency spectrum characteristics of each IMF $_j$ ($j = 2, 3, \dots, L$), the parameters of band-pass filter corresponding to each IMF component are constructed as follows:

$$\begin{aligned} f_{jp1} &= \frac{(f_j + f_{j+1})}{2}, \\ f_{js1} &= f_{jp1} - 10, \\ f_{jp2} &= \frac{(f_j + f_{j-1})}{2}, \\ f_{js2} &= f_{jp2} + 10, \end{aligned} \quad (6)$$

$$j = 2, 3, \dots, L,$$

where f_{js1} and f_{js2} are the stopband lower limit frequency and upper limit frequency, respectively, and f_{jp1} and f_{jp2} are the passband lower limit frequency and upper limit frequency, respectively. And the stopband attenuation and passband attenuation of the constructed filters are 60 dB and 0.1 dB, respectively. f_j is the center frequency of the j th IMF component. It is worth noting that the parameters of filters in (5) and (6) are the reference values, which can be appropriately adjusted according to the specific requirements on the premise of ensuring the orthogonality of adjacent filters as much as possible. In Figure 1, the amplitude-frequency curves of the constructed filters corresponding to the IMF components are denoted by the red dotted line. Obviously, these constructed filters preserve good orthogonality at

boundaries, and the filter processing for the L IMFs with the constructed filters can effectively improve the orthogonality of adjacent IMFs.

3.3. Multiscale Reconstruction with Noise Intensity of Different Scales. The L IMFs obtained by the modified EEMD algorithm are filtered with the constructed filters and their corresponding filtered results R_j ($j = 1, 2, \dots, L$) are adjusted with a controllable parameter α to realize the noise control of different scales and then employed to reconstruct the input signal. The reconstruction formula is as follows:

$$\tilde{x} = \sum_{j=1}^L (2^{\alpha+j} \cdot R_j) \quad j = 1, 2, \dots, L, \quad (7)$$

where \tilde{x} is the reconstructed signal, α is a controllable parameter, which denotes the noise intensity of different scales, and R_j is the filtered j th IMF component.

3.4. Algorithm Flow of Adaptive Multiscale Noise Control Enhanced SR Based on Modified EEMD. The algorithm procedure of adaptive multiscale noise control enhanced SR based on modified EEMD is illustrated in Figure 2. The concrete steps of this algorithm are as follows.

(1) *Vibration Signal Preprocessing.* The vibration signals collected from bearings are demodulated by the Hilbert envelop analysis, and the obtained envelope signals are used as the input of modified EEMD algorithm.

(2) *Parameter Initialization.* Set the SR system parameters $a = 1$ and $b = 1$, the searching range of noise intensity, the initial parameters of genetic algorithm, and the standard deviation of added white noise and the ensemble number of modified EEMD algorithm.

(3) *Multiscale Decomposition and Filter Processing.* The envelop signal obtained by preprocessing is adaptively decomposed into $L + 1$ IMFs by using modified EEMD algorithm and then by using the frequency spectrum distribution characteristics of $L + 1$ IMFs to construct the corresponding filters, which are applied to filter the L IMFs.

(4) *Adaptive Multiscale Noise Control SR.* Equation (7) is applied to reconstruct the input signal with the filtered L IMFs, and then the reconstructed signal \tilde{x} is entered into the rescaling SR system. The genetic algorithm is employed to adaptively select and optimize noise intensity α , whose fitness function is constructed by using the signal-to-noise ratio (SNR) of resonance output. And based on the maximum of resonance output SNR, the optimal noise intensity α of different scales can be obtained, thus realizing the resonance detection of input signal. The calculation formula of SNR is as follows:

$$\text{SNR} = 10 \log_{10} \left(\frac{A_d}{A_n} \right), \quad (8)$$

where A_d is the amplitude value corresponding to the driving signal frequency and A_n is the sum of all the amplitude values except A_d in the amplitude spectrum.

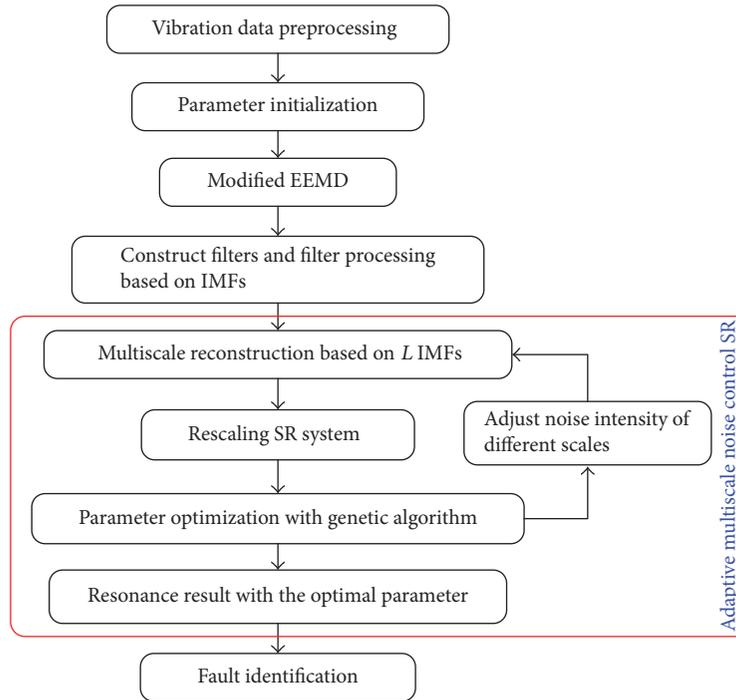


FIGURE 2: Flow diagram of adaptive multiscale noise control enhanced SR algorithm based on modified EEMD.

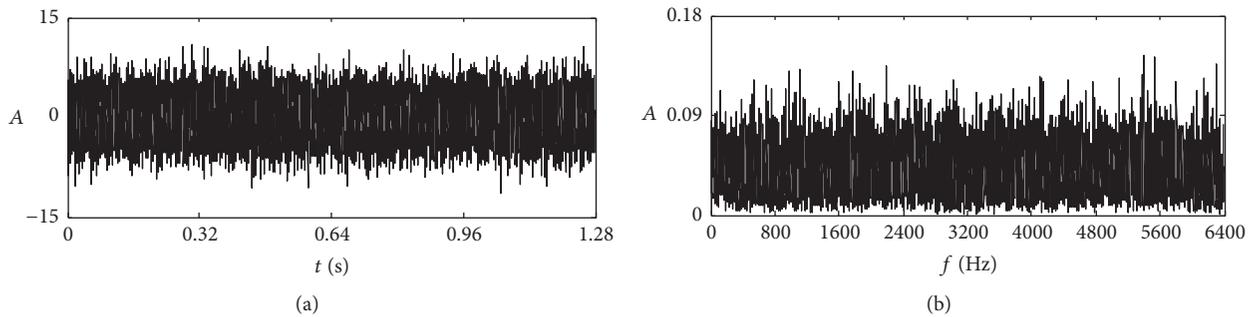


FIGURE 3: Simulation signal: (a) the original waveform; (b) the frequency spectrum.

(5) *Fault Identification.* Combined with the mechanical equipment parameters information, the running status of bearing can be identified and diagnosed with SR detection results.

4. Experiment and Engineering Application

4.1. Numerical Simulation. In this subsection, the proposed method is applied to extract weak periodic signal from strong noise and analyze the important role of filter processing in the proposed algorithm. The simulation signal is defined as $x(t) = 0.09 \sin(100\pi t) + n(t)$, and $n(t)$ is Gaussian white noise with zero mean and standard deviation of 3. The sample frequency is $f_s = 12800$ Hz, and the length of data is $N = 16384$. The time waveform of simulation signal $x(t)$ is shown in Figure 3(a), and Figure 3(b) is the corresponding frequency spectrum. It can be found that the

periodic component is completely submerged by strong noise and cannot be distinguished.

The simulation signal is processed by the proposed method in this study. The searching range of noise intensity α is $[0.1, 10]$, the standard deviation of added white noise is 0.08, and the ensemble number is 50. The genetic algorithms parameters settings are as follows: the number of initial population is 50, the maximum number of generations is 25, the precision of the variables is $1e-8$, and so forth. Based on the maximum of output SNR of SR system, the optimal noise intensity obtained by genetic algorithm is $\alpha = 5.6$, and the corresponding detection results are displayed in Figure 4. Figure 4(a) indicates the time waveform of resonance output of SR system, and the noise is reduced significantly. Furthermore, the spectrum peak at 50 Hz is very prominent in the frequency spectrum, as shown in Figure 4(b), thus realizing the effective extraction of weak periodic component from the strong background noise.

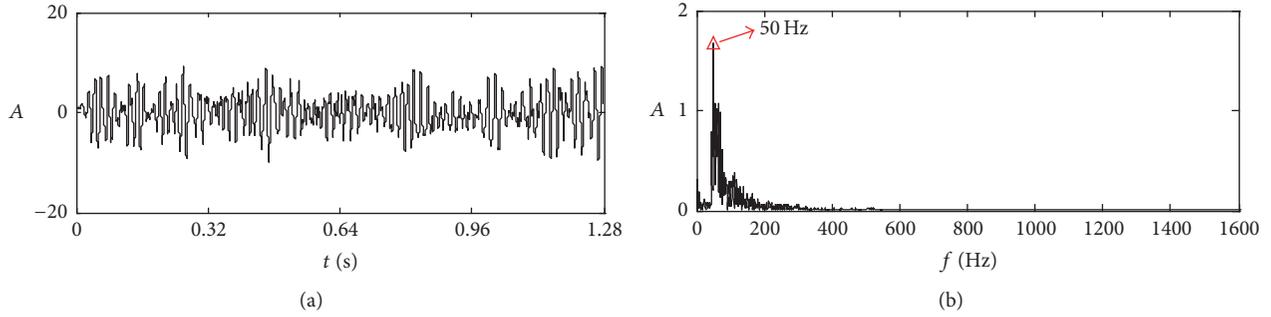


FIGURE 4: Detection result: (a) the output waveform of SR system; (b) the frequency spectrum of resonance output.

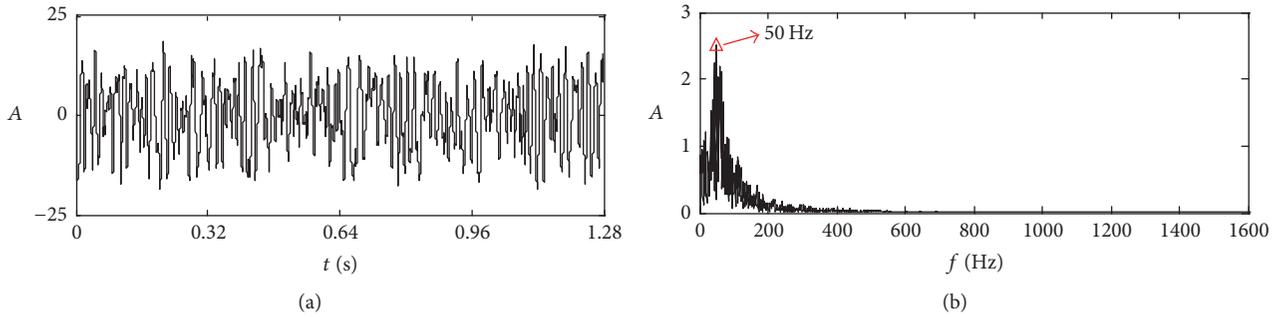


FIGURE 5: Detection result: (a) the output waveform of SR system; (b) the frequency spectrum of resonance output.

In order to illustrate the important role of the filtering processing in the proposed method, a comparison result is provided in Figure 5. In the algorithm flow of the proposed method shown in Figure 2, the part of construction of filters and filtering processing is removed; namely, the obtained L IMFs are used directly to reconstruct the input signal with (7). Figure 5 displays the time waveform of detection result and its corresponding frequency spectrum. Although the periodic component submerged in strong noise can be extracted, there still exists more interference nearby the spectrum peak at 50 Hz, as shown in Figure 5(b), which is less obvious than that of Figure 4. In consequence, the comparison results indicate that through constructing the corresponding filters and filter processing of the L IMFs, the proposed method effectively eliminates the interference of useless components and achieves the enhancement and extraction of feature information in the weak signal detection.

4.2. Experiment Verification. To verify the effectiveness and efficiency of the proposed method in engineering practice, some vibration data collected from bearings are analyzed with the proposed method shown in Figure 2.

(a) Rolling Element Fault Detection of Bearing. In view of the structure features of rolling bearing, when damage occurs in rolling elements, the useful information features that are indicative of structural defects are weak and submerged in strong noise, which makes it difficult to extract. Therefore, fault diagnosis for rolling elements of bearings has remained a major challenge. In this paper, vibration data of bearings that are from the Case Western Reserve University Bearing

Data Center website [29] are analyzed by the proposed method. Fan end bearing was used as the analysis object, and the rolling element fault was generated by electrodischarge machining with fault diameter of 0.178 mm and depth of 0.279 mm. The type of fan end bearing was 6203-2RS JEM SKF, the motor speed was 1777 r/min, and the sample frequency was 12000 Hz. According to the bearing parameters, the fault characteristic frequency of rolling element is 118.09 Hz, and the length of data is 16384.

Figure 6 displays the rolling element fault signal and its corresponding frequency spectrum. It is observed that the time waveform is disorderly, and the frequency components are abundant and useful features relevant to rolling element fault can hardly be identified in the frequency spectrum. In the envelop spectrum, as shown in Figure 6(c), the frequency component 117.9 Hz that is consistent with the fault characteristic frequency of rolling element can be found, but it is weak and not obvious, making it hard to identify the fault efficiently.

The proposed method in the paper is applied to process the vibration signal caused by rolling element fault. The algorithm parameters are the same as in the previous section except for the reaching range of noise intensity α , which is set to [0.1, 20]. The detection results obtained by the proposed method are displayed in Figure 7. The processing results indicate that the high-frequency signal energy is concentrated to target signal through SR system, and some useless low-frequency interferences are eliminated, which makes the spectrum peak at 117.9 Hz more outstanding in the frequency spectrum. Accordingly, the weak fault feature of rolling element of bearing is enhanced and extracted

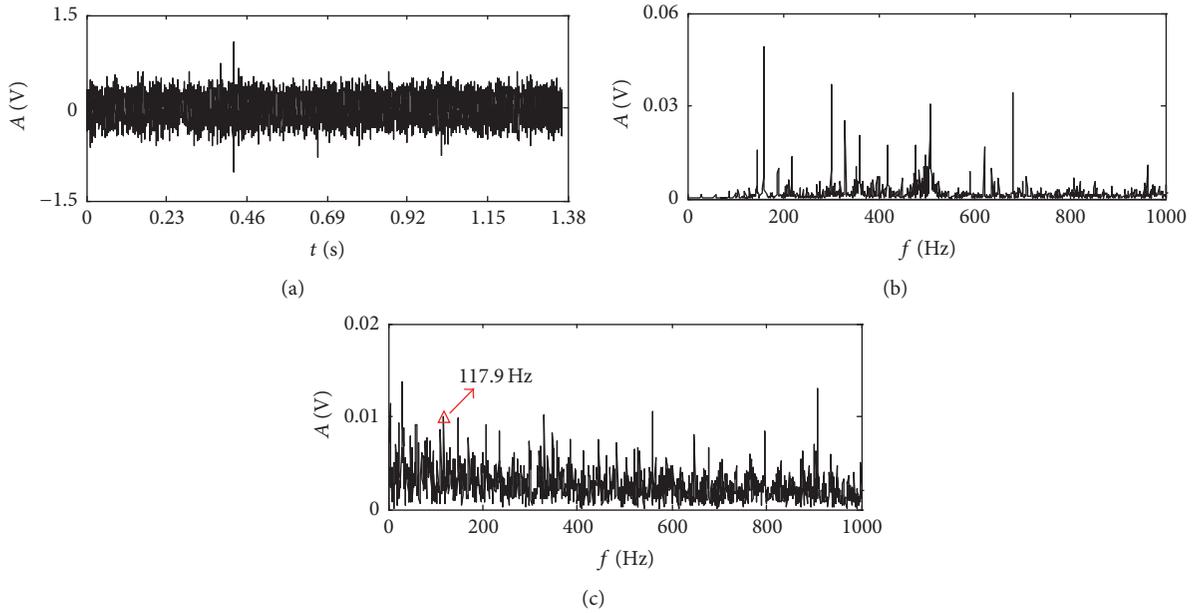


FIGURE 6: Vibration signal of bearing: (a) the time waveform; (b) the frequency spectrum; (c) the envelop spectrum.

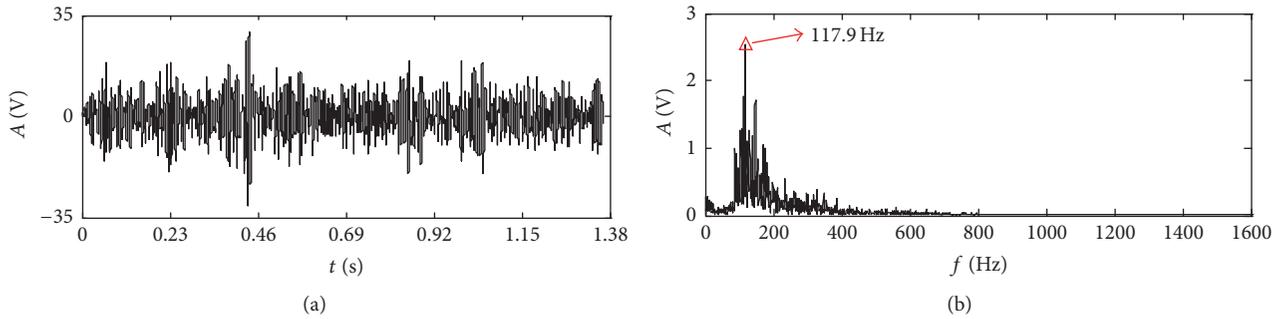


FIGURE 7: Detection results of the proposed method: (a) the output waveform of SR system; (b) the frequency spectrum.

effectively by using the proposed method to improve the accuracy of fault diagnosis.

Besides, the analyzed results of EEMD method with the parameters of the number of ensembles being 50 and the standard deviation of the added white noise being 0.08 are illustrated in Figure 8. Obviously, the EEMD method fails to provide the effective decomposition results, and it is difficult to find the characteristic frequency 117.9 Hz in each IMF and its corresponding envelop spectrum. Therefore, the detection results of the proposed method are superior to those of the envelop spectrum analysis of EEMD.

(b) Inner Race Fault Detection of Rolling Bearing. The vibration data of bearing fault were collected from a fault motor of Spectra Quest Inc.'s mechanical fault simulator with acceleration sensors, as shown in Figure 9. In the experiment, the employed bearing with inner race defect was the deep groove ball bearing with the type of 6203 SKF, and the motor rotating speed was 1433 r/min. The sample frequency was 6400 Hz, and the length of data was 16384. Combining the information of bearing parameters and its corresponding rotating

frequency, it can be derived that the fault characteristic frequency of inner race of bearing is 117.52 Hz.

Figure 10(a) shows the time waveform of bearing signal, and it is difficult to identify the feature information relevant to inner race fault. In Figure 10(b), the frequency components are abundant, and apart from the rotating frequency, the prominent spectrum peak fails to be found at the fault characteristic frequency of bearing inner race. And in the Hilbert envelop spectrum, as Figure 10(c) shows, besides the rotating frequency of 23.83 Hz, the inconspicuous spectrum peak at 117.2 Hz can be found, which is approximately consistent with the fault characteristic frequency of bearing inner race. However, due to the influence of interference frequency components, the feature information of inner race fault is still not easy to be identified to detect faults.

The proposed method is applied to analyze the vibration signal. The algorithm parameters are the same as in the previous section except for the reaching range of noise intensity α , which is set to [0.1, 20]. And the corresponding detection results obtained by the proposed method are displayed in

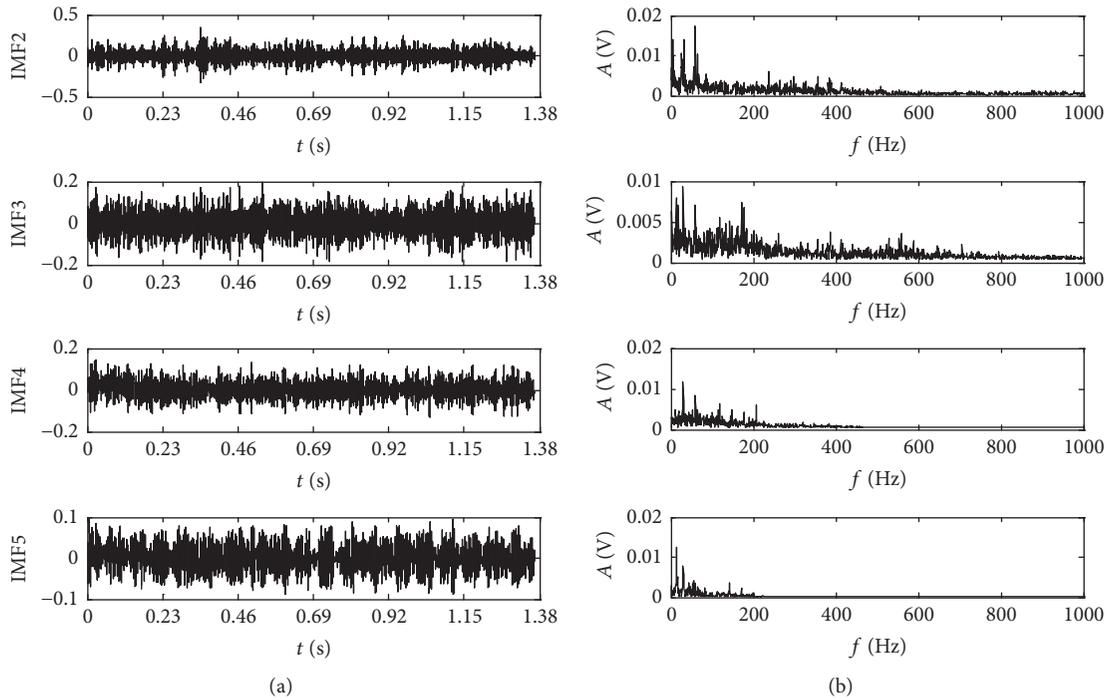


FIGURE 8: Analyzed results of EEMD: (a) IMFs; (b) the envelop spectrum of IMFs.



FIGURE 9: SQI's mechanical fault simulator.

Figure 11. Obviously, a lot of background noise is eliminated, and the spectrum peak at 117.2 Hz is very prominent in the frequency spectrum as shown in Figure 11(b). Therefore, the proposed method in this study can effectively remove useless interferences and enhance the weak feature information submerged in strong noise, thus improving the accuracy of fault identification.

In addition, the method of spectral kurtosis [30] is also adopted to analyze the bearing vibration signal showed in Figure 10(a). The Kurtogram and the resulting signals of spectral kurtosis are illustrated in Figure 12. However, it is difficult to find useful features information relevant to the inner race fault of bearing. Therefore, the two experiments and corresponding comparison results confirm that the proposed method has a better performance for enhancing the weak diagnostic information.

4.3. Engineering Application. With the shortage of fossil fuels and serious environment problems, wind energy sources have drawn intense attention from various countries around the world. However, due to the harsh operation environment, wind turbine, as the important equipment converting wind energy into electrical energy, is often subjected to various loads like wind gust impaction and irregular alternating loads, making some critical components prone to failure. Because of the influence of noisy work environments and strong electromagnetic interference, vibration signals acquired from the generator of wind turbine contain a large amount of noise and interferences to decrease the SNR, which makes it more difficult to detect the fault of bearing at an early stage. Therefore, the proposed method is used for the fault diagnosis of generator bearing in wind turbine and achieves the effective extraction of the feature information from strong background noise.

Through the regular detection of wind turbines in a wind farm, it was found that the vibration response of the front bearing of generator of number 8 wind turbine increased. The installation location of the acceleration transducer is shown in Figure 13. To find out the reasons, the vibration signal needs to be further analyzed. The signals were sampled at 12800 Hz, and the length of data was 16384. The average rotational speed of the generator was 1406 r/min, and the bearing parameters of generator are displayed in Table 1.

Figure 14(a) shows the original vibration signal of the front bearing of the generator. The inconspicuous impulse components can be found in the waveform of original signal. In Figure 14(b), apart from the rotational frequency of 23.44 Hz, there is not useful feature information relevant

TABLE 1: Generator bearing parameters of wind turbine.

Type	Inner diameter D_i/mm	Outer diameter D_o/mm	Roller diameter d/mm	Number of roller	Contact angle $\beta/(\circ)$
6324	120	280	41.275	8	0

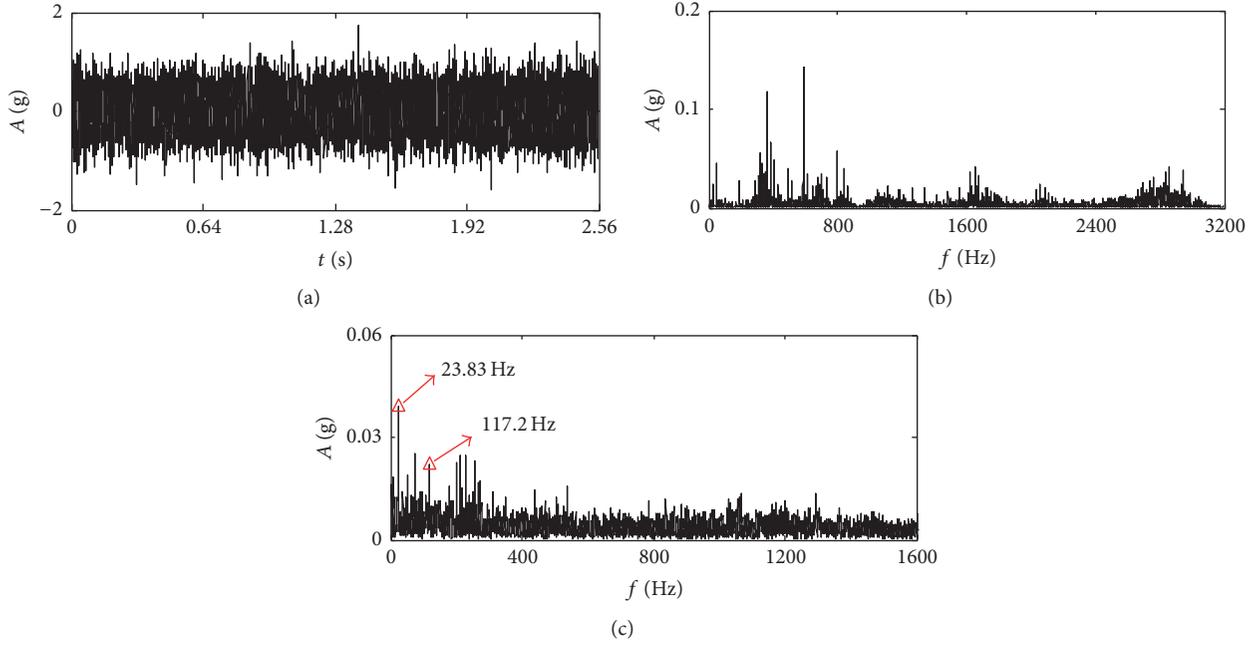


FIGURE 10: Vibration signal of bearing: (a) the time waveform; (b) the frequency spectrum; (c) the envelope spectrum.

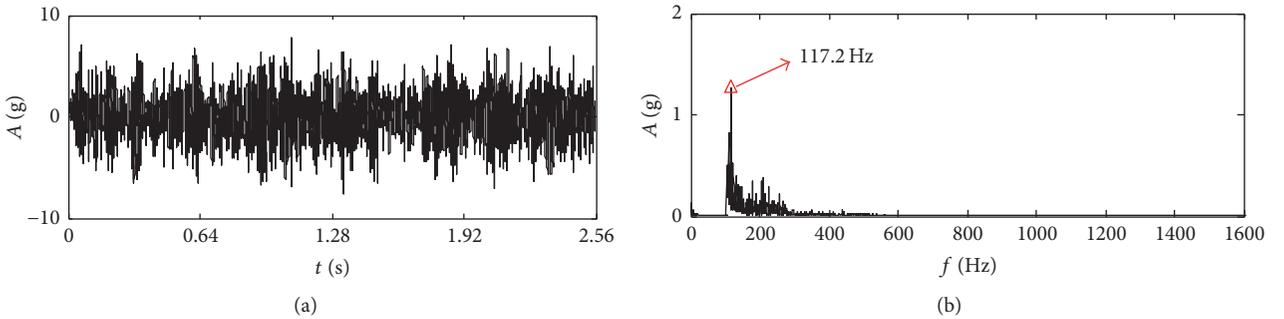


FIGURE 11: Detection results of the proposed method: (a) the output waveform of SR system; (b) the frequency spectrum.

to bearing fault in the frequency spectrum. And in the Hilbert envelop spectrum, as Figure 14(c) shows, besides the rotational frequency, the spectrum peak at 113.3 Hz can be found, which consists with the fault characteristic frequency of bearing inner race. However, because of the influence of strong background noise, features of the useful signal are not evident enough to detect faults.

For the purpose of detecting the performance of the generator, the proposed method is applied to process the bearing vibration signal. The algorithm parameters are the

same as in the previous section except for the reaching range of noise intensity α , which is set to $[0.1, 10]$. And the optimal noise intensity obtained by genetic algorithm is $\alpha = 6.7$, and the corresponding detection results are displayed in Figure 15. Obviously, a lot of background noise is eliminated, and the spectrum peak at 113.3 Hz is very highlighted in the spectrum as shown in Figure 15(b), useless interferences are greatly decreased, thereby improving the accuracy of fault identification. The decibel value corresponding to the fault characteristic frequency of bearing inner race calculated by using Shock

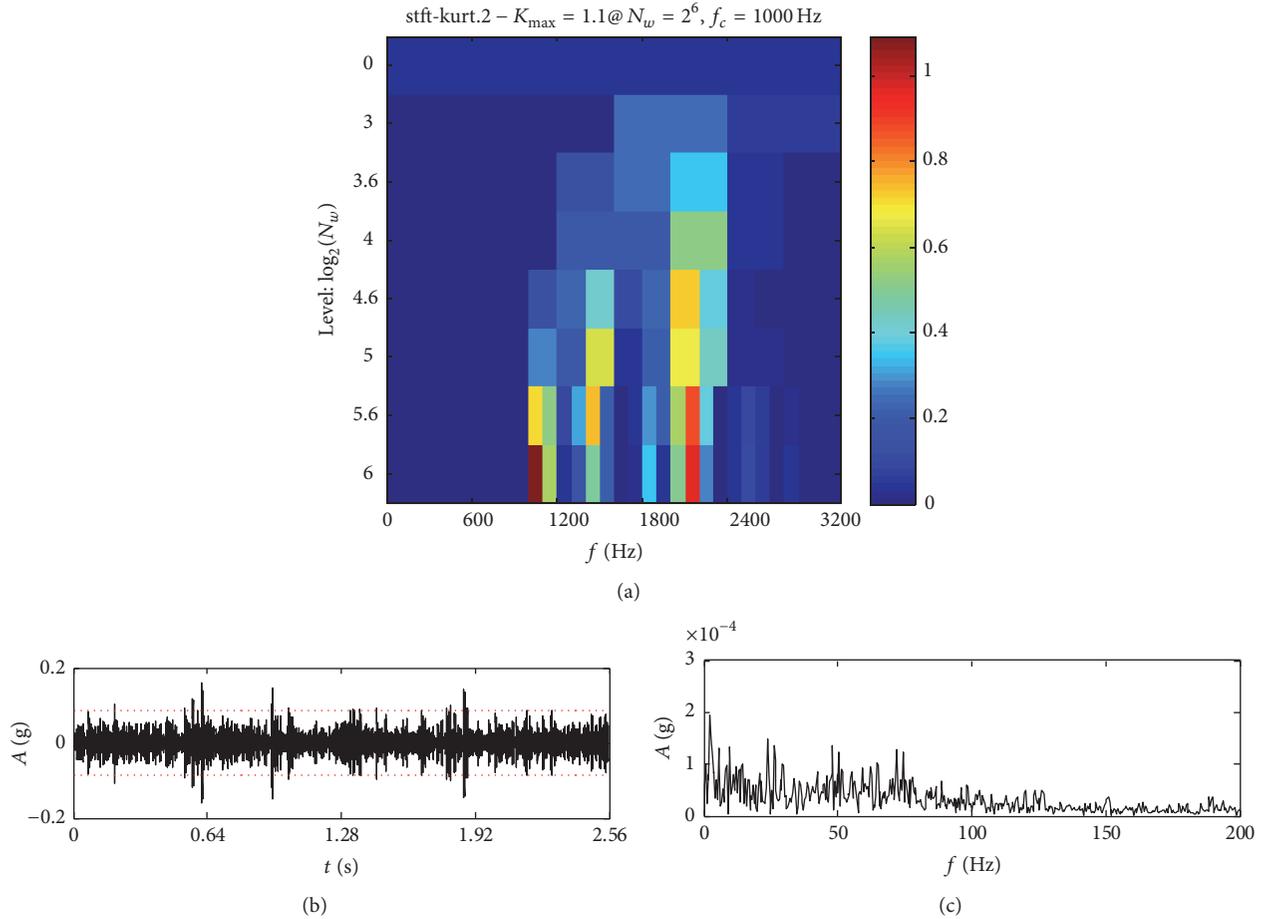


FIGURE 12: Analyzed results of spectral kurtosis: (a) the Kurtogram; (b) the purified signal; and (c) the envelop spectrum.



FIGURE 13: The installation location of the acceleration transducer on the front bearing of generator.

Pulse Method is 21.64 dB, which indicates there is an early damage on the bearing inner race. And the bearing of generator should be paid more attention to and detected regularly.

5. Conclusion

Aiming at the complexity of vibration signals and the shortcomings existing in multiscale noise tuning based on wavelet

decomposition, such as the selection of optimal wavelet bases, a new adaptive multiscale noise control enhanced SR method based on modified EEMD has been investigated in the paper. Compared with wavelet decomposition, EEMD can adaptively decompose the signal into different scales according to the oscillation characteristics of signal itself. And the IMFs at different scales are determined by the signal itself, instead of predetermined kernels, thus strengthening the adaptive capability of the proposed method. And according

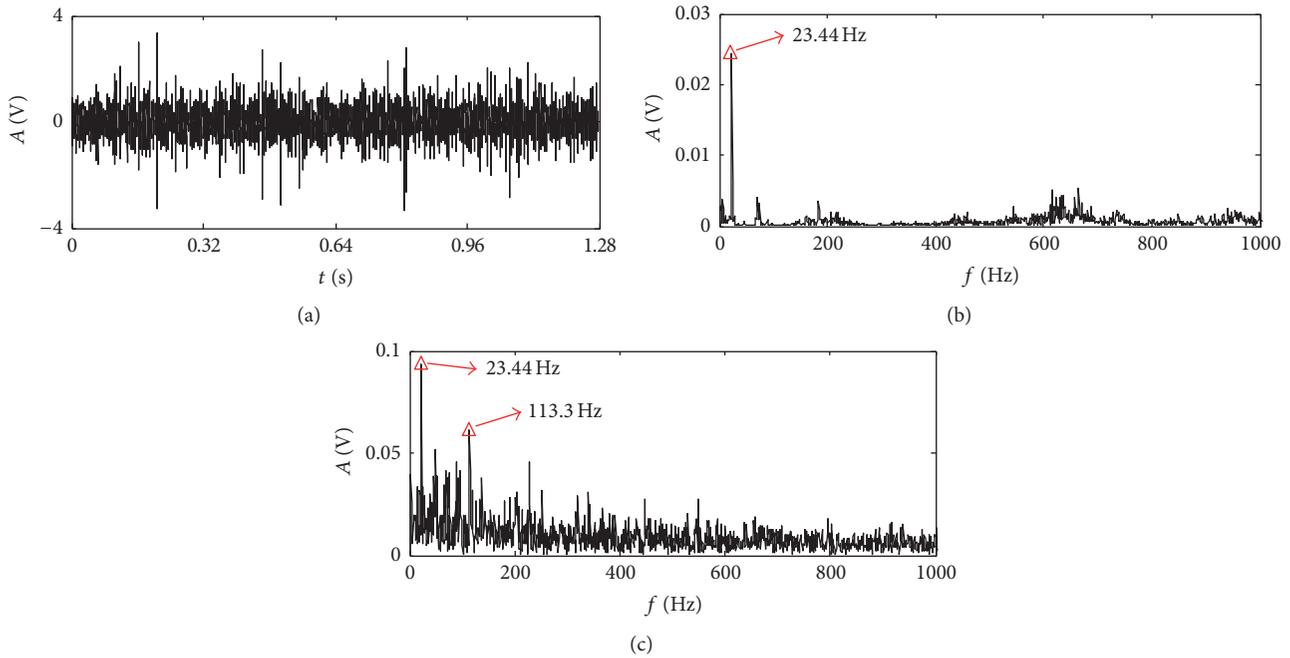


FIGURE 14: Vibration signal of bearing: (a) the time waveform; (b) the frequency spectrum; (c) the envelop spectrum.

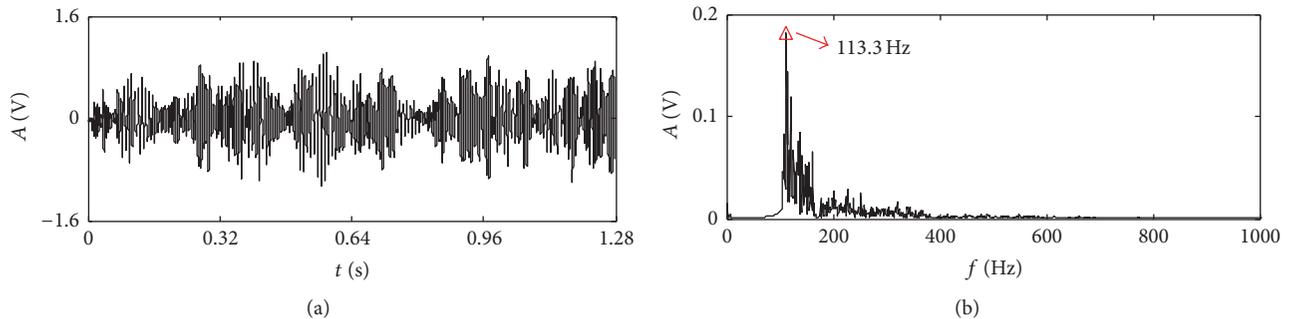


FIGURE 15: Detection results of the proposed method: (a) the output waveform of SR system; (b) the frequency spectrum.

to the requirements of multiscale noise control algorithm, a modified EEMD algorithm is proposed to effectively reduce the decomposition levels and improve the overall calculation efficiency of the proposed method. By using the algorithm of modified EEMD, not only can the SR be realized by adjusting the noise intensity at different scales with genetic algorithm, but also the orthogonality of adjacent IMFs can be improved to enhance the extraction of weak features from strong noise. Therefore, the adaptive multiscale noise control enhanced SR method based on modified EEMD is conducive to the detection of weak periodic signal in mechanical fault diagnosis. The proposed method has been applied to analyze the bearing vibration signals carrying fault information which are taken from some experiments and wind turbine. The results indicate that the proposed method is able to enhance and extract the fault characteristic information and identify the faults effectively.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 51505415), China Postdoctoral Science Foundation Funded Project (Grant no. 2015M571279), and Science and Technology Support Project of Qinhuangdao City (Grant no. 201502A008).

References

- [1] X. W. Zhang, R. X. Gao, R. Q. Yan, X. F. Chen, C. Sun, and Z. B. Yang, "Multivariable wavelet finite element-based vibration

- model for quantitative crack identification by using particle swarm optimization,” *Journal of Sound and Vibration*, vol. 375, pp. 200–216, 2016.
- [2] R. N. Liu, B. Y. Yang, X. L. Zhang, S. B. Wang, and X. F. Chen, “Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis,” *Mechanical Systems and Signal Processing*, vol. 75, no. 15, pp. 345–370, 2016.
- [3] Y. K. Akilu, J. K. Sinha, and K. Elbhah, “An improved data fusion technique for faults diagnosis in rotating machines,” *Measurement*, vol. 58, pp. 27–32, 2015.
- [4] Y. Wang, R. Markert, J. Xiang, and W. Zheng, “Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system,” *Mechanical Systems and Signal Processing*, vol. 60, pp. 243–251, 2015.
- [5] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.
- [6] S. Wang, X. Chen, G. Li, X. Li, and Z. He, “Matching demodulation transform with application to feature extraction of rotor rub-impact fault,” *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 5, pp. 1372–1383, 2014.
- [7] C. Li and M. Liang, “A generalized synchrosqueezing transform for enhancing signal time-frequency representation,” *Signal Processing*, vol. 92, no. 9, pp. 2264–2274, 2012.
- [8] G. Cai, X. Chen, and Z. He, “Sparsity-enabled signal decomposition using tunable Q-factor wavelet transform for fault feature extraction of gearbox,” *Mechanical Systems and Signal Processing*, vol. 41, no. 1-2, pp. 34–53, 2013.
- [9] J. Yuan, C. J. Wei, B. Zou et al., “A comparative study on multiwavelet construction methods and customized multiwavelet library for mechanical fault detection,” *Shock and Vibration*, vol. 2015, Article ID 963528, 12 pages, 2015.
- [10] H. R. Cao, F. Fan, K. Zhou, and Z. J. He, “Wheel-bearing fault diagnosis of trains using empirical wavelet transform,” *Measurement*, vol. 82, pp. 439–449, 2016.
- [11] Y. Lei, D. Han, J. Lin, and Z. He, “Planetary gearbox fault diagnosis using an adaptive stochastic resonance method,” *Mechanical Systems and Signal Processing*, vol. 38, no. 1, pp. 113–124, 2013.
- [12] R. Benzi, A. Sutera, and A. Vulpiani, “The mechanism of stochastic resonance,” *Journal of Physics A: Mathematical and General*, vol. 14, no. 11, pp. L453–L457, 1981.
- [13] C. Nicolis and G. Nicolis, “Stochastic aspects of climatic transitions—additive fluctuations,” *Tellus*, vol. 33, no. 3, pp. 225–234, 1981.
- [14] R. Zhao, R. Q. Yan, and R. X. Gao, “Dual-scale cascaded adaptive stochastic resonance for rotary machine health monitoring,” *Journal of Manufacturing Systems*, vol. 32, no. 4, pp. 529–535, 2013.
- [15] X. F. Zhang, N. Q. Hu, L. Hu, and Z. Cheng, “Multi-scale bistable stochastic resonance array: a novel weak signal detection method and application in machine fault diagnosis,” *Science China Technological Sciences*, vol. 56, no. 9, pp. 2115–2123, 2013.
- [16] Y. Qin, Y. Tao, Y. He, and B. P. Tang, “Adaptive bistable stochastic resonance and its application in mechanical fault feature extraction,” *Journal of Sound and Vibration*, vol. 333, no. 26, pp. 7386–7400, 2014.
- [17] D. Y. Han, P. Li, S. J. An, and P. M. Shi, “Multi-frequency weak signal detection based on wavelet transform and parameter compensation band-pass multi-stable stochastic resonance,” *Mechanical Systems and Signal Processing*, vol. 70-71, pp. 995–1010, 2016.
- [18] J. Wang, Q. B. He, and F. R. Kong, “Adaptive multiscale noise tuning stochastic resonance for health diagnosis of rolling element bearings,” *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 2, pp. 564–577, 2015.
- [19] X.-H. Chen, G. Cheng, X.-L. Shan, X. Hu, Q. Guo, and H.-G. Liu, “Research of weak fault feature information extraction of planetary gear based on ensemble empirical mode decomposition and adaptive stochastic resonance,” *Measurement*, vol. 73, pp. 55–67, 2015.
- [20] Z.-H. Lai and Y.-G. Leng, “Generalized parameter-adjusted stochastic resonance of duffing oscillator and its application to weak-signal detection,” *Sensors*, vol. 15, no. 9, pp. 21327–21349, 2015.
- [21] J. Li, X. Chen, Z. Du, Z. Fang, and Z. He, “A new noise-controlled second-order enhanced stochastic resonance method with its application in wind turbine drivetrain fault diagnosis,” *Renewable Energy*, vol. 60, pp. 7–19, 2013.
- [22] F. B. Duan, F. Chapeau-Blondeau, and D. Abbott, “Double-maximum enhancement of signal-to-noise ratio gain via stochastic resonance and vibrational resonance,” *Physical Review E*, vol. 90, no. 2, Article ID 022134, 2014.
- [23] N. E. Huang, Z. Shen, S. R. Long et al., “The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proceedings of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [24] Z. Wu and N. E. Huang, “Ensemble empirical mode decomposition: a noise-assisted data analysis method,” *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.
- [25] Y. Lei, Z. He, and Y. Zi, “Application of the EEMD method to rotor fault diagnosis of rotating machinery,” *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1327–1338, 2009.
- [26] Q. He, J. Wang, Y. Liu, D. Dai, and F. Kong, “Multiscale noise tuning of stochastic resonance for enhanced fault diagnosis in rotating machines,” *Mechanical Systems and Signal Processing*, vol. 28, pp. 443–457, 2012.
- [27] B. Hu and B. Li, “A new multiscale noise tuning stochastic resonance for enhanced fault diagnosis in wind turbine drivetrains,” *Measurement Science and Technology*, vol. 27, no. 2, Article ID 025017, 2016.
- [28] J. J. Zhang and T. Zhang, “Parameter-induced stochastic resonance based on spectral entropy and its application to weak signal detection,” *Review of Scientific Instruments*, vol. 86, no. 2, Article ID 025005, pp. 1–6, 2015.
- [29] <http://csegroups.case.edu/bearingdatacenter/home>.
- [30] Y. X. Wang, J. W. Xiang, R. Markert, and M. Liang, “Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: a review with applications,” *Mechanical Systems and Signal Processing*, vol. 66-67, pp. 679–698, 2016.

Research Article

Bearing Performance Degradation Assessment Using Lifting Wavelet Packet Symbolic Entropy and SVDD

Jianmin Zhou, Huijuan Guo, Long Zhang, Qingyao Xu, and Hui Li

School of Mechatronic Engineering, East China Jiaotong University, Nanchang 330013, China

Correspondence should be addressed to Jianmin Zhou; hotzjm@163.com

Received 16 June 2016; Accepted 28 September 2016

Academic Editor: Ganging Song

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

Bearing performance degradation assessment is of great significance for proactive maintenance and near-zero downtime. For this purpose, a novel assessment method is proposed based on lifting wavelet packet symbolic entropy (LWPSE) and support vector data description (SVDD). LWPSE is presented for feature extraction by jointing use of lifting wavelet packet transform and symbolic entropy. Firstly, the LWPSEs of bearing signals from normal bearing condition are extracted to train an SVDD model by fitting a tight hypersphere around normal samples. Then, the relative distance from the LWPSEs of testing signals to the hypersphere boundary is calculated as a quantitative index for bearing performance degradation assessment. The feasibility and efficiency of the proposed method were validated by the life-cycle data obtained from NASA's prognostics data repository and the comparison with Hidden Markov Model (HMM). Finally, the assessment results were verified by the envelope spectrum analysis method based on empirical mode decomposition and Hilbert envelope demodulation.

1. Introduction

Bearings play an important role in rotating machinery. The performance of bearings generally affects the operation reliability of whole equipment directly [1]. Actually, bearings always undergo a degenerative process from normal states to failure states. There is possibility to make some proper maintenance strategies to prevent the performance deterioration of bearings, if we can monitor the degenerative process of bearings in time [2]. Therefore, degradation assessment of bearings is significant to help reduce production downtime and save maintenance costs [3].

Vibration signals are always being used to monitor the running state of mechanical system because they carry a lot of information which indicates the health condition of mechanical equipment [4]. Based on vibration signal analysis, lots of fault diagnosis methods combining advanced signal processing and pattern recognition techniques have been proposed for bearings [5]. Although bearing fault diagnosis can be helpful for the indication of condition-based maintenance (CBM), they cannot reveal the degenerative trends of

bearings. Given this problem, bearing performance degradation assessment has been a subject of extensive research in recent years, and more and more attention has been received for its benefits in implementing CBM strategies [6].

Generally speaking, performance degradation assessment of bearings mainly includes two steps, namely, feature extraction and degradation assessment. Until now, a lot of work covering these two aspects has been done. Conventional monitor indexes, such as root mean square (RMS) and kurtosis, have been frequently used as features for degradation assessment of bearings [7, 8]. Pan et al. proposed spectral entropy as a complementary index to evaluate the degradation state of bearings and the results of both simulations and experiments showed that spectral entropy can effectively reflect the degradation process of bearings [9]. Yu proposed a method of locality preserving projection for feature extraction and used a complementary index, namely, negative log-likelihood probability-based exponential weighted moving average statistic (NLLP-EWMA) to assess the performance degradation process of bearings [10]. Hong et al. proposed a method of wavelet packet-empirical mode decomposition for

feature extraction and then used self-organization mapping (SOM) for performance degradation assessment of bearings [3]. Nelwamondo et al. utilized GMM and HMM to diagnose fault in rolling bearings, based on extracted features using Multiscale Fractal Dimension (MFD), Mel Frequency Cepstral Coefficients, and kurtosis. However, the major drawback of HMM classifier is that it is costly and complex [11]. Guo et al. investigated the Hilbert envelope spectrum and SVM method to diagnose REB with ball fault, inner race fault, and outer race fault [12]. The results show that the proposed method provides accurate diagnosis and good diagnostic resolution. Zhang et al. utilized particle swarm optimization-support vector machine (PSO-SVM) to realize the classification of fault location and degradation degree of rolling bearings [13]. Dong and Luo used principal component analysis (PCA) to reduce the dimension of original features and then constructed the LS-SVM model for degradation evaluation of bearings [14]. Sun et al. employed a kernel locality preserving projection-based method to obtain an index for evaluating the degradation degree of a bearing [15]. Liu et al. extracted zero crossing features of vibration signals and then employed a coupled Hidden Markov Model for assessing the performance degradation of bearings quantitatively [16]. Wang and Chen used bilateral spectrum for feature extraction as it could reveal nonlinearities and nonstationary characteristic and then employed a support vector data description (SVDD) model to assess the degradation process of bearings [17].

Though a lot of work has been done, there are still many challenges in assessing the performance degradation of bearings effectively. One of the challenges is how to extract more consistent features for degradation assessment since different features may only be useful at certain service stage or be applicable for specific degradation modes of bearings. For example, RMS can correlate well with the fault development of bearings but it is not sensitive to incipient faults, while Kurtosis Factor is sensitive to impulse faults (especially incipient faults) but it shows the poor stability as the damage grows [7, 8]. Therefore, a composite index which can satisfy both the conditions of sensitivity and stability simultaneously is necessary. Another challenge is how to build an intelligent assessment model based on the extracted features.

To solve the abovementioned problems, a novel assessment method combining lifting wavelet packet symbolic entropy (LWPSE) and support vector data description (SVDD) is proposed in this paper. LWPSE is the combination of lifting wavelet packet transform (LWPT) and symbolic entropy. LWPT has some advantages over classical wavelet packet transform, including performing integer-to-integer wavelet transform and less computation and memory. Besides, LWPT can be reconstructed no matter how the prediction and update operators are designed. Recently, the applications of LWPT in fault diagnosis have achieved some good results [18, 19]. Symbolic time series analysis (STSA) provides a coarse grained description of a dynamical system based on a set of symbols. It is not sensitive to the measure noise, and hence it has a good robustness. Recently, it has been used for motoring degradation and fatigue damage using ultrasonic signals [20, 21]. SVDD is a single value classification method developed from SVM

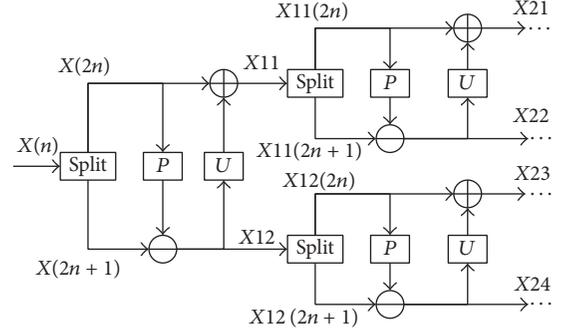


FIGURE 1: Block diagram of the forward transform of LWPT.

and the model can be established using normal data only. Moreover, it possesses the advantages of high computing efficiency and robustness. Therefore, SVDD has been widely used in degradation assessment of bearings [8, 17, 22].

The following paper is organized as follows. Sections 2 and 3 are dedicated to the theories of lifting wavelet packet symbolic entropy and SVDD, respectively. Section 4 presents the degradation assessment method based on LWPSE and SVDD and HMM. Experimental validation and the related analysis are provided in Section 5. The conclusions are provided in Section 6.

2. Lifting Wavelet Packet Symbolic Entropy

2.1. Lifting Wavelet Packet Transform. Lifting wavelet packet transform (LWPT) is realized through the lifting scheme. It no longer relies on Fourier transform, and all computations are implemented in time domain. Moreover, it is possible to obtain the wavelets with some special properties through the design of predictor and updater [23]. The forward transform process of LWPT is shown in Figure 1. We can run the lifting scheme backwards to derive the inverse transform from the forward transform. The process of lifting scheme mainly includes three steps: split, predict, and update. The detail procedure is as follows [24].

(1) *Split.* Split the original signal $X(n) = \{x(n), n = 0, 1, \dots, N\}$ into two subsets, namely, even samples $X(2n) = \{x(2n), n = 0, 1, \dots, N/2\}$ and odd samples $X(2n+1) = \{x(2n+1), n = 0, 1, \dots, N/2\}$, where N is the length of $X(n)$.

(2) *Predict and Update.* In this step, each subband coefficient is calculated at level j :

$$\begin{aligned}
 X_{j,1} &= X_{j-1,1}(2n) + U(X_{j-1,1}(2n+1)) \\
 X_{j,2} &= X_{j-1,1}(2n+1) - P(X_{j,1}(2n)) \\
 &\vdots \\
 X_{j,2^{j-1}} &= X_{j-1,2^{j-1}}(2n) + U(X_{j-1,2^{j-1}}(2n+1)) \\
 X_{j,2^j} &= X_{j-1,2^{j-1}}(2n+1) - P(X_{j-1,2^{j-1}}(2n)),
 \end{aligned} \tag{1}$$

where P is a prediction operator and U is an update operator. Here, P and U are designed by interpolation subdivision principle [25]. In the context of interpolation subdivision principle, the orders of P and U decide their values. Therefore, selecting different orders is equivalent to selecting different bilateral orthogonal wavelet filters with different vanishing moments. In the current study, the orders of P and U are both selected as 12, since this can structure a wavelet function which is closely similar to mechanical impulses [24]. Then, we can obtain the values of P and U .

2.2. Symbolic Entropy. Symbolic time series analysis (STSA) aims at symbolizing the time series data. The basic process of STSA is converting the original time series signals into sequences of discrete symbols via partition function, and then we can use statistical features of the symbols to describe the dynamic statuses of a system [26]. STSA includes the symbolization of original time series and the quantitative analysis of symbolic series.

Consider an original time series $X = \{x(i), i = 1, 2, \dots, N\}$. Converting X into a binary symbol series $S = \{s(k), k = 1, 2, \dots, N\}$ is the most commonly used method in the symbolization of original time series. In this process, partitioning is the first step, and it is also a key step. Here, the partition function is obtained by finite difference method. It is defined as follows:

$$s(k) = \begin{cases} 0 & x(i+1) - x(i) \leq 0 \\ 1 & x(i+1) - x(i) > 0, \end{cases} \quad (2)$$

where $x(i)$ ($i = 1, 2, \dots, N$) is the element of original time series signals X and $s(k)$ is the binary symbol series which is equivalent to 0 or 1. After the symbolization, a binary-coded symbol series S that only contains "0" and "1" can be got. And the next thing to do is dividing S into decimal sequences with length L , namely, words. When the delay time T and the length of word L are given, S can be cut into short symbol sequences:

$$P(k) = (s(k), s(k+T), \dots, s(k+(L-1)T)), \quad (3)$$

$$k = 1, 2, \dots, N - (L-1)T,$$

where N is the length of S . Then, $P(k)$ needs to be transformed to the word $D(k)$.

$$D(k) = \sum_{j=1}^L 2^{L-j} P_j(k) \quad (4)$$

$$P_j(k) = s(k + (j-1)T).$$

Based on the work above, a series of words can be obtained. Here, in order to reveal the intrinsic complexity of the original time series, the histogram of symbol sequence as the basis for quantitative statistics is introduced to measure the occurrence number of each word in all the words [27]. For example, a series of words derived from a binary-coded symbol series S (the length of S is 124) are 0, 1, 2, ..., 7, and the occurrence numbers of each word are 27, 16, 2, 18, 15, 4,

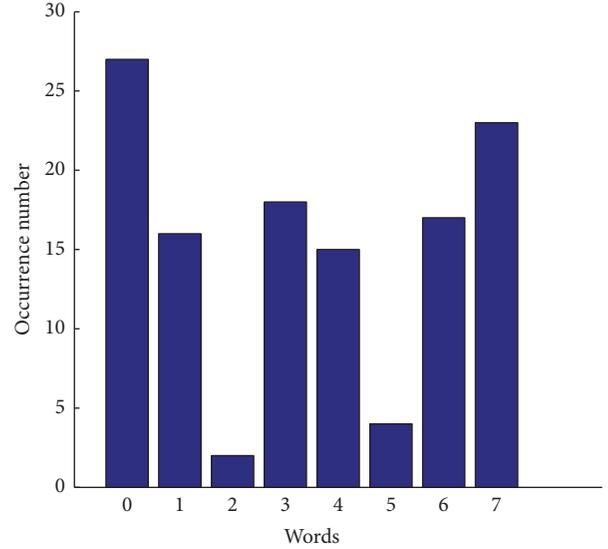


FIGURE 2: Histogram of symbol sequence word.

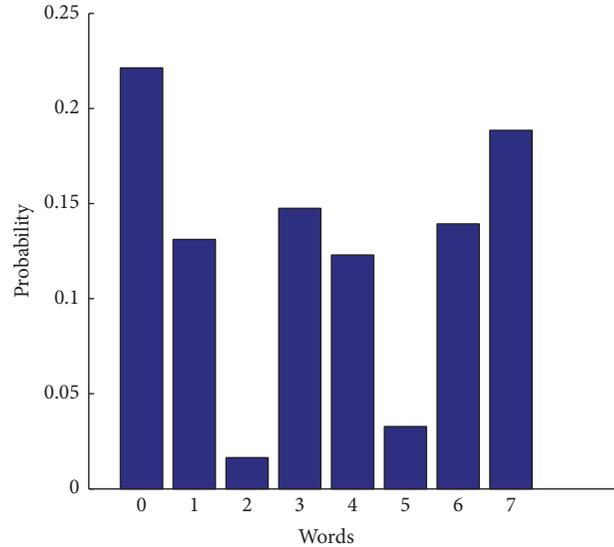


FIGURE 3: Probability of each word.

and 17; then, the histogram of symbol sequence can be shown in Figure 2, and the probability of each word (the normalization of Figure 2) can be shown in Figure 3. From the histogram of symbol sequence, we can use the modified Shannon entropy to describe the complexity of symbol sequence. The modified Shannon entropy is defined as

$$H(T, L) = -\frac{1}{\log M} \sum p_i \log p_i, \quad (5)$$

where p_i is the probability of the i th word and M is the number of all the symbol sequences. In the end, symbolic entropy is equivalent to the modified Shannon entropy of STSA.

Symbolic entropy is a complexity measurement; it can measure the complexity of signals. The vibration signals will

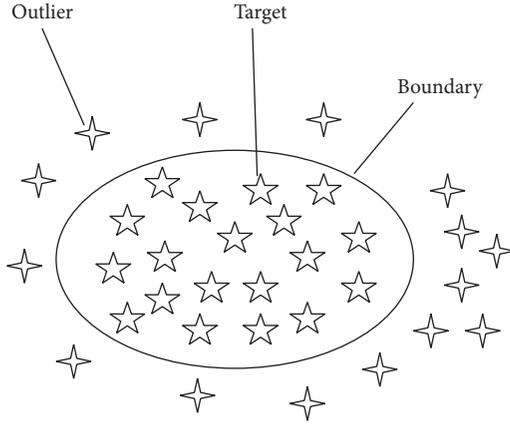


FIGURE 4: Schematic of the two-dimensional SVDD.

be more and more complex with the increase of fault severity, thus causing the increase of the entropy. Therefore, symbolic entropy can reveal the complexity of vibration signals.

2.3. Feature Extraction Using Lifting Wavelet Packet Symbolic Entropy. The basic steps in lifting wavelet packet symbolic entropy (LWPSE) based feature extraction are as follows:

- (1) Employ LWPT to decompose bearing vibration signals. It has been illustrated in Section 2.1 that the best choices of the orders of P and U are both 12. And in order to describe the features of bearing vibration signals more subtly, the decomposition level is defined as 4. Then, the decomposition coefficients can be obtained.
- (2) Choose the decomposition coefficients of each junction in the last level to reconstruct the original signals. Then, we can obtain the reconstructed signals.
- (3) Symbolize the reconstructed signals. Here, the values of T and L are suggested to be 5 and 15 based on a lot of trials since there is no theoretical rule to determine them. Then, the symbolization of each reconstructed signal can be done and the histogram of symbol sequence can be obtained.
- (4) Compute symbolic entropy of each reconstructed signal by formula (5). Then, a 16-dimensional feature vector $V = [H_1, H_2, \dots, H_{16}]$ of bearing vibration signals can be obtained.

3. Support Vector Data Description

Support vector data description (SVDD) proposed by Tax and Duin is inspired by the theory of support vector machine (SVM) proposed by Vapnik [28]. The main idea of SVDD is to find an optimal hypersphere with minimal volume containing all or most targets, as shown in Figure 4.

Consider a training set $\{x_i, i = 1, 2, \dots, n\}$, n is the total number of samples. We try to find the optimal hypersphere which contains all or most normal samples. This hypersphere

is described by center c and radius R and satisfies the following optimization function:

$$\begin{aligned} \min \quad & L(R, c, \xi) = R^2 + C \sum_{i=1}^n \xi_i, \\ \text{s.t.} \quad & (x_i - c)^T (x_i - c) \leq R^2 + \xi_i \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, n, \end{aligned} \quad (6)$$

where C is a penalty parameter which controls the tradeoff between the volume of hypersphere and errors and ξ_i are slack variables which permit a few training data to be outside the hypersphere.

Generally speaking, (6) is solved by introducing Lagrange multipliers and it can be transformed into the following maximizing function L with respect to the Lagrange multipliers α_i :

$$\begin{aligned} \max \quad & L = \sum_{i=1}^n \alpha_i x_i \cdot x_i - \sum_{i,j=1}^n \alpha_i \alpha_j x_i \cdot x_j \\ \text{s.t.} \quad & \sum_{i=1}^n \alpha_i = 1, \quad 0 \leq \alpha_i \leq C \quad \forall i. \end{aligned} \quad (7)$$

Since the data in the input space are not always linearly predicted, we introduce a kernel function $K(x_i, x_j) = (\Phi(x_i) \cdot \Phi(x_j))$ to replace the inner product $(x_i \cdot x_j)$, where K is a Mercer kernel. The kernel function $K(x_i, x_j)$ can map the data into a high-dimensional feature space and transform the nonlinear problem to a linear model. Any function meeting Mercer's theorem can be employed as kernel function, but not all of them are useful for SVDD. Gaussian kernel is the most commonly used function. It is defined as follows:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \quad (8)$$

where σ is the width parameter. Since Gaussian kernel can restrain the growing distances for large feature spaces for describing the target data more compactly, we employ it as $K(x_i, x_j)$. Then, (7) becomes

$$\max L = \sum_{i=1}^n \alpha_i K(x_i, x_i) - \sum_{i,j=1}^n \alpha_i \alpha_j K(x_i, x_j). \quad (9)$$

All α_i are got by solving (9) and only a few of them are nonzero. The samples with $\alpha_i > 0$ are called support vectors. Then, the radius R is obtained by any support vector x_{sv} :

$$\begin{aligned} R^2 = & K(x_{sv} \cdot x_{sv}) + \sum_{i,j=1}^n \alpha_i \alpha_j K(x_i \cdot x_j) \\ & - 2 \sum_{i=1}^n \alpha_i K(x_i \cdot x_{sv}). \end{aligned} \quad (10)$$

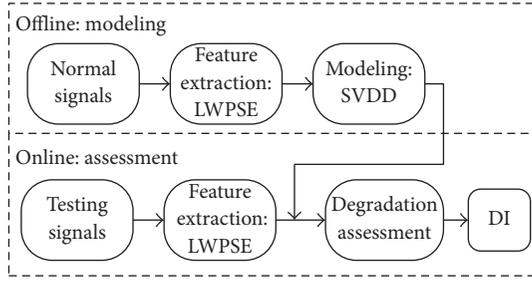


FIGURE 5: Procedure of performance degradation assessment.

For a new sample x_w , its distance to the center c can be described as follows:

$$R_w^2 = K(x_w \cdot x_w) + \sum_{i,j=1}^n \alpha_i \alpha_j K(x_i \cdot x_j) - 2 \sum_{i=1}^n \alpha_i K(x_w \cdot x_i). \quad (11)$$

Here, we introduce the theory of SVDD into bearing performance degradation assessment. Then, the relative distance between the new sample x_w and the hypersphere boundary can be used as the degradation value (DV) of x_w . It is defined as follows:

$$DV = \frac{(R_w - R)}{R}. \quad (12)$$

If $DV \leq 0$, x_w is accepted as a target which indicates that the bearing runs in a normal state. Otherwise, it is an outlier which indicates that the bearing runs in a degradation state.

4. Degradation Assessment

4.1. Degradation Assessment Based on LWPSE and SVDD. In this paper, we proposed a bearing performance degradation assessment method based on LWPSE and SVDD. The framework for performance degradation assessment is shown in Figure 5, which includes two parts, namely, offline modeling and online assessment. The steps of the proposed method are illustrated as follows:

- (1) The historical normal signals of a rolling bearing are collected and their feature vectors are extracted using LWPSE.
- (2) The feature vectors of normal state obtained from step (1) are used as training samples to establish an SVDD model. Then, the radius R can be obtained by (10).
- (3) As for the new testing signal x_w , its feature vectors are extracted using LWPSE. Then, the generalized distance is calculated using (11) which is related to the model established in step (2).
- (4) The degradation value (DV) of x_w is calculated using (12). Then, the degradation index with a series of DVs of testing signals can be obtained. Here, we set all the values of DV to 0 when $DV \leq 0$.

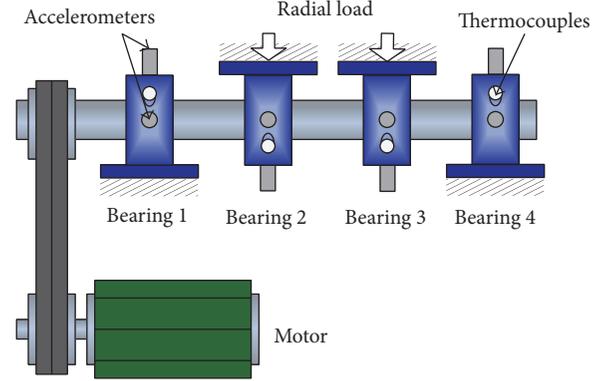


FIGURE 6: Bearing run-to-failure test rig.

From the degradation index, we can know that the bearing runs in a normal state while $DV \leq 0$. Otherwise, it runs in a degradation state. Moreover, DV also reflects the degree of fault severity of a bearing, namely, the larger DV means the larger degree of fault severity.

4.2. Degradation Assessment Based on Hidden Markov Model (HMM). Hidden Markov model can be written as $\lambda = (N, M, \pi, A, B)$ [29]. Where N is the state number of Markov chain. Normal, initial failure, deterioration, deep degradation, and failure are the five states of bearings. M is observation that may occur in each state. In the paper, M is 11 and π is the vector of initial probability distribution. A is the model of the state transition probability matrix. Parameters A and π are generally uniform or selected randomly. B is the model of observation probability matrix.

After model initialization is completed, this paper uses the first 200 sets of data to train model. It is set when two adjacent output of log-likelihood probability value are less than 10^{-3} and model training is completed. Then, enter the 986 group under testing data into the model trained above. When the program stops iterating, the likelihood probability output value of rolling bearing is got. Because the similar probability of their output is relatively small, this paper uses the log-likelihood probability $\log P$ to represent the index of performance degradation. After getting its log-likelihood probability value, the paper uses plot function to draw performance degradation curve of rolling bearing.

5. Experimental Validation

5.1. Description of the Experiment. The experimental data were obtained from NASA's prognostics data repository [30]. The schematic of bearing run-to-failure test rig is shown in Figure 6. During the experiment, four Rexnord ZA-2115 double row bearings were tested on one shaft. The bearings have 16 rollers in each row. The roller diameter is 8.407 mm. The pitch diameter is 71.501 mm. And the tapered contact angle is 15.17° . The shaft was driven by a motor. The rotating speed was kept constant at 2000 rpm, and a 6000 lb radial load was added to the shaft and bearings. All the bearings were lubricated. The test stopped when the debris adhered to

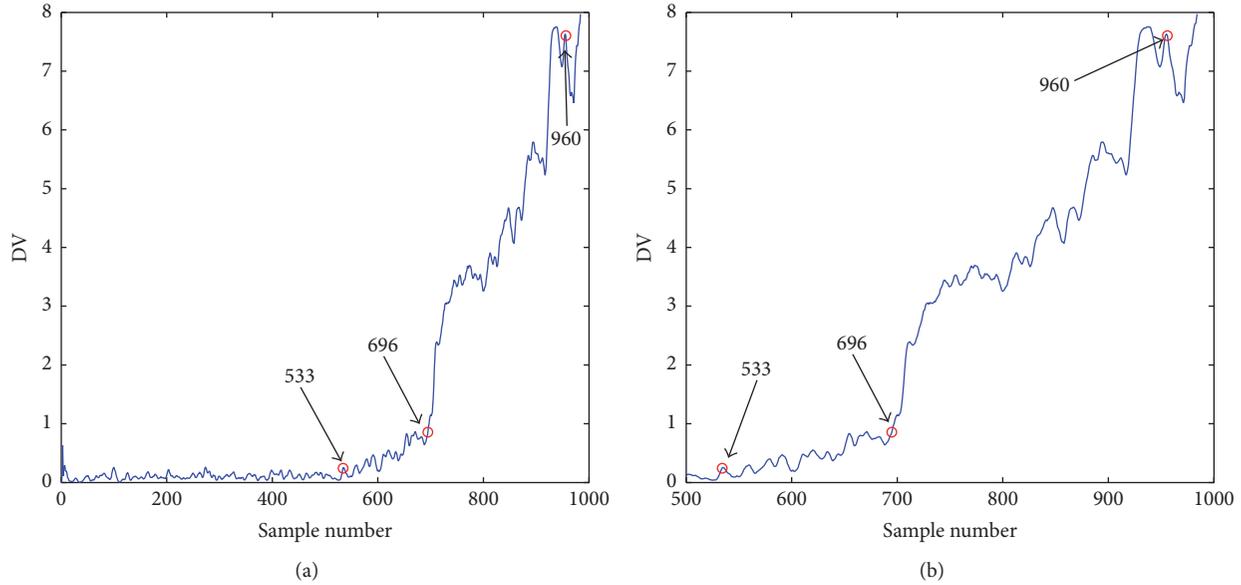


FIGURE 7: Assessment result using LWPSE and SVDD. (a) Life-cycle assessment result of bearing 1. (b) Local enlargement of (a).

the magnetic plug exceeded a certain level. On each bearing, two accelerometers PCB 353B33 were installed to collect the accelerations generated by the vibration signals, with a sampling rate of 20 kHz. The interval time of each collection was 10 minutes and the data length of each collection was 20480 points. The test was carried out for about 163 hours and it ended up with an outer race defect in bearing 1. Therefore, the life-cycle data of bearing 1 was employed to assess the performance degradation of a bearing in this paper. From the geometric parameters of ZA-2115 bearings, we can know that its ball pass frequency of the outer race (BPFO) is about 236.4 Hz.

5.2. Experimental Results Analysis

5.2.1. Assessment Results Using the Proposed Method. The first 200 sets of the life-cycle data of bearing 1 were used as normal signals, and their feature vectors were extracted using the feature extraction method proposed in Section 2.3. Then, we obtained a 200×16 matrix, and the matrix was used to establish an SVDD model. After the model was established, all the life-cycle data were used as testing signals and their feature vectors were extracted. Then, the feature vectors were inputted to the trained SVDD model. Finally, the degradation index DV of bearing 1 was obtained. From the theory of time series analysis, we can know that to certain extent the output of a system at any time is determined by the output of the previous few moments. According to this theory, we can deal with the degradation index by five-point smoothing, namely, $DV(t) = \text{mean}(DV(t-4:t))$, where $\text{mean}(DV)$ means getting mean values. The DV after five-point smoothing is shown in Figure 7. It is seen that the bearing's performance degradation process is clearly revealed by the DV. Before 5330 (the sample number is 533) minutes, the DV is approximately equal to 0 which indicates that the rolling bearing runs in

the normal stage. At 5330 minutes, the DV has an obvious increase and it indicates that the initial fault begins to occur. After 6960 minutes, the increasing trend of the DV is more obvious and it indicates that the rolling bearing runs in the fault progression stage. After 9600 minutes, the DV has no more dramatic increase which indicates that the fault of rolling bearing is developing to failure. Also, sometimes the DV has some abrupt fluctuation since the edge of crack may be smoothed and rounded rapidly after their occurrence [31]. In sum, the performance degradation of bearing 1 could be divided into four periods successfully: normal stage from 0 to 5330 minutes, slight fault stage from 5330 to 6960 minutes, fault progression stage from 6960 to 9600 minutes, and developing-to-failure stage from 9600 minutes to the end.

In the proposed feature extraction method, there are two parameters to be predetermined, namely, the delay time T and the length of word L which is described in the third step of Section 2.3. Figure 8 plots the DV when the parameter T is equal to different values. It can be seen that the increasing trend of the DV is the best when T is equal to 5 as it can describe the degradation process more clearly. Figure 9 plots the DV when the parameter L is equal to different values. It can be observed that the DV is the steadiest when L is equal to 15. Therefore, suitable T and L are meaningful for feature extraction.

5.2.2. Assessment Results Using RMS. RMS is one of the frequently used monitoring indexes. The RMS of bearing 1 is shown in Figure 10. It can be seen that the increasing trend of RMS amplitudes is not obvious before 6990 minutes which indicates the difference of RMS in the normal stage and slight fault stage is not evident, while the proposed degradation index is evident. In addition, it decreases after 7040 minutes which is not consistent with the degradation process.

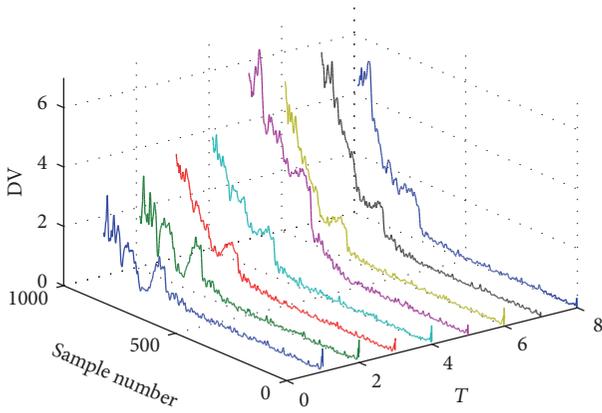


FIGURE 8: Assessment result with different values of T .

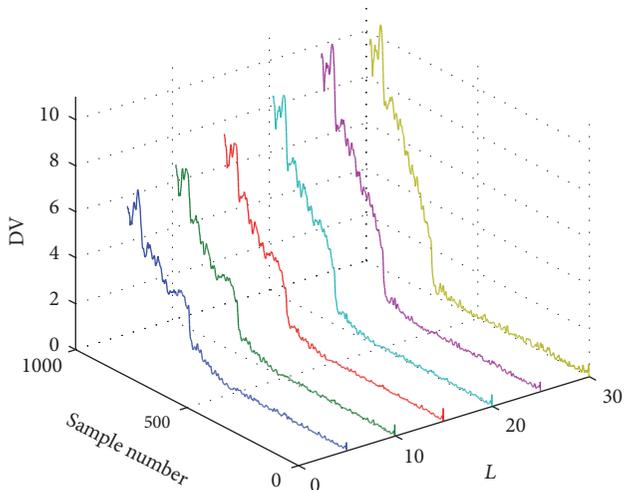
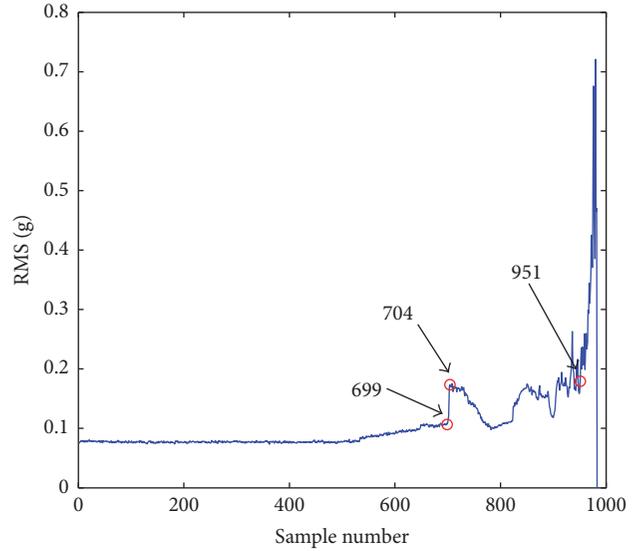


FIGURE 9: Assessment result with different values of L .

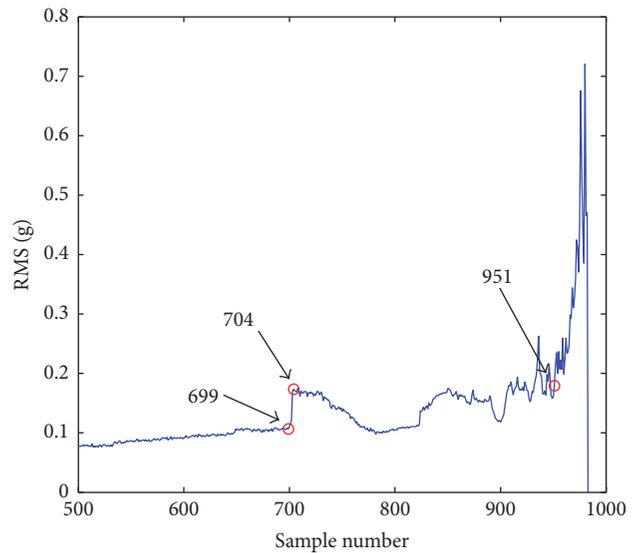
5.2.3. Assessment Results Using Hidden Markov Model. Performance degradation curve of the rolling bearing using HMM is shown in Figure 11.

The figure shows that probably in the 539th sample, the performance of the rolling bearing began to decline and early failure appeared. Probably in the 699th sample, the rolling bearing degraded. After the 820th sample, the rolling bearing exacerbated deeply. After the 963rd sample, the rolling bearing failed completely. We also can see that after about the 737th sample the performance degradation curve is inconsistent with its degree of fault. The overall trend of degradation curve is falling, but the curve still has relatively large fluctuations.

5.2.4. Assessment Results Using Lifting Wavelet Packet Entropy and SVDD. For a comparison between different feature extraction methods, the lifting wavelet packet entropy (LWPE) was used as feature to assess the performance degradation of bearing 1. LWPE is the combination of lifting wavelet packet transform and energy entropy. The basic steps in LWPE are similar to the steps in LWPSE, but



(a)



(b)

FIGURE 10: Root mean square (RMS). (a) Life-cycle RMS of bearing 1. (b) Local enlargement of (a).

LWPE extracts the energy entropy of each reconstructed signal obtained by lifting wavelet packet transform [32]. The assessment result based on LWPE and SVDD is shown in Figure 12. It is observed that the DV has an obvious increase at 5350 minutes and the increasing trend is more obvious after 7000 minutes, which indicates the description of normal stage and slight fault stage is similar to the assessment result based on LWPSE and SVDD. But the curve increases after 7930 minutes, which is not consistent with the degradation process. By comparison among Figures 7, 10, and 12, it is obvious that the degradation index proposed in this paper can reflect the bearing performance degradation process more effectively than the RMS and the degradation index based on LWPE and SVDD.

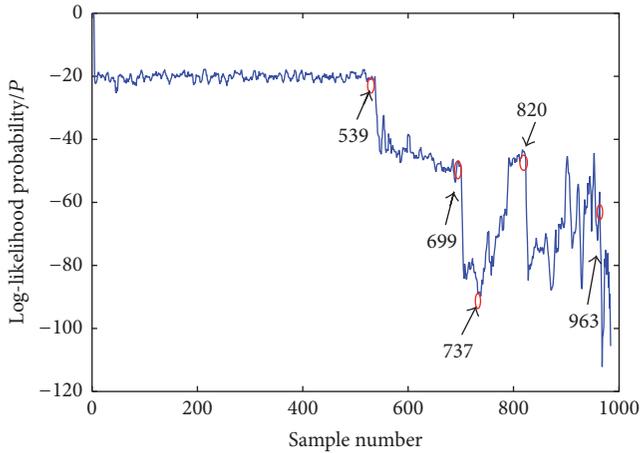
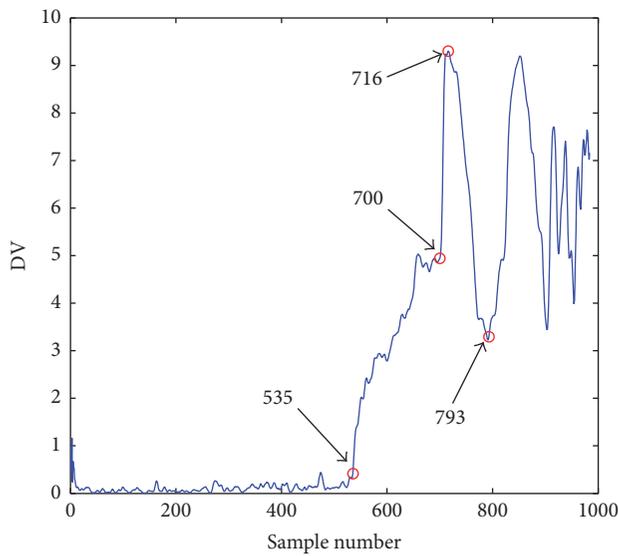
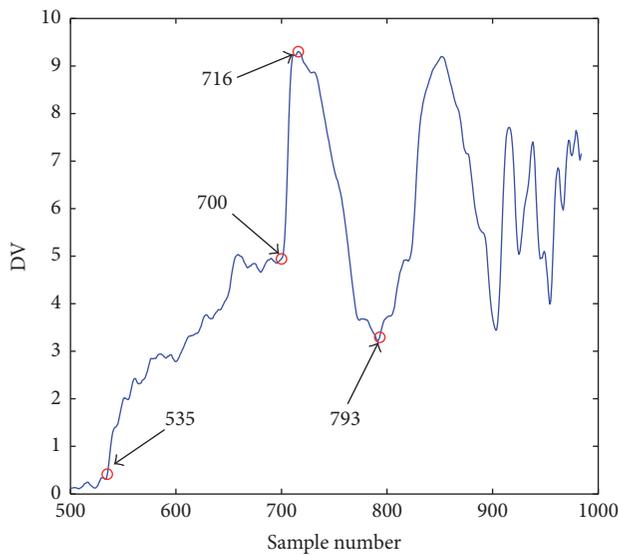


FIGURE 11: Assessment result using HMM.



(a)



(b)

FIGURE 12: Assessment result using LWPE and SVDD. (a) Life-cycle assessment result of bearing 1. (b) Dramatic local enlargement of (a).

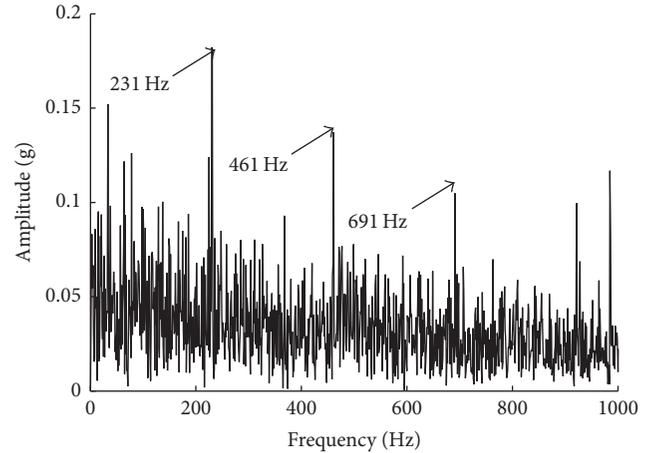


FIGURE 13: Demodulation result of the 533rd sample.

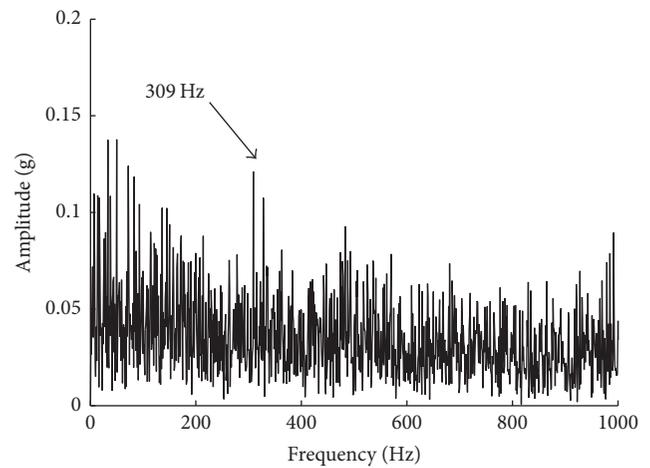


FIGURE 14: Demodulation result of the 532nd sample.

5.3. Correctness Validation of Assessment Results. To validate the correctness of assessment results that the initial fault occurs at 5330 minutes, the envelope spectrum analysis method based on empirical mode decomposition (EMD) and Hilbert envelope demodulation was employed in this paper. Here, we analyze the 533rd sample and the 986th sample. It was decomposed into several simple intrinsic mode functions (IMFs) by EMD first of all [33]. Then, we applied the Hilbert envelope demodulation analysis on IMF1 due to the fact that it is the highest frequency signal which includes the most detailed information of vibration signals [34]. The demodulation result is shown in Figure 13. It can be seen that there is an obvious spectrum peak at the frequency with 231 Hz, which is close to the BPFO with 236.4 Hz. Also, there exists obvious harmonic frequency characteristic. Moreover, the envelope spectrum of 532nd sample depicted in Figure 14 shows no obvious spectrum peak at the frequency close to the BPFO (the samples before the 532nd sample show the same results). The envelope spectrum of 986th sample depicted in Figure 15 shows that there is an obvious spectrum peak at the frequency with 231 Hz, which is close to the BPFO

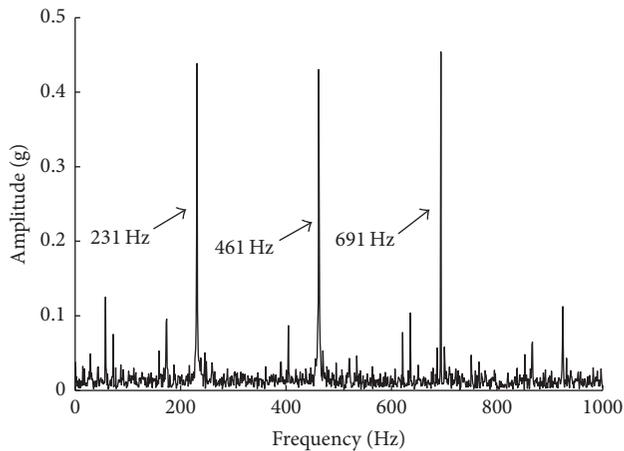


FIGURE 15: Demodulation result of the 986th sample.

with 236.4 Hz. Also, there exists obvious harmonic frequency characteristic. It thus can be inferred that the initial fault of the outer race occurs at 5330 minutes. The analysis results are consistent with the previous assessment results.

6. Conclusions

In this paper, the degradation index based on LWPSE and SVDD is proposed for bearing performance degradation assessment. LWPSE is used to extract feature vectors and SVDD is employed to obtain the assessment results. The efficiency and validity of the proposed method are verified by the life-cycle data obtained from NASA's prognostics data repository. Analysis results show that, compared with the RMS and the degradation index based on LWPE and SVDD, the proposed degradation index is more sensitive to initial fault and it has a consistent increasing trend with the development of bearing fault. The method of HMM can show the overall performance degradation but its consistency of performance degradation is relatively poor. Further analysis shows that the degradation index is affected by the parameters T and L of the symbolic entropy; thus, suitable T and L are meaningful for feature extraction. Moreover, the correctness of assessment results is verified by the envelope spectrum analysis method based on EMD and Hilbert envelope demodulation. It may be concluded that the proposed method is better than HMM in the paper and it is of great significance in guiding the maintenance of rotating machinery.

Competing Interests

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

Acknowledgments

This work was funded under the Natural Science Foundation of China, Grant no. 51205130. The authors are grateful to all study participants.

References

- [1] W. B. Xiao, J. Chen, G. M. Dong, Y. Zhou, and Z. Y. Wang, "A multichannel fusion approach based on coupled hidden Markov models for rolling element bearing fault diagnosis," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 226, no. 1, pp. 202–216, 2012.
- [2] Y. N. Pan, J. Chen, and G. M. Dong, "A hybrid model for bearing performance degradation assessment based on support vector data description and fuzzy c -means," *Proceedings of the Institution of Mechanical Engineers Part C: Journal of Mechanical Engineering Science*, vol. 223, no. 11, pp. 2687–2695, 2009.
- [3] S. Hong, Z. Zhou, E. Zio, and K. Hong, "Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method," *Digital Signal Processing*, vol. 27, no. 1, pp. 159–166, 2014.
- [4] S. Hong, B. Wang, G. Li, and Q. Hong, "Performance degradation assessment for bearing based on ensemble empirical mode decomposition and gaussian mixture model," *Journal of Vibration and Acoustics*, vol. 136, no. 6, article 061006, 2014.
- [5] I. El-Thalji and E. Jantunen, "A summary of fault modelling and predictive health monitoring of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 60, pp. 252–272, 2015.
- [6] B. Zhang, L. Zhang, J. Xu, and P. Wang, "Performance degradation assessment of rolling element bearings based on an index combining SVD and information exergy," *Entropy*, vol. 16, no. 10, pp. 5400–5415, 2014.
- [7] B. Tao, L. M. Zhu, H. Ding, and Y. Xiong, "An alternative time-domain index for condition monitoring of rolling element bearings—a comparison study," *Reliability Engineering & System Safety*, vol. 92, no. 5, pp. 660–670, 2007.
- [8] Z. Shen, Z. He, X. Chen, C. Sun, and Z. Liu, "A monotonic degradation assessment index of rolling bearings using fuzzy support vector data description and running time," *Sensors*, vol. 12, no. 8, pp. 10109–10135, 2012.
- [9] Y. N. Pan, J. Chen, and X. L. Li, "Spectral entropy: a complementary index for rolling element bearing performance degradation assessment," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 223, no. 5, pp. 1223–1231, 2009.
- [10] J. B. Yu, "Local and nonlocal preserving projection for bearing defect classification and performance assessment," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 5, pp. 2363–2376, 2012.
- [11] F. V. Nelwamondo, T. Marwala, and U. Mahola, "Early classifications of bearing faults using hidden Markov models, Gaussian mixture models, mel-frequency cepstral coefficients and fractals," *International Journal of Innovative Computing, Information & Control*, vol. 2, no. 6, pp. 1281–1299, 2006.
- [12] L. Guo, J. Chen, and X. I. Li, "Rolling bearing fault classification based on envelope spectrum and support vector machine," *Journal of Vibration and Control*, vol. 15, no. 9, pp. 1349–1363, 2009.
- [13] Y. Zhang, H. Zuo, and F. Bai, "Classification of fault location and performance degradation of a roller bearing," *Measurement*, vol. 46, no. 3, pp. 1178–1189, 2013.
- [14] S. Dong and T. Luo, "Bearing degradation process prediction based on the PCA and optimized LS-SVM model," *Measurement*, vol. 46, no. 9, pp. 3143–3152, 2013.

- [15] C. Sun, Z. Zhang, Z. He, Z. Shen, B. Chen, and W. Xiao, "Novel method for bearing performance degradation assessment—a kernel locality preserving projection-based approach," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 228, no. 3, pp. 548–560, 2014.
- [16] T. Liu, J. Chen, and G. Dong, "Zero crossing and coupled hidden Markov model for a rolling bearing performance degradation assessment," *Journal of Vibration and Control*, vol. 20, no. 16, pp. 2487–2500, 2014.
- [17] H. Wang and J. Chen, "Performance degradation assessment of rolling bearing based on bispectrum and support vector data description," *Journal of Vibration and Control*, vol. 20, no. 13, pp. 2032–2041, 2014.
- [18] Y. Huang, C. Liu, X. F. Zha, and Y. Li, "An enhanced feature extraction model using lifting-based wavelet packet transform scheme and sampling-importance-resampling analysis," *Mechanical Systems and Signal Processing*, vol. 23, no. 8, pp. 2470–2487, 2009.
- [19] Z. Wang, S. Bian, M. Lei, C. Zhao, Y. Liu, and Z. Zhao, "Feature extraction and classification of load dynamic characteristics based on lifting wavelet packet transform in power system load modeling," *International Journal of Electrical Power and Energy Systems*, vol. 62, pp. 353–363, 2014.
- [20] R. A. Gupta, A. K. Wadhvani, and S. R. Kapoor, "Early estimation of faults in induction motors using symbolic dynamic-based analysis of stator current samples," *IEEE Transactions on Energy Conversion*, vol. 26, no. 1, pp. 102–114, 2011.
- [21] D. S. Singh, S. Gupta, and A. Ray, "In-situ fatigue damage monitoring using symbolic dynamic filtering of ultrasonic signals," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, vol. 223, no. 6, pp. 643–653, 2009.
- [22] X. Zhu, Y. Zhang, and Y. Zhu, "Bearing performance degradation assessment based on the rough support vector data description," *Mechanical Systems and Signal Processing*, vol. 34, no. 1-2, pp. 203–217, 2013.
- [23] Y. N. Pan, J. Chen, and X. L. Li, "Bearing performance degradation assessment based on lifting wavelet packet decomposition and fuzzy C-means," *Mechanical Systems and Signal Processing*, vol. 24, no. 2, pp. 559–566, 2010.
- [24] L. Zhang, G. L. Xiong, H. S. Liu, H. Zou, and W. Guo, "Fault diagnosis based on optimized node entropy using lifting wavelet packet transform and genetic algorithms," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 224, no. 5, pp. 557–573, 2010.
- [25] W. Sweldens, "The lifting scheme: a custom-design construction of biorthogonal wavelets," *Applied and Computational Harmonic Analysis*, vol. 3, no. 2, pp. 186–200, 1996.
- [26] R. Li, A. Mita, and J. Zhou, "Abnormal state detection of building structures based on symbolic time series analysis and negative selection," *Structural Control and Health Monitoring*, vol. 21, no. 1, pp. 80–97, 2014.
- [27] T.-W. Chen and W.-D. Jin, "Feature extraction of radar emitter signals based on symbolic time series analysis," in *Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR '07)*, pp. 1277–1282, Beijing, China, November 2007.
- [28] D. M. J. Tax and R. P. W. Duin, "Support vector domain description," *Pattern Recognition Letters*, vol. 20, no. 11–13, pp. 1191–1199, 1999.
- [29] L. Tao, C. Jin, and D. Guangming, "The rolling bearing fault diagnosis based on KPCA and coupled hidden Markov model," *Vibration and Shock*, vol. 21, pp. 85–89, 2014.
- [30] J. Lee, H. Qiu, and G. Yu, "NASA Ames Prognostics Data Repository-Bearing Data Set," <https://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository>.
- [31] R. Rubini and U. Meneghetti, "Application of the envelope and wavelet transform analyses for the diagnosis of incipient faults in ball bearings," *Mechanical Systems and Signal Processing*, vol. 15, no. 2, pp. 287–302, 2001.
- [32] W.-H. Li, B.-X. Dai, and S.-H. Zhang, "Bearing performance degradation assessment based on Wavelet packet entropy and Gaussian mixture model," *Journal of Vibration and Shock*, vol. 32, no. 21, pp. 35–40, 2013.
- [33] J. B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Applied Acoustics*, vol. 89, pp. 16–27, 2015.
- [34] J. Ma, J. Wu, Y. Fan, and X. Wang, "The rolling bearing fault feature extraction based on the LMD and envelope demodulation," *Mathematical Problems in Engineering*, vol. 2015, Article ID 429185, 13 pages, 2015.

Research Article

Stator Vibration Characteristic Identification of Turbogenerator among Single and Composite Faults Composed of Static Air-Gap Eccentricity and Rotor Interturn Short Circuit

Yu-Ling He,¹ Wei-Qi Deng,¹ Bo Peng,¹ Meng-Qiang Ke,¹ Gui-Ji Tang,¹
Shu-Ting Wan,¹ and Xiang-Yu Liu²

¹Department of Mechanical Engineering, North China Electric Power University, Baoding 071003, China

²State Grid Hebei Electric Power Research Institute, Shijiazhuang 050021, China

Correspondence should be addressed to Yu-Ling He; heyuling1@163.com

Received 8 May 2016; Accepted 14 August 2016

Academic Editor: Lu Chen

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

This paper investigates the radial stator vibration characteristics of turbogenerator under the static air-gap eccentricity (SAGE) fault, the rotor interturn short circuit (RISC) fault, and the composite faults (CFs) composed of SAGE and RISC, respectively. Firstly, the impact of the faulty types on the magnetic flux density (MFD) is analyzed, based on which the detailed expressions of the magnetic pull per unit area (MPPUA) on the stator under different performing conditions are deduced. Then, numerical FEM simulations based on Ansoft and an experimental study are carried out, taking the SDF-9 type fault simulating generator as the study object. It is shown that SAGE will increase the stator vibration at $2f$ (f is the electrical frequency) which already exists even in normal condition, while RISC and CF will bring in stator vibrations at f , $2f$, $3f$, and $4f$ at the same time. The vibration amplitudes under CF are larger than those under RISC. As SAGE increases, the vibration amplitudes of each harmonic component under CF will all be increased, while the development of RISC will decrease the 2nd harmonic vibration but meanwhile increase the 4th harmonic vibration. The achievements of this paper are beneficial for fault identification and condition monitoring of the turbogenerator.

1. Introduction

SAGE and RISC are common faults of turbogenerator. When the eccentricity degree is more than 10%, severe vibrations, stator core deformations, and even winding damage will be caused [1]. When RISC fault develops progressively, serious consequences, such as field current increment, reactive power output decrease, and bearing vibration aggravation, will be produced. It may also develop into an earth fault and magnetize the rotor. More severely, the shaft journal and the bearing bush may also be burned out [2].

By far, scholars have paid much attention to either SAGE or RISC, while few of them studied the CF considering these two faults occurring at the same time. Achievements obtained for SAGE mainly focus on the three-phase current or voltage change [3–6], the circulating current inside parallel branches (CCPB) and its corresponding diagnosis method [7–9],

the unbalanced magnetic pull (UMP) calculation via FEM [10, 11], the winding inductance variation based on improved winding function theory [12], and so forth.

For RISC, the primary monitor performances are taken by means of installing extra search coils inside the generator [13]. Meanwhile, the detective methods based on the shaft voltage [14, 15] and the CCPB characteristics [16] have also been found as effective tools for this fault. However, the search coil method and the CCPB-based method need to install extra components such as the search coils and the current transformers inside the generator, which is actually not welcomed by the practical performers and needs a higher cost. Relatively, methods based on the stator and rotor vibration characteristics [17, 18] and the electromagnetic torque (EMT) [19] are much more convenient due to the easy accomplishment on vibration and EMT data obtaining. The vibrations can be easily obtained by the velocity and acceleration sensors

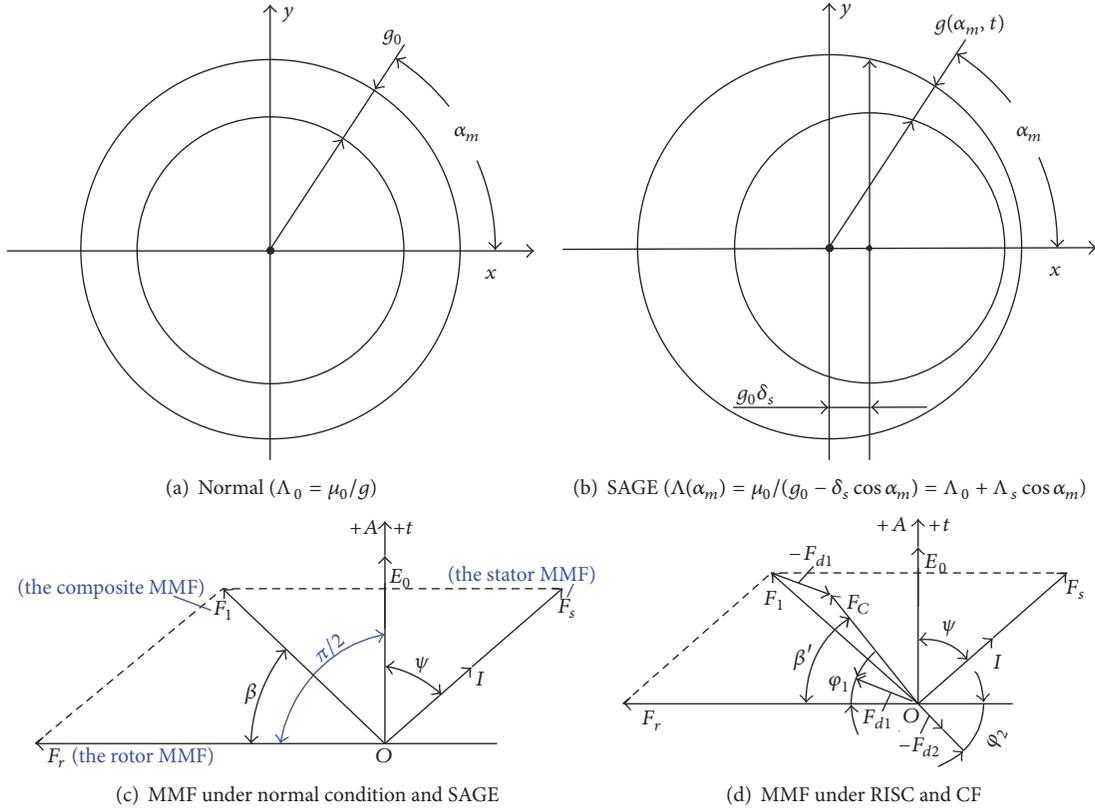


FIGURE 1: Air gap and MMFs of turbogenerator under different conditions.

which have usually been already set up inside the generator or by setting sensors outside the generator conveniently, while the EMT can be calculated by using the phase voltages, currents, and the rotating speed of the rotor which can be easily obtained without installing extra sensors [19].

Currently, achievements obtained by scholars have set up a good basis for condition monitoring and diagnosis for these two faults. However, since SAGE exists in almost every generator, there is actually a CF occurring when RISC takes place. Previous studies on this kind of CF found that the failure characteristics are different from those of the single SAGE and RISC faults [20]. Moreover, it has been also found that the rotor vibration has very similar features under many different faults. For example, the mass unbalance of the rotor, the RISC fault, and the stator interturn short circuit (SISC) will all cause rotor vibrations at f [20–22]. However, the stator vibration properties are found to be much more distinguishable. The stator will vibrate at f , $2f$, $3f$, and $4f$ under RISC [20, 21] and $2f$, $4f$, and $6f$ under SISC [23, 24], respectively, while under the mass unbalance fault the stator vibration will change very little. Therefore, comparably, the stator vibration has a more sensitive effect to identify the failure types, though most scholars have paid their primary attention to the rotor vibrations but not the stator vibration characteristics [18, 20–22, 24].

As an improvement, this paper investigates the stator vibration characteristics under SAGE, RISC, and CF

composed of SAGE and RISC, respectively, in order to obtain a significant identification criterion for these three faults. The whole work is taken based on the theoretical analysis, the numerical FEM simulation, and the experiment study.

2. Theoretical Analysis

2.1. Magnetic Flux Density. The air-gap magnetic flux density (MFD) is composed of the magnetomotive force (MMF) and the permeance per unit area (PPUA) through a multiplying operation. Generally, RISC mainly affects MMF, while SAGE primarily affects PPUA. Thus, MFD will be affected by either of these two faults. Typically, the air gap and MMFs under different conditions are indicated in Figure 1 [20].

PPUAs and MMFs for each case can be written as

$$\Lambda_{\text{Normal}}(\alpha_m, t) = \frac{\mu_0}{g_0} = \Lambda_0$$

(Normal condition, RISC),

$$\Lambda_{\text{SGE}}(\alpha_m, t) = \frac{\mu_0}{g(\alpha_m, t)} = \frac{\mu_0}{g_0(1 - \delta_s \cos \alpha_m)}$$

$$\approx \Lambda_0(1 + \delta_s \cos \alpha_m)$$

$$= \Lambda_0 + \Lambda_s \cos \alpha_m \quad (\text{SAGE, CF}),$$

(1)

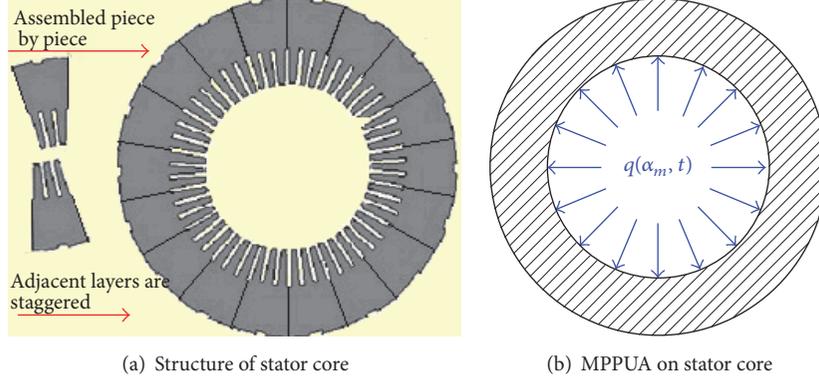


FIGURE 2: Structure and magnetic force of stator core.

$$\begin{aligned}
 f_{\text{Normal}}(\alpha_m, t) &= F_s \cos(\omega t - \alpha_m - \psi - 0.5\pi) \\
 &\quad + F_r \cos(\omega t - \alpha_m) \\
 &= F_1 \cos(\omega t - \alpha_m - \beta) \\
 &\quad \text{(Normal condition, SAGE)},
 \end{aligned}$$

$$\begin{aligned}
 f_{\text{RISC}}(\alpha_m, t) &= F_1 \cos(\omega t - \alpha_m - \beta) \\
 &\quad - F_{d1} \cos(\omega t - \alpha_m - \varphi_1) \\
 &\quad - F_{d2} \cos 2(\omega t - \alpha_m - \pi - \varphi_2) \\
 &= F_C \cos(\omega t - \alpha_m - \beta') \\
 &\quad - F_{d2} \cos(2\omega t - \alpha_m - \pi - \varphi_2) \\
 &\quad \text{(RISC, CF)},
 \end{aligned} \tag{2}$$

where

$$\begin{aligned}
 F_1 &= \sqrt{(F_r - F_s \sin \psi)^2 + (F_s \cos \psi)^2}, \\
 \beta &= \arctan \frac{F_s \cos \psi}{F_r - F_s \sin \psi}, \\
 F_C &= \sqrt{(F_r - F_s \sin \psi - F_{d1} \cos \varphi_1)^2 + (F_s \cos \psi - F_{d1} \sin \varphi_1)^2}, \\
 \beta' &= \arctan \frac{F_s \cos \psi - F_{d1} \sin \varphi_1}{F_r - F_s \sin \psi - F_{d1} \cos \varphi_1}.
 \end{aligned} \tag{3}$$

Correspondingly, MFDs under different conditions are

$$\begin{aligned}
 B(\alpha_m, t) &= F_1 \Lambda_0 \cos(\omega t - \alpha_m - \beta) \\
 &\quad \text{(Normal condition)},
 \end{aligned} \tag{4a}$$

$$\begin{aligned}
 B(\alpha_m, t) &= F_1 (\Lambda_0 + \Lambda_s \cos \alpha_m) \cos(\omega t - \alpha_m - \beta) \\
 &\quad \text{(SAGE)},
 \end{aligned} \tag{4b}$$

$$\begin{aligned}
 B(\alpha_m, t) &= [F_C \cos(\omega t - \alpha_m - \beta') \\
 &\quad - F_{d2} \cos 2(\omega t - \alpha_m - \pi - \varphi_2)] \Lambda_0 \quad \text{(RISC)},
 \end{aligned} \tag{4c}$$

$$\begin{aligned}
 B(\alpha_m, t) &= (\Lambda_0 + \Lambda_s \cos \alpha_m) F_C \cos(\omega t - \alpha_m - \beta') \\
 &\quad - (\Lambda_0 + \Lambda_s \cos \alpha_m) F_{d2} \cos 2(\omega t - \pi - \alpha_m - \varphi_2) \\
 &\quad \text{(CF)},
 \end{aligned} \tag{4d}$$

where F_1 and F_C are the composite MMFs of 1st harmonic under normal condition and CF, respectively ($F_1 > F_C$). F_r and F_s are, respectively, the rotor MMF and the stator MMF at the fundamental frequency. F_{d1} and F_{d2} are the reverse MMFs of 1st and 2nd harmonic produced by RISC. Λ_0 is the constant component of PPUA, while $\Lambda_s = \Lambda_0 \delta_s$ is a mechanical angle to indicate the circumferential position of the air gap.

As indicated in (4a)–(4d), MFDs under normal condition and SAGE are mainly composed of 1st harmonic components, with the amplitudes of $F_1 \Lambda_0$ and $F_1 (\Lambda_0 + \Lambda_s \cos \alpha_m)$, respectively. Actually, the occurrence of SAGE will generally increase the permeance, because the full expansion of the power series (the second formula in (1)) is $\Lambda_0 (1 + \Lambda_s \cos \alpha_m + 0.5 \Lambda_s^2 + 0.5 \cos 2\alpha_m + \dots)$, which is obviously larger than the normal permeance Λ_0 . Consequently, as SAGE is increased, the faulty MFD amplitude will be enlarged.

However, in the cases of RISC and CF, besides the 1st harmonic components, 2nd harmonic components also exist. As RISC develops, the inverse MMFs of 1st harmonic and 2nd harmonic will both be increased, resulting in the decrease of the 1st MFD but meanwhile the increment of the 2nd MFD (see Figure 1(d)).

2.2. Exciting Characteristics of MPPUA

2.2.1. Stator Structure and Essential Exciting Force.

The generator stator is made up of the stator core, the windings, the frame, the end covers, and so forth. Stator vibrations are primarily produced by the magnetic forces acting on the stator core. This kind of magnetic force is actually a unit magnetic pull performing on the whole inner surface of the stator core, usually called magnetic pull per unit area (MPPUA), as indicated in Figure 2. Since the structure of the stator core is similar to a hollow shell, its radial rigidity is small. Even if the resultant force of MPPUA over the whole inner face is zero, the stator core will still have a periodic shrinking-expanding

deformation under the excitation of the pulsating MPPUA. Typically, this periodic shrinking-expanding deformation is actually the radial vibration, which has been well known by people.

MPPUA mainly acts on the stator core surface, and the composite magnetic pulls can be obtained by integrating MPPUA. Because the stator core is a shell structure that is made up of silicon steel sheets, the radial rigidity is small and the composite magnetic pull is zero (shown in Figure 2). However, under the periodic MPPUA, the stator core also can generate the radial vibration, such as the periodic expansion-contraction. Therefore, MPPUA is the nature exciting force of stator vibration.

Therefore, due to the specific structure of the stator core, the essential exciting force for the stator vibration is MPPUA, which can be deduced based on

$$q(\alpha_m, t) = \frac{B^2(\alpha_m, t)}{2\mu_0}. \quad (5)$$

2.2.2. *Exciting Characteristics of MPPUA.* Substituting (4a)–(4d) into (5), the MPPUA expressions, respectively, under normal condition, SAGE, RISC, and CF can be obtained:

$$q(\alpha_m, t) = \frac{F_1^2 \Lambda_0^2}{4\mu_0} [1 + \cos(2\omega t - 2\alpha_m - 2\beta)] \quad (6a)$$

(Normal condition),

$$q(\alpha_m, t) = \frac{F_1^2}{8\mu_0} \left[(2\Lambda_0^2 + \Lambda_s^2) + (4\Lambda_0\Lambda_s \cos \alpha_m) \right. \\ \left. + (\Lambda_s^2 \cos 2\alpha_m) + 0.5\Lambda_s^2 \cos(2\omega t - 2\beta) + 2\Lambda_0\Lambda_s \right. \\ \left. \cdot \cos(2\omega t - \alpha_m - 2\beta) + (2\Lambda_0^2 + \Lambda_s^2) \right. \\ \left. \cdot \cos(2\omega t - 2\alpha_m - 2\beta) + 2\Lambda_0\Lambda_s \right. \\ \left. \cdot \cos(2\omega t - 3\alpha_m - 2\beta) + 0.5\Lambda_s^2 \right. \\ \left. \cdot \cos(2\omega t - 4\alpha_m - 2\beta) \right] \quad (6b)$$

(SAGE),

$$q(\alpha_m, t) = \frac{\Lambda_0^2}{4\mu_0} \left[F_C^2 + F_{d2}^2 - 2F_C F_{d2} \right. \\ \left. \cdot \cos(\omega t - \alpha_m + \beta_1 - 2\varphi_2) + F_C^2 \right. \\ \left. \cdot \cos 2(\omega t - \alpha_m - \beta) - 2F_C F_{d2} \right. \\ \left. \cdot \cos(3\omega t - 3\alpha_m - \beta_1 - 2\varphi_2) + F_{d2}^2 \right. \\ \left. \cdot \cos 4(\omega t - \alpha_m - \varphi_2) \right] \quad (6c)$$

(RISC),

$$q(\alpha_m, t) = \frac{1}{8\mu_0} \left\{ [(F_C^2 + F_{d2}^2)(2\Lambda_0^2 + \Lambda_s^2) \right. \\ \left. + 4(F_C^2 + F_{d2}^2)\Lambda_0\Lambda_s \cos \alpha_m \right. \\ \left. + (F_C^2 + F_{d2}^2)\Lambda_s^2 \cos 2\alpha_m] \right.$$

$$\left. + F_C F_{d2} \left[-4\Lambda_0\Lambda_s \cos(\omega t + \beta_1 - 2\varphi_2) \right. \right. \\ \left. - 2(2\Lambda_0^2 + \Lambda_s^2) \cos(\omega t - \alpha_m + \beta_1 - 2\varphi_2) \right. \\ \left. - \Lambda_s^2 \cos(\omega t + \alpha_m + \beta_1 - 2\varphi_2) \right. \\ \left. - 4\Lambda_0\Lambda_s \cos(\omega t - 2\alpha_m + \beta_1 - 2\varphi_2) \right. \\ \left. - \Lambda_s^2 \cos(\omega t - 3\alpha_m + \beta_1 - 2\varphi_2) \right] \\ \left. + F_C^2 \left[0.5\Lambda_s^2 \cos(2\omega t - 2\beta_1) \right. \right. \\ \left. + 2\Lambda_0\Lambda_s \cos(2\omega t - \alpha_m - 2\beta_1) \right. \\ \left. + (2\Lambda_0^2 + \Lambda_s^2) \cos(2\omega t - 2\alpha_m - 2\beta_1) \right. \\ \left. + 2\Lambda_0\Lambda_s \cos(2\omega t - 3\alpha_m - 2\beta_1) \right. \\ \left. + 0.5\Lambda_s^2 \cos(2\omega t - 4\alpha_m - 2\beta_1) \right] \\ \left. + F_C F_{d2} \left[-\Lambda_s^2 \cos(3\omega t - \alpha_m - \beta_1 - 2\varphi_2) \right. \right. \\ \left. - 4\Lambda_0\Lambda_s \cos(3\omega t - 2\alpha_m - \beta_1 - 2\varphi_2) \right. \\ \left. - 2(2\Lambda_0^2 + \Lambda_s^2) \cos(3\omega t - 3\alpha_m - \beta_1 - 2\varphi_2) \right. \\ \left. - 4\Lambda_0\Lambda_s \cos(3\omega t - 4\alpha_m - \beta_1 - 2\varphi_2) \right. \\ \left. - \Lambda_s^2 \cos(3\omega t - 5\alpha_m - \beta_1 - 2\varphi_2) \right] \\ \left. + F_{d2}^2 \left[0.5\Lambda_s^2 \cos(4\omega t - 2\alpha_m - 4\varphi_2) \right. \right. \\ \left. + 2\Lambda_0\Lambda_s \cos(4\omega t - 3\alpha_m - 4\varphi_2) \right. \\ \left. + (2\Lambda_0^2 + \Lambda_s^2) \cos(4\omega t - 4\alpha_m - 4\varphi_2) \right. \\ \left. + 2\Lambda_0\Lambda_s \cos(4\omega t - 5\alpha_m - 4\varphi_2) \right. \\ \left. + 0.5\Lambda_s^2 \cos(4\omega t - 6\alpha_m - 4\varphi_2) \right] \left. \right\} \quad (6d)$$

(CF).

As indicated in (6a)–(6d), MPPUA has different pulsating components under different performing conditions. Since the stator vibration is actually the response of the exciting MPPUA, the vibration spectrum should have the same harmonic components as MPPUA. Obviously, under normal condition and SAGE, the stator vibration has only 2nd harmonic components (the DC component force will not bring in vibrations to the stator but will induce a radial deformation tendency after a long action). The occurrence of SAGE will increase the vibration amplitude at $2f$. However, when RISC or CF takes place, besides the 2nd harmonic components, 1st, 3rd, and 4th harmonic components will also be produced. To further study the exciting characteristics of MPPUA and the stator vibration response features, the upper amplitude expressions of each MPPUA component under different conditions are shown in Table 1, where I_f is the exciting current of the generator, f_{sd} is the shorted degree of RISC, and δ_s is the relative SAGE value.

TABLE 1: Upper amplitude expressions of MPPUA for different conditions.

Component	Performing condition	Amplitude formulas	Influential factors
DC component	Normal condition	$F_1^2 \Lambda_0^2 / 4\mu_0$	I_f
	SAGE	$F_1^2 (\Lambda_0 + \Lambda_s)^2 / 4\mu_0$	I_f, δ_s
	RISC	$(F_C^2 + F_{d2}^2) \Lambda_0^2 / 4\mu_0$	I_f, f_{sd}
	CF	$(F_C^2 + F_{d2}^2) (\Lambda_0 + \Lambda_s)^2 / 4\mu_0$	I_f, δ_s, f_{sd}
1st harmonic	Normal condition	—	—
	SAGE	—	—
	RISC	$-F_C F_{d2} \Lambda_0^2 / 2\mu_0$	I_f, f_{sd}
	CF	$F_C F_{d2} (\Lambda_0 + \Lambda_s)^2 / 2\mu_0$	I_f, δ_s, f_{sd}
2nd harmonic	Normal condition	$F_1^2 \Lambda_0^2 / 4\mu_0$	I_f
	SAGE	$F_1^2 (\Lambda_0 + \Lambda_s)^2 / 4\mu_0$	I_f, δ_s
	RISC	$F_C^2 \Lambda_0^2 / 4\mu_0$	I_f, f_{sd}
	CF	$F_C^2 (\Lambda_0 + \Lambda_s)^2 / 4\mu_0$	I_f, δ_s, f_{sd}
3rd harmonic	Normal condition	—	—
	SAGE	—	—
	RISC	$-F_C F_{d2} \Lambda_0^2 / 2\mu_0$	I_f, f_{sd}
	CF	$F_C F_{d2} (\Lambda_0 + \Lambda_s)^2 / 2\mu_0$	I_f, δ_s, f_{sd}
4th harmonic	Normal condition	—	—
	SAGE	—	—
	RISC	$F_{d2}^2 \Lambda_0^2 / 4\mu_0$	I_f, f_{sd}
	CF	$F_{d2}^2 (\Lambda_0 + \Lambda_s)^2 / 4\mu_0$	I_f, δ_s, f_{sd}

As shown in Table 1, SAGE has larger MPPUA magnitude at $2f$, while RISC has smaller MPPUA magnitude at $2f$ than normal condition. Specifically, CF has larger MPPUA amplitudes at f , $2f$, $3f$, and $4f$ than RISC. Consequently, the occurrence of SAGE will increase the 2nd harmonic vibration of stator, while RISC will decrease the 2nd harmonic vibration amplitude. The stator vibration amplitudes at f , $2f$, $3f$, and $4f$ under CF condition will be more intense than under RISC condition. This can be used for the fault identification among the single and the composite faults composed of SAGE and RISC.

3. Numerical Simulation and Experiment Study

3.1. Object and Method. The numerical FEM simulation and the experiment study are taken for a SDF-9 type nonsalient pole fault simulating generator (see Figure 3(a)) in the State Key Lab of New Energy and Electric Power System, China, whose primary parameters are indicated in Table 2.

The rotor of the generator is fixed by the bearing block, while the stator can be moved along the horizontally radial direction by adjusting the four screws on the generator to simulate SAGE faults. The movement performance can be controlled by the two dial indicators, as shown in Figure 3(b). Outside the generator, there is a plate with shorted taps of the exciting windings, as indicated in Figure 3(c). RISC can be simulated by connecting different taps. Also, CF conditions with different faulty degrees can be applied via setting different SAGE values and connecting different taps at the same time. During the experiment, the exciting current is set to

TABLE 2: Primary parameters of SDF-9 type generator.

Parameter	Value
Rated capacity	7.5 kVA
Rated voltage	400 V
Power factor	0.8
Rated speed	$n_r = 3000$ rpm
Number of pole pairs	$p = 1$
Radial air-gap length	$g_0 = 0.8$ mm
Stator series conductors per phase	$N_s = 100$
Axial length	$L_0 = 100$ mm
Number of stator slots	$Z_1 = 24$
Ratio of pitch to polar distance	$K_y = y/\tau = 0.83$
Pitch-shortening value	$k_p = 0.966$
Distribution coefficient	$k_d = 0.958$
Number of parallel branches	$\alpha = 2$
Number of exciting turns for each pole	$N_f = 480$

0.8 A, the rotating speed of the rotor is set to 3000 rpm (the rotary period is 20 ms and the frequency is 50 Hz), and the line voltage is 380 V. The experimental testing system is indicated in Figure 3(d).

Objectively, there are some differences to set the velocity sensors in different positions. The general rigidities in the horizontal direction and the vertical direction are different (the vertical direction has a larger rigidity because the foundation is in this direction and therefore sensors here will sample smaller vibration amplitudes). What is more, since

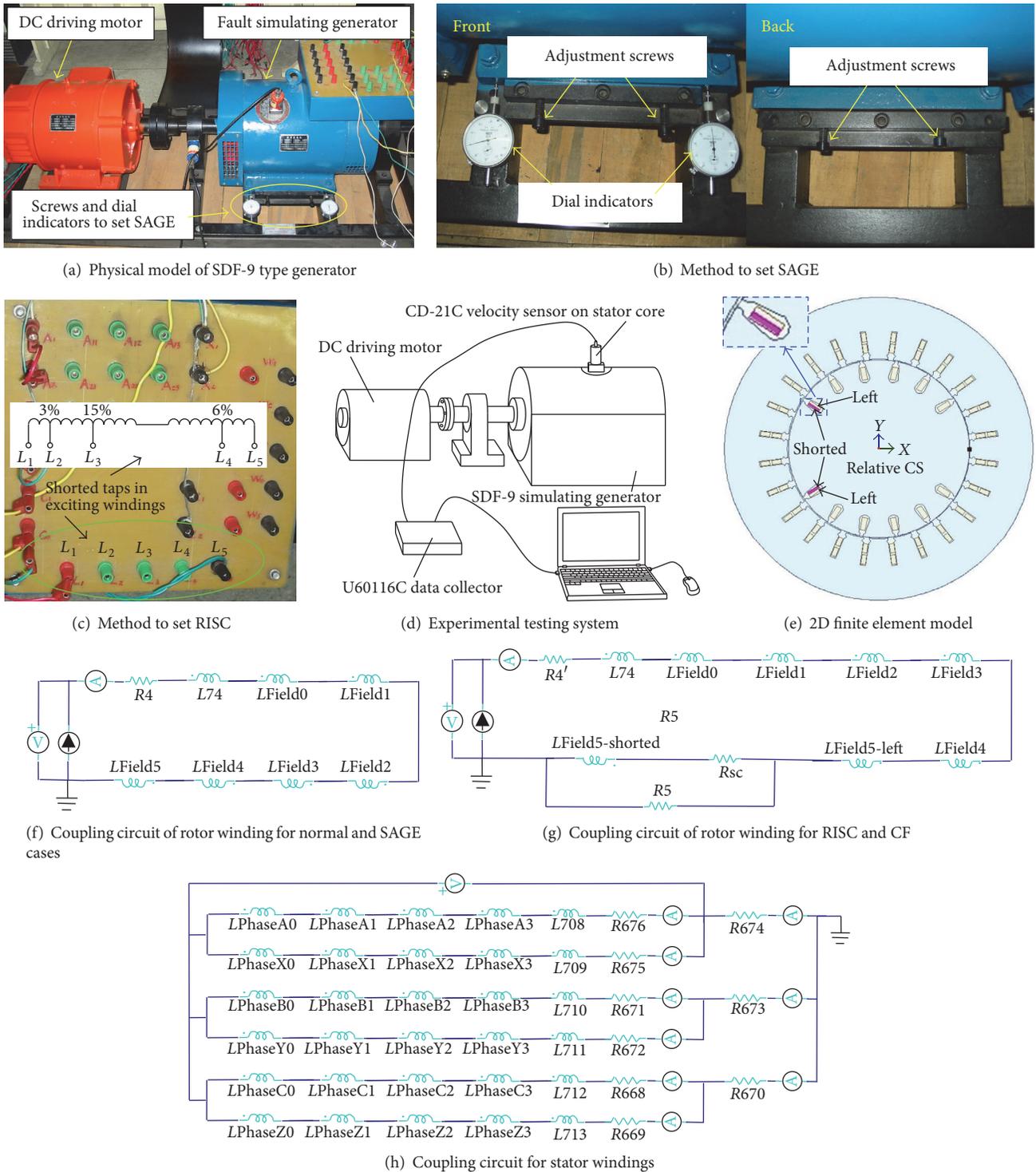


FIGURE 3: Study object of SDF-9 type generator.

the stator is moved along the horizontal direction, the magnetic flux density is larger on one side in this direction but smaller on the other side. So, the MPPUA density is actually larger on the side which has the minimum air-gap length but smaller on the opposite side. Consequently, the sensors set on different positions will have varied vibration amplitudes.

However, no matter where the sensors are set, the developing tendency of the stator vibration will be the same due to the uniform exciting force indicated in (6a)–(6d).

The finite element model is built up, setting the performing parameters the same as the experiment, as indicated in Figure 3(e). The corresponding external coupling circuits for

the exciting windings and the stator windings are illustrated in Figures 3(f), 3(g), and 3(h).

The experiment work and the FEM simulation study are gradually taken by following the four steps.

(1) *Normal Performance.* The generator performs in normal condition without setting SAGE or RISC, and the stator vibration data is collected as a comparing reference. In addition, the MPPUA in normal condition on the stator is also calculated through the FEM simulation.

(2) *SAGE Cases.* 0.1 mm (12.5%), 0.2 mm (25%), and 0.3 mm (37.5%) SAGEs are set each time, respectively, both in the experiment performance and in the FEM simulation. For the simulation, the rotor and the exciting windings are shifted along the x -axis to simulate different SAGE conditions (see Figure 3(e)). Experimental stator vibration signals and calculated MPPUA data are collected for further study.

(3) *RISC Cases.* 3%, 6%, and 12% RISC are, respectively, set in the experiments by connecting L_1L_2 , L_4L_5 , and L_2L_3 . Also, the same RISC degrees are set in the FEM simulation by adjusting the values of $L_{\text{Field5-shorted}}$ and $L_{\text{Field5-left}}$ indicated in Figure 3(g). Correspondingly, the conductors in the shorted slot of the FEM model are divided into two parts. One is the shorted part, and the other is the left part, as indicated in Figure 3(e).

(4) *CF Cases.* Firstly, 0.1 mm SAGE is set up, and 3%, 6%, and 12% RISC are set each time, respectively, both in the experiments and in the simulations. Secondly, 3% RISC is set up, and 0.1 mm, 0.3 mm, and 0.5 mm SAGE conditions are applied each time, respectively. By doing this, the impacts of SAGE and the effects of RISC on MPPUA and the stator vibrations under CF conditions can be obtained.

3.2. Results and Discussion

3.2.1. Pass-Band Results. MPPUA distributions on stator under different performing conditions obtained by FEM calculation are indicated in Figure 4. It is shown that the district near the pole has a much more intense MPPUA, while at other positions MPPUA has a very little value. This is because the magnetic lines near the poles have the strongest intensity. In normal condition, the MPPUA distribution is of symmetry, while the occurrence of SAGE, RISC, or CF will break this symmetry. As SAGE increases, MPPUA will be increased, as indicated in Figures 4(a)–4(d). It is also indicated that the very place which has the largest MPPUA is generally in accordance with the position where the radial air-gap length is the minimum. This is caused due to the max permeance in this position; see (1) and Figure 1(b).

However, when RISC takes place, MPPUA will be decreased, as indicated in Figures 4(a) and 4(e). This is because RISC decreases the effective exciting ampere turns. This can be also explained by comparing Figure 1(c) with Figure 1(d). The primary MMF F_C under RISC (see Figure 1(d)) is actually smaller than F_1 in normal condition (see Figure 1(c)) due to the effect of the induced inverse MMF F_{d1} . The more

serious RISC is, the more MPPUA will be decreased, as indicated in Figures 4(e)–4(g).

Comparing Figure 4(h) with Figures 4(b) and 4(e), it is obvious that MPPUA under CF is smaller than that under SAGE but meanwhile larger than that under RISC. Besides, as SAGE increases, MPPUA under CF will also be increased, as indicated in Figures 4(h), 4(k), and 4(l). However, the development of RISC will decrease MPPUA in general, as illustrated in Figures 1(h)–1(j). This means RISC will generally decrease MFD while SAGE will generally increase MFD, which follows the previous theoretical analysis in Section 2.1.

To show the developing tendency of MPPUA more clearly, the time domain curves for a stable period obtained from FEM calculations for each performing condition are indicated in Figure 5. It is shown that, under the single SAGE fault, the increment of SAGE will raise the value of MPPUA, as shown in Figure 5(a). On the contrary, the occurrence and the development of RISC will generally bring down the value of MPPUA (the mean value and the RMS value are both decreased), as indicated in Figure 5(b).

For the CF cases, the increment of SAGE will increase MPPUA (see Figure 5(c)), while the development of RISC will decrease MPPUA (see Figure 5(d)). This developing trend is in accordance with the single faults. Comparatively, MPPUA under CF is larger than that under the single RISC fault, as indicated in Figure 5(c). However, MPPUA under CF is smaller than that under the single SAGE fault, as illustrated in Figure 5(d). This also means that SAGE will generally increase MFD while RISC will primarily decrease MFD, which again confirms the theoretical analysis in Section 2.1.

Correspondingly, the stator vibration intensity, which is computed via (7), is shown in Figure 6. In (7), y_i is the vibration speed, m is the total number of sampling points, and y is the vibration intensity. Hence,

$$y = \sqrt{\frac{\sum_{i=1}^m y_i^2}{m}}. \quad (7)$$

As indicated in Figure 6, it is clear that whether under the single SAGE fault or CF, the increment of SAGE will increase the stator vibration intensity. However, the development of RISC will decrease the stator vibration intensity. In addition, it is also suggested from Figure 6 that the vibration intensity under CF is smaller than that under the single SAGE fault but meanwhile larger than that under the single RISC fault. This is in accordance with the qualitative result obtained in the theoretical analysis and the MPPUA data presented above.

3.2.2. Special-Frequency Component Results. To further study the impact of different faulty kinds on the stator vibration characteristics, special-frequency components of MPPUA and vibration, including the 1st, 2nd, 3rd, and 4th harmonic components, are investigated as well. The MPPUA amplitudes for each harmonic component under different performing conditions are indicated in Figure 7, while the stator vibration spectra are shown in Figure 8.

It is suggested from Figure 7(a) that, in normal condition, MPPUA mainly has a DC component and a 2nd harmonic component, which well follows the theoretical result

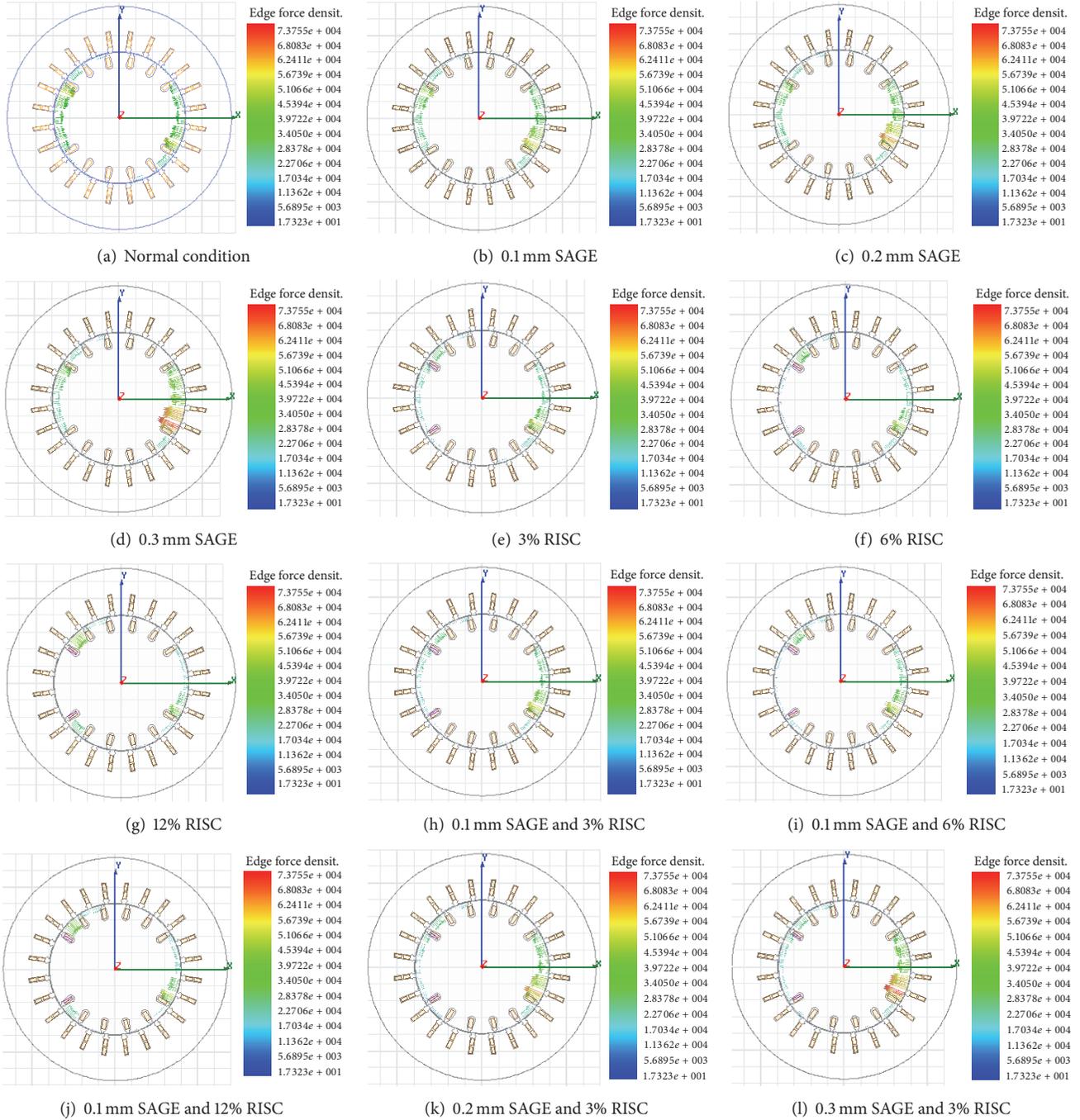


FIGURE 4: MPPUA distribution under different conditions.

indicated in (6a)–(6d). Actually, there is still a tiny 4th harmonic component. The very reason for this extra 4th harmonic component is because MPPUA is the square result of MFD (see (5)) which has not only the fundamental-frequency component but also the other odd harmonics. Consequently, MPPUA should theoretically have each even harmonic component due to the square operation, with the even harmonic components' amplitudes decreased as the harmonic order increases. Equations (6a)–(6d) only consider the 1st harmonic but ignore the higher harmonics due to their tiny

values. For the sake of convenience, in this paper, the higher order harmonics are not taken into account. Comparing Figure 7(a) with Figures 7(b)–(d), it is shown that as SAGE happens and increases, both the DC component and the 2nd harmonic component of MPPUA will be increased. This is consistent with the previous analysis.

As illustrated in Figures 7(a) and 7(e)–(g), the occurrence and development of RISC will generally increase the 1st, 3rd, and 4th harmonic components but meanwhile decrease the 2nd harmonic component. This can be qualitatively

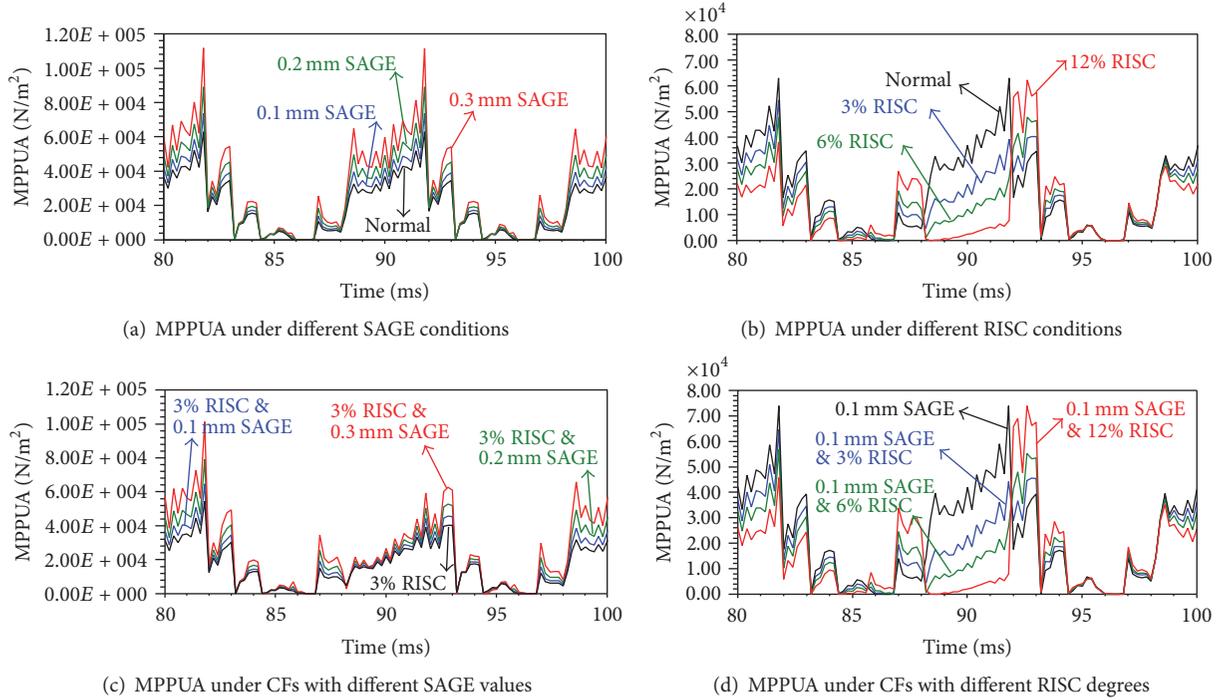


FIGURE 5: Time domain curves of MPPUA under different conditions.

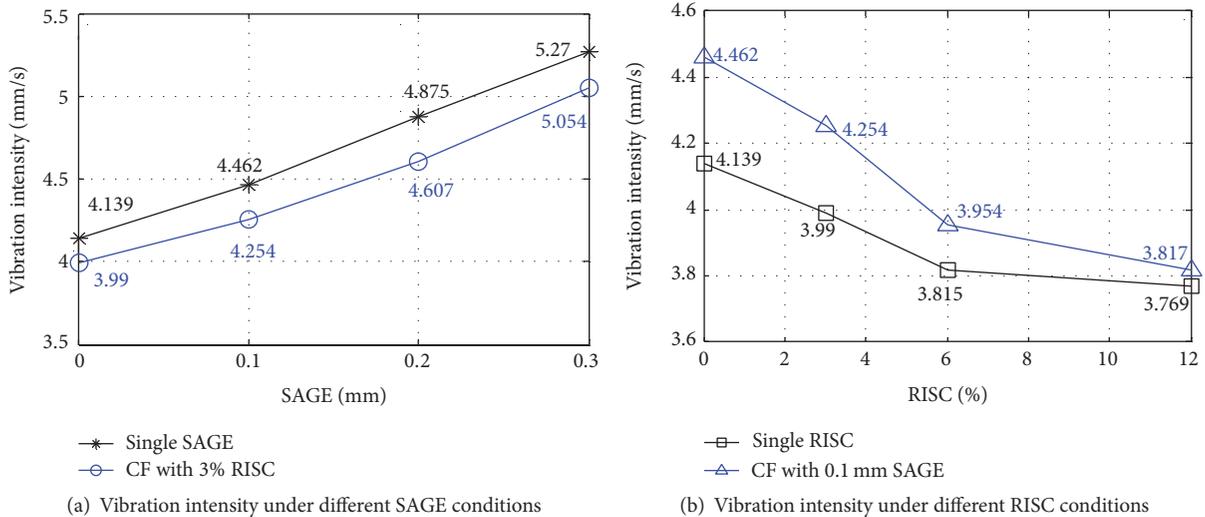


FIGURE 6: Vibration intensity under different performing conditions.

explained by Table 1. As indicated in Table 1, normally, there are no 1st, 3rd, and 4th harmonic components, while under RISC the 1st, 2nd, 3rd, and 4th harmonic components appear. Therefore, the new components brought in by RISC will be increased. Moreover, the MPPUA amplitude in normal condition is $F_1^2 \Lambda_0^2 / 4\mu_0$, while under RISC the amplitude is $F_C^2 \Lambda_0^2 / 4\mu_0$. As illustrated in Figures 1(c) and 1(d), it is obvious that F_C is smaller than F_1 , resulting in the decrease of the 2nd harmonic component of MPPUA.

Comparing Figure 7(h) with Figure 7(b), Figure 7(k) with Figure 7(c), and Figure 7(l) with Figure 7(d), it can be found

that MPPUA under CF has a smaller value at $2f$ (f is the electrical frequency) but meanwhile has a larger value at f , $3f$, and $4f$ than under the single SAGE fault. This again confirms the conclusion presented in the previous paragraph that RISC will increase the 1st, 3rd, and 4th harmonic components but meanwhile decrease the 2nd harmonic component. In addition, comparing Figure 7(h) with Figure 7(e), Figure 7(i) with Figure 7(f), and Figure 7(j) with Figure 7(g), it is shown that the 1st to the 4th harmonic components of MPPUA under CF are all larger than those under RISC. This means SAGE will generally increase MFD, which has been found out

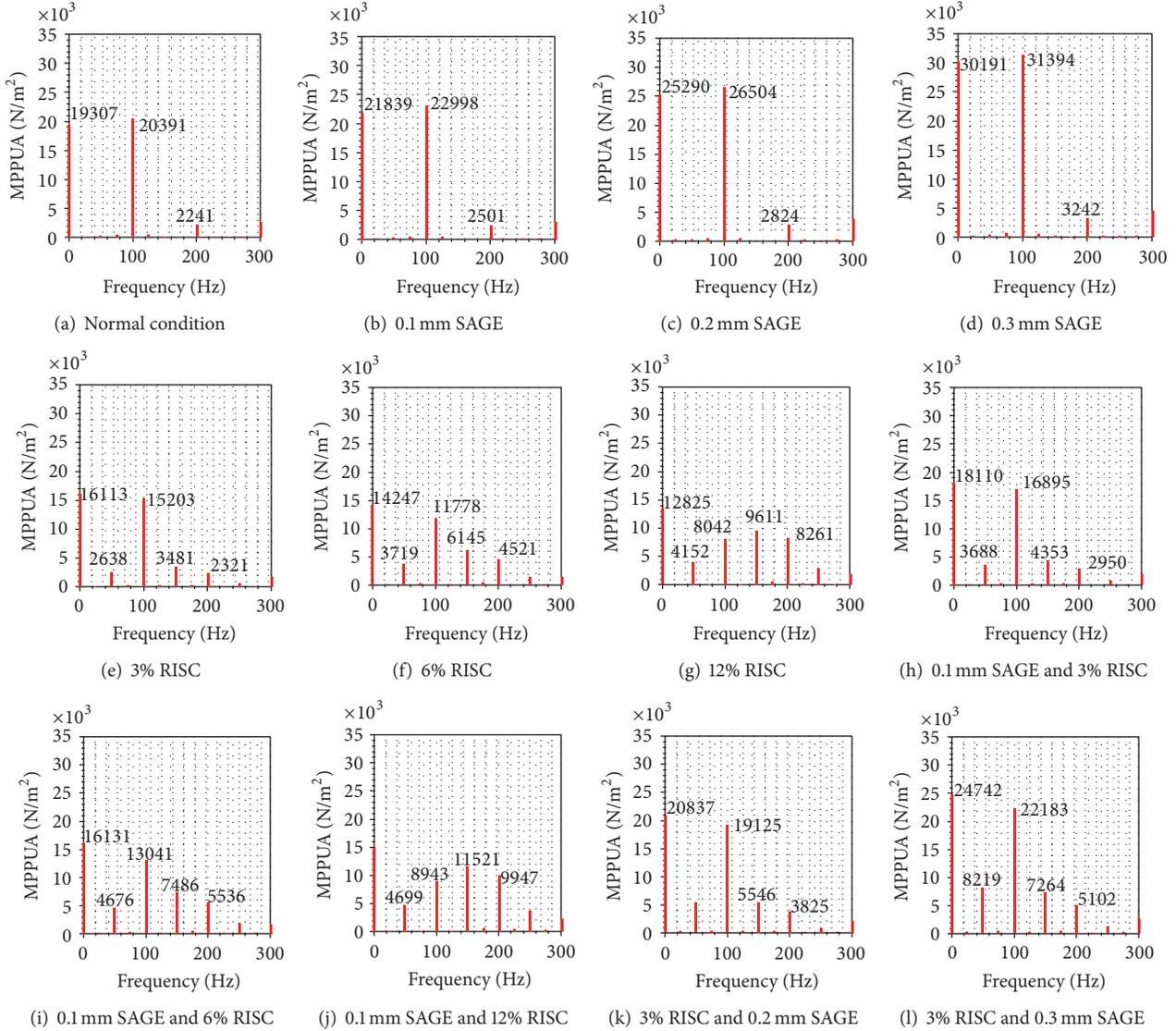


FIGURE 7: Spectra of MPPUA under different conditions.

previously. Actually, the amplitude formulas indicated in Table 1 can also explain this well, for the permeance under CF is $(\Lambda_0 + \Lambda_s)$, while the permeance under RISC is only Λ_0 (see Table 1).

Taking MPPUA as the exciting force, the stator vibration can be treated as the response. As indicated in Figure 8, the stator vibration tendency at each harmonic frequency well follows the developing trend of MPPUA. Specifically, Figure 8(a) shows that the stator vibration in normal condition not only has each even harmonic component (primarily the 2nd and the 4th harmonics) but also includes each odd harmonic component (primarily the 1st, 3rd, and 5th harmonics). In theory, as analyzed previously, there should be only even harmonic components existing (primarily the 2nd harmonic). The extra odd harmonic components of the stator vibration are mainly transferred from the rotor vibration (the rotor and the stator both located on the basis; see Figure 3(a))

and caused by the initial asymmetry inside the generator. To study the essential effect of each kind of fault on the stator vibration, the vibration amplitudes of each harmonic in normal condition are treated as the null drifts of the experiment system, and the faulty vibration amplitudes are subtracted by these null drifts. For example, comparing with the normal condition, vibration amplitude at $2f$ for 0.1 mm, 0.2 mm, and 0.3 mm SAGE increases by 32.9% (0.886 mm/s), 50.5% (1.359 mm/s), and 70.9% (1.909 mm/s), respectively. This indicates that these increments of the vibration amplitude are the essential impact caused by the three SAGE fault conditions.

To better show the relation between the input excitation force and the output vibration response, the line charts of MPPUA and the stator vibration at each frequency under different running conditions are illustrated in Figure 9. It is shown that the developing tendency of the tested stator

TABLE 3: Fault identification criterion based on stator vibration characteristics.

Stator vibration characteristics	Performing condition
Having only soft 2nd harmonic vibrations	Normal condition
The vibration at $2f$ is increased, while the other components generally keep stable	SAGE exists
The vibration at $2f$ is decreased, but meanwhile vibrations at f , $3f$, and $4f$ are increased	RISC exists
The vibration at $2f$ is firstly decreased and then increased; meanwhile, vibrations at f , $3f$, and $4f$ are increased	CF exists

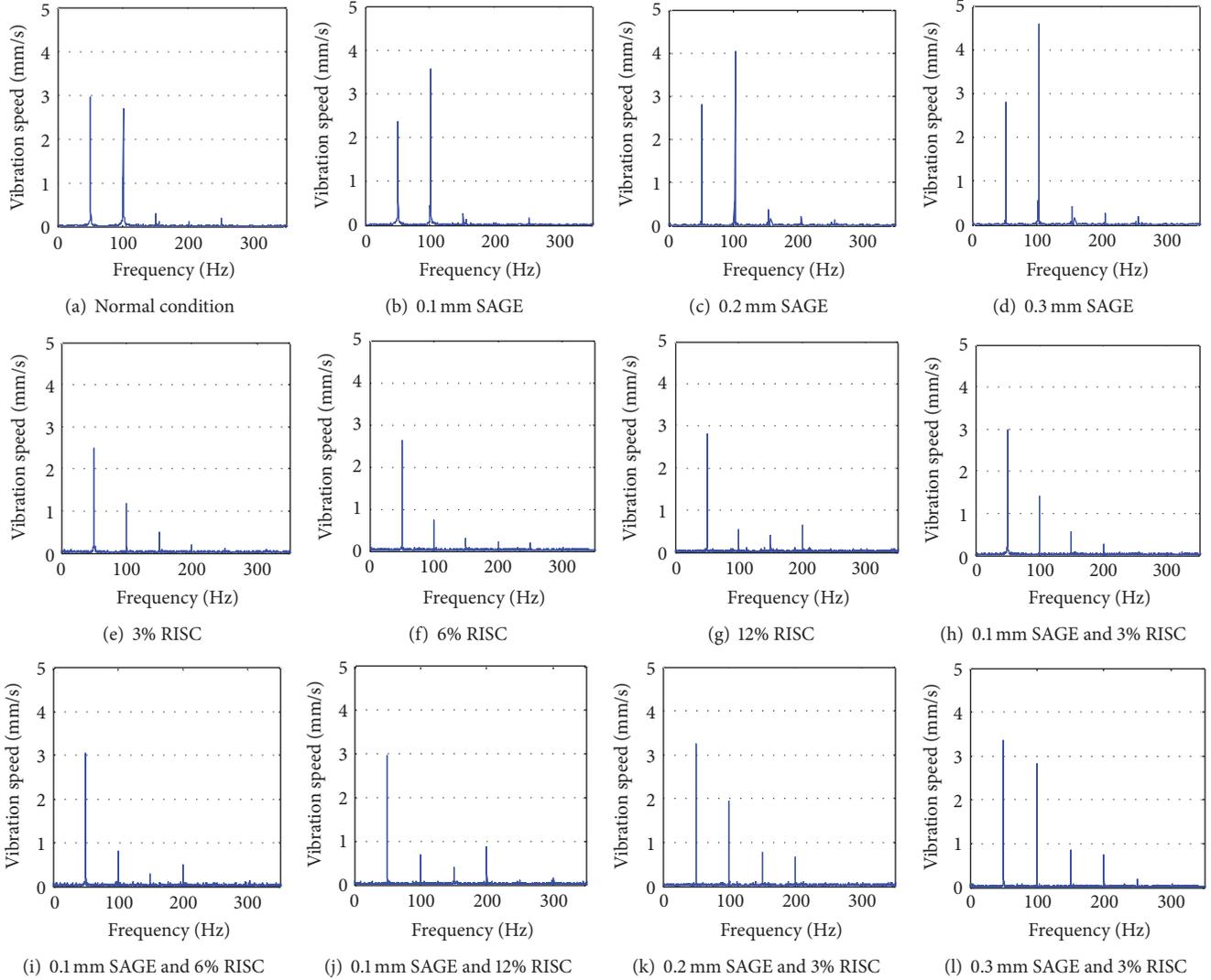


FIGURE 8: Tested stator vibration spectra under different conditions.

vibration amplitudes at each harmonic frequency generally accords with the changing trend of MPPUA. Meanwhile, it is also clear that as SAGE increases, both MPPUA and the vibration under CF have a larger value at f , $3f$, and $4f$ but in the meantime have a smaller value at $2f$ than those under the single SAGE fault. However, as RISC develops, both MPPUA and the stator vibration will have a larger value under the CF condition than under the single RISC fault. These results are in good accordance with the conclusions obtained previously, confirming that both the FEM calculation and the experiment data match the previous analysis. In addition, this

also indicates that the stator vibration response is consistent with the exciting MPPUA.

3.2.3. Fault Identification Criterion. Based on the theoretical analysis, the numerical FEM calculation, and the experiment study, it is found that the stator vibration characteristic varies under different faulty conditions. This can be employed to develop an identification criterion for the single and the composite faults composed of SAGE and RISC. More details can be found in Table 3.

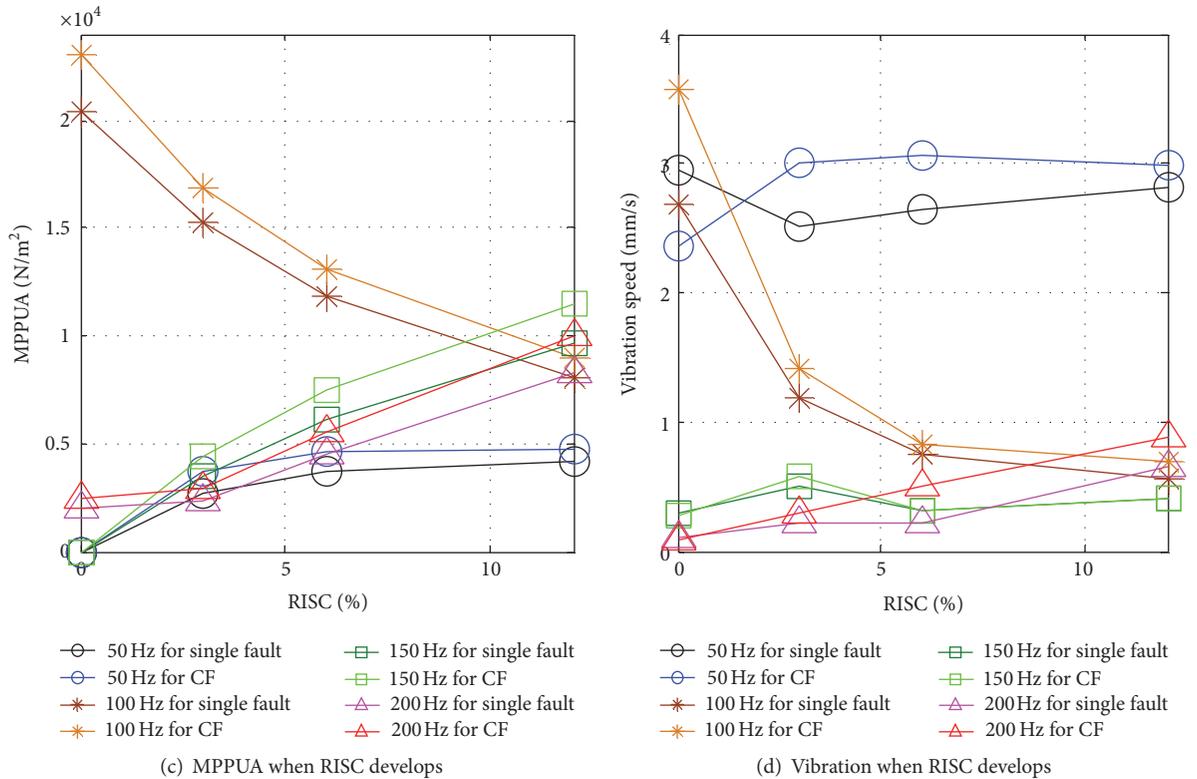
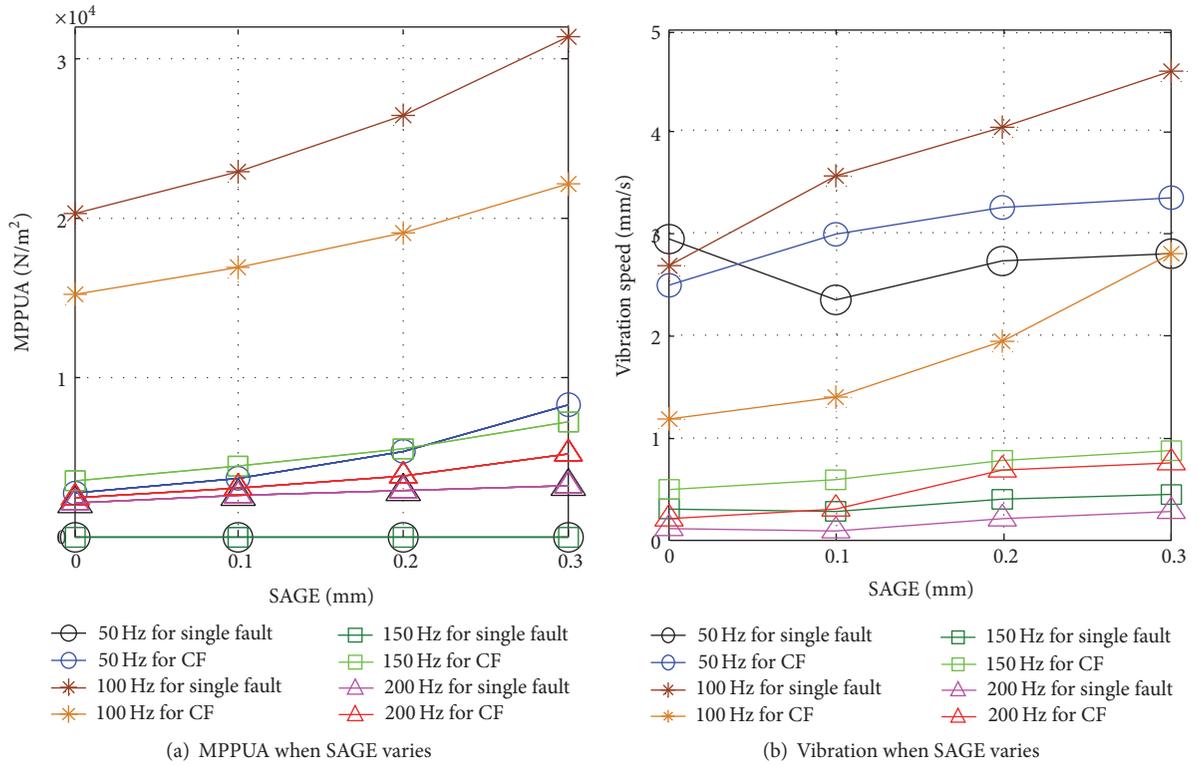


FIGURE 9: Comparison between MPPUA and tested stator vibration for different conditions.

4. Conclusions

This paper investigates the stator vibration difference among SAGE, RISC, and the composite faults composed of SAGE and RISC. Based on the proposed theoretical analysis, the numerical FEM calculation, and the experimental study, primary conclusions can be generally drawn as follows:

- (1) Normally, stator vibration has only the 2nd harmonic component in theory, while actually it also has the 1st harmonic component which is transferred from the rotor vibration. The occurrence of SAGE will increase the 2nd harmonic component.
- (2) As RISC and CF take place in the generator, extra stator vibrations at f , $2f$, $3f$, and $4f$ will be produced. As RISC develops, the 1st, 3rd, and the 4th harmonic components will be increased, while the 2nd harmonic component will be decreased.
- (3) In the case of CF, the increment of SAGE will raise the 1st, 2nd, 3rd, and the 4th harmonic components of the stator vibration at the same time, while under the single SAGE fault only the 2nd harmonic component will be increased.
- (4) As RISC develops, the stator vibration under CF at f , $2f$, $3f$, and $4f$ will have a larger intensity than under the single RISC fault. However, as SAGE increases, the stator vibration under CF will be more intensive at f , $3f$, and $4f$ but meanwhile have weaker amplitude at $2f$ than under the single SAGE fault.

Potentially, these sensitive conclusions can be applied to the practical online monitoring on the generators. Since it is very convenient to test the stator vibration with the industrial velocity sensors, further development of monitoring systems and utilities, which will be beneficial for the fault identification among SAGE, RISC, and the composite fault composed of SAGE and RISC, is probably to be carried out.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (51307058), the Natural Science Foundation of Hebei Province, China (E2015502013, E2015502008), and the Chinese Fundamental Research Funds for the Central Universities (2015ZD27).

References

- [1] C. Bruzzese, "Diagnosis of eccentric rotor in synchronous machines by analysis of split-phase currents—part II: experimental analysis," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 8, pp. 4206–4216, 2014.
- [2] W. Yucai and L. Yonggang, "Diagnosis of rotor winding inter-turn short-circuit in turbine generators using virtual power," *IEEE Transactions on Energy Conversion*, vol. 30, no. 1, pp. 183–188, 2015.
- [3] C. Bruzzese and G. Joksimovic, "Harmonic signatures of static eccentricities in the stator voltages and in the rotor current of no-load salient-pole synchronous generators," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 1606–1624, 2011.
- [4] M. Blödt, J. Regnier, and J. Faucher, "Distinguishing load torque oscillations and eccentricity faults in induction motors using stator current wigner distributions," *IEEE Transactions on Industry Applications*, vol. 45, no. 6, pp. 1991–2000, 2009.
- [5] C. Bruzzese, "Diagnosis of eccentric rotor in synchronous machines by analysis of split-phase currents—part II: experimental analysis," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 8, pp. 4206–4216, 2014.
- [6] X. Y. Zhao and W. T. Ge, "Simulation research of fault model of detecting rotor dynamic eccentricity in brushless DC motor based on motor current signature analysis," *Proceedings of the Chinese Society of Electrical Engineering*, vol. 31, no. 36, pp. 124–130, 2011.
- [7] S. Yuguang, H. Liangliang, and W. Xiangheng, "Steady-state currents harmonic characteristics of synchronous machine with inter-turn short circuits of field windings," *Proceedings of the CSEE*, vol. 33, no. 33, pp. 51–57, 2010.
- [8] J. Li and Y. Cho, "Dynamic reduction of unbalanced magnetic force and vibration in switched reluctance motor by the parallel paths in windings," *Mathematics and Computers in Simulation*, vol. 81, no. 2, pp. 407–419, 2010.
- [9] Y.-L. He, G.-J. Tang, S.-T. Wan, and Y.-C. Wu, "Effect of the static air-gap eccentricity & stator inter-turn short circuit composite fault on the generator circulating current characteristics," *International Review of Electrical Engineering*, vol. 7, no. 5, pp. 5442–5450, 2012.
- [10] D. Žarko, D. Ban, I. Vazdar, and V. Jarić, "Calculation of unbalanced magnetic pull in a salient-pole synchronous generator using finite-element method and measured shaft orbit," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 6, pp. 2536–2549, 2012.
- [11] L. J. Wu, Z. Q. Zhu, J. T. Chen, and Z. P. Xia, "An analytical model of unbalanced magnetic force in fractional-slot surface-mounted permanent magnet machines," *IEEE Transactions on Magnetics*, vol. 46, no. 7, pp. 2686–2700, 2010.
- [12] S. Nadarajan, S. K. Panda, B. Bhangu, and A. K. Gupta, "Hybrid model for wound-rotor synchronous generator to detect and diagnose turn-to-turn short-circuit fault in stator windings," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 3, pp. 1888–1900, 2015.
- [13] D. Shah, S. Nandi, and P. Neti, "Stator-interturn-fault detection of doubly fed induction generators using rotor-current and search-coil-voltage signature analysis," *IEEE Transactions on Industry Applications*, vol. 45, no. 5, pp. 1831–1842, 2009.
- [14] H.-M. Li, Y.-C. Wu, and Y.-G. Li, "Influence of rotor windings inter-turn short circuit fault on electric machine shaft voltage," *Proceedings of the Chinese Society of Electrical Engineering*, vol. 29, no. 36, pp. 96–100, 2009.
- [15] Y. C. Wu, Y. G. Li, and H. M. Li, "Diagnosis of non-salient pole synchronous generator rotor's typical faults based on shaft voltage," *Transactions of China Electrotechnical Society*, vol. 25, no. 6, pp. 178–184, 2010.
- [16] L. Yong-gang, Z. Yan-jun, C. Lei, and J. Xuan, "Rotor winding inter-turn short circuit fault diagnosis based on the circulating current in stator winding parallel branches," *High Voltage Engineering*, vol. 35, no. 5, pp. 1014–1019, 2009.

- [17] J. G. Cintron-Rivera, S. N. Foste, and E. G. Strangas, "Mitigation of turn-to-turn faults in fault tolerant permanent magnet synchronous motors," *IEEE Transactions on Energy Conversion*, vol. 30, no. 2, pp. 465–475, 2015.
- [18] Y.-C. Wu, Y.-G. Li, W.-Z. Feng, and W.-J. Zhang, "Analysis on unbalanced magnetic pull generated by turn-to-turn short circuit of rotor windings within turbine generator," *Electric Machines and Control*, vol. 19, no. 3, pp. 37–44, 2015.
- [19] L. Hao, J. Wu, and Y. Zhou, "Theoretical analysis and calculation model of the electromagnetic torque of nonsalient-pole synchronous machines with interturn short circuit in field windings," *IEEE Transactions on Energy Conversion*, vol. 30, no. 1, pp. 110–121, 2015.
- [20] Y.-L. He, F.-L. Wang, M.-Q. Ke, and G.-J. Tang, "Rotor vibration difference among the single and the combined faults composed by static air-gap eccentricity and rotor interturn short circuit," in *Proceedings of the 9th IFToMM International Conference on Rotor Dynamics*, vol. 21 of *Mechanisms and Machine Science*, pp. 637–648, 2015.
- [21] L. Hao, J. Wu, Z. Chen, and H. Song, "Mechanism of effects of inter-turn short circuits in field windings on large turbo-generator vibration," *Automation of Electric Power Systems*, vol. 38, no. 4, pp. 25–50, 2014.
- [22] H.-W. Fang, C.-L. Xia, and G.-P. Li, "Analysis of synchronous generator electro-magnetic torque and vibration with armature winding fault," *Journal of Tianjin University*, vol. 42, no. 4, pp. 322–326, 2009.
- [23] P. C. M. Lamim Filho, R. Pederiva, and J. N. Brito, "Detection of stator winding faults in induction machines using flux and vibration analysis," *Mechanical Systems and Signal Processing*, vol. 42, no. 1-2, pp. 377–387, 2014.
- [24] Y.-L. He, M.-Q. Ke, F.-L. Wang, G.-J. Tang, and S.-T. Wan, "Effect of static eccentricity and stator inter-turn short circuit composite fault on rotor vibration characteristics of generator," *Transactions of the Canadian Society for Mechanical Engineering*, vol. 39, no. 4, pp. 767–781, 2015.

Research Article

Fault Diagnosis of a Hydraulic Pump Based on the CEEMD-STFT Time-Frequency Entropy Method and Multiclass SVM Classifier

Wanlin Zhao,^{1,2} Zili Wang,^{1,2} Jian Ma,^{1,2} and Lianfeng Li^{1,2}

¹Science & Technology Laboratory on Reliability & Environmental Engineering, Beijing 100191, China

²School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China

Correspondence should be addressed to Jian Ma; 09977@buaa.edu.cn

Received 27 April 2016; Accepted 23 August 2016

Academic Editor: Ganging Song

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

The fault diagnosis of hydraulic pumps is currently important and significant to ensure the normal operation of the entire hydraulic system. Considering the nonlinear characteristics of hydraulic-pump vibration signals and the mode mixing problem of the original Empirical Mode Decomposition (EMD) method, first, we use the Complete Ensemble EMD (CEEMD) method to decompose the signals. Second, the time-frequency analysis methods, which include the Short-Time Fourier Transform (STFT) and time-frequency entropy calculation, are applied to realize the robust feature extraction. Third, the multiclass Support Vector Machine (SVM) classifier is introduced to automatically classify the fault mode in this paper. An actual hydraulic-pump experiment demonstrates the procedure with a complete feature extraction and accurate mode classification.

1. Introduction

Hydraulic systems have been widely used in aeronautics, astronautics, automobiles, shipping, and so on. As the heart of a hydraulic system, the performance of the hydraulic pump significantly affects the entire hydraulic system [1]. Thus, achieving real-time fault diagnosis of the hydraulic pump is essential and urgent to maintain the entire system [2]. For the hydraulic pump, its structure is complex, the relationship among its internal parameters is highly nonlinear, and there are strong couplings among various fault features. As a result, an accurate mathematical model is difficult to establish. Therefore data-driven diagnostic methods are commonly used for hydraulic pumps based on their vibration signals. Generally, the entire fault diagnosis process can be considered a pattern identification problem that mainly includes two important procedures: feature extraction and mode classification.

Many data-driven feature extraction methods have emerged in recent years that are different from the traditional

time-domain analysis and frequency-domain analysis methods. The Empirical Mode Decomposition (EMD), which was developed by Huang et al., is a time-frequency analysis method and has advantages in addressing nonlinear and nonstationary signals [3]. The EMD can decompose any signal into intrinsic mode functions (IMFs) based on the local timescale of the data, without using a priori basis [4]. However, the EMD faces a serious problem, “mode mixing,” where a notably disparate amplitude in a mode oscillates or notably similar oscillations occur in different modes. Because of this problem, a new method was proposed: Ensemble Empirical Mode Decomposition (EEMD), which performs the EMD over an ensemble of the signal plus Gaussian white noise to obtain more regular modes. However, the EEMD also created new difficulties. The reconstructed signal contains residual noise, and different realizations of signal plus noise may produce different numbers of modes. To overcome these difficulties, another EMD method has been proposed and successfully applied to vibration signal analysis, complete EEMD (CEEMD), which provides an exact reconstruction

of the original signal and a better spectral separation of the modes [5, 6]. Han and van der Baan used CEEMD to analyze the synthetic and real seismic data and obtained a good result [7]. In our study, CEEMD is selected to adaptively decompose signals into a small number of IMFs or modes, and the Short-Time Fourier Transform (STFT) algorithm and time-frequency entropy analysis method are simultaneously used to obtain the fault feature vectors composed by multi-scale time-frequency entropy. This feature extraction method is defined as the CEEMD-STFT time-frequency entropy method.

After the fault feature is extracted, a classifier is exploited to automatically achieve mode classification. Support Vector Machine (SVM) is a powerful machine learning method based on the statistical learning theory and structural risk minimization principle that has been successfully applied to fault diagnosis and satisfactorily solved the overfitting and local optimal solution problem [8]. However, there are no elegant approaches to solve multiclass problems. A better alternative is provided by the construction of multiclass SVM [9], which is inherently two-class SVM classifiers. In this paper, we build a multiclass SVM classifier to classify the fault mode over the feature vectors, whose dimensions have been compressed using the Principal Component Analysis (PCA) algorithm because the original feature vectors are always too large, complex, and variable for postprocessing.

This paper is organized as follows: Section 2 introduces the relevant feature extraction and mode classification methodology, which includes the CEEMD, STFT, time-frequency entropy, and multiclass SVM method; Section 3 describes the case study to validate the entire method; Section 4 presents the conclusions of this paper.

2. Methodology

As shown in Figure 1, the complete fault diagnosis scheme has three elements: data preprocessing, fault feature extraction, and fault mode classification. More details are provided in the following parts.

2.1. Feature Extraction Based on the CEEMD-STFT Time-Frequency Entropy Method

2.1.1. Complete Ensemble Empirical Mode Decomposition (CEEMD)

(A) *Empirical Mode Decomposition (EMD)*. The EMD is an adaptive signal analysis method based on the signal characteristic local extrema, which separate a signal into a certain number of IMF components. To be considered an IMF, a signal must satisfy two conditions: (1) the number of extrema and the number of zero-crossings must be equal or differ at most by one; (2) the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero at any data location [10]. Assume that $x(t)$ is the signal to be decomposed, the concrete steps of the EMD are shown as follows.

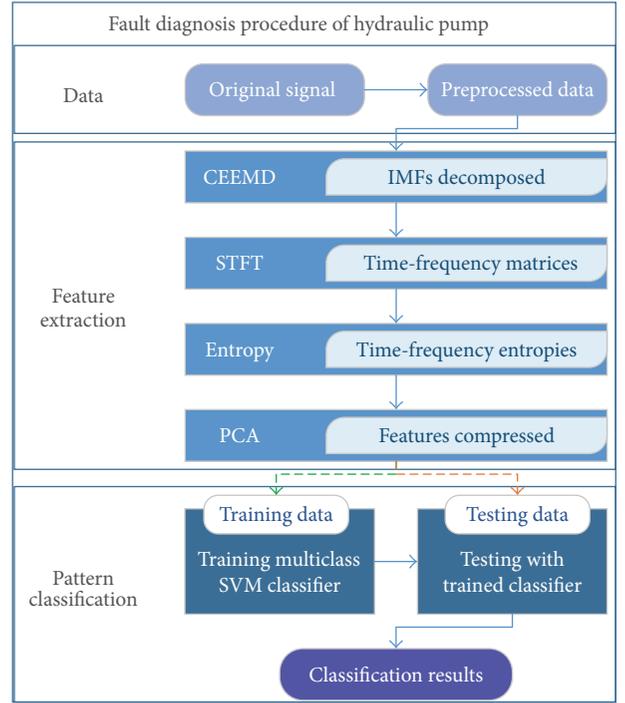


FIGURE 1: Entire fault diagnosis procedure.

Step 1 (Initialization). Set $k = 0$, where k indicates the mode, and the residual component $r_0(t) = x(t)$.

Step 2. Calculate the mean envelope $m(t)$

$$m(t) = \frac{[e_{\max}(r_k(t)) + e_{\min}(r_k(t))]}{2}, \quad (1)$$

where $e_{\max}(r_k(t))$ and $e_{\min}(r_k(t))$ are the upper and lower envelopes, respectively, which are obtained through cubic-spline interpolation on the local maxima and minima of $r_k(t)$.

Step 3. Compute the IMF candidate

$$c_{k+1}(t) = r_k(t) - m(t). \quad (2)$$

Step 4. Does $c_{k+1}(t)$ satisfy the IMF properties?

- (i) Yes. Set $\text{IMF}_{k+1} = c_{k+1}(t)$ and $r_{k+1}(t) = r_k(t) - \text{IMF}_{k+1}$; let $k = k + 1$, and go to Step 2.
- (ii) No. Set $c_{k+1}(t)$ as $r_k(t)$, and go to Step 2.

Step 5. Continue Steps 2–4 until $r_k(t)$ becomes a monotonic function.

Each obtained IMF through the EMD contains different frequency components of the signal from high to low frequencies and represent the inherent mode characteristics of the signal.

(B) *Ensemble EMD (EEMD)*. To solve the “mode mixing” problem, Wu and Huang proposed the Ensemble EMD (EEMD) method [11], which defines the “true” modes as

the mean of the corresponding IMFs that are obtained via EMD over an ensemble of the original signal plus different realizations of finite variance white noise [12]. Considering x as an example signal, the EEMD algorithm can be described as follows.

Step 1. Generate

$$x^{(i)} = x + \beta w^{(i)}, \quad (3)$$

where $w^{(i)}$ ($i = 1, \dots, I$) are different realizations of white Gaussian noise.

Step 2. Each $x^{(i)}$ ($i = 1, \dots, I$) is fully decomposed by EMD to obtain the modes $\text{IMF}_k^{(i)}$, where $k = 1, \dots, K$ indicates the mode.

Step 3. Assign $\overline{\text{IMF}}_k$ as the k th mode of x , which is obtained by averaging the corresponding modes:

$$\overline{\text{IMF}}_k = \frac{1}{I} \sum_{i=1}^I \text{IMF}_k^{(i)}. \quad (4)$$

However, the EEMD method has some disadvantages: (1) the decomposition is not complete; (2) different realizations of signal plus white noise may generate different numbers of modes.

(C) *Complete EEMD (CEEMD).* To address the aforementioned reconstruction error, complete EEMD (CEEMD) was proposed by Torres et al. in 2011 [5]. The procedure of CEEMD is described as follows.

Step 1. The first IMF is calculated in the identical method to EEMD. First, add white noise to the original signal and obtain the first EMD component of the data with noise. Repeat the decomposition by adding different noise realizations, and compute the ensemble average to define it as the first IMF_1 of the original signal x , that is,

$$\text{IMF}_1 = \frac{1}{I} \sum_{i=1}^I E_1(x + \varepsilon_0 w^i), \quad (5)$$

where $E_j(\cdot)$ is defined as an operator and the j th mode can be computed through EMD when it meets a new signal. x is the raw signal, w^i is the different white noise, and ε_0 is a ratio coefficient.

Step 2. Calculate a unique first residue as

$$r_1 = x - \text{IMF}_1. \quad (6)$$

Then set $r_1 + \varepsilon_1 E_1(w^i)$ ($i = 1, \dots, I$) as the new signal for decomposition. When the first IMF component has been obtained, we must calculate the ensemble average as the second component IMF_2 :

$$\text{IMF}_2 = \frac{1}{I} \sum_{i=1}^I E_1(r_1 + \varepsilon_1 E_1(w^i)). \quad (7)$$

Step 3. Repeat Steps 1-2, and we can obtain the $(k+1)$ th IMF component IMF_{k+1} :

$$\text{IMF}_{k+1} = \frac{1}{I} \sum_{i=1}^I E_1(r_k + \varepsilon_k E_k(w^i)). \quad (8)$$

Step 4. Finally, obtain the last residual function R , until the residue cannot be decomposed. Then, $R = x - \sum_{k=1}^K \text{IMF}_k$, where K is the total number of IMF. The signal is described as

$$x = \sum_{k=1}^K \text{IMF}_k + R. \quad (9)$$

The last step makes the proposed decomposition complete and provides an exact reconstruction of the original signal.

2.1.2. Short-Time Fourier Transform (STFT). The time and frequency information in each IMF relates to the sampling frequency and changes with the signal itself, so research on the time-frequency domain characterization of signals has been a key component of signal analysis [13]. The STFT is a popular method to analyze nonstationary signals. The STFT of the signal $x(t)$ is defined as

$$X(t, f) = \int_{-\infty}^{\infty} x(\tau) h(\tau - t) \cdot e^{-j2\pi f\tau} d\tau, \quad (10)$$

where $h(t)$ should be a low-pass filter, and $\|h\|^2 = 1$. Note that $h(\tau - t) \cdot e^{j2\pi f\tau}$ has its energy concentrated at time t and frequency f . Thus, $|X(t, f)|^2$ can be considered the energy in $x(t)$ at frequency f and time t . Generally, one displays the energy at each time and frequency pair, that is,

$$P(t, f) = |X(t, f)|^2. \quad (11)$$

$P(t, f)$ is known as the spectrogram (SP) of $x(t)$ [14].

The spectrogram algorithm is an analysis algorithm that produces a two-dimensional image representation of vibration signals. The Power Spectrum Density (PSD) function $P(t, f)$ is expressed as a Pseudo Color Map (PCM), which is a spectrogram with a time axis and a frequency axis. This time-frequency spectrum, which can be called the ‘‘visual language,’’ shows the modulation characteristics of the signals.

2.1.3. Time-Frequency Entropy. The time-frequency distribution of the vibration signal obtained through the STFT method presents modulation characteristic, that is, the energy distribution changes at different moments. Therefore, a fault can be detected by comparing the energy distribution of the signals with and without fault conditions in the time-frequency domain, which indicates that the energy variation in the time-frequency plane may indicate a fault occurrence [15]. Because the spectrogram can provide an accurate energy-frequency-time distribution, the information entropy theory, which measures the uniformity of the probability distribution, can be introduced to the time-frequency distribution to quantitatively describe the divergence in different operating conditions [16].

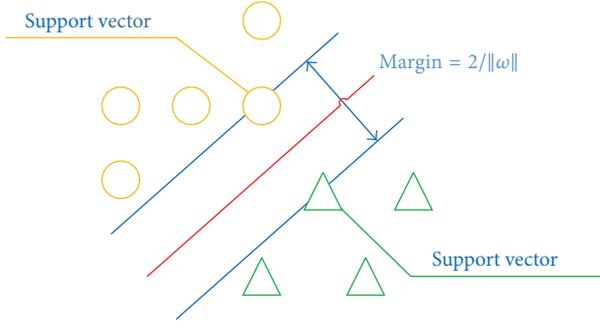


FIGURE 2: Illustration for data classification using a 2-class SVM.

Let a time-frequency plane have N blocks with equal areas, where the information source for the entire plane is A and for each block is W_i ($i = 1, \dots, N$), so the probability that each information source appears in the entire system is

$$\sum_{i=1}^N q_i = 1, \quad q_i = \frac{W_i}{A}, \quad (i = 1, \dots, N). \quad (12)$$

According to the information entropy calculation [17], the time-frequency entropy is defined as

$$s(q) = -\sum_{i=1}^N q_i \ln q_i. \quad (13)$$

Now, we can consider the time-frequency entropy of each IMF as the extracted feature vectors, which will be the input of the mode classification.

2.2. Mode Classification Based on the Multiclass SVM

2.2.1. Support Vector Machine (SVM). Support Vector Machine (SVM), which originated from the statistical learning theory and an optimal separating hyper-plane in the case of linear separation, was developed by Cortes and his coworker [18]. Through some nonlinear mapping functions, the original mode space is mapped into the high-dimensional feature space \mathbb{Z} . Then the optimal separating hyper-plane is constructed in the feature space. Consequently, the nonlinear problem in the low-dimensional space corresponds to the linear problem in the high-dimensional space.

2.2.2. Two-Class SVM. SVMs are primarily designed for 2-class classification problems. To illustrate the basic principle, a schematic diagram of 2-class SVM is shown in Figure 2, where two different classes (circles and triangles) are classified by a linear boundary H , and the distance between the boundary and the nearest data point in each class is maximal.

Assume that the input vector is x , which is mapped into high-dimensional space \mathbb{Z} through the nonlinear mapping function $\varphi(x)$, and the linear function $(\omega \cdot \varphi(x)) + b = 0$ in the high-dimensional feature space can be used to construct the optimal classification hyper-plane. The training data are set as $\{x_i, y_i\}$, $i = 1, 2, \dots, l$; $x_i \in R^n$, $y_i \in \{-1, +1\}$, y_i is the corresponding label of x_i . Then, ω is a weight vector, and

the margin is $1/\|\omega\|$. The following constraint optimization problem is the solution of maximizing the margin $1/\|\omega\|$:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \xi_i \\ \text{s.t.} \quad & y_i (\omega \cdot \varphi(x_i) + b) \geq 1 - \xi_i, \quad i = 1, \dots, l \\ & \xi_i \geq 0, \quad i = 1, \dots, l, \end{aligned} \quad (14)$$

where coefficient C is a penalty factor and ξ_i is a slack factor [16].

In addition, using the duality theory of optimization theory and Kernel function, the final decision function is described by

$$\begin{aligned} f(x) &= \text{sgn} \left(\sum_{x_i \in SVs} y_i \alpha_i (\varphi(x_i) \cdot \varphi(x)) + b \right) \\ &= \text{sgn} \left(\sum_{x_i \in SVs} y_i \alpha_i K(x_i, x) + b \right), \end{aligned} \quad (15)$$

where $K(x_i, x)$ is the kernel function, which satisfies Mercer condition; the constants α_i are named Lagrange multipliers and are determined in the optimization procedure. The typical kernel functions are the polynomial kernel, Radial Basis Function (RBF) kernel, sigmoid kernel, and linear kernel. In many practical applications, the RBF kernel has the highest classification accuracy rate compared to the other kernel functions, so we mainly consider the RBF kernel in this paper.

The SVM was originally designed for binary classification and had good performance, but it still faced many difficulties in addressing multiclass classification problems. The SVM is not sufficient to handle a practical situation.

2.2.3. Multiclass SVM. Currently, several methods based on the SVM have been proposed for multiclass classification, such as “one-against-all,” “one-against-one,” and Directed Acyclic Graph (DAG). Experiments indicate that the “one-against-one” and DAG-SVM methods are most suitable for practical situation. In this paper, the “one-against-one” method is selected for classification [19].

Let us suppose that the training data set is $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$, $x_i \in R^n$ and the “one-against-one” method constructs $C_k^2 = k(k-1)/2$ classifiers, each of which is trained using the data from two classes. For example, we should solve the following binary classification problem for the training data from the i th and j th classes:

$$\begin{aligned} \min_{\omega^{i,j}, b^{i,j}, \xi_t^{i,j}} \quad & \frac{1}{2} \|\omega^{i,j}\|^2 + C \left[\sum_{i=1}^l \xi_t^{i,j} \right] \\ \text{s.t.} \quad & [(\omega^{i,j} \cdot x_t) + b^{i,j}] - 1 + \xi_t^{i,j} \geq 0, \quad \text{if } y_t = i \\ & [(\omega^{i,j} \cdot x_t) + b^{i,j}] + 1 - \xi_t^{i,j} \leq 0, \\ & \text{if } y_t = j, \quad \xi_t^{i,j} \geq 0. \end{aligned} \quad (16)$$

TABLE 1: Six-dimensional fault features.

Fault pattern	No.	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6
Normal	1	4.9681	3.9921	4.0155	3.6082	3.2670	2.6576
	2	4.9732	3.9936	4.0145	3.6031	3.2395	2.6541
	...						
	20	4.7565	3.9406	4.1075	3.6274	3.3763	2.6567
Rotor wear	21	3.8933	4.2478	3.5822	3.7066	3.2937	2.5966
	22	3.8923	4.1903	3.5846	3.7407	3.2990	2.5675
	...						
	40	3.8991	3.9456	3.5417	3.6743	3.3190	2.5426
Swash plate wear	41	4.1123	3.4824	3.4975	3.6294	3.2926	2.5332
	42	4.1285	3.4886	3.5151	3.6196	3.3310	2.5820
	...						
	60	4.1782	3.4949	3.5132	3.5932	3.3496	2.6182

When testing is performed for the unknown sample x , we construct all $k(k-1)/2$ classifiers to realize the class discrimination and make decisions using the following voting strategy: if $\text{sgn}((\omega^{i,j} \cdot x_t) + b^{i,j})$, x is in the i th class, then, the vote for the i th class is increased by one; otherwise, the j th class is increased by one. Finally, we predict that x is in the class with the largest vote [20].

3. Experimental Verification

3.1. Experiment Setup. The plunger pump test-rig is shown in Figure 3; from this test-rig, the original vibration signals were obtained to verify the proposed method. The vibration data were obtained from the front side of the hydraulic pump with a stabilized motor speed of 528 r/min and a sampling rate of 1000 Hz. In this experiment, two commonly occurring faults were set: swash plate wear and rotor wear. Under three conditions (two faulty conditions and the normal state), 20 groups of samples (1024 sampling points for each group) were selected for the analysis.

3.2. Model for Fault Diagnosis of Hydraulic Pumps

3.2.1. Feature Extraction Based on the CEEMD-STFT and Time-Frequency Entropy

(1) *CEEMD Model.* The parameters of the CEEMD model were set as follows: the noise standard deviation (Nstd) was 0.2, the Number of Realizations (NR) was 600, and the maximum number of sifting iterations allowed (MaxIter) was 5000. The original signals of each state were decomposed into a series of IMFs; the first six IMFs were selected for further analysis, as shown in Figure 4.

(2) *Procedure of the STFT and Time-Frequency Entropy Acquisition.* The parameters of STFT were selected as follows: the length of the window, number of overlaps, and sampling frequency (fs) were 256, 254, and 1000, respectively, and the length of the discrete Fourier transforms was equal to



FIGURE 3: Plunger pump test-rig.

the window length. Then, the time-frequency matrices or spectrograms of each state were obtained in Figure 5.

The time-frequency entropy of each state can be calculated based on the time-frequency matrices. The time-frequency block was set as length = width = 64, and both the lateral and longitudinal slip steps were 32. Then, a six-dimensional time-frequency entropy was obtained for each group, which is one of the fault feature vectors. All of the fault features are listed in Table 1.

(3) *Feature Dimension Reduction Based on PCA.* To improve the accuracy and robustness of the fault diagnosis, dimension reduction is necessary for the high dimensional fault feature vectors. PCA, which is an important and powerful methods to extract the most significant information from data and compress the size of the data [21], was used to acquire the three-dimensional feature vectors in Table 2.

The clustering result of the fault features is visually displayed in Figure 6, which obviously shows a good performance of the hydraulic-pump fault mode classification.

3.2.2. Fault Mode Classification Based on Multiclass SVM. The extracted fault feature sets were divided into training data and testing data (the first ten groups were set as the training data and the remainder was set as the testing data for every state). First, the training multiclass SVM classifier was

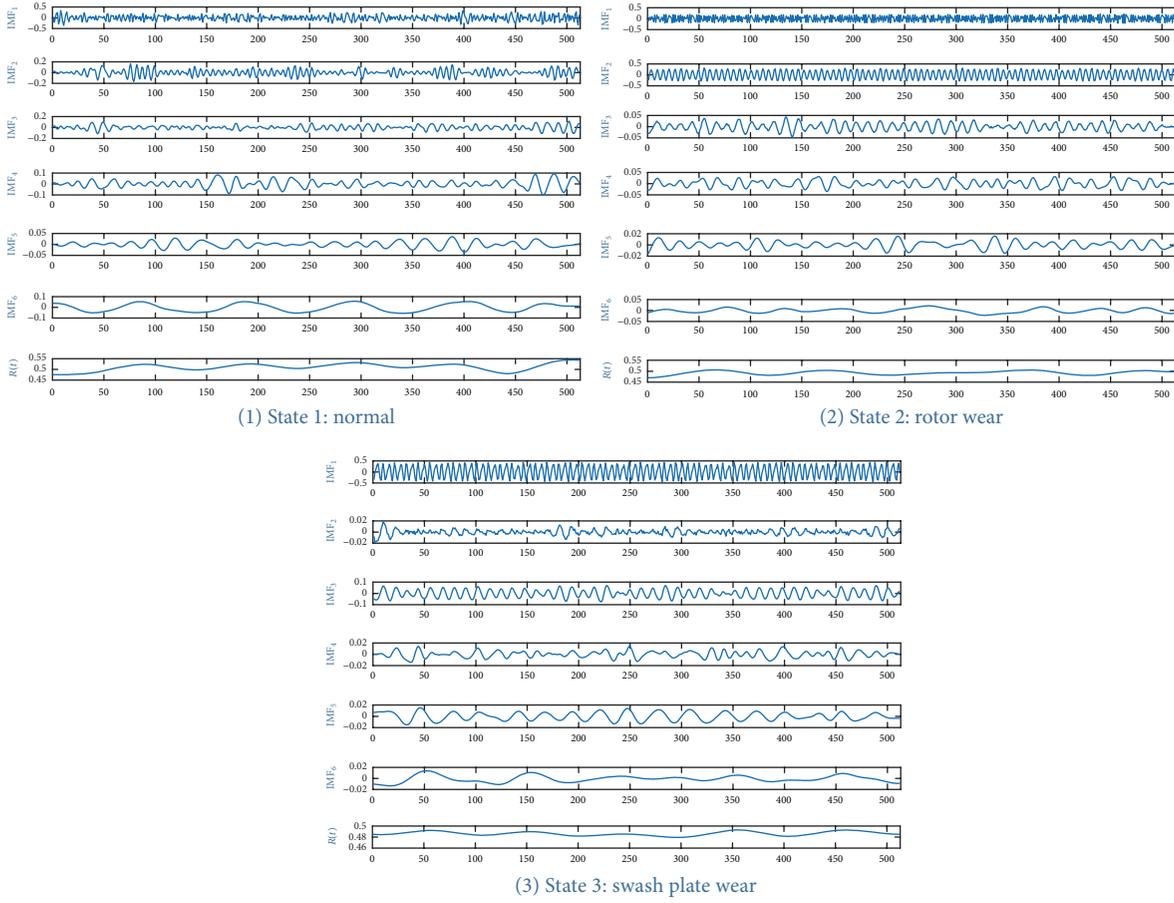


FIGURE 4: First 6 IMFs of each state.

TABLE 2: Feature vector set after dimension reduction.

Fault pattern	No.	Feature 1	Feature 2	Feature 3
Normal	1	4.1284	2.9741	5.6483
	2	4.1327	2.9729	5.6553
	...			
Rotor wear	20	3.9954	2.9829	5.5353
	21	3.7289	2.0520	5.2589
	22	3.6789	2.0772	5.2629
Swash plate wear	...			
	40	3.5266	2.1772	5.1467
	41	3.2972	2.4613	5.1078
	42	3.3079	2.4822	5.1103
...				
60	3.3408	2.5041	5.1157	

trained as previously proposed with the training data. Then, the trained classifier was used to classify the fault mode of the testing data and calculate the recognition accuracy. The classification results of the testing data are shown in Table 3 and Figure 7. These testing results verify that the recognition

performance is absolutely good, and the multiclass SVM method is notably effective for mode classification.

Combining the clustering figure and multiclass SVM classification results, the effectiveness and feasibility of this method for hydraulic-pump fault diagnosis were proven, and a high classification performance was also obviously obtained.

4. Conclusion

An effective method for the feature extraction and mode classification of vibration signals has been performed in this paper, and this algorithm was successfully verified on practical signals from a hydraulic pump. The CEEMD model, which is an improvement of EMD and can solve the “mode mixing” problem, was combined with the STFT analysis method and time-frequency entropy calculation to extract the robust and significant fault feature. Meanwhile, the multiclass SVM classifier was selected to process the small sample and multiple-fault situation, and it obtained a perfect classification result. Then, the accuracy and feasibility of this hydraulic-pump fault diagnosis method were demonstrated. Future work will concentrate on the application of this method to other objects or fields for signal analysis and fault diagnosis.

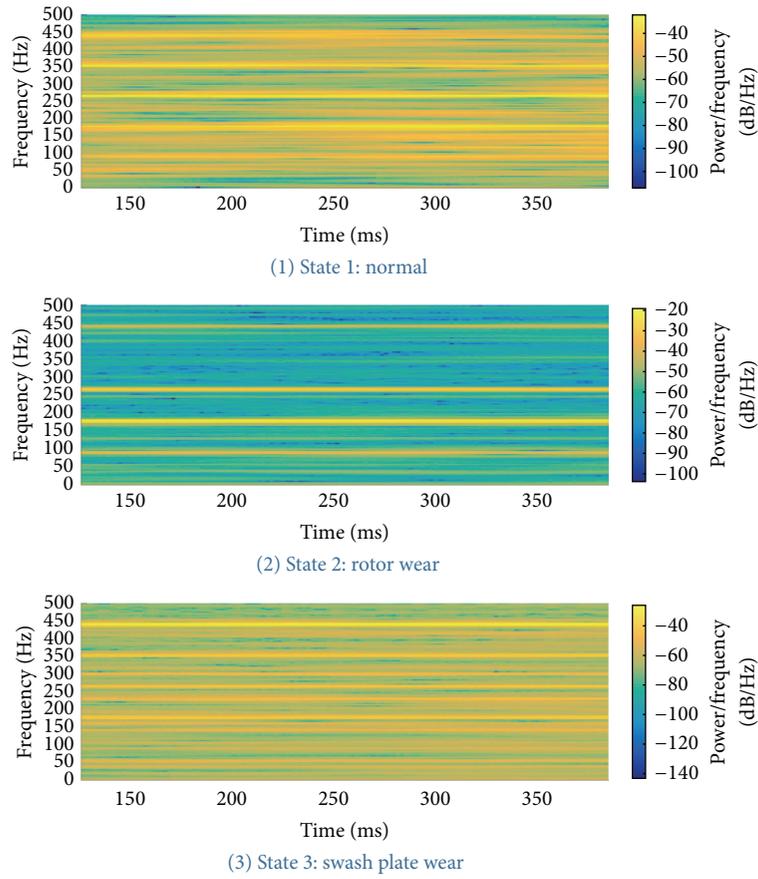


FIGURE 5: Spectrograms of the first IMF of each state.

TABLE 3: Classification results of the testing data.

Fault pattern	Actual label	Index	No.	Feature 1	Feature 2	Feature 3	Predicted label
Normal	1	1	11	4.0906	2.9178	5.5591	1
		2	12	4.0961	2.9253	5.5036	1
					1
		10	20	3.9954	2.9829	5.5353	1
Rotor wear	2	11	31	3.7442	2.0107	5.1790	2
		12	32	3.6416	2.1110	5.1027	2
					2
Swash plate wear	3	20	40	3.5266	2.1772	5.1467	2
		21	51	3.3093	2.5163	5.1968	3
		22	52	3.3018	2.4982	5.2170	3
					3
		30	60	3.3408	2.5041	5.1157	3

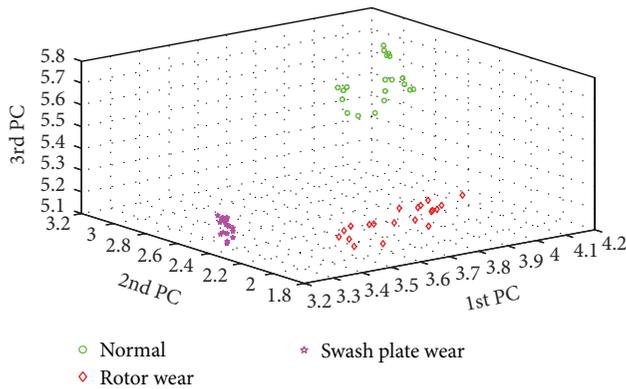


FIGURE 6: Clustering result of the fault features.

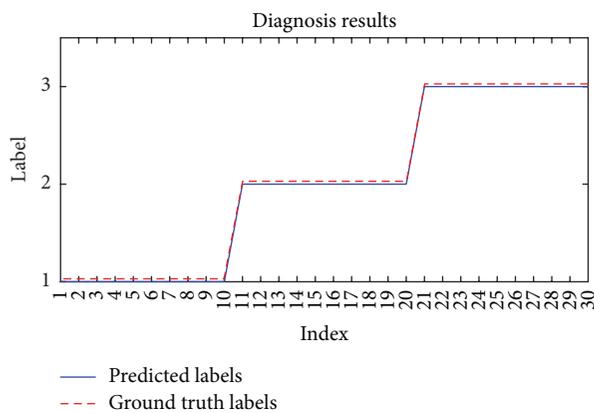


FIGURE 7: Classification results of the testing data.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant nos. 51105019 and 51575021), as well as the Technology Foundation Program of National Defense (Grant no. Z132013B002).

References

- [1] Z. Ruixiang, L. Tingqi, H. Jianding, and Y. Dongchao, "Fault diagnosis of airplane hydraulic pump," in *Proceedings of the 4th World Congress on Intelligent Control and Automation*, Shanghai, China, June 2002.
- [2] T. Zhang and N. Zhang, "Vibration modes and the dynamic behaviour of a hydraulic plunger pump," *Shock and Vibration*, vol. 2016, Article ID 9679542, 7 pages, 2016.
- [3] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. The Royal Society*, vol. 454, no. 1971, pp. 903–995, 1998.
- [4] T. Han, D. Jiang, and N. Wang, "The fault feature extraction of rolling bearing based on EMD and difference spectrum of singular value," *Shock and Vibration*, vol. 2016, Article ID 5957179, 14 pages, 2016.
- [5] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *Proceedings of the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '11)*, pp. 4144–4147, IEEE, Prague, Czech Republic, May 2011.
- [6] L. Zhao, W. Yu, and R. Yan, "Gearbox fault diagnosis using complementary ensemble empirical mode decomposition and permutation entropy," *Shock and Vibration*, vol. 2016, Article ID 3891429, 8 pages, 2016.
- [7] J. Han and M. van der Baan, "Empirical mode decomposition for seismic time-frequency analysis," *Geophysics*, vol. 78, no. 2, pp. O9–O19, 2013.
- [8] E. Mayoraz and E. Alpaydin, "Support vector machines for multi-class classification," in *Engineering Applications of Bio-Inspired Artificial Neural Networks*, pp. 833–842, Springer, Berlin, Germany, 1999.
- [9] J. Yang, Y. Zhang, and Y. Zhu, "Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension," *Mechanical Systems and Signal Processing*, vol. 21, no. 5, pp. 2012–2024, 2007.
- [10] M. A. Colominas, G. Schlotthauer, and M. E. Torres, "Improved complete ensemble EMD: a suitable tool for biomedical signal processing," *Biomedical Signal Processing and Control*, vol. 14, no. 1, pp. 19–29, 2014.
- [11] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.
- [12] J. Li, C. Liu, Z. Zeng, and L. Chen, "GPR signal denoising and target extraction with the CEEMD method," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 8, pp. 1615–1619, 2015.
- [13] S. H. Nawab, T. F. Quatieri, and J. S. Lim, "Signal reconstruction from short-time fourier transform magnitude," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 31, no. 4, pp. 986–998, 1983.
- [14] H. K. Kwok and D. L. Jones, "Improved instantaneous frequency estimation using an adaptive short-time Fourier transform," *IEEE Transactions on Signal Processing*, vol. 48, no. 10, pp. 2964–2972, 2000.
- [15] L. Mingliang, W. Keqi, S. Laijun, and Z. Jianju, "Applying empirical mode decomposition (EMD) and entropy to diagnose circuit breaker faults," *Optik—International Journal for Light and Electron Optics*, vol. 126, no. 20, pp. 2338–2342, 2015.
- [16] X. Zhang, Y. Liang, J. Zhou, and Y. Zang, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164–179, 2015.
- [17] D. Yu, Y. Yang, and J. Cheng, "Application of time-frequency entropy method based on Hilbert–Huang transform to gear fault diagnosis," *Measurement*, vol. 40, no. 9–10, pp. 823–830, 2007.
- [18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [19] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.
- [20] Z. Cai, Q. Ding, and M. Wang, "Study on automated incident detection algorithms based on PCA and SVM for freeway," in *Proceedings of the International Conference on Optoelectronics and Image Processing (ICOIP '10)*, Haikou, China, November 2010.
- [21] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1–3, pp. 37–52, 1987.

Research Article

An Approach to Fault Diagnosis for Gearbox Based on Image Processing

Yang Wang^{1,2} and Yujie Cheng^{1,2}

¹*School of Reliability and Systems Engineering, Beihang University, Beijing, China*

²*Science & Technology on Reliability and Environmental Engineering Laboratory, Beijing, China*

Correspondence should be addressed to Yujie Cheng; chengyujie@buaa.edu.cn

Received 2 June 2016; Accepted 3 August 2016

Academic Editor: Minvydas Ragulskis

Copyright © 2016 Y. Wang and Y. Cheng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The gearbox is one of the most important parts of mechanical equipment and plays a significant role in many industrial applications. A fault diagnostic of rotating machinery has attracted attention for its significance in preventing catastrophic accidents and beneficially guaranteeing sufficient maintenance. In recent years, fault diagnosis has developed in the direction of multidisciplinary integration. This work addresses a fault diagnosis method based on an image processing method for a gearbox, which overcomes the limitations of manual feature selection. Differing from the analysis method in a one-dimensional space, the computing method in the field of image processing in a 2-dimensional space is applied to accomplish autoextraction and fault diagnosis of a gearbox. The image-processing-based diagnostic flow consists of the following steps: first, the vibration signal after noise reduction by wavelet denoising and signal demodulation by Hilbert transform is transformed into an image by bispectrum analysis. Then, speeded up robustness feature (SURF) is applied to automatically extract the image feature points of the bispectrum contour map, and the feature dimension is reduced by principal component analysis (PCA). Finally, an extreme learning machine (ELM) is introduced to identify the fault types of the gearbox. From the experimental results, the proposed method appears to be able to accurately diagnose and identify different types of faults of the gearbox.

1. Introduction

In modern industrial production, a gearbox is one of the most important transmission components of mechanical equipment because of its large power transmission capacity in a compact structure. However, the gearbox is acted upon by the impulse load, environmental corrosion, and fluctuating circulating stress in the work that is performed, which leads to a high failure rate. As an effective component of condition-based maintenance [1], the fault diagnosis has become prominent to guarantee the safe operation of gearboxes.

Gearbox conditions can be reflected by measurements of vibratory [2], acoustic [3], thermal [4], electrical [5], and oil-based signals [6]. Vibration signals are extremely sensitive to the existence of a gearbox, and a fault diagnosis based on the vibration signal is one of the most widely used methods. Generally, gearboxes are operated within harsh environments with severe noises and interferences [7]. The signal analysis

and fault characteristic parameters are rather complicated in the fault diagnosis of the gearbox. Generally, the field staff do not have enough professional knowledge to recognize the spectrum of the vibration signal and do not understand the principle of time domain frequency domain analysis. Therefore, it is difficult to diagnose the fault of the gearbox effectively. If an intelligent diagnosis of the fault of the gearbox can be achieved without extensive professional knowledge and feature extraction can be performed automatically, then the performance and efficiency of the diagnostic method will be greatly improved. Currently, fault diagnosis methods based on multiple disciplines have become the leading direction of development in the field of fault diagnosis. For example, a genetic algorithm of natural evolution theory (genetic algorithm), machine learning (support vector for the machine) [8], and bionics (ant colony algorithm) [9] have been applied to fault diagnosis in rotating machinery.

However, the calculation method of image processing has scarcely been applied to fault diagnosis for gearboxes. Fault diagnosis for a gearbox is essentially a process of fault pattern recognition, which is the same as the image classification process, belonging to the category of pattern recognition. Because the application of image feature extraction technology has been successful, this technology has high feasibility and a probable reference value for introducing the calculation method of image processing to the field of fault diagnosis.

In this paper, we present a novel approach that applies image feature extraction techniques to gearbox fault diagnosis. The method can accomplish a fully automatic feature extraction procedure for gearboxes with high accuracy and strong robustness.

The image expression of a vibration signal is a key step in the proposed method. A high-order spectrum method is one of the modern signal processing methods developed in recent years, and the high-order spectrum method plays an important role in non-Gaussian, nonlinear, noncausal, nonminimum phase and nonstationary signal processing [10, 11]. The bispectrum [12–14] is a subset of the higher order spectrum. Bispectrum preserves the phase information of the signals and can theoretically restrain the Gaussian noise. To avoid the interference of the working noise, a wavelet transform is applied to reduce the noise of the original vibration signal. Simultaneously, there are modulation components in the vibration signal spectrum, which cause difficulties in the identification of the characteristic frequency of the fault. It is therefore essential to demodulate the signal. This paper proposes an image conversion method based on bispectrum and Hilbert transform, and the images generated are used as inputs for feature extraction techniques.

In the development of the image automatic feature extraction technique in recent decades, the scale invariant feature transform (SIFT) method has become one of the most widespread image processing methods, with good robustness and high accuracy [15, 16]. However, the shortcomings of high resource consumption, high time complexity, and large computational time requirements constitute major limitations of SIFT. In 2006, Bay et al. proposed speeded up robust features (SURF) [17], which not only maintain the advantages of the high accuracy of the scale invariant feature transform (SIFT) algorithm but also overcome the shortcomings of its slow speed. Herein, SURF is employed to extract the feature points of the bispectrum contour map. The feature points extracted by SURF are described in terms of the 64-dimensional descriptor. To avoid the waste of resources for subsequent calculations, PCA is employed to reduce the dimension of the feature points.

The extreme learning machine (ELM), as an intelligent technology, has shown good performance in regression applications as well as in large datasets and multilabel classification applications [18]. Moreover, ELM has been proven to require less human intervention and less running time than most other pattern recognition methods. In this study, ELM was introduced to accomplish the state classification of the gearbox.

This paper is organized as follows: Section 2 introduces the related algorithms, Section 3 describes the case study

performed to validate the method, and Section 4 presents the conclusions and related future work.

2. Methodology

The procedure contains three major steps as follows: (1) image transformation of the vibration signal based on the bispectrum and Hilbert transform, (2) feature point extraction based on SURF-PCA, and (3) fault diagnosis based on ELM as illustrated in Figure 1.

2.1. Bispectrum Analysis Based on the Wavelet Transform Domain. Wavelet transform can be utilized to analyze a non-stationary signal with obvious time-frequency localization and multiresolution analysis ability, which can effectively enhance the transient information hidden in the mechanical signal. A bispectrum analysis is a powerful tool to analyze a non-Gaussian signal. A bispectrum analysis can characterize random signals from a higher probability structure, and, theoretically, it can suppress Gaussian noise completely. In a high-order spectrum, a bispectrum analysis has the minimal order but with all of the characteristics of a high-order spectrum. The analysis procedure of the bispectrum analysis based on the wavelet transform domain is shown in Figure 2.

In Figure 2, $n(t)$ is the interfering noise, which can be a Gaussian or a non-Gaussian noise. When $n(t)$ is a Gaussian noise, we can further process the signal without wavelet denoising. However, when $n(t)$ is a non-Gaussian noise, a bispectrum analysis is powerless to analyze the signal, and spectrum characteristics may be covered by the non-Gaussian noise. To eliminate non-Gaussian noise disturbance, the wavelet transform is employed to remove the noise. Some important parts of the gearbox, such as the gear and bearing, are typical rotating components. When the centralized or distributed fault occurred on these components, the vibration signal always shows strong nonstationary characteristics. However, these nonstationary vibration signals often show a strong characteristic of modulation. Therefore, it is necessary to demodulate the signal before the bispectrum analysis. The Hilbert transform is one of the most widely used demodulation methods [19]. Zero mean normalization is applied to the signal, denoising the signal, and then Hilbert transform is implemented on the signal. Finally, a bispectrum analysis is utilized to accomplish the conversion between the analytical signal and the image. The estimated method for a bispectrum analysis includes a parametric model and a nonparametric model. Compared with the indirect method, less computation is required by the direct method. Therefore, the direct estimation method is adopted in this paper. The flow of the algorithm is described as follows:

(1) By assuming that the observation data are finite in length, the sampling frequency is f_s . In the bispectrum domain, the number of the points is ω_1 and ω_2 ; so, the frequency sampling interval is $\Delta_0 = f_s/N_0$. Dividing $x(t)$ into the k th segment, each segment contains M points; that is, $N = KM$; then, subtract the mean of each sample.

(2) The discrete Fourier transform (DFT) is undertaken for the j th data segment; that is,

$$X^j(\lambda) = \frac{1}{M} \sum_{i=1}^M x^j(i) \exp\left(-j \frac{2\pi}{M} i\lambda\right). \quad (1)$$

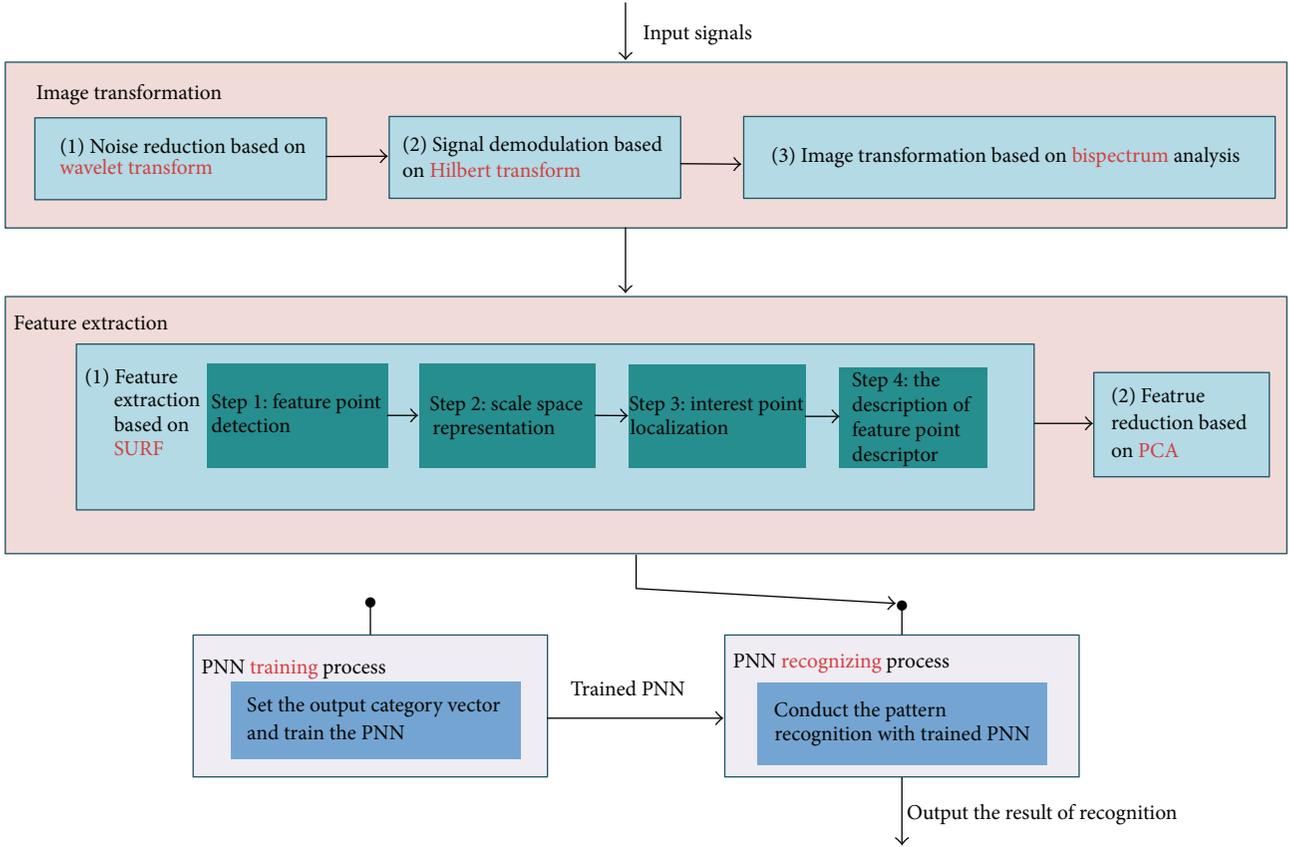


FIGURE 1: Scheme for fault diagnosis.

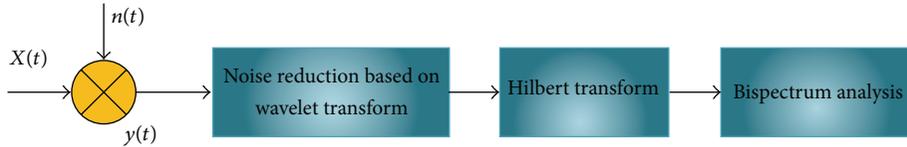


FIGURE 2: Bispectrum analysis process in wavelet transform.

(3) According to the coefficient of the DFT, calculate the bispectrum estimation of each segment:

$$B_x^j(\lambda_1, \lambda_2) = \frac{1}{\Delta_0^2} \sum_{k_1=-L_1}^L \sum_{k_2=-L_1}^L X^j(\lambda_1 + k_1) X^j(\lambda_2 + k_2) \cdot X^j(\lambda_1 + k_1 + \lambda_2 + k_2). \quad (2)$$

(4) The bispectrum estimation is the mean value of K segment data; that is,

$$B_x(\omega_1, \omega_2) = \frac{1}{K} \sum_{j=1}^k B_x^j(\omega_1, \omega_2), \quad (3)$$

where $\omega_1 = [2\pi f_s/N_0]\lambda_1$ and $\omega_2 = [2\pi f_s/N_0]\lambda_2$.

2.2. SURF Descriptor

2.2.1. Interest Point Detection. A Hessian matrix is utilized to detect interest points by SURF, and the use of an integral image can greatly reduce the amount of calculation.

(1) **Integral Image.** The integral image can greatly increase the efficiency of box-type convolution filters. For a point X , $X = (x, y)$ in the integral image. The value of the point can be described as follows: the sum of all pixel values in a region formed by the origin point and point X is as follows:

$$I_\Sigma(X) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j). \quad (4)$$

Once an image is converted to an integral image, the sum of the gray levels in a rectangular region can be

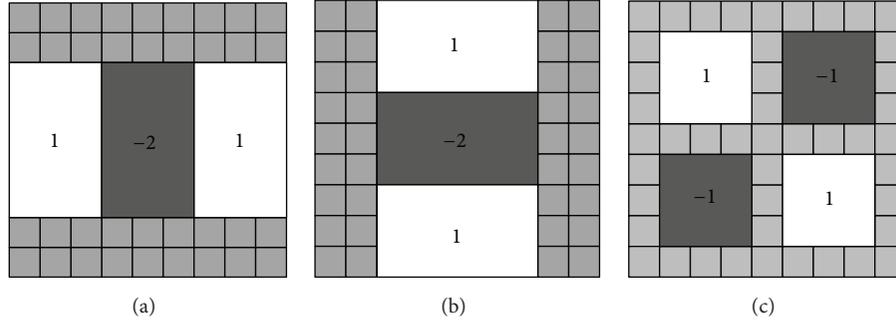


FIGURE 3: (a) x direction, (b) y direction, and (c) xy direction.

calculated by 3 plus-minus calculations. We can see in the following discussion that the convolution templates used by SURF are all box templates, which can greatly improve the computational efficiency.

(2) *Approximate Hessian Matrix: H_{approx} .* The interest point detection of SURF is based on the Hessian matrix, which relies on the local maximum of the Hessian matrix determinant. When the Hessian matrix determinant appears to have a local maximum, the detection result is a blob-like structure, that is, a region brighter or darker than the surrounding area. Here, the register refers to the point and its small neighborhood. For a point $X = (x, y)$, the Hessian matrix is defined as

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix}, \quad (5)$$

where $L_{xx}(x, \sigma)$, $L_{xy}(x, \sigma)$, and $L_{yy}(x, \sigma)$ are second-order partial derivatives and the two-dimension convolution of point (x, y) of the image I , respectively.

Instead of using a second-order Gaussian filter, a block filter is employed to approximate the two-order partial derivative of the Gaussian function to construct a fast-Hessian matrix. The approximate convolution template is utilized to process the integral image, with obvious benefits. The template is composed of simple rectangles, with computation independent of the size of the template. A 9×9 box filter is shown in Figure 3. In the original image, an image pyramid of a different size is formed by expanding the size of the box filter, and an integral image is applied to speed up the image convolution. Therefore, the approximate Hessian matrix can be obtained as

$$|H| = D_{xy}(X)D_{xy}(X) - [0.9D_{xy}(X)]^2, \quad (6)$$

where D_{xx} , D_{yy} , and D_{xy} are the computational results of the box filter and the integral image. The extreme points are determined by the determinant and the eigenvalue of the matrix. If the determinant of H_{approx} matrix is positive and the two eigenvalues are different, the point is denoted as an extreme point.

According to the fast-Hessian matrix, the extremum of (x, y, σ) in the scale image can be obtained. First, nonmaximum suppression is carried out in the $3 \times 3 \times 3$ stereo neighborhood. Only the biggest of all points smaller than the

adjacent 26 points can be selected as the interest point. To locate (subpixel location) the candidate feature points, interpolation can be applied in a scale space and image space, and stable feature points and the scale value can be obtained [20].

2.2.2. *Interest Point Location.* To keep the characteristics of the rotation invariance of the interest point descriptor, the direction of the interest point should be determined first. Next, construct a Harr wavelet response in the direction of x -axis and y -axis, which is centered on an interest point, covering a 6σ radius. Give this response a different Gaussian weight coefficient; the closer the distance between feature points is, the larger the weight coefficient is. The sum of the Harr wavelet response is in the direction of x -axis and y -axis within 60° to form a local direction vector. Traverse the entire circular area, and the direction of the longest vector is selected as the principal direction of the interest point as shown in Figure 4.

Construct a window region with the center as a feature point and a side length of 20σ . Then, divide the window region into 4×4 subregions. The 5×5 sampling point is obtained from the subregion. Seek the wavelet response in the direction of x -axis and y -axis of each sampling point, designated as dx and dy , respectively. The Gaussian filter is performed as dx and dy in each subregion. The filter center is seen as the feature point, and a four-dimensional eigenvector $(\sum dx, \sum dy, \sum |dx|, \sum |dy|)$ is formed by summing dx , dy , $|dx|$, and $|dy|$ in each subregion. The four-dimensional eigenvector makes the 4 dimensions of the descriptor. Each descriptor consists of 4 dimensions. Thus, the $4 \times 4 \times 4 = 64$ dimensions are obtained, which are the SURF descriptor.

2.3. *Extreme Learning Machine (ELM).* ELM, proposed by Huang et al., was originally developed for single-hidden-layer feedforward neural networks and then extended to “generalized” single-hidden-layer feedforward networks (SLFNs). ELM is a novel learning algorithm with a faster learning speed and better generalization performance [21, 22]. The details of the ELM algorithm can be found in [18, 23–25].

The architecture of an SLFN can be described as shown in Figure 5, where x_i is the input sample, $\omega_i = [\omega_{1i}, \omega_{2i}, \dots, \omega_{ni}]$ represents the vector of link weights between all nodes in the input layer, $g(x)$ represents the activation function of neurons in the hidden layer, b_i represents the threshold of neurons in

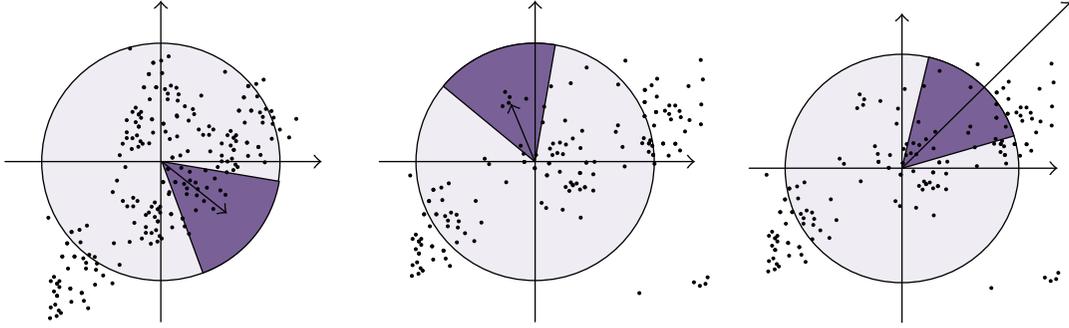


FIGURE 4: Determine the main direction of the feature point.

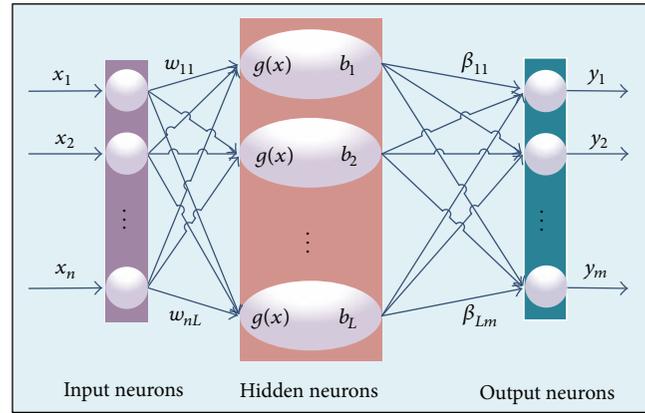


FIGURE 5: Architecture of an SLFN.

the hidden layer, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ represents the vector of link weights between the i th node in the hidden layer and all nodes in the output layer, y_j represents the outputs of the network, n represents the number of nodes in the input layer, L represents the number of nodes in the hidden layer, m represents the number of nodes in the output layer, and $j = 1, 2, 3, \dots, m$. Then, the mathematical model of ELM can be described as

$$y_i = \sum_{j=1}^L \beta_j g_i(x_j) = \sum_{j=1}^L \beta_j g(\omega_j \cdot x_j + b_j). \quad (7)$$

Compared with SVM, whose kernel function greatly affects the results [21], ELM is not as sensitive to the activation function $g(x)$ as SVM, and almost all nonlinear piecewise continuous functions satisfying the ELM universal approximation capability theorems can be selected as the activation function [26], such as the sigmoid function, hard-limit function, or multiquadric function. The sigmoid function is the major activation function applied to the feedforward neural networks, and ELM with the hard-limit and multiquadric functions also shows good performance [23]. Thus, in this study, $g(x)$ is selected as a sigmoid function, and the process of ELM can be concluded as follows:

- (1) Determine the number, L , of neurons in the hidden layer and the activation function $g(x)$ and randomly assign ω_i , b_i , and β_i .

- (2) Calculate the output vector of the hidden layer.

- (3) Calculate the output weight $\hat{\beta}$.

3. Case Study

3.1. Experimental Facilities. The experimental data were obtained from the 2009 PHM data challenge competition. Data were collected from a two-class standard cylinder spur gear reducer. The reducer contains an input shaft, idler shaft, and an output shaft. The first stage reduction gear ratio is 1.5, and the second stage reduction gear ratio is 1.667. There are 32 teeth in the input shaft and 80 teeth in the output shaft. The two gears on the idler shaft both have 48 teeth. Figure 6 is a schematic of the gearbox used to collect the data and Figure 7 is the physical map of two-stage reducer.

The data were acquired using input shaft speeds of 35 Hz, 45 Hz, and 50 Hz. The sampling frequency is 66.7 Hz, and the sampling time is set to 66.7 kHz. The fault was detected as shown in Table 1.

3.2. Feature Extraction Based on Bispectrum and SURF. In this section, bispectrum and SURF were utilized to extract the feature vector, which is employed to process the two-class gear reducer signal. Considering the influence of environmental noise, wavelet denoising was performed on the gearbox signal, and then the signal was demodulated by Hilbert transform to obtain the analytic signal. Then,

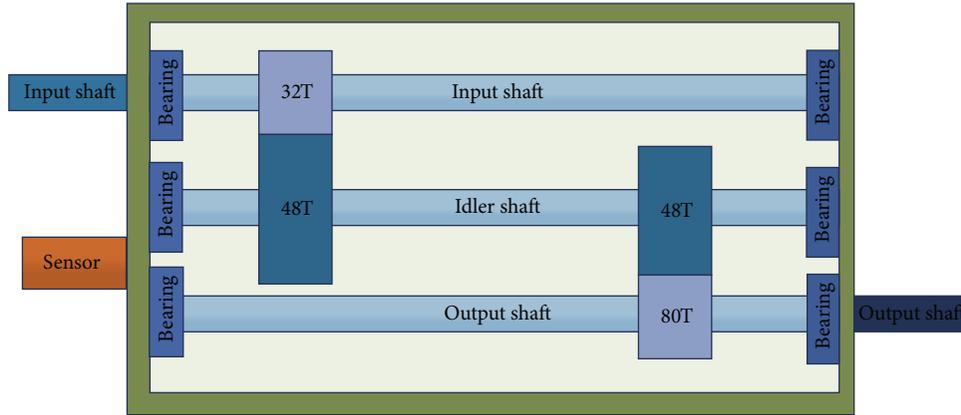


FIGURE 6: Schematic of the apparatus.

TABLE 1: Fault information of gearbox.

		Normal	Fault 1	Fault 2	Fault 3	Fault 4
Gear	32T	Good	Chipped	Good	Good	Good
	96T	Good	Good	Good	Good	Good
	48T	Good	Eccentric	Eccentric	Good	Good
	80T	Good	Good	Good	Broken	Good
Bearing	IS:IS	Good	Good	Good	Inner	Inner
	ID:IS	Good	Good	Good	Ball	Good
	OS:IS	Good	Good	Good	Outer	Good
	IS:OS	Good	Good	Good	Good	Good
	ID:OS	Good	Good	Good	Good	Good
	OS:OS	Good	Good	Good	Good	Good
Shaft	Input	Good	Good	Good	Imbalance	Good
	Output	Good	Good	Good	Good	Keyway sheared

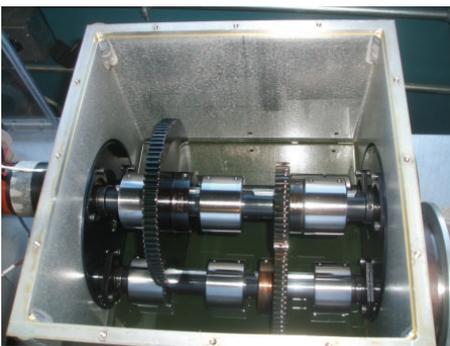


FIGURE 7: Physical map of two-stage reducer.

bispectrum was employed to analyze the analytic signal. Herein, the direct method was adopted, and the bispectrum contour map of different fault modes is shown below.

Four fault modes were collected in this section, namely, fault 1, fault 2, fault 3, and normal condition. Specific information is shown in Table 1. The length of the data is 5000 in each group. Figures 8–12 are bispectrum contour maps of faults

1–4, respectively. Three datasets have been selected for the comparison. From the contour map, we can see that different failure modes have different frequency distributions, which can reflect the differences between the different fault modes.

In the experiment, the SURF descriptor extracted feature points with 64 dimensions. Then, principal component analysis (PCA) was utilized to reduce the dimensions to 20 dimensions. The first three principal components are presented in a three-dimensional space as shown in Figure 13, from which we can conclude that the features separately show a strong ability. However, to identify the failure modes accurately, the feature must be combined with a classifier.

(3) *Fault Diagnosis Results Based on ELM.* The input eigenvector of ELM was extracted by bispectrum and SURF. For each fault mode, 240 sets of data were collected. The 80 sets were selected randomly for training the extreme learning machine. The other 160 sets were selected as test data. Usually, fault diagnosis accuracy is selected as the metric of the effectiveness of fault diagnosis method, as in [7, 27, 28]. In this study, fault diagnosis accuracy is also selected to verify the proposed method. To display the results visually, the

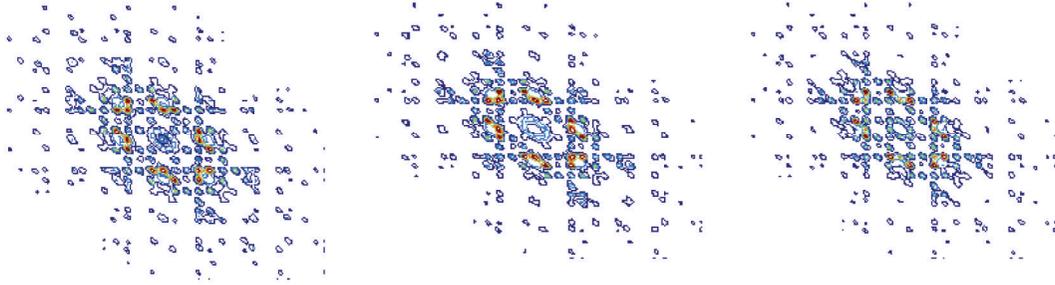


FIGURE 8: Bispectrum contour map of normal condition.

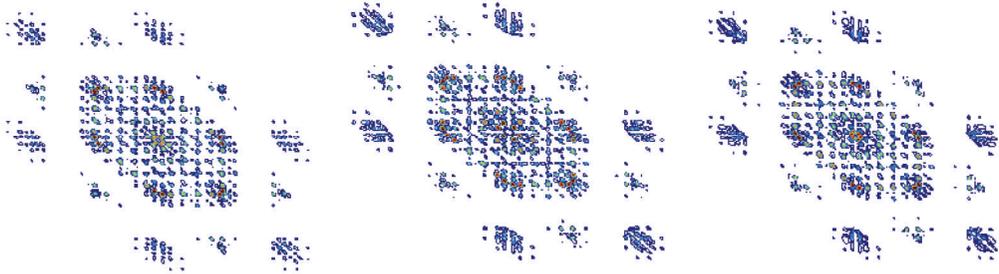


FIGURE 9: Bispectrum contour map of fault 1.

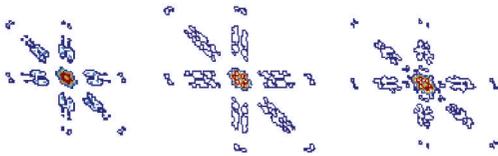


FIGURE 10: Bispectrum contour map of fault 2.

classification result was contrasted with the actual result. The accuracy rate is defined as the odds ratio of the right result and the total result. The diagnosis result is shown in Figure 14(a).

For comparison, the proposed method is compared with traditional diagnosis method based on empirical mode decomposition (EMD). EMD, as an effective decomposition method for nonlinear and nonstationary signals, has been applied to fault diagnosis of gearbox. However, the problem of mode mixing which existed in the EMD makes Figures 14(a) and 14(b) unable to fully decompose a signal with intermittent components, which will affect the accuracy of fault diagnosis. Herein, EMD is used to decompose the origin signal into several intrinsic mode functions (IMFs) and the singular value is used as feature vector. Then put them into the ELM classifier. The fault diagnosis result is shown in Figure 14(b).

In Figure 14 the horizontal ordinate represents the normal condition and 4 fault conditions, while the vertical ordinate represents the numbers of correct diagnosis classifications. The diagnosis accuracy of each fault mode using both diagnosis method based on image processing and traditional EMD is listed in Table 2 in detail.

From Table 2, we can see that, in the proposed fault diagnosis method based on image processing, the diagnosis accuracy of normal condition and 4 fault conditions reaches 96.25%, 98.75%, 97.5%, 87.50%, and 97.5%, respectively, among which the highest diagnosis accuracy reaches 98.75%, while the lowest is 87.5%. The average diagnosis accuracy of the five conditions is 95.5%, indicating the high effectiveness of the proposed method based on image processing.

The diagnosis accuracy using traditional method based on EMD under each fault mode is 90.625%, 92.5%, 91.25%, 79.375%, and 91.25%, respectively. The highest accuracy, 92.5%, is 6.25% lower than the proposed diagnosis method based on image processing. While the lowest accuracy, 79.375%, is 8.125% lower than the proposed method. The average accuracy of the five fault modes is 89%, 6.5% lower than the proposed method.

The diagnosis accuracy results are listed in Figure 15 in histogram for a clearer comparison. From Figure 15, it can be seen that, for each fault mode, the fault diagnosis based on bispectrum and SURF has higher accuracy, demonstrating the superiority of the proposed method.

4. Conclusions

In this paper, a novel fault diagnosis method for a gearbox based on the image processing technology is first proposed. The method contains the following three main steps. (1) The first step is image generation process: a bispectrum analysis is applied to transform the vibration after wavelet denoising and Hilbert demodulation. (2) The second step is feature

TABLE 2: Fault diagnosis result.

Fault types	Testing samples	Correct diagnosis results		Accuracy	
		Bispectrum and SURF	EMD	Bispectrum and SURF	EMD
Normal	160	154	145	96.25%	90.625%
Fault 1	160	158	148	98.75%	92.50%
Fault 2	160	156	146	97.50%	91.25%
Fault 3	160	140	127	87.50%	79.375%
Fault 4	160	156	146	97.50%	91.25%
Total	800	764	712	95.5%	89.00%

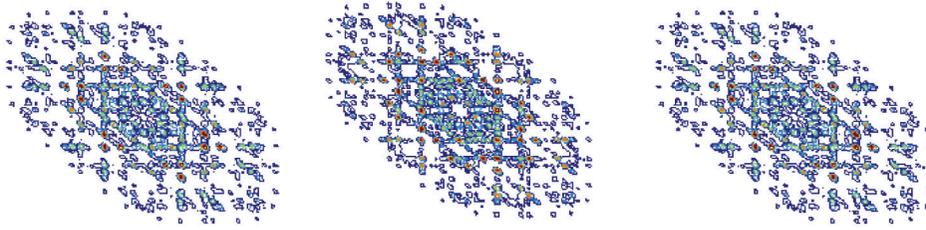


FIGURE 11: Bispectrum contour map of fault 3.

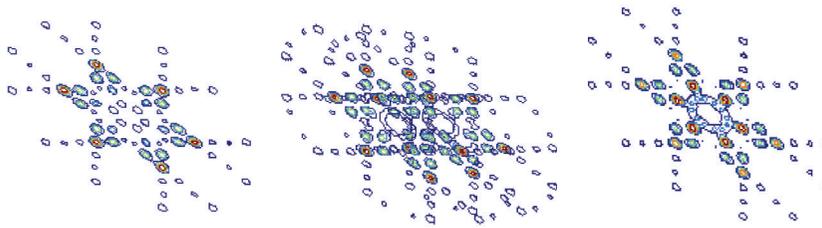


FIGURE 12: Bispectrum contour map of fault 4.

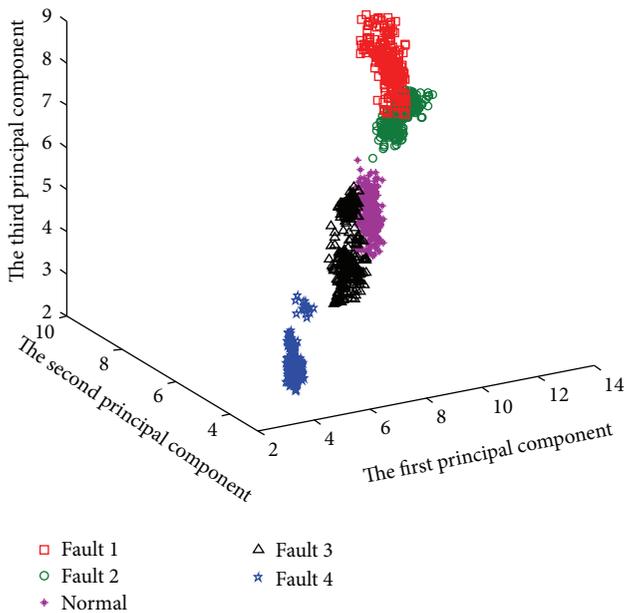


FIGURE 13: The first three features extracted using PCA.

extraction process: SURF is introduced to extract the feature points of the bispectrum contour map automatically. (3) The third step is fault pattern recognition process: based on the features extracted by SURF, ELM is then used to identify the fault types of the gearbox. Differing from the traditional fault diagnosis method of vibration analysis in a one-dimensional space, the method applying the computing method to a field of image processing in a 2-dimensional space accomplished autoextraction and fault diagnosis of the gearbox. Almost no manual intervention was needed for the whole diagnosis process, which avoids the limitations of other methods that require a large amount of expert knowledge. Simultaneously, the results of the case study fully demonstrated the effectiveness and high diagnostic accuracy of the image-processing-based method for the gearbox.

Our subsequent work will be focused on the following:

- (1) In view of the complex structure of the gearbox, more fault modes should be taken into account. Therefore, we should apply this method to more fault types of the gearbox.

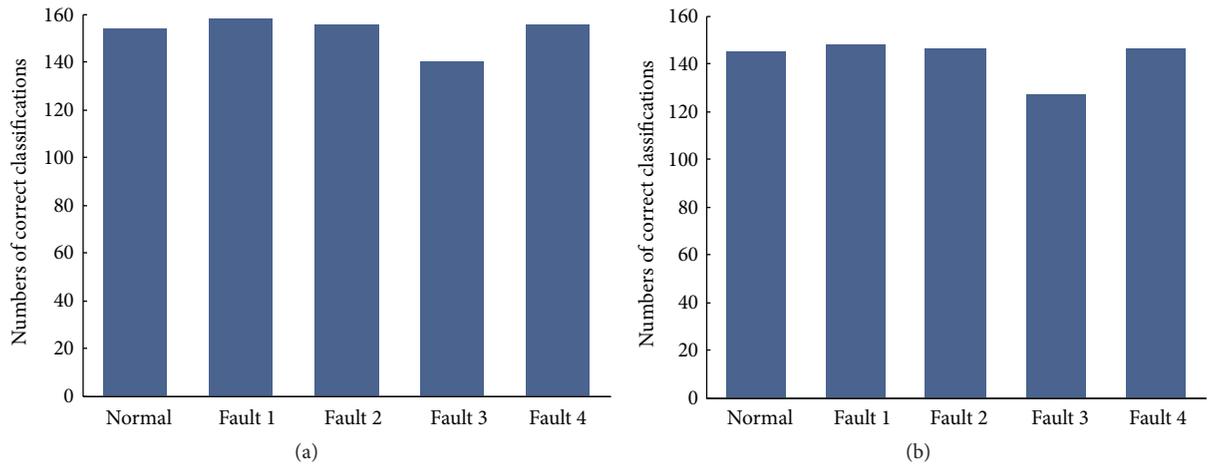


FIGURE 14: (a) Fault diagnosis result of ELM based on EURF. (b) Fault diagnosis result of ELM based on EMD.

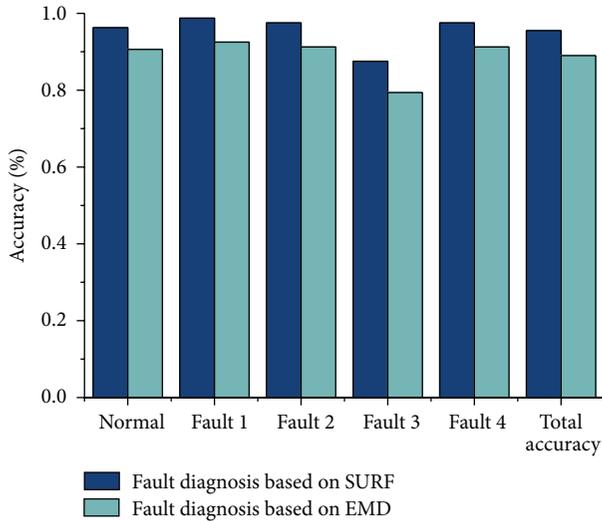


FIGURE 15: Result comparison of EMD and SURF.

- (2) We should improve the computing speed while maintaining diagnostic accuracy.

Competing Interests

The authors declare that there are no potential competing interests in this research.

Acknowledgments

This study was supported by the Fundamental Research Funds for the Central Universities (Grant no. YWF-16-BJ-J-18), the National Natural Science Foundation of China (Grant no. 51575021), and the Technology Foundation Program of National Defense (Grant no. Z132013B002).

References

- [1] D. Galar, A. Thaduri, M. Catelani, and L. Ciani, "Context awareness for maintenance decision making: a diagnosis and prognosis approach," *Measurement*, vol. 67, pp. 137–150, 2015.
- [2] Y. Lei, N. Li, J. Lin, and Z. He, "Two new features for condition monitoring and fault diagnosis of planetary gearboxes," *JVC/Journal of Vibration and Control*, vol. 21, no. 4, pp. 755–764, 2015.
- [3] J. Hou, W. Jiang, and W. Lu, "Application of a near-field acoustic holography-based diagnosis technique in gearbox fault diagnosis," *Journal of Vibration and Control*, vol. 19, no. 1, pp. 3–13, 2013.
- [4] A. M. D. Younus and B.-S. Yang, "Intelligent fault diagnosis of rotating machinery using infrared thermal image," *Expert Systems with Applications*, vol. 39, no. 2, pp. 2082–2091, 2012.
- [5] J. R. Ottewill and M. Orkisz, "Condition monitoring of gearboxes using synchronously averaged electric motor signals," *Mechanical Systems and Signal Processing*, vol. 38, no. 2, pp. 482–498, 2013.
- [6] C. Li and M. Liang, "Extraction of oil debris signature using integral enhanced empirical mode decomposition and correlated reconstruction," *Measurement Science and Technology*, vol. 22, no. 8, Article ID 085701, 2011.
- [7] C. Li, R. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez, "Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals," *Mechanical Systems and Signal Processing*, vol. 76–77, pp. 283–293, 2016.
- [8] L. Lu, J. Yan, and C. W. De Silva, "Dominant feature selection for the fault diagnosis of rotary machines using modified genetic algorithm and empirical mode decomposition," *Journal of Sound and Vibration*, vol. 344, pp. 464–483, 2015.
- [9] X. Zhang, W. Chen, B. Wang, and X. Chen, "Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization," *Neurocomputing*, vol. 167, pp. 260–279, 2015.
- [10] K. M. Hock, "Narrowband weak signal detection by higher order spectrum," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 874–879, 1996.

- [11] X. Zhang, *Time Series Analysis-Higher Order Statistics Method*, vol. 1, Tsinghua University Press, Beijing, China, 1996.
- [12] J. Treerong, J. K. Sinha, F. Gu, and A. Ball, "Bispectrum of stator phase current for fault detection of induction motor," *ISA Transactions*, vol. 48, no. 3, pp. 378–382, 2009.
- [13] F. E. H. Montero and O. C. Medina, "The application of bispectrum on diagnosis of rolling element bearings: a theoretical approach," *Mechanical Systems and Signal Processing*, vol. 22, no. 3, pp. 588–596, 2008.
- [14] L. Saidi, F. Fnaiech, H. Henao, G.-A. Capolino, and G. Cirrincione, "Diagnosis of broken-bars fault in induction machines using higher order spectral analysis," *ISA Transactions*, vol. 52, no. 1, pp. 140–148, 2013.
- [15] S. Yue, Q. Wang, and R. Zhao, "Robust wide baseline feature point matching based on scale invariant feature descriptor," in *Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues*, pp. 329–336, Springer, 2008.
- [16] Y. Wang, X. Ban, J. Chen, B. Hu, and X. Yang, "License plate recognition based on SIFT feature," *Optik—International Journal for Light and Electron Optics*, vol. 126, no. 21, pp. 2895–2901, 2015.
- [17] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: speeded up robust features," in *Computer Vision—ECCV 2006*, A. Leonardis, H. Bischof, and A. Pinz, Eds., vol. 3951 of *Lecture Notes in Computer Science*, pp. 404–417, Springer, Berlin, Germany, 2006.
- [18] G.-B. Huang, X. Ding, and H. Zhou, "Optimization method based extreme learning machine for classification," *Neurocomputing*, vol. 74, no. 1–3, pp. 155–163, 2010.
- [19] W. Teng, X. Ding, X. Zhang, Y. Liu, and Z. Ma, "Multi-fault detection and failure analysis of wind turbine gearbox using complex wavelet transform," *Renewable Energy*, vol. 93, pp. 591–598, 2016.
- [20] J. Li, Y. Wang, and Y. Wang, "Visual tracking and learning using speeded up robust features," *Pattern Recognition Letters*, vol. 33, no. 16, pp. 2094–2101, 2012.
- [21] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, vol. 2, pp. 985–990, July 2004.
- [22] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.
- [23] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 2, pp. 513–529, 2012.
- [24] G.-B. Huang, "An insight into extreme learning machines: random neurons, random features and kernels," *Cognitive Computation*, vol. 6, no. 3, pp. 376–390, 2014.
- [25] V. Malathi, N. S. Marimuthu, S. Baskar, and K. Ramar, "Application of extreme learning machine for series compensated transmission line protection," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 5, pp. 880–887, 2011.
- [26] G.-B. Huang, L. Chen, and C.-K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 879–892, 2006.
- [27] X. Jiang, S. Li, and Y. Wang, "A novel method for self-adaptive feature extraction using scaling crossover characteristics of signals and combining with LS-SVM for multi-fault diagnosis of gearbox," *Journal of Vibroengineering*, vol. 17, no. 4, pp. 1861–1878, 2015.
- [28] L. Zhao, W. Yu, and R. Yan, "Gearbox fault diagnosis using complementary ensemble empirical mode decomposition and permutation entropy," *Shock and Vibration*, vol. 2016, Article ID 3891429, 8 pages, 2016.

Research Article

Intelligent Analysis Method of Gear Faults Based on FRWT and SVM

Hongfang Chen, Yanqiang Sun, Zhaoyao Shi, and Jiachun Lin

Beijing Engineering Research Center of Precision Measurement Technology and Instruments, Beijing University of Technology, Pingleyuan No. 100, Chaoyang District, Beijing, China

Correspondence should be addressed to Hongfang Chen; chf0302@126.com

Received 5 April 2016; Revised 3 August 2016; Accepted 4 August 2016

Academic Editor: Minvydas Ragulskis

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

An intelligent analysis method for gear faults based on fractional wavelet transform (FRWT) and support vector machine (SVM) is proposed. Based on this method, FRWT is used to eliminate noise from the gear vibration signal, and the vibration signal after noise elimination is carried through wavelet packet decomposition and reconstruction. A sequence corresponding to the signal is constructed consisting of the module with the highest level wavelet coefficients after decomposition and feature vectors corresponding to the energy sequence which were obtained by calculation. Then, a particle optimization method is used to optimize SVM parameters, and the feature vectors as training samples are input into SVM for training while the test samples are input for fault recognition. Experimental results show that the gear fault analysis method proposed in this paper is able to effectively extract the weak fault signal. The accuracy rate for identification of the type of gear fault reached 96.7%.

1. Introduction

As an indispensable component in modern mechanical structure to transmit power and motion, the gear is widely used in various equipment due to its large load capacity, high transmission accuracy, and constant transmission power. As power-transmission equipment develops toward larger scale, greater complexity, automation, and continuous use, gear faults may induce increasingly large losses. Therefore, it is necessary to develop an intelligent gear fault analysis method that can improve the efficiency and accuracy of detection and reduce cost [1–3].

Gear faults are always found in the vibration signal produced by meshing gears. This vibration signal also contains a large number of nonstationary signal components. Using classical spectral analysis methods will result in averaged spectrum values across the whole time scale, which inadvertently conceals some details of mutations. In contrast, a time-frequency method can be used effectively for analysis of nonstationary signals, which can fill the gap in the standard FFT-based method and is developing rapidly as a new approach in the fault analysis field. Wavelet transform (WT) [4–6]

is considered as an ideal time-frequency method capable of providing localized information of time and frequency domains at the same time, which can help identify the mutations in the vibration signal. But wavelet transform has limitations, since WT of various scales is similar to a group of frequency-domain band-pass filters [7] which are limited in only time and frequency domain, which makes WT an imperfect choice for chirp type signals whose energy is not optimally gathered in the frequency domain and which exist widely in nature and in various artificial situations. Therefore, analysis methods like fractional Fourier transform (FRFT) [8–10], short-time fractional Fourier transform [11], Radon-Wigner transform [12], and fractional wavelet transform (FRWT) [13, 14] are proposed one after another. FRFT has some unique characteristics but cannot present localized features. Short-time fractional Fourier transform has the fault of low resolution. Radon-Wigner transform is based on quadratic transformations, which have cross terms. FRWT was originally proposed by Mendlovic in 1997, who applied FRFT to the signal to obtain a fractional spectrum, which was then transformed by traditional WT. As a generalized form of WT, FRWT merged many advantages of WT and FRFT

such as no cross terms, linear characteristics, multiresolution analysis, and fractional domain presentation. Thus, FRWT has attracted more attention as a potential application in signal processing field.

Taking the gear vibration signal as the researched object, two key problems are studied in the practical engineering application, including the gear fault feature extraction and the pattern recognition. An intelligent gear fault analysis method based on a combination of FRWT and SVM is proposed. FRWT makes use of the WT's multiresolution analysis theory and FRFT's time-frequency focusing properties to filter the gear vibration signal. So it can ensure the consistency and effectiveness of time-frequency resolution. Then wavelet packet transform was applied to the denoised signal extracting energy values in the frequency band to build feature vectors, which can be used for training and learning by SVM to further realize intelligent identification of the faulty gear. The SVM method has the advantage of identifying any sample according to the limited samples information. It overcomes the shortcoming of poor classification accuracy caused by insufficient samples in faults of detection of gears. So it improves the efficiency and accuracy of faults of detection of gears.

Combining FRWT and SVM, the method can quickly achieve the detection of finished gears. Then qualities and faults of gears are determined according to the detection results. All in all, this method will provide a powerful analysis means for the diagnosis of gears quality.

2. Fundamental Principles

2.1. Fractional Fourier Transform. As a generalized form of Fourier Transform, fractional Fourier transform (FRFT) merges signal information in both the time and frequency domains, which is suitable for time-varying nonstationary signals. The p order fractional Fourier transform of signal $x(t)$ is defined as

$$X_\alpha(u) = \int_{-\infty}^{+\infty} x(t) K_\alpha(t, u) dt. \quad (1)$$

Here, $K_\alpha(t, u)$ is the transforming kernel function:

$$\sqrt{\frac{1 - j \cot \alpha}{2\pi}} \exp\left(j \frac{t^2 + u^2}{2} \cot \alpha - tu \csc \alpha\right), \quad \alpha \neq n\pi \quad (2)$$

$$\delta(t - u), \quad \alpha = 2n\pi$$

$$\delta(t + u), \quad \alpha = (2n + 1)\pi.$$

Here, p is the order of FRFT, which is an arbitrary real number. α represents a rotation angle; $\alpha = p\pi/2$. n is an integer. $\delta(\cdot)$ is an impulse function. The inverse transformation of FRFT is

$$x(t) = \int_{-\infty}^{+\infty} X(u) K_{-\alpha}(t, u) du. \quad (3)$$

2.2. Fractional Wavelet Transform. Fractional wavelet transform combines wavelet transform and fractional theory, which promotes multiresolution wavelet analysis theory to a time domain and generalized frequency domain. Meanwhile, it merges characteristics of wavelet transform and fractional theory, which makes it more powerful in signal processing. Mendlovic first proposed the definition of fractional wavelet transform [13, 15] and applies it to signal $f(x)$:

$$W^{(p)}(a, b) = \iint_{-\infty}^{+\infty} B_p(x, x') f(x') h_{ab}^*(x) dx' dx. \quad (4)$$

Here, $h_{ab}^*(x)$ is wavelet basis function and $B_p(x, x')$ is kernel function and can be presented as follows:

$$B_p(x, x') = \sqrt{2} \exp[-\pi(x^2 + x'^2)] \cdot \sum_{n=0}^{\infty} \frac{i^{-pn}}{2^n n!} H_n(\sqrt{2\pi}x) H_n(\sqrt{2\pi}x'). \quad (5)$$

Here, H_n is n order Hermite polynomials, $\phi = \pi p/2$, and p is the order of fractional wavelet transform $W^{(p)}(a, b)$, in its value ranging from 0 to 1.

The reconstruction formula of fractional wavelet transform is

$$f(x) = \frac{1}{C} \iiint_{-\infty}^{+\infty} \frac{1}{a^3} W^{(p)}(a, b) B_{-p}(x, x') \times h\left(\frac{x' - b}{a}\right) da db dx'. \quad (6)$$

We can see that the combination and decomposition of fractional wavelet transform are actually the consistence transform in the time-frequency domain, which accurately regulates information within time domain and frequency domain.

2.3. Support Vector Machine. The advantage of support vector machine (SVM) [16–18] applied to gear vibration signal analysis is that it fits for a small sample of the decision and can maximally excavate the classification knowledge using limited feature information, which has a great practical significance for gear fault analysis. Therefore, we use the SVM in this study to establish the gear fault model by the identification and analysis of the gear fault type from the filtered signal.

The main idea of SVM is to create a hyperplane as a decision interface that has the ability to correctly classify all the training samples and maximize the distance from the nearest point in the training sample to the classification face. The geometry of the hyperplane in the linearly separable two-dimensional input pattern space is shown in Figure 1. Here, the pentacle and rectangles represent the two types of samples, P is the classification lines, and straight lines P_1 and P_2 are the nearest samples to the corresponding classification line and parallel to the classification lines. The distance between P_1 and P_2 is classification interval. The sample on P_1 and P_2 is the SVM.

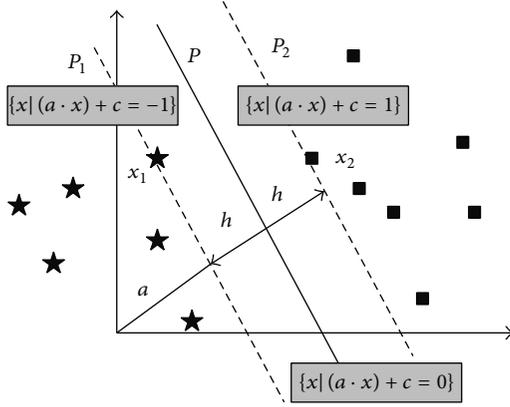


FIGURE 1: Hyperplane in the linearly separable pattern.

Assume that the training sample is $\{(x_i, y_i)\}_{i=1}^n$, in which input is $x_i \in R^d$ and output is $y_i \in \{-1, 1\}$. The hyperplane equation is $w^T x + b = 0$. w is the adjustable weight vector, x is weight function, and b is the offset. And the constraint condition of the optimization problem is

$$y_i (w^T x_i + b) \geq 1, \quad i = 1, \dots, n. \quad (7)$$

The unknown problem is

$$\min: \quad \Phi(w, b) = \frac{1}{2} \|w\|^2. \quad (8)$$

This formula is a mathematical inequality problem solved through constrained optimization. The point that makes the equation balance is called support vector. In order to solve the mathematical optimization problem, Lagrange multipliers are introduced to convert it into the dual problem to solve.

Namely, the constraint condition is

$$\sum_{i=1}^m \alpha_i y_i = 0, \quad (9)$$

$$\alpha_i \geq 0, \quad i = 1, 2, \dots, m. \quad (10)$$

And the unknown problem is

$$\max: \quad \Phi(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j \cdot y_i y_j x_i^T y_j. \quad (11)$$

To solve the optimal coefficients above \hat{a} , \hat{w} , \hat{b} , the optimal classification function is

$$f(x) = \text{sgn}(\hat{w}^T x + \hat{b}) = \text{sgn}\left(\sum_{i=1}^n \hat{\alpha}_i y_i x_i^T + \hat{b}\right). \quad (12)$$

The certain nonlinear mapping can map the sample space into a high-dimensional feature space to convert the liner nonseparable pattern which becomes linearly separable and to construct the optimal classification hyperplane in the high-dimensional feature space. The nonlinear mapping function is called a kernel function. Therefore, the choice

of appropriate inner product kernel function can map the low-dimensional to high-dimensional space to achieve linear classification without increasing computational complexity. Compared with linearly separable patterns, the constraints (10) become (parameter c which is used to control the degree of penalty to the wrong sample points)

$$0 \leq \alpha_i \leq c, \quad i = 1, \dots, n. \quad (13)$$

The optimal classification function becomes

$$f(x) = \text{sgn}(\hat{w}^T x + \hat{b}) = \text{sgn}\left(\sum_{i=1}^n \hat{\alpha}_i y_i K(x_i, x) + \hat{b}\right). \quad (14)$$

The Gauss radial primary kernel function $K(x_i, x_j) = \exp(-g\|x_i - x_j\|^2)$ is used in this paper, where g is a constant. As the particle swarm optimization method is faster and more efficient than the grid search method [19], this particle swarm optimization method is used in this paper to optimize the SVM classification model.

The mathematical description of a particle swarm optimization algorithm is as follows: assuming that in a D dimensional optimization space there are a group of particles m , where the speed for the i th particle is $v_i = (v_{i1}, v_{i2}, \dots, v_{iD})$, whose location is $x_i = (x_{i1}, x_{i2}, \dots, x_{iD})$, and then the optimal position for the i th particle is $p_i = (p_{i1}, p_{i2}, \dots, p_{iD})$, and the optimal position of the whole population [20] is (the particle update formula is as follows)

$$v_{id}(t+1) = \omega v_{id}(t) + c_1 r_1 (p_{id}(t) - x_{id}(t)) + c_2 r_2 (p_{gd}(t) - x_{id}(t)), \quad (15)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1).$$

If $v_{id} > V_{\max}$, $v_{id} = V_{\max}$; if $v_{id} < -V_{\max}$, $v_{id} = -V_{\max}$, where $i = 1, 2, \dots, m$ and $d = 1, 2, \dots, D$. t is the number of iterations, c_1, c_2 are accelerating constant, r_1, r_2 are a random number in $[0, 1]$, and the ω is inertia weight.

3. Intelligent Analysis Method of Gear Fault Detection

3.1. Noise Removal Based on Fractional Wavelet Transform. Fractional Fourier transform is used to filter the gear vibration signal, whose model is $x(n) = s(n) + d(n)$, where $x(n)$ is the gear vibration signal in the mixed noise, $s(n)$ is a useful component, and $d(n)$ is the background noise. Fractional wavelet transform is applied to both sides of the following equation:

$$X_p(v_k) = S_p(v_k) + D_p(v_k). \quad (16)$$

Here, $X_p(v_k)$, $S_p(v_k)$, and $D_p(v_k)$ are separately the discrete fractional wavelet transform of $x(n)$, $s(n)$, and $d(n)$, and p is the transform order. The basic steps are as follows:

- (1) Use a sensor to collect the gear vibration signal $x(n)$ in the acceleration stage, select the range and step order, and apply the fractional Fourier transform to signal

acquisition, forming a two-dimensional distribution of the signal energy in the fractional Fourier domain (p, u) .

- (2) In the plane formed in step one, peak points are found through two-dimensional search, where the maximum peaks' corresponding FRFT domain is taken as the best FRFT domain while the optimal order p_{opt} is determined according to (17), and p_{opt} order fractional Fourier transform is applied to signal $x(n)$:

$$(p_{\text{opt}}, u_{\text{opt}}) = \arg \max_{p, u} |X_p(u)|^2. \quad (17)$$

- (3) For the signal in the fractional Fourier domain, its wavelet basis function is determined, with proper decomposition layers N selected. After the N th wavelet decomposition, the fractional wavelet signal is obtained, including the N th low frequency coefficients ca_N and high frequency coefficients from 1st to N th layer cd_1, cd_2, \dots, cd_N .
- (4) The signal was decomposed to 1st to N th fractional wavelet domain signal, where adaptive threshold filter was applied to the corresponding high frequency coefficients (i.e. cd_1, cd_2, \dots, cd_N) to eliminate noise.
- (5) Wavelet reconstruction was applied to the processed signal, where the basis function and layer number should be ensured the same way as that for wavelet decomposition. Then, wavelet reconstruction is applied to the N layers' frequency coefficients (ca_N) and high frequency coefficients (cd_1, cd_2, \dots, cd_N) obtained in the steps above.
- (6) $-p_{\text{opt}}$ order fractional Fourier transform was applied to the wavelet reconstruction signal to get the time-domain signal after noise removal.

3.2. Energy Feature Extraction for Wavelet Packet Frequency Band. Wavelet packet transform is applied to multiple decompositions and coefficient reconstruction for the filtered signal of the fractional Fourier transform. As the energy projection sequence of gear vibration signal in various frequency bands is a product of the gear transmission mode, analysis of gear vibration signals in different frequency bands can help inconspicuous signal frequency features stand out in a more obvious way as energy change in several subspaces of various resolution. Signal energy after wavelet packet decomposition in each frequency band in various scale spaces is lined up according to their scale as feature vectors. Steps for extracting wavelet packet signal features are as follows:

- (1) The k -layer wavelet packet and coefficient reconstruction are carried out by filtered vibration signal after FRWT, of which the feature signals $x_{j,m}(i)$ are decomposition coefficients of filtered vibration signal on the j scale: $j = 1, 2, \dots, k = 1, 2, \dots, j, m = 0, 1, 2, \dots, 2^{k-1}$. Here i is the time-domain location; data length of original signal is N ; data length of feature signal is $n = N/2^k$.

- (2) Solve the energy E_{km} on k th layer within various frequency bands:

$$E_{km} = \frac{1}{n-1} \sum_{i=1}^n |x_{k,m}(i)|^2. \quad (18)$$

- (3) Solve the total energy of feature signals. A group of energy sequences $\{E_{km}, m = 1, 2, \dots, 2^k\}$ corresponding to the signal are obtained after wavelet packet decomposition, where the total energy is

$$E_t = \sum_{m=1}^{2^k} E_{km}. \quad (19)$$

- (4) Construct the feature vector. As the gear faults will have a great impact on the energy of the signal in each frequency band and E_{km} may result in some difficulty in data analysis when the value is large, the feature vector is constructed after frequency band energy is normalized:

$$T = \frac{[E_{k1}, E_{k2}, \dots, E_{k2^k}]}{E_t}. \quad (20)$$

3.3. Intelligent Recognition for SVM. The normalized feature vectors in Section 3.2 are used as the sample inputs to support vector machines for training and recognition. In this paper, an open source SVM pattern recognition and regression package LIBSVM is used, which utilizes a particle swarm optimization method to optimize the LIBSVM classification model, and then tests the penalty factor c and kernel function parameter g finally obtaining proper SVM correction parameters. The main steps for particle swarm optimization (PSO) to optimize the parameters of SVM for intelligent fault identification are as follows:

- (1) Use an initialized particle swarm algorithm, in the search space. Twenty (20) particles are randomly generated, randomly determining each particle's initial position and initial velocity. Initialize error penalty parameters c and Gaussian kernel parameters g of support vector machines, where the optimal ranges of c and g are $[0.1 \sim 100]$ and $[0.1 \sim 1000]$, respectively.
- (2) The initialized SVM parameter will be used in the SVM algorithm to build a corresponding model, which is used to predict and classify the test sample. The fitness value of each particle will be calculated according to the fitness function.
- (3) The initial particle fitness value is used as the model's initial optimal solution, which is compared with the global optimization objective function value. If the initial fitness value of particles is better than the optimal value of the objective function, the former shall be used as the current optimal objective function value, going on to search for the global optimal solution.

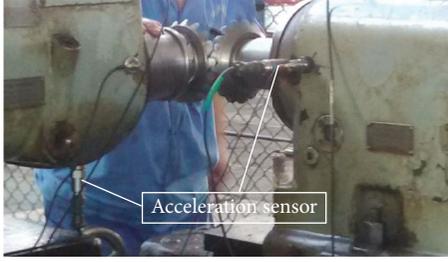


FIGURE 2: Layout of acceleration sensors in the experimental stage.



FIGURE 3: Tooth surface pitting.

- (4) Update formula of the particle velocity and position is used to update and obtain the best position p_{best} of each particle itself. After comparing all the best positions of the particles, the best position g_{best} of all particles is obtained.
- (5) Check whether the iteration conforms to the ending condition. If a prescribed error requirement or a certain number of iterations are reached, the iteration will be stopped. Otherwise, go back to step (2).
- (6) The penalty parameter c and Gauss kernel parameter g obtained through optimization are used as SVM parameters. Input training samples for training and respective category labels are established according to different types of gear faults. Meanwhile, “one against rest” SVC is used for training to ultimately obtain some certain training parameter information (including support Vector, pull-down factor, and network bias.).
- (7) The test samples are input into support vector machines for recognition, outputting the test results, which are used to determine gear fault types.

4. Experiments

In this paper, the rolling-gear inspection machine is used to carry out the experiment, in which the arrangement of acceleration sensor used to collect gear vibration signal is shown in Figure 2. The gear drive is a one-level drive, whose driving wheel and a driven wheel share a gear tooth number of 17, with a transmission ratio of 1:1. The driving wheel rotation speed is 750–1000 rpm. The sampling frequency of the vibration signal sound is 10 kHz. The gear vibration signals were collected in three cases of normal gear, shaft unbalance fault, and pitting fault.

When gears are in states of normal meshing, shaft unbalance fault, and pitting fault (Figure 3), respectively, curves of measured vibration signal in the time domain and frequency domain are shown in Figures 4(a), 4(b), and 4(c) separately, where obvious modulation phenomena occur in the time-domain signals for all of three types of gears mentioned above. From the FFT frequency spectrum diagram in Figure 3, background noise has masked the noise generated by gear meshing, which makes the time and frequency-domain curves of the three types of gears so similar that

it is difficult to distinguish between them based only on a comparison of the curves in either the time domain or in the frequency domain.

A gear fault analysis method based on a combination of FRWT and SVM proposed in this paper is used to analyze the three kinds of gear vibration signals mentioned in Figure 4.

Step 1. Fractional wavelet transform is applied to the collected gear vibration signal to eliminate background noise, extracting the useful signal. Fractional Fourier transform is applied to the gear vibration signal with a transform order of p . According to the peak search and calculation, an optimal order can be determined, by the basis on which a narrowband filter is constructed to mask the optimal FRFT domain signal and further to eliminate the energy of background noise while retaining that of useful vibration signal. Therefore, the gear vibration signal after denoising based on the fractional wavelet transform is obtained. Figure 5 is the 2-dimensional distribution of signal energy before filtering and after narrow band filtering in the fractional Fourier transform domain in three cases, showing that the filtering method proposed in this article can achieve better noise suppression effects.

Step 2. Fractional Fourier transform is applied to the signal filtering to get useful gear vibration signal $W(t)$, which is decomposed and reconstructed by db6 3-layer wavelet packet and to obtain the decomposition energy spectrum. Frequency of gear vibration signal ranges from 0 to 2000 Hz, and sampling frequency f_s is 10 kHz. Therefore, when k -layer wavelet packet decomposition is applied to the gear vibration signal, the node bandwidth at the k th layer is

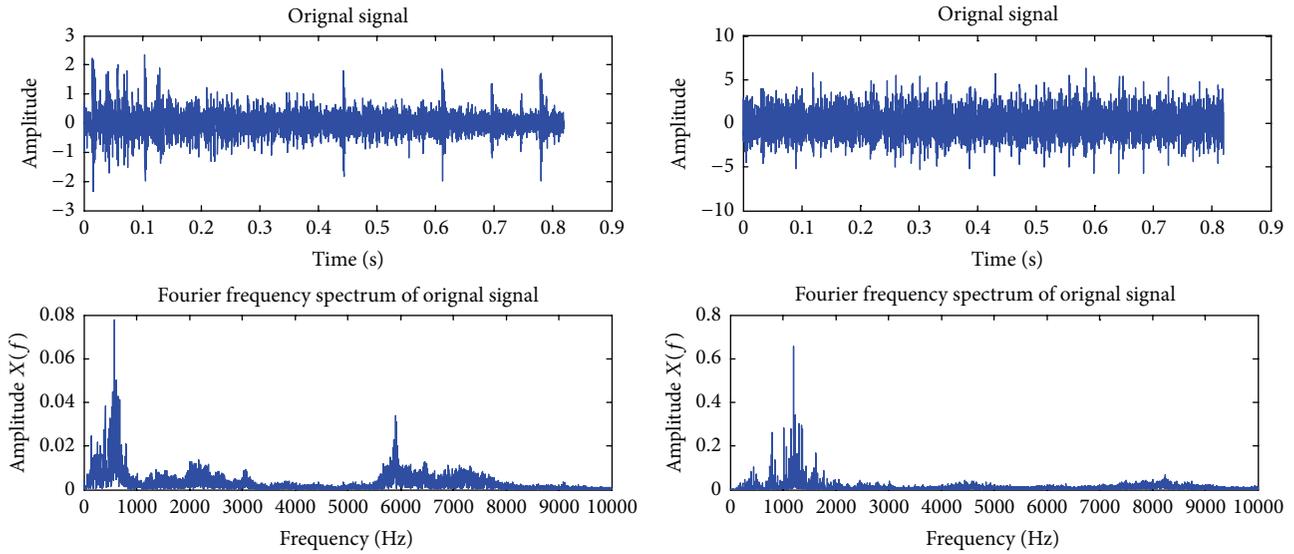
$$f_{\text{BW}} = \frac{f_s}{2^{k+1}}, \quad (21)$$

where frequency range for each node is

$$mf_{\text{BW}} \sim (m+1)f_{\text{BW}}, \quad m = 0, 1, 2, \dots, 2^k - 1. \quad (22)$$

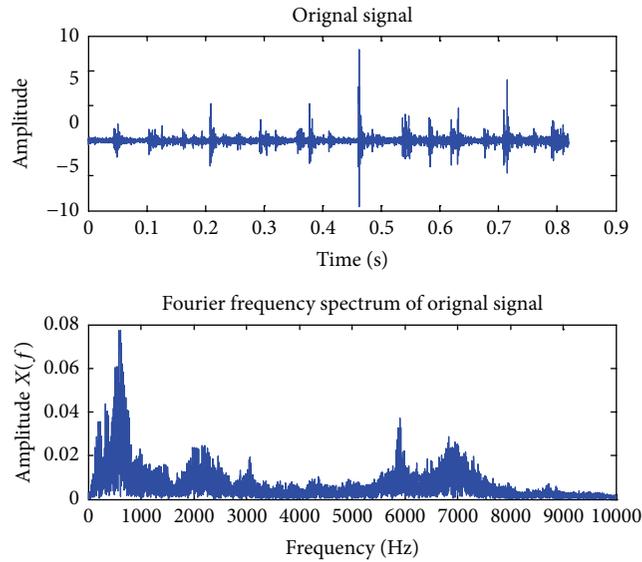
Here, m represents the node number at the k th layer in wavelet decomposition. Wavelet packet frequency band and frequency correspondence are shown in Table 1.

A sequence corresponding to the signal which consists of the module of the highest level wavelet coefficients after decomposition and feature vectors corresponding to the



(a) Gear vibration signal of normal gear meshing

(b) Gear vibration signal of shaft unbalance fault



(c) Gear vibration signal of pitting fault

FIGURE 4: Gear vibration signal.

energy sequence were obtained according to (20). Figure 6 shows the histogram for the feature vectors of the three cases of normal gear meshing, shaft unbalance fault, and pitting fault.

As is shown in Figure 6, in vibration signal wavelet energy spectrum for different types of gear fault, energy is concentrated mainly within the 1st band because of the vibration due to stiffness stimulation and load variation. When the gear fault exists, the amplitude of the meshing frequency and its frequency band tend to produce large growth. However, it is difficult to directly observe the effects of background noise in the FFT spectrum.

As can be seen clearly in the wavelet energy spectrum diagram, the characteristic value of the 1st frequency band for shaft unbalance gear fault is 2% lower than that of normal gear. At the same time, the characteristic value of the 3rd frequency band is increased by about 90% and the characteristic value of the 6th frequency band is decreased by about 99%.

As for the pitting gear fault, the characteristic value in the 1st band decreases by 29% compared with the normal gear operation. Meanwhile, the characteristic value of the 3rd frequency band is increased by about 97% and the characteristic value of the 6th frequency band is increased by

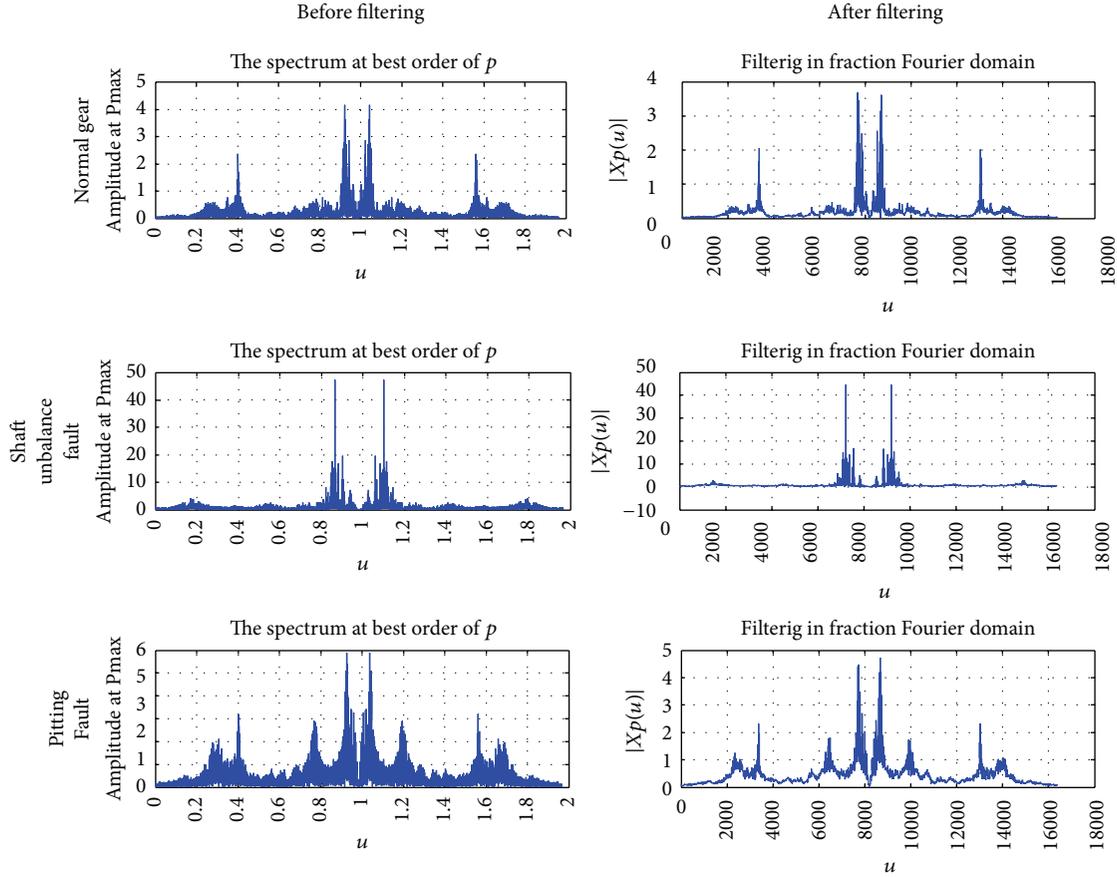


FIGURE 5: 2-dimensional distribution of signal energy before filtering and after narrow band filtering on fractional Fourier transform domain in three cases.

TABLE 1: A wavelet packet frequency band-frequency tablet.

Band serial number	Sample point	Frequency range (Hz)
1	$W(3,0)$	0~625
2	$W(3,1)$	625~1250
3	$W(3,2)$	1250~1875
4	$W(3,3)$	1875~2500
5	$W(3,4)$	2500~3125
6	$W(3,5)$	3125~3750
7	$W(3,6)$	3750~4375
8	$W(3,7)$	4375~5000

about 40%. The characteristic values of the other bands also appear to increase in varying degrees.

The above analysis shows that there are marked differences in characteristic values of the three cases of normal gear, shaft unbalance fault, and pitting fault, which can be used as feature vectors for gear fault type recognition.

Step 3. Use SVM to study the feature vectors of multiple sets of gear vibration signal data. The sample consists of 60

groups of gears, where the 20 sets of gear vibration signal samples are collected, respectively, in three cases of normal gear, shaft unbalance fault, and pitting fault. The 60 sets of feature vectors are obtained by calculation, according to first and second steps of the mentioned method. Divide the feature vectors into two groups, training samples and test samples, each containing 10 groups of normal gears, 10 groups of shaft unbalance fault gears, and 10 groups of pitting fault gears. The training samples are optimized by the particle optimization method, and the optimization results are shown in Figure 7.

After the completion of training, test samples are input into SVM for recognition, the results of which are shown in Figure 8. In Figure 8, vertical coordinates 1, 2, and 3 represent the normal gear, shaft unbalance fault, and pitting fault, respectively. The circle represents the actual test set classification of the input data while the asterisk represents predictive test set classification of the input data. As is directly shown, the results of practical training and prediction for three kinds of gear fault type using multiclassification SVM can reach an accuracy of 96.7%.

Experimental results prove that application of the fractional Fourier transform proposed in this paper can effectively extract useful signals reflecting the fault feature under

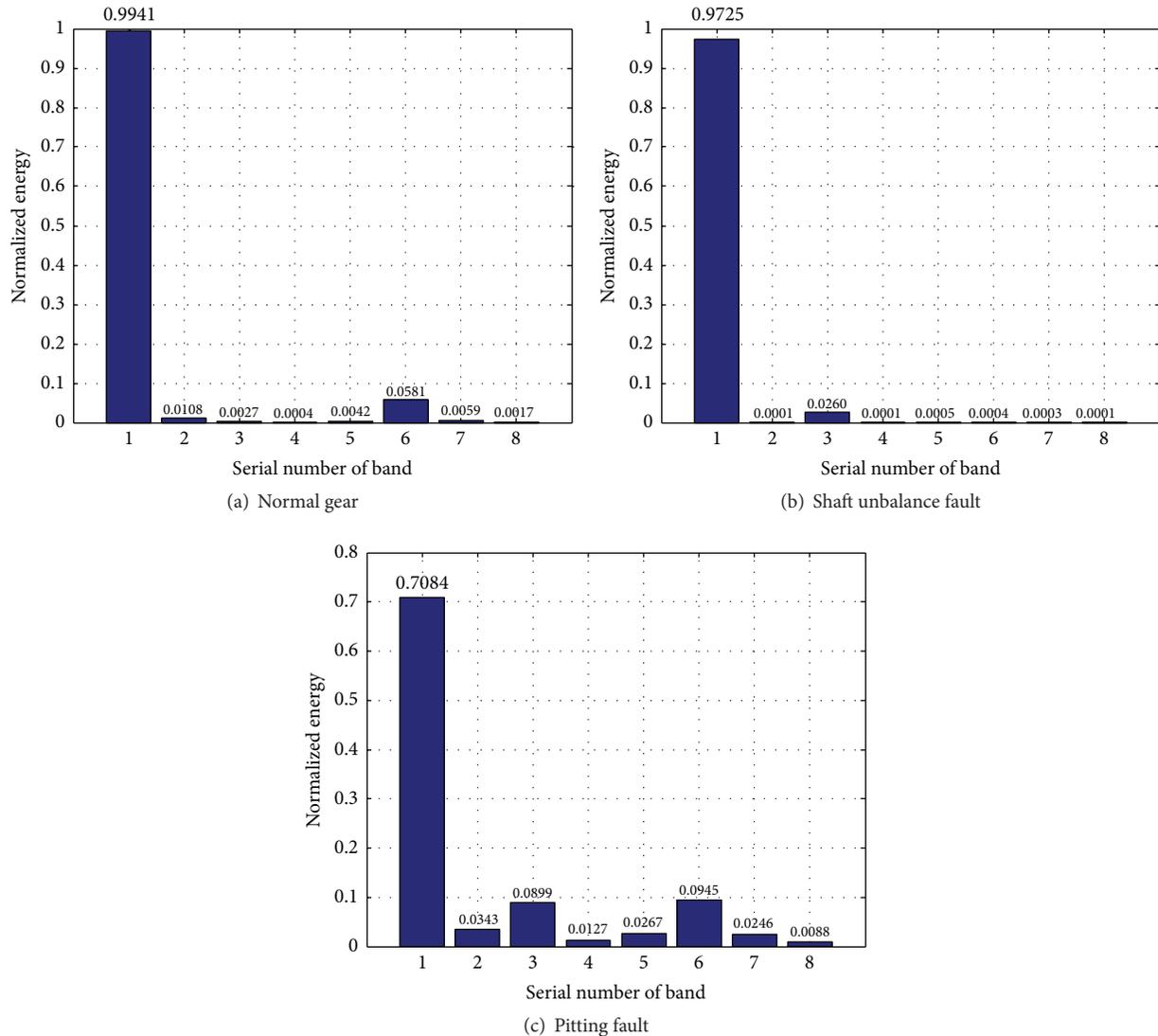


FIGURE 6: Histogram of the feature vector.

a background condition of powerful noise. SVM can intelligently recognize gear fault types without the artificial manipulation.

5. Conclusions

This paper studies a gear fault detection method based on the feature vector of gear vibration signals. For the collected vibration signal, the noise is eliminated by the FRWT method. Using the wavelet packet transform, the feature vectors are constituted by the frequency band energy. On this basis, the feature vectors are trained and classified by support vector machines. Through the experiments, the FRWT can extract the useful signals which reflect the gear fault features, achieving good effects in elimination of background noise. And the automatic recognition of gear fault is realized through the clustering and classification of vibration signals

vector by SVM, so that it can solve the problem of faults recognition in time and frequency domain which is caused by the interference of the background noise and the complexity of the side band. Therefore, it is intelligent, efficient, and accurate for the gear vibration analysis method based on FRWT and SVM.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The research is supported by national major scientific instruments and equipment development projects (No. 2013YQ17053904), national science and technology major

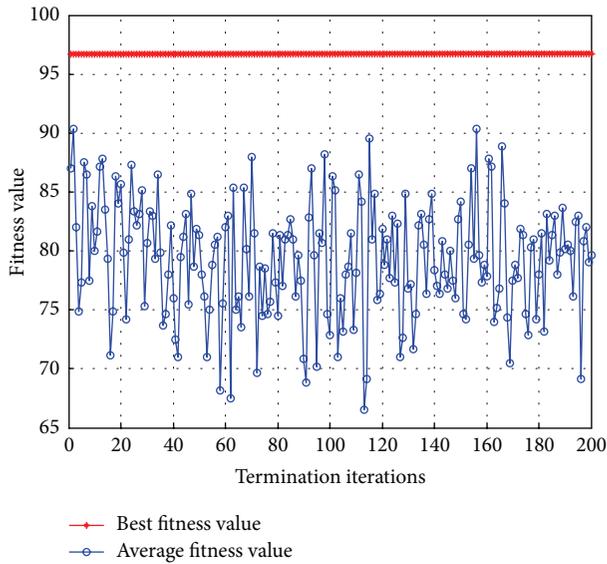


FIGURE 7: Optimal results of SVM via particle optimal method.

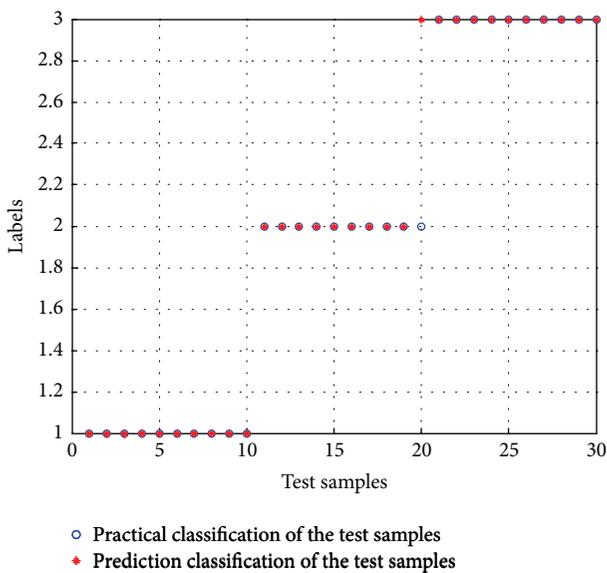


FIGURE 8: Recognized results of gear fault.

special projects (No. 2010ZX04014-091), and project of national science foundation of china (No. 50805003).

References

[1] M. Amarnath and I. R. Praveen Krishna, "Empirical mode decomposition of acoustic signals for diagnosis of faults in gears and rolling element bearings," *IET Science, Measurement & Technology*, vol. 6, no. 4, pp. 279–287, 2012.

[2] Z. Y. Shi, H. Lin, and J. Lin, "Current status and trends of large gears metrology," *Journal of Mechanical Engineering*, vol. 49, no. 10, pp. 35–44, 2013.

[3] B. Samanta, "Artificial neural networks and genetic algorithms for gear fault detection," *Mechanical Systems and Signal Processing*, vol. 18, no. 5, pp. 1273–1282, 2004.

[4] R. Bajric, N. Zuber, G. A. Skrmpas, and N. Mijatovic, "Feature extraction using discrete wavelet transform for gear fault diagnosis of wind turbine gearbox," *Shock and Vibration*, vol. 2016, Article ID 6748469, 10 pages, 2016.

[5] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[6] Z. K. Peng and F. L. Chu, "Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography," *Mechanical Systems and Signal Processing*, vol. 18, no. 2, pp. 199–221, 2004.

[7] C. Gargour, M. Gabrea, V. Ramachandran, and J.-M. Lina, "A short introduction to wavelets and their applications," *IEEE Circuits and Systems Magazine*, vol. 9, no. 2, pp. 57–68, 2009.

[8] L. B. Almeida, "The fractional fourier transform and time-frequency representations," *IEEE Transactions on Signal Processing*, vol. 42, no. 11, pp. 3084–3091, 1994.

[9] L. Yuan, "Wavelet-fractional Fourier transforms," *Chinese Physics B*, vol. 17, no. 1, pp. 170–176, 2008.

[10] V. A. Narayanan and K. M. M. Prabhu, "The fractional Fourier transform: theory, implementation and error analysis," *Microprocessors and Microsystems*, vol. 27, no. 10, pp. 511–521, 2003.

[11] R. Tao, Y.-L. Li, and Y. Wang, "Short-time fractional Fourier transform and its applications," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2568–2580, 2010.

[12] J. C. Wood and D. T. Barry, "Linear signal synthesis using the Radon-Wigner transform," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 2105–2111, 1994.

[13] D. Mendlovic, Z. Zalevsky, D. Mas, J. García, and C. Ferreira, "Fractional wavelet transform," *Applied Optics*, vol. 36, no. 20, pp. 4801–4806, 1997.

[14] L. Chen and D. Zhao, "Optical image encryption based on fractional wavelet transform," *Optics Communications*, vol. 254, no. 4–6, pp. 361–367, 2005.

[15] Y. Huang and B. Suter, "The fractional wave packet transform," *Multidimensional Systems and Signal Processing*, vol. 9, no. 4, pp. 399–402, 1998.

[16] D. Wang, K.-L. Tsui, P. W. Tse, and M. J. Zuo, "Principal components of superhigh-dimensional statistical features and support vector machine for improving identification accuracies of different gear crack levels under different working conditions," *Shock and Vibration*, vol. 2015, Article ID 420168, 14 pages, 2015.

[17] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2560–2574, 2007.

[18] B. Samanta, K. R. Al-Balushi, and S. A. Al-Araimi, "Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection," *Engineering Applications of Artificial Intelligence*, vol. 16, no. 7-8, pp. 657–665, 2003.

[19] R. Eberhart and J. Kennedy, "New optimizer using particle swarm theory," in *Proceedings of the 6th International Symposium on Micro Machine and Human Science*, pp. 39–43, Nagoya, Japan, October 1995.

[20] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1948, Perth, Australia, November–December 1995.

Research Article

Investigations of Dynamic Behaviors of Face Gear Drives Associated with Pinion Dedendum Fatigue Cracks

Zhengminqing Li, Shuang Liu, Rupeng Zhu, and Xiaodong Xu

Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China

Correspondence should be addressed to Zhengminqing Li; lzmq_cmee@nuaa.edu.cn

Received 10 March 2016; Revised 31 July 2016; Accepted 1 August 2016

Academic Editor: Ganging Song

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

Face gear dynamics is addressed by many scholars. However, dynamic behaviors of face gear drives associated with pinion dedendum fatigue cracks are yet to be investigated. Thus, in the study, a calculation solution of static transmission errors (STE) of face gear drives associated with pinion dedendum fatigue cracks is constructed, based on the proposed equivalent face gear drives, and a four-degree-of-freedom (DOF) dynamic model of face gear drives is formulated. The dynamic behaviors of an example case of face gear drives associated with pinion dedendum fatigue cracks are investigated. The simulation results indicate natural frequencies of face gear drives would be decreased with pinion dedendum fatigue crack extensions; the relationship between accelerations and velocities of face gear drives, dynamic mesh forces, and vibration accelerations at bearings would also be changed. However, pinion dedendum fatigue cracks of face gear drives are detected difficultly, due to simulation result similarities. Therefore, a determination solution for pinion dedendum fatigue cracks of face gear drives is proposed. The result shows a trend of vibration accelerations at bearings, namely, a first decreasing and then increasing tendency, which could be beneficial to detecting pinion dedendum fatigue cracks of face gear drives. These contributions would improve engineering applications of face gear drives in the future.

1. Introduction

A face gear drive, which is a kind of intersection gear drives with an involute spur pinion, is one of the significant gear drives due to its insensitive characteristics of manufacture and alignment errors versus spiral bevel gear drives and is addressed by scholars. There are a vast number of manuscripts discussing face gear drives in the past few years. Litvin et al. investigated tooth modeling solutions, stresses, and tooth contact analyses (TCA) of face gear drives [1–4]. Li et al. constructed an equivalent face gear tooth modeling solution and discussed influences of geometry parameters on strengths of face gear drives [5]. Guingand et al. tested tooth bending stresses of face gear drives under a quasistatic experiment [6]. A research team as core of Litvin, according to their study achievements, formed a design handbook of face gear drives [7] and suggested face gear drives to be used in first-stage gear drives of helicopter main gear boxes [8], namely, input-stage gear drives, which occupy high rotation speed characteristics. Due to the suggestion, face gear dynamics

becomes one of the study focuses of face gear drives. Li et al. discussed dynamic behavior differences among three version face gear drives associated with high contact ratios [9]. Hu et al. evaluated impacts of mesh stiffness on dynamic behaviors of face gear drives [10]. Jin et al. established a nonlinear dynamic model of face gear drives [11]. Yang et al. assessed vibration and bifurcation characteristics of face gear drives [12, 13]. Li et al. studied influences of sliding frictions on dynamic behaviors of face gear drives [14]. Wang et al. probed load sharing effects of a face gear split torque transmission system [15]. However, according to the limited published issues, solutions of introducing pinion dedendum fatigue cracks into dynamic behavior analyses of face gear drives are not to be constructed and dynamic behaviors of face gear drives associated with pinion dedendum fatigue cracks are yet to be investigated. Thus, in the study, a calculation solution of static transmission errors (STE) of face gear drives associated with pinion dedendum fatigue cracks, based on the proposed equivalent face gear drives, is constructed, a four-degree-of-freedom (DOF) dynamic model of face gear drives is

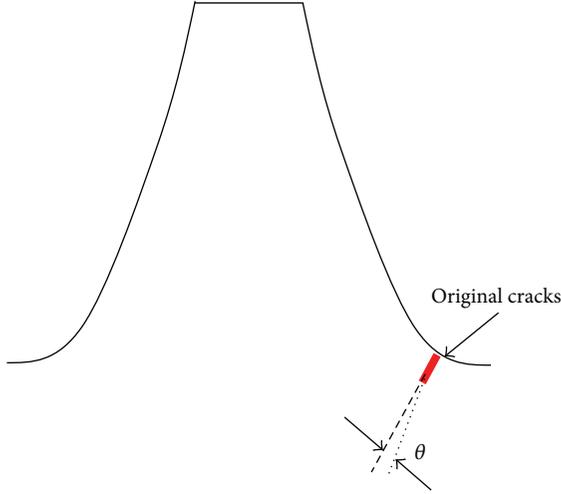


FIGURE 1: A sketch of pinion dedendum fatigue crack extension directions.

established, and the natural frequencies, dynamic behaviors, dynamic mesh forces, and vibration accelerations at bearings of an example case of face gear drives without and with pinion dedendum fatigue cracks are simulated. The results indicate that natural frequencies would be decreased with pinion dedendum fatigue crack extensions, and dynamic mesh forces, vibration accelerations at bearings, and relationships between accelerations and velocities of face gear drives would be changed with pinion dedendum fatigue crack extensions, but these dynamic behavior changes caused by pinion dedendum fatigue crack extensions are detected hardly, due to simulation result similarities. Therefore, a determination solution of pinion dedendum fatigue cracks of face gear drives is proposed and verified. These contributions would be beneficial to improving engineering applications of face gear drives in the future.

2. Constructed Analysis Solutions

2.1. STE Calculation Solution Associated with Pinion Dedendum Fatigue Cracks. Original pinion dedendum fatigue crack generations are random, due to influences of materials, manufactures, and operating conditions. Otherwise, pinion dedendum fatigue crack extension directions, as shown in Figure 1, can be predicted.

As shown in Figure 1, according to [17–19], crack extension angle θ can be expressed as

$$\theta = \cos^{-1} \left(\frac{3k_{II}^2 + \sqrt{k_I^4 + 8k_I^2 k_{II}^2}}{k_I^2 + 9k_{II}^2} \right), \quad (1)$$

where k_I is an opening-mode stress intensity factor and k_{II} is a sliding-mode stress intensity factor and, according to

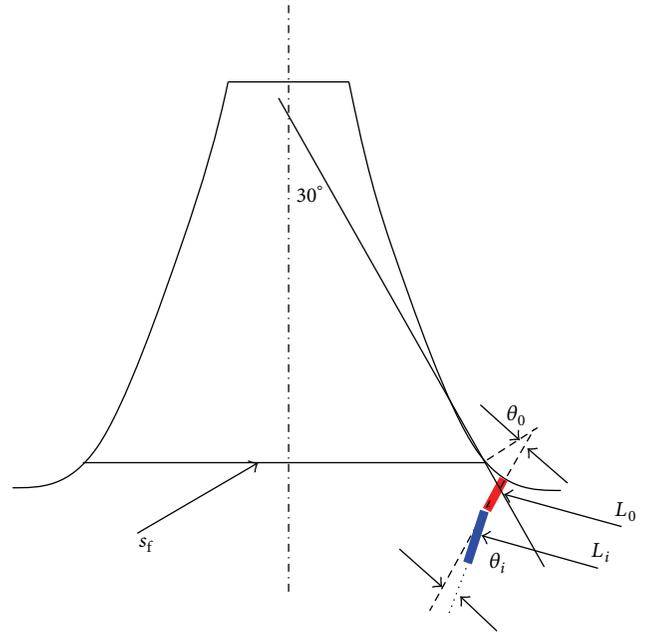


FIGURE 2: A sketch of influences of crack extensions on dedendum minimum life sections.

[17], both factors, based on finite element methods, can be calculated by

$$k_I = \frac{2G}{(3-4\gamma)+1} \sqrt{\frac{\pi}{2L}} (4V_d - V_e - 4V_b + V_c), \quad (2)$$

$$k_{II} = \frac{2G}{(3-4\gamma)+1} \sqrt{\frac{\pi}{2L}} (4U_d - U_e - 4U_b + U_c),$$

where G is a shear modulus, γ is Poisson ratio, L is a finite element length, and U and V are symbols of finite element node displacements.

Meanwhile, an influence sketch of dedendum fatigue crack extensions on dedendum minimum life sections is shown in Figure 2.

As illustrated in Figure 2, θ_0 is an original crack angle, θ_i is a crack extension angle of “ i ” stage, L_0 is an original crack length, L_i is a crack extension length of “ i ” stage, and s_f is a dedendum minimum life section. After dedendum fatigue crack extensions, s_f becomes s_{ff} , namely, a dedendum minimum life section associated with dedendum fatigue crack extensions, and based on the geometry relationship between fatigue crack extensions and dedendum minimum life sections, as shown in Figure 2, s_{ff} can be derived by

$$s_{ff} = s_f - L_0 \cos(30 + \theta_0) - \sum_{i=1}^n L_i \cos(30 + \theta_0 + \theta_i), \quad (3)$$

where symbol n means stages and the dedendum minimum life section s_f can be calculated, according to [16].

Based on Ishikawa model and [20], pinion tooth flexibilities associated with dedendum fatigue cracks can be deduced as listed in Table 1.

TABLE 1: Pinion tooth flexibilities associated with dedendum fatigue cracks.

Symbols	Equations
q_{Br}	$\frac{12\cos^2(\omega_x)}{Es_{ff}^3} \left(h_x^2 h_r - h_x h_r^2 + \frac{h_r^3}{3} \right)$
q_{Bt}	$\frac{6\cos^2(\omega_x)}{Es_{ff}^3} (h_i - h_r)^3 \left[\frac{h_i - h_x}{h_i - h_r} \left(4 - \frac{h_i - h_x}{h_i - h_r} \right) - 2 \ln \left(\frac{h_i - h_x}{h_i - h_r} \right) - 3 \right]$
q_G	$\frac{24h_x^2 \cos^2(\omega_x)}{\pi Es_{ff}^2}$
q_s	$\frac{2(1+\gamma)\cos^2(\omega_x)}{Es_{ff}} \left[h_r + (h_i - h_r) \ln \left(\frac{h_i - h_x}{h_i - h_r} \right) \right]$

q_{Br} is a rectangle bending flexibility, q_{Bt} is a trapezoid bending flexibility, q_s is a share flexibility, q_G is a flexibility caused by tooth base rotations, and E is an elastic modulus. Moreover, h_i , h_x , and h_r are geometry parameters of pinions and ω_x is an acting angle. Both geometry parameters and acting angle can be calculated, according to [16].

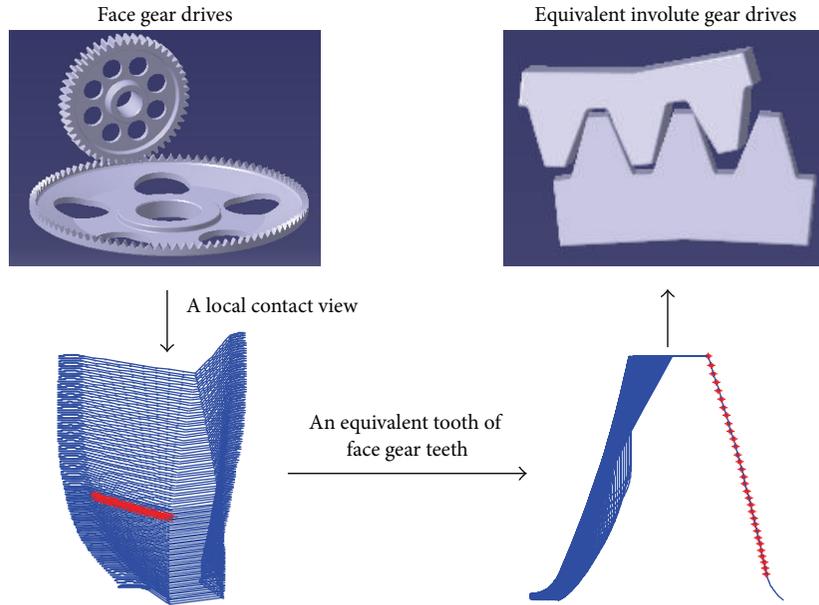


FIGURE 3: An evolution process of equivalent face gear drives.

Thus, based on the deduced pinion tooth flexibilities associated with dedendum fatigue cracks, as listed in Table 1, and according to the relationship between flexibilities and deformations of gear teeth, a comprehensive deformation of pinion teeth associated with dedendum fatigue cracks can be obtained as

$$D_p = F(q_{Br} + q_{Bt} + q_G + q_s + q_{c1}), \quad (4)$$

where F is a pinion tooth normal load and q_{c1} is a contact flexibility and can be expressed as [16]

$$q_{c1} = \frac{\left(\left((1 - \gamma_1^2) / E_1 \right) + \left((1 - \gamma_2^2) / E_2 \right) \right)}{\pi}, \quad (5)$$

where subscripts 1 and 2 mean pinions and gears, respectively.

Based on the proposed solutions, pinion dedendum fatigue cracks could be introduced into involute gear tooth deformations, while face gear drives are not equal to involute gear drives. Thus, in order to introduce the proposed

solutions into face gear drives, an equivalent face gear drive solution is presented, according to face gear tooth geometry characteristics.

A face gear drive can be equivalent as an involute gear drive in contact viewpoints, as shown in Figure 3, because a face gear tooth can be considered as a sequence in which modified involute gears are superimposed along its face width.

Meanwhile, STE of face gear drives, as shown in Figure 4, can be defined as a displacement deviation between actual contact points and theoretical contact points.

As given in Figure 4, r_{b1} and r_{b2} are base circle radii of driving and driven gears, respectively, and STE of face gear drives e can be written in

$$e = D_f - D_p - \Lambda, \quad (6)$$

where D_f is an equivalent face gear tooth deformation, which can be calculated according to (4), and Λ is a comprehensive

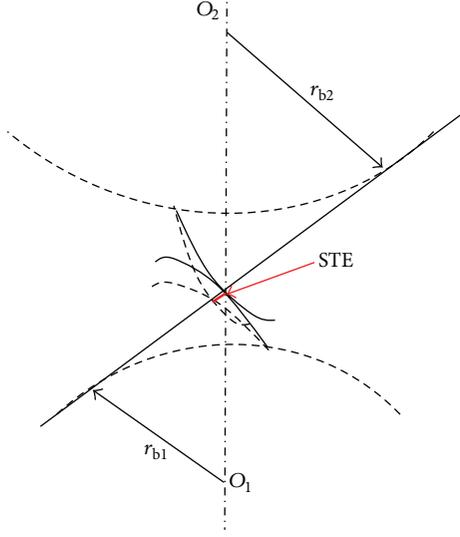


FIGURE 4: A sketch of STE of equivalent face gear drives.

meshing error caused by manufacture and alignment errors of face gear drives.

2.2. Dynamic Model. In order to investigate dynamic behaviors of face gear drives associated with pinion dedendum fatigue cracks, a four-DOF dynamic model of face gear drives is formulated, as shown in Figure 5.

As shown in Figure 5, according to Newtonian mechanics, mathematic equations of the four-DOF dynamic model can be derived by

$$\begin{aligned}
 m_p s_p'' + c_p s_p' + k_p s_p &= -F_m, \\
 m_f s_f'' + c_f s_f' + k_f s_f &= F_m, \\
 I_p \theta_p'' + F_m R_{bp} &= T_p, \\
 I_f \theta_f'' + F_m R_{bf} &= -T_f,
 \end{aligned} \tag{7}$$

where F_m can be deduced as

$$\begin{aligned}
 F_m &= k_m \sin(\gamma) (s_p - s_f + R_{bp} \theta_p - R_{bf} \theta_f - e) \\
 &+ c_m \sin(\gamma) (s_p' - s_f' + R_{bp} \theta_p' - R_{bf} \theta_f' - e'),
 \end{aligned} \tag{8}$$

where R_b is a base circle radius, θ is a torsion degree of freedom, s is a bending degree of freedom, T is a torsion, k is a bending stiffness, c is a bending damping, m is a quality, I is a moment of inertia, γ is a shaft angle, $'$ is first derivative, $''$ is second derivative, and subscripts f and p express a face gear and a pinion, respectively. In addition, k_m is mesh stiffness and c_m is mesh damping, which can be calculated, according to [16].

3. Simulation and Analysis

3.1. Pinion Dedendum Fatigue Crack Extension Simulations. In order to discuss impacts of pinion dedendum fatigue

TABLE 2: Parameters and operating conditions.

	Names	Values	Units
Geometry parameters	Modulus	5	mm
	Pressure angle	25	°
	Tooth number of the pinion	25	—
	Tooth width of the pinion	40	mm
	Tooth number of the face gear	125	—
Geometry parameters	Internal radius of the face gear	295	mm
	External radius of the face gear	330	mm
	Addendum coefficient	1	—
Material parameters	Clearance coefficient	0.25	—
	Elastic modulus	210000	Mpa
Material parameters	Poisson ratio	0.3	—
	Operating conditions	Power	500
Input rotation speed		20900	r/min

TABLE 3: Crack lengths and extension angles.

Step	Crack length L_i (mm)	Crack extension angles θ_i (°)
1	0.4	1.592
2	0.8	3.326
3	1.2	2.982
4	1.6	2.735

cracks on dynamic behaviors of face gear drives, an example case of face gear drives associated with pinion dedendum fatigue cracks is investigated, and geometry and material parameters and operating conditions of the example case are listed in Table 2.

In the simulation, an intersection point between minimum life sections and fillets is taken as the original point, and the crack length step is set as 0.4 mm, which are determined by gear material characteristics and engineering experiences. Meanwhile, according to (1) and using 2D finite element method (FEM), whose setting is shown in Figure 6, the original fatigue crack is simulated, as shown in Figure 7.

In Figure 6, q is an original point direction and n is a normal direction to q . Based on the original fatigue crack, as shown in Figure 7, the fatigue crack extensions of the example case are simulated by FEM, as shown in Figure 8.

In the case of Figure 8 and according to the relationship between θ_i and θ_0 , as shown in Figure 2, the fatigue crack extension angles of the example case can be extracted, as listed in Table 3.

3.2. Dynamic Behavior Simulations. Employing (3) and the equations as listed in Table 1 and according to [16], the proposed equivalent face gear drives, as given in Figure 3, and the parameters as listed in Table 2, mesh stiffness of the example case of face gear drives associated with pinion dedendum fatigue cracks can be calculated. Then, introducing the mesh stiffness into (7), the natural frequencies of the example case are simulated, as shown in Figure 9.

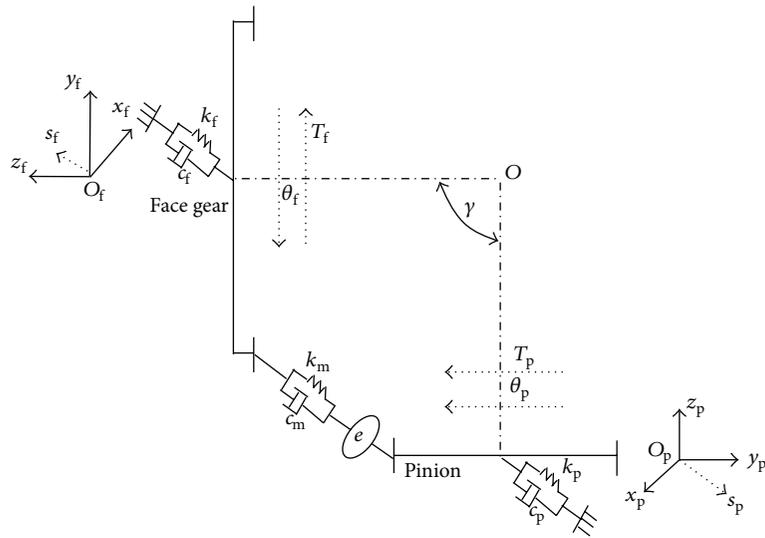
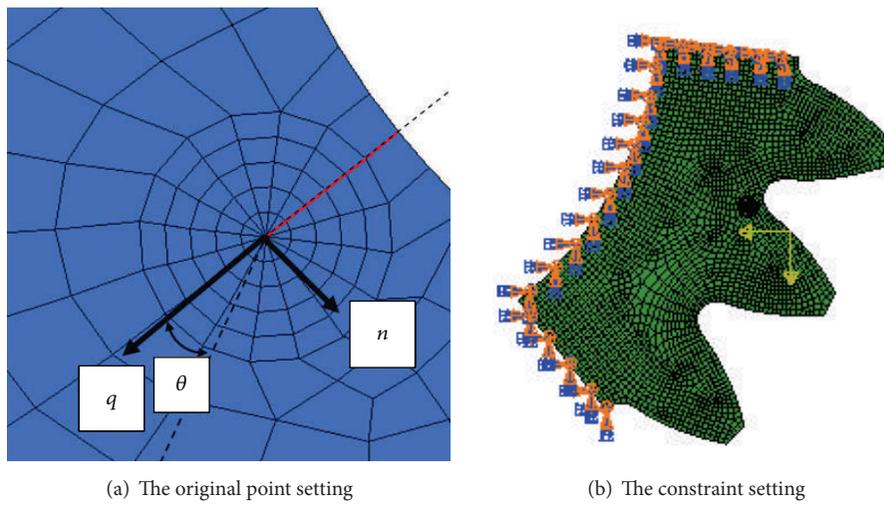


FIGURE 5: A four-DOF dynamic model of face gear drives.



(a) The original point setting

(b) The constraint setting

FIGURE 6: The 2D FEM settings.

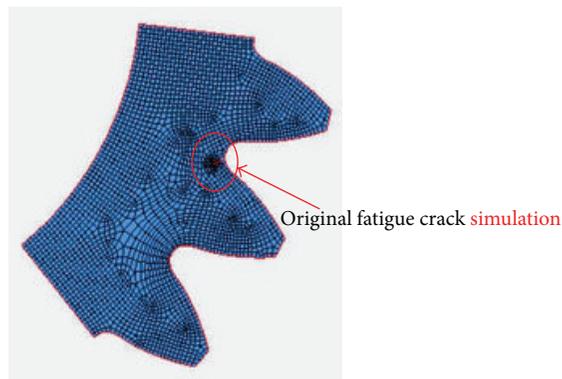


FIGURE 7: The simulation of the original fatigue crack of the example case.

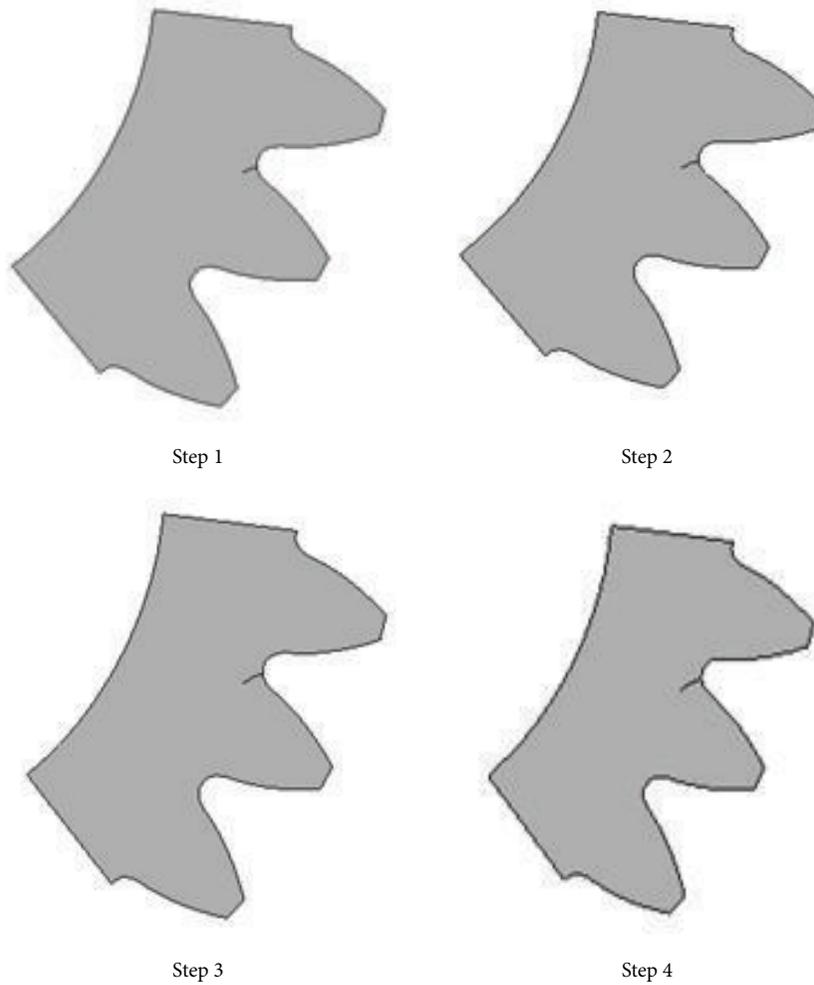


FIGURE 8: The simulations of fatigue crack extensions of the example case.

As illustrated in Figure 9, natural frequencies of face gear drives would be decreased with pinion dedendum fatigue crack extensions.

According to (6), parameters as listed in Table 2, and the results of fatigue crack extensions, as listed in Table 3, the STE without and with pinion dedendum fatigue cracks of the example case is simulated, as shown in Figure 10.

Introducing the simulated STE, as shown in Figure 10, into (7), the dynamic behaviors of the example case of face gear drives associated with pinion dedendum fatigue cracks are given in Figures 11–13.

As shown in Figure 11, the relationship between accelerations and velocities of the example case of face gear drives associated with pinion dedendum fatigue cracks would be changed; that is, the escape phenomenon of small loops becomes more and more obvious. However, due to similar simulation results at any instance, the influence of pinion dedendum fatigue cracks on dynamic behaviors of face gear drives is not to be detected easily.

In the case of Figures 12 and 13, the dynamic mesh forces versus mesh frequency and the vibration accelerations at

pinion remote bearings would be changed with the pinion dedendum fatigue cracks. However, whatever dynamic mesh forces or accelerations at bearings, the absolute values of the differences between drives without and with pinion dedendum fatigue cracks are very small, which also allowed discovering pinion dedendum fatigue cracks difficultly. Thus, based on the accelerations at pinion remote bearings and according to logarithm characteristics, namely, holding data properties and relationships, and compressing data scale, a pinion dedendum fatigue crack determination solution is constructed as

$$d_p = 10 \log \left(\frac{\sqrt{\sum_{i=1}^t A_i^2}}{t} \right) \quad [1, t] \in \text{one cycle}, \quad (9)$$

where symbol t is a number of peaks and valleys of one cycle and A_i is an amplitude of peaks or valleys of one cycle. According to (9) and Figure 13, d_p of the example case is simulated, as given in Figure 14.

As illustrated in Figure 14, with the crack extensions, the symbol d_p would first drop and then increase rapidly, which

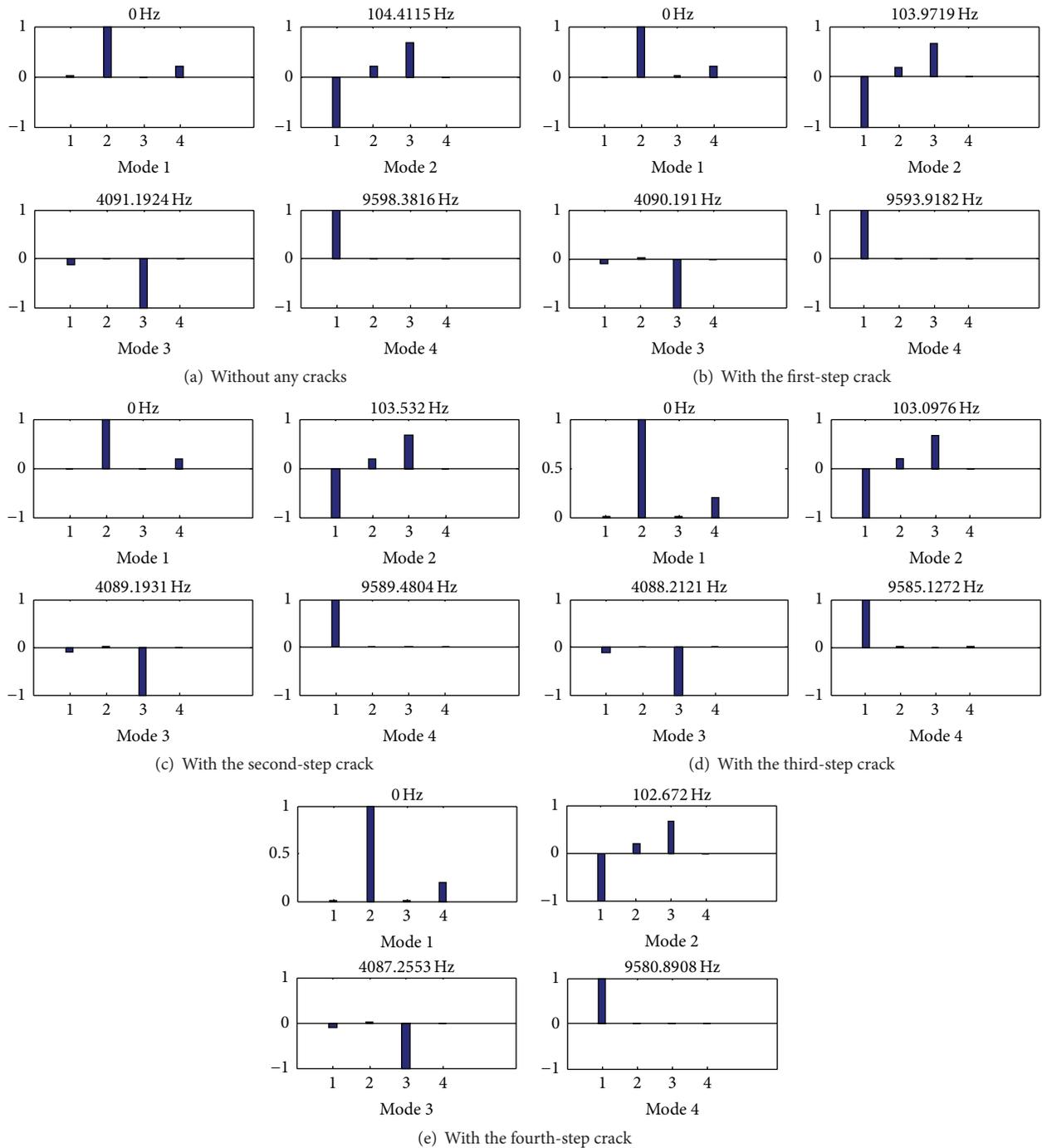


FIGURE 9: The simulations of natural frequencies of the example case.

would help to detect pinion dedendum fatigue cracks of face gear drives.

4. Conclusions

In the study, three important works can be extracted as follows:

- (1) A study solution of face gear dynamics considering pinion dedendum fatigue crack extensions is constructed, based on the proposed STE calculation solution of face gear drives associated with pinion dedendum fatigue cracks.
- (2) Dynamic behaviors of face gear drives associated with pinion dedendum fatigue cracks are investigated by

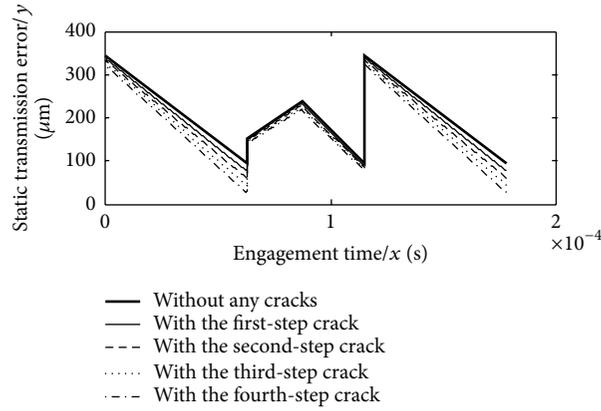


FIGURE 10: The simulation of the STE of the example case.

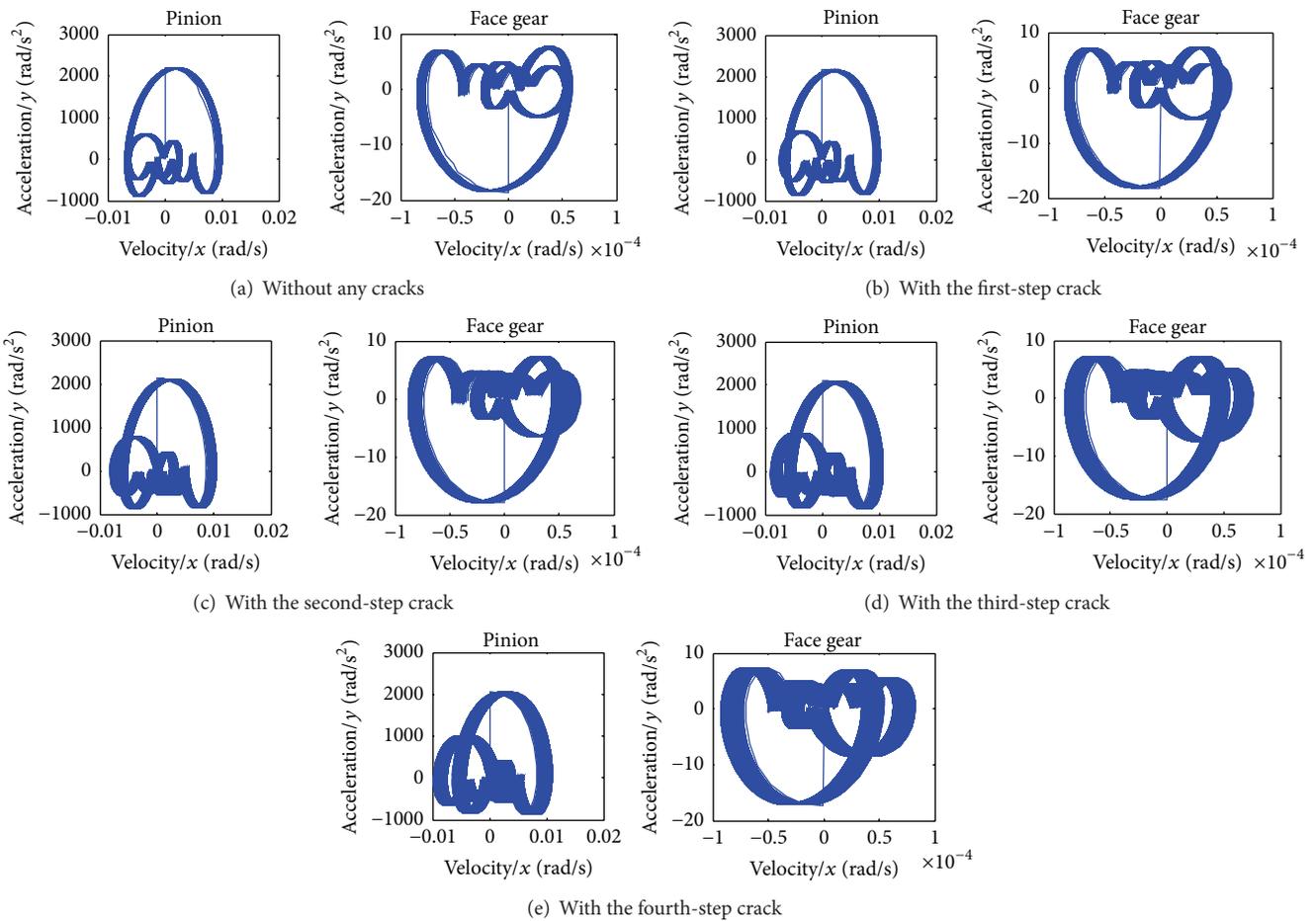


FIGURE 11: The simulations of the dynamic behaviors of the example case.

an example case simulation. The results indicate natural frequencies of face gear drives would be reduced with pinion dedendum fatigue crack extensions and the relationship between accelerations and velocities of face gear drives would be changed; that is, the escape phenomenon of small loops becomes obvious. Meanwhile, due to simulation result similarities at any

instance, pinion dedendum fatigue cracks of face gear drives are not detected easily.

- (3) A pinion dedendum fatigue crack determination solution is proposed and verified by the example case simulation. The results indicate the presented determination solution is accepted.

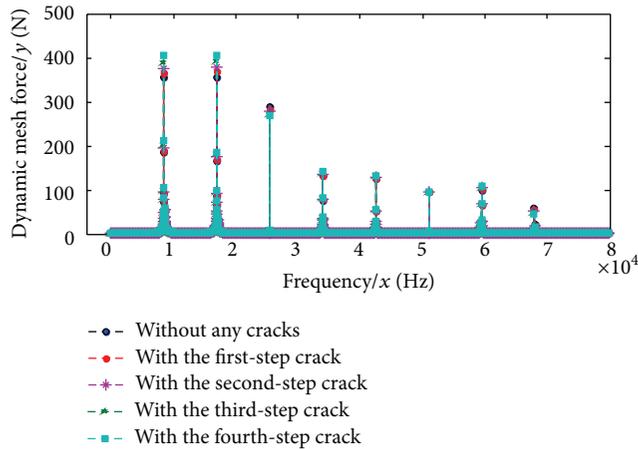


FIGURE 12: The simulation of dynamic mesh forces of the example case.

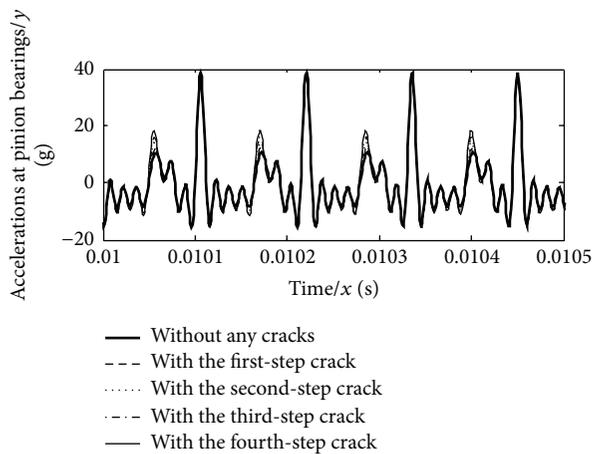


FIGURE 13: The simulation of pinion remote bearing accelerations of the example case.

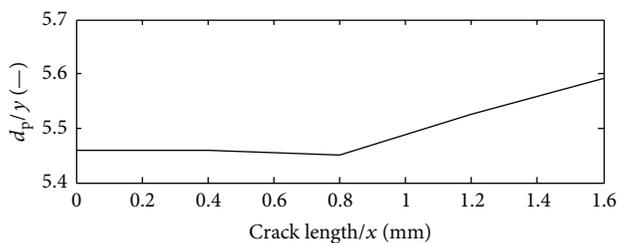


FIGURE 14: The simulation of d_p of the example case.

These contributions would improve engineering applications of face gear drives in the future.

Competing Interests

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

Acknowledgments

The authors are grateful for the financial support provided by the National Natural Science Foundation of China under Grant nos. 51105194 and 51375226 as well as the Fundamental Research Funds for the Central Universities under Grant nos. NS2015049, NS2015097, and NS2015054.

References

- [1] F. L. Litvin, Y. Zhang, J.-C. Wang, R. B. Bossler Jr., and Y.-J. D. Chen, "Design and geometry of face-gear drives," *Journal of Mechanical Design*, vol. 114, no. 4, pp. 642–647, 1992.
- [2] F. L. Litvin, A. Fuentes, and M. Howkins, "Design, generation and TCA of new type of asymmetric face-gear drive with modified geometry," *Computer Methods in Applied Mechanics and Engineering*, vol. 190, no. 43-44, pp. 5837–5865, 2001.
- [3] F. L. Litvin, A. Fuentes, C. Zanzi, and M. Pontiggia, "Design, generation, and stress analysis of two versions of geometry of face-gear drives," *Mechanism and Machine Theory*, vol. 37, no. 10, pp. 1179–1211, 2002.
- [4] F. L. Litvin, I. Gonzalez-Perez, A. Fuentes, D. Vecchiato, B. D. Hansen, and D. Binney, "Design, generation and stress analysis of face-gear drive with helical pinion," *Computer Methods in Applied Mechanics and Engineering*, vol. 194, no. 36-38, pp. 3870–3901, 2005.
- [5] Z. Li, H. Wu, and R. Zhu, "Influence predictions of geometric parameters on face gear strength," *Advances in Mechanical Engineering*, vol. 7, no. 3, pp. 1–7, 2015.
- [6] M. Guingand, J.-P. De Vaujany, and C.-Y. Jacquin, "Quasi-static analysis of a face gear under torque," *Computer Methods in Applied Mechanics and Engineering*, vol. 194, no. 39-41, pp. 4301–4318, 2005.
- [7] F. L. Litvin, A. Egelja, J. Tan, D. Chen, and G. Heath, *Handbook on Face Gear Drives with a Spur Involute Pinion*, DTIC Document, 2000.
- [8] F. Litvin, R. Bossler, Y.-J. Chen, D. Lewicki, G. Heath, and J.-C. Wang, "Application of face-gear drives in helicopter transmissions," *Journal of Mechanical Design*, vol. 116, no. 3, pp. 672–676, 1994.
- [9] Z. Li, X. Liu, and R. Zhu, "Comparison of dynamic behaviors and strength among three versions of face gear drives with high contact ratios," *Journal of Vibroengineering*, vol. 17, no. 1, pp. 125–136, 2015.
- [10] Z. Hu, J. Tang, S. Chen, and D. Lei, "Effect of mesh stiffness on the dynamic response of face gear transmission system," *Journal of Mechanical Design, Transactions of the ASME*, vol. 135, no. 7, Article ID 071005, 2013.
- [11] G.-H. Jin, R.-P. Zhu, and H.-Y. Bao, "Nonlinear dynamical characteristics of face gear transmission system," *Journal of Central South University (Science and Technology)*, vol. 5, no. 41, pp. 1807–1813, 2010 (Chinese).
- [12] Z. Yang, S.-M. Wang, Y.-S. Fan, and H.-X. Liu, "Bifurcation characteristics of face-gear transmission system," *Journal of Harbin Institute of Technology*, vol. 43, no. 3, pp. 107–110, 2011 (Chinese).
- [13] Z. Yang, S.-M. Wang, Y.-S. Fan, and H.-X. Liu, "Vibration characteristics of face-gear transmission system with parametric excitation," *Journal of Chongqing University*, vol. 35, no. 1, pp. 26–35, 2011 (Chinese).

- [14] X.-Z. Li, R.-P. Zhu, Z.-M. Li, and G.-H. Jin, "Influences of frictional coefficient on vibration characteristic of face-gear transmission system," *Journal of Vibration Engineering*, vol. 27, no. 4, pp. 583–588, 2014 (Chinese).
- [15] R. Wang, N. Zhao, L. Tao, Q. Jia, and H. Guo, "Floating shaft load sharing method for face gear split torque transmission system," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 5, no. 12, pp. 3386–3392, 2013.
- [16] X.-L. Zhu and E. Zhongkai, *Analysis of Load Capacity of Gears*, Higher Education Press, Beijing, China, 1992 (Chinese).
- [17] J. Kramberger, M. Šraml, S. Glodež, J. Flašker, and I. Potrč, "Computational model for the analysis of bending fatigue in gears," *Computers & Structures*, vol. 82, no. 23–26, pp. 2261–2269, 2004.
- [18] J. Chen and S.-S. Zhao, *Fracture Mechanics*, Science Press, Beijing, China, 2006 (Chinese).
- [19] A. T. Zehnder, *Fracture Mechanics*, Springer, Berlin, Germany, 2012.
- [20] J. L. Shi, X. G. Ma, C. L. Xu, and S. J. Zang, "Meshing stiffness analysis of gear using the Ishikawa method," *Applied Mechanics and Materials*, vol. 401–403, pp. 203–206, 2013.

Research Article

Vestas V90-3MW Wind Turbine Gearbox Health Assessment Using a Vibration-Based Condition Monitoring System

A. Romero,^{1,2} Y. Lage,¹ S. Souza,¹ B. Wang,² and T.-H. Gan^{1,2}

¹Condition and Structural Health Monitoring, TWI Ltd, Cambridge CB21 6AL, UK

²School of Engineering and Design, Brunel University, Uxbridge UB8 3PH, UK

Correspondence should be addressed to A. Romero; antonio.romero@affiliate.twi.co.uk

Received 15 March 2016; Revised 20 June 2016; Accepted 11 July 2016

Academic Editor: Lu Chen

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

Reliable monitoring for the early fault diagnosis of gearbox faults is of great concern for the wind industry. This paper presents a novel approach for health condition monitoring (CM) and fault diagnosis in wind turbine gearboxes using vibration analysis. This methodology is based on a machine learning algorithm that generates a baseline for the identification of deviations from the normal operation conditions of the turbine and the intrinsic characteristic-scale decomposition (ICD) method for fault type recognition. Outliers picked up during the baseline stage are decomposed by the ICD method to obtain the product components which reveal the fault information. The new methodology proposed for gear and bearing defect identification was validated by laboratory and field trials, comparing well with the methods reviewed in the literature.

1. Introduction

Despite the current global economic situation the wind turbine industry is still growing worldwide [1]. However, the wind turbine industry experiences premature turbine component failures, which leads to increase in the operation and maintenance (O&M) costs and loss of productivity due to turbine downtime [2, 3]. The O&M costs constitute a sizeable share of the total annual costs of a wind turbine. For a new turbine, O&M costs may easily make up 20%–25% of the total levelised cost per kWh produced over the lifetime of the turbine [4]. Unpredictable failure of certain wind turbine components, such as turbine blades, tower, gearbox, generator, braking system, and yaw system, can lead to substantially higher maintenance costs and reduced availability of wind turbines. As a result, O&M costs are attracting greater attention, as there is a need for the industry to reduce the turbine downtime, increase reliability, and decrease the cost of the energy (COE).

Wind turbine gearboxes tend to fail more prematurely than those in any other applications. Gearboxes do not always achieve their desired 20-year design life, failing prematurely from 2 to 11 years. Although it is well documented in literature

that the failure rate of mechanical components is lower in comparison with the electrical subsystems, mechanical failures still create unplanned maintenance, long turbine downtimes due to gearbox replacement, and rebuild which increases the warranty reserves. Based on the data compiled in the Wind Stats Newsletter which covers from 2003 to 2009, gearboxes are the subsystem that causes the highest downtimes (Figure 1) [5, 6]. Wind industry is experiencing higher gearbox failure rates than other due to underestimation of true operating loads, unexpected overloads due to unusual operating conditions, defective design of gearbox components, and poor maintenance, all of which have clear effects on gearbox reliability. From the O&M point of view, the main concerns are the following: high replacement costs after failure, high costs of the removal and reinstallation works because a crane is needed, and high revenue losses due to the long downtime [7]. All the previous factors trigger an increase in the cost of the energy as previously stated. In order to decrease this cost, turbine uptime has to be increased, a goal which can be reached by improving gearboxes reliability [8–13]. For this reason, the Vestas V90 gearbox is the main targeted subsystem of this paper.

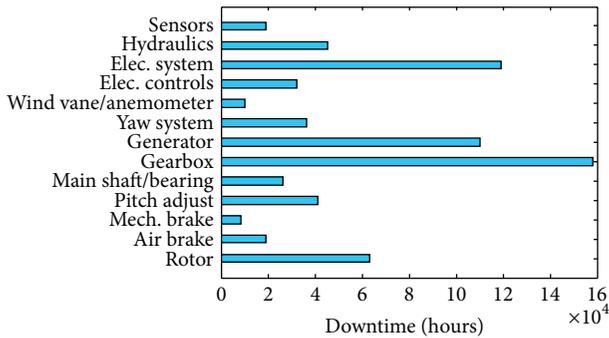


FIGURE 1: Wind Stats: 2003–2009 aggregated downtime per turbine subsystem.

Vibration-based monitoring of all drive train components is the most developed and widely applied condition monitoring technique by far [14–17]. A survey carried out by the UK Supergen Wind Energy Technologies Consortium [18] shows that 14 out of 20 commercially available wind turbine condition monitoring system (CMS) provide gearbox vibration monitoring. Another survey carried out by Durham University [19] shows that 27 out of 36 widely available CMS are based on drivetrain vibration analysis. Oil-based analysis CMS are also used for gearbox monitoring purposes. This technology is in an early development stage as regards sensor technology and the validation of its capabilities for fault detection [20]. Acoustic Emission is becoming a typical drivetrain CM practice. A recent study on wind turbine CM shows a design of a new continuous condition monitoring system with automated warnings based on a combination of vibration and AE analysis [21, 22]. AE can detect faults earlier than vibration [23, 24] but implementation is more complicated due to equipment noise which reduces the quality of the output signals [25]. SCADA systems are also used for condition monitoring of wind turbines although they are not a standalone solution. They store 10 minute-average data which is valuable to help other CMS achieve better fault detection performances [7].

The technique applied in this paper is vibration analysis using accelerometers mounted in specific locations on the machine. The reason vibration analysis is used instead of one of the aforementioned methods is the advantages this technique offers. In comparison to oil analysis, vibration analysis can detect failures inside and outside the gearbox and can be used for online monitoring without incurring high costs [16]. Additionally, small defects generate intense vibration signatures which are detectable by the accelerometer. However they do not necessarily cause changes in the chemical composition [26].

The majority of existing commercially available vibration-based CMS use time domain, frequency domain, or time-frequency analysis methods. Vibrations in the time domain have been widely studied. Statistical and parameter-based methods have shown to be useful indicators of the system condition [27]. Minimum and maximum value, peak to peak value, root mean square (RMS), kurtosis, skewness, crest factor are all examples of time domain features. Comparisons

between the pristine status and the current have been carried out using trend analysis [28]. Features reflect component deterioration only when working under changeless operating conditions. Time synchronous averaging (TSA) and planet separation methods resample the vibration data synchronously with the shaft rotation for the sake of extracting periodic waveform from noisy data [29]. Amplitude demodulation has proven to be appropriate for detecting defects that produce impacting [9]. Most of the commercially available vibration-based CMS focus on the frequency domain analysis based on Fast Fourier Transform (FFT) [30–32], order analysis, cepstrum analysis [33], envelope analysis, and so forth, which demonstrated to be effective at detecting gear faults. FFT is suitable for analysing steady-state or quasi-steady vibrations with slowly varying frequencies [34, 35]. Therefore FFT is not appropriate for wind turbine monitoring due to their stochastic operation behavior (variable aerodynamic loads and extreme environmental conditions which make the running speed change continuously). Vibrations coming from the wind turbine's gear or bearing are considered as nonstationary and nonlinear. To deal with nonstationary and nonlinear signals, time-frequency methods such as the Short-Time Fourier Transform (STFT) [36], Wigner-Ville distribution (WVD) [37], or wavelet transform (WT) are required [38, 39]. Wavelet analysis is probably the most popular technique. Continuous wavelet transform, discrete wavelet transform, and harmonic wavelet transform have been validated as methods for wind turbine condition monitoring [40–43]. Wavelet transform methods have the drawback that the basic functions of the decompositions are fixed and do not necessarily match the varying nature of the signals [44]. Signal decomposition methods have been applied to the practical gear and rolling bearing fault diagnosis. Empirical Mode Decomposition (EMD) has been presented as a more advanced condition monitoring technique to ensure the high availability of wind turbines [45]. The main downsides of EMD include the use of cubic spline to interpolate the local extrema of the signal which experiences end effect, mode mixing, overshoot, and undershoot problems [46, 47]. Local Mean Decomposition (LMD) is an improved version of EMD which also suffers from mode mixing, distorted components, and time consuming decomposition. LMD can be more effective than EMD in capturing local features of the nonlinear and nonstationary signals [48]. The Intrinsic Time-scale Decomposition (ITD) method overcomes the limitations of EMD listed earlier, as well as those previously mentioned, and is associated with more classical approaches such as Fourier and wavelets [49]. ITD decompose the nonstationary signal into a sum of proper rotation components (PRs) which are not real monocomponents. This causes erroneous instantaneous characteristics.

The existing work has some limitations. Most of the existing methods applied for fault detection in rotating machinery are signal processing based and they consist of finding the faulty peak in the frequency spectrum. Fault detection is carried out by comparison of the pristine and faulty frequency spectrum which is sensitive to noise methodology [30]. The task of identifying faulty peaks has to be done offline which falls into higher costs of labour and computational

complexity. In addition, offline analysis of the data gathered does not provide any benefit as faults are identified after they occurred. As is well known, practical vibrations are nonlinear and nonstationary due to the complexity of rotating machinery systems [48]. Fast Fourier Transform (FFT) did not show good result when it was applied to that type of complex vibrations. Nonadaptive time-frequency methods do not provide with a meaningful interpretation either [50–52]. Consequently, a novel methodology for automated fault diagnosis in rotating machinery is needed to avoid loss of time in planning and carrying out unnecessary operation and maintenance (O&M) tasks, reduce machinery downtime, increase reliability, and reduce the cost of energy (COE).

The algorithms and methodology presented in this paper highlight some of the limitations of vibration analysis when it is applied for fault diagnosis. The approach uses a combination of time and time-frequency adaptive techniques to develop an automated solution for condition monitoring of wind turbines which contributes to the state of the art. Automation is a must in industrial applications such as wind turbine monitoring to reduce the manual inspection of data. These algorithms are based on unsupervised learning that uses statistical process control (SPC) charts for defining the healthy status of the machinery. Signals considered as faulty are processed using the intrinsic characteristic-scale decomposition (ICD) method. ICD is a relatively novel adaptive method which has been first proposed by Li et al. [53]. With ICD, any nonlinear and nonstationary signal can be decomposed into a collection of product components (PCs). The envelope spectrum of these PCs can identify the faulty frequencies in a gearbox. This proposed methodology is explained in this paper.

2. Novel Software Concept for Fault Detection in Wind Turbines

The application of vibration analysis is well developed in the field of rotating machinery. In spite of the fact that industrial machinery is monitored using vibration analysis, the industry still experiences premature failures. Failure in machinery implies an increase of the downtime and subsequently an increase of the cost of the energy (COE). The current systems used to have full remote control of the individual wind turbines or even an entire wind farm health are SCADA (Supervisory Control And Data Acquisition) which offer full remote control and supervision of individual wind turbines. SCADA systems give an overview of all the relevant features to monitor any health changes of wind turbines such as temperature, pitch angle, electrical parameters, and rotor speed. These data gathering systems provide real time and historical information. The real time and historical data are compared to assess whether or not there is a failure arising or happening.

A new methodology which combines operating and vibration data for condition monitoring of wind turbines in rotating machinery is showcased in this paper. The aim is to propose a novel solution that helps the wind industry improve the reliability and availability of wind turbines. The objectives

to reach with the development and implementation of this novel methodology are as follows:

- (i) Establishment of a baseline representative of the healthy status of the wind turbine.
- (ii) Setting of operational limits of the machine being studied.
- (iii) Identification of deviations (outliers) from the baseline which could be indicative of machinery failure or malfunctions which are being developed within it.
- (iv) Application of a time-frequency domain technique for discarding any false alarm and identifying the source of defect if any.
- (v) Database creation for condition monitoring management purposes.

The algorithm developed to reach the aforementioned objectives can be split into the following four modules: baseline generation, forcing frequencies calculation, intrinsic-scale decomposition (ICD), and data base.

Figure 2 depicts the flow chart novel methodology for condition monitoring of wind turbines gearboxes proposed in this paper. The algorithm is recognized as two lines which work in parallel, coming together in a single endpoint. The proposed method starts gathering vibration data and takes 10-minute averages of the speed of the low speed shaft (LSS) and power output continuously. Then, vibrational data passes through a signal conditioning stage and the rotor speed is used for calculating the forcing frequencies of the machine being studied. It is important to know the speed of the LSS as some wind turbines and other pieces of rotating machinery work at different speeds. Thus, the method selected for storing the data gathered is based on bins which are related to the speed of the LSS at the time the measurement is taken. The number of bins is selected by the operator whereas the bin width is established according to the maximum speed of the machine and the number of bins previously mentioned.

Research in the laboratory was carried out to select the feature that depicts better and more reliable changes in the operational status of gearboxes. Counts parameter, root mean square (RMS), peak value, and crest factor were studied. Peak value showed an increase with the onset of defect in gearboxes but sometimes that feature was badly affected by single noise spikes. For that reason, peak value and crest factor were considered not reliable indicators of defect. Counts parameter was not considered reliable either as it has a strong dependence on the threshold value selected for its calculation. RMS levels increased with the increase of defect within the gearbox studied and results were not affected by single noise spikes as much as the peak value. Therefore, RMS was selected as the feature to be binned. Depending on the application, different features might contain more relevant information for processing with monitoring systems [54]. Then the method depicted in Figure 2 is able to generate a baseline representative of the normal operation conditions of a wind turbine gearbox and use that baseline to assess the current health of the gearbox. At this stage, deviations from the baseline may or may not be an indication of defect. Accordingly, a novel time-frequency domain technique is applied to the data which

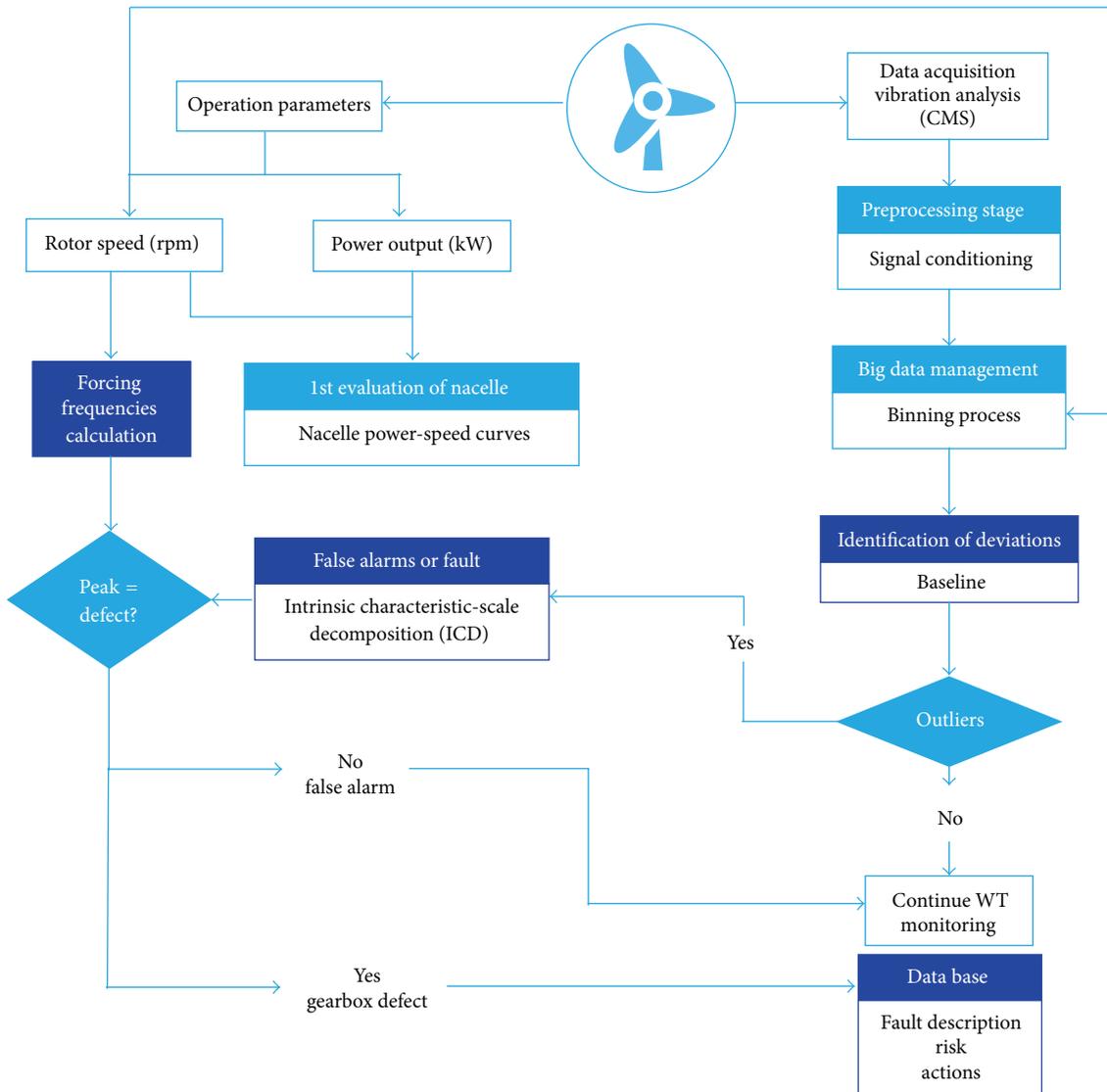


FIGURE 2: Flow chart of a novel condition monitoring algorithm for gearbox fault detection.

is causing deviations, that is, to the outliers. The technique, which is called intrinsic characteristic-scale decomposition (Figure 4), demodulates the signals until the faulty frequency is easy to identify. The output of the ICD is compared with the forcing frequencies previously calculated to conclude whether or not there is a defect. If there is a match between the ICD and the frequencies calculated, the source of defect is identified and the information is recorded in a database to evaluate the risk and the actions which need to be taken. On the other hand, if there is no match, the outlier is considered as a false alarm and the monitoring process continues.

The main parts within the algorithm described in Figure 2, the baseline generation and the demodulation technique (ICD), are explained extensively in this section.

2.1. Identification of Deviations from Normal Operation Conditions Based on a Baseline. The baseline is a statistical process control that can be used when a large number of similar files

are being produced, such as data coming from the wind turbine. Wind turbine monitoring is subject to variability. The variability present when a process is running well is called short term or inherent variability and is usually measured by the standard deviation. The vibration signatures files generated will have a target value. When defects appear, this value increases or decreases depending on the feature selected for the baseline generation. This method falls within the machine learning field of research. It is classified as an unsupervised parametric learning method [55] because the source of defect is unknown and it is based on the statistical representation of vibration features. It acts as a unary classification; that is, it establishes discrimination criteria by learning from a training set in order to identify whether the data belongs to pristine condition of the method or to a novel class.

The purpose of generating a baseline with the wind turbine data is to produce an alarm when the process values have moved away from the target. A second purpose is to generate

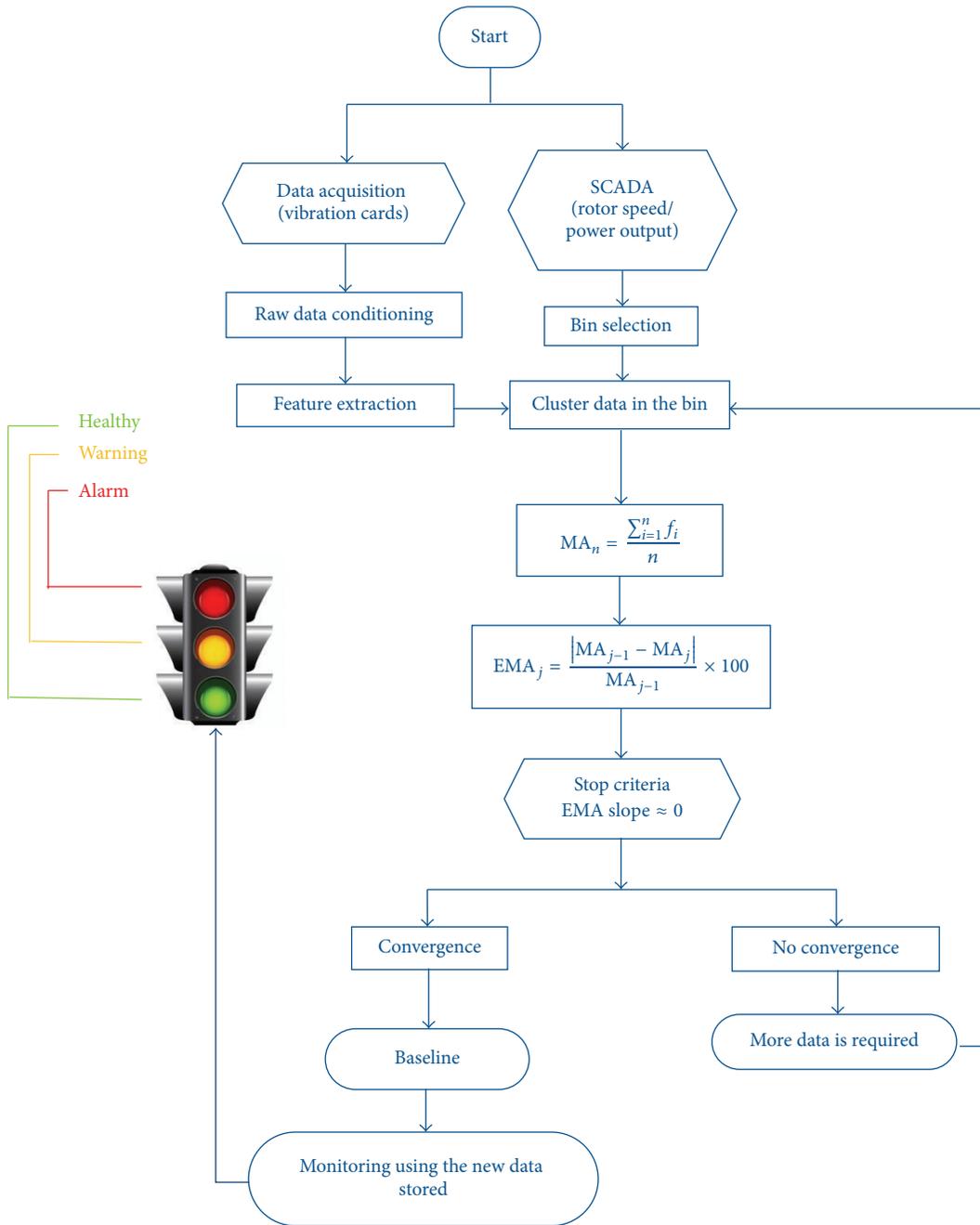


FIGURE 3: Flow chart for the baseline generation algorithm.

a signal when item to item variability has increased. In either case appropriate action must be taken by a machine operator or an engineer. Statistics can only give the signal; the on action relies on other skills. Setting the alarm limits based on statistics enables all alarm levels on a complex machine, such as wind turbine with many transducers, to be calculated in seconds instead of taking many hours to view the data and set the levels manually. Thus, the monitoring process can be carried out more economically. Automation of the alarm setting process can greatly improve the efficiency of the monitoring process. Figure 3 shows a more detailed flow chart

of the baseline generation method. After having binned the features extracted in the suitable bin according to the rotor speed, the method needs to establish when the data available is enough to determine the normal operation limits of the wind turbine gearbox. Obviously, the more data the better to establish reliable alarm limits. However, it entails spending more time learning from the machine which is not the best practice when early fault detection is a goal. Therefore, a compromise between reliability and time efficiency is needed. A convergence criterion to ensure that the normal operation limits are calculated using the minimum number of files is

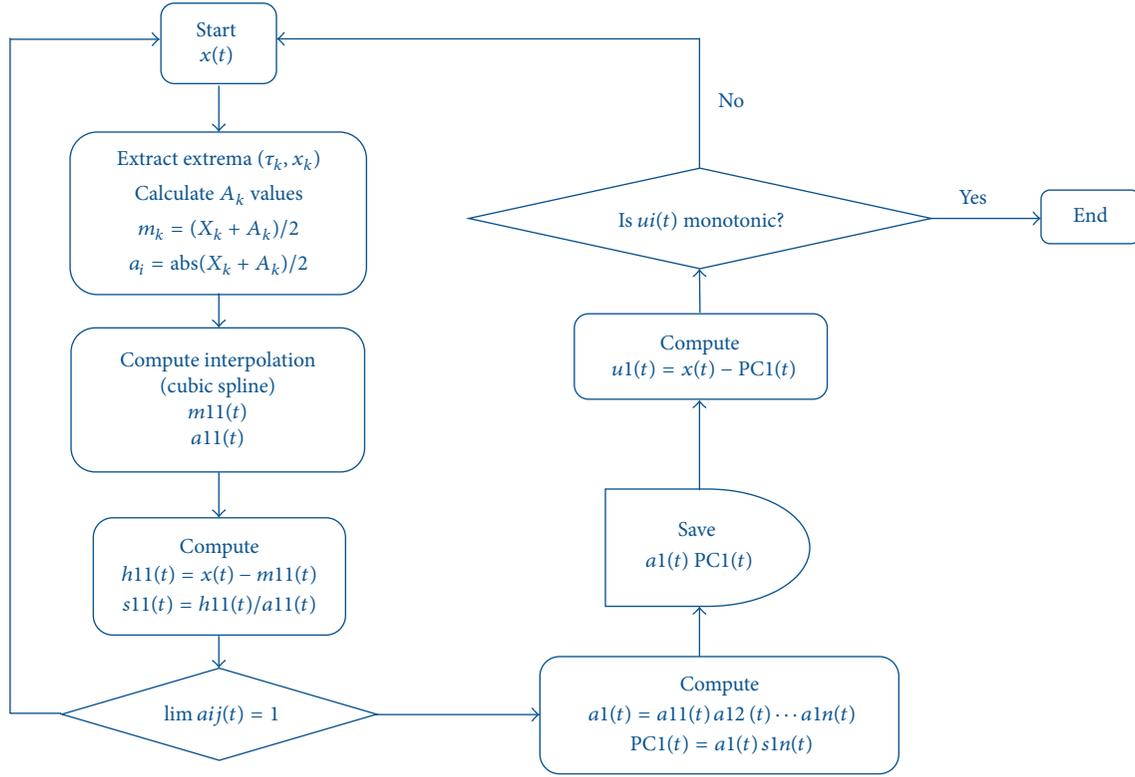


FIGURE 4: Flow chart of the ICD method.

presented in this paper. In this way, the wind turbine gearbox health can start to be assessed as soon as possible. Due to different variability of the data in different bins, the number of files required will vary for each bin. Therefore, convergence is studied separately in every single bin.

The process starts calculating the moving average (MA) of the data stored in the bin being studied. MA is a trustworthy parameter which indicates when the addition of more data will not have any effect on the calculation of the alarm limits. The moving average error (EMA) between consecutive files is calculated and it establishes the termination conditions. Considering the coordinates of the last EMA value calculated (x_n, y_n) and another one calculated earlier (x_{n-m+1}, y_{n-m+1}) , where m indicates the number of EMA points included between the last two, the slope of the straight-line connecting both can be figured. The slope is not calculated using the last two measurements as that would be more a short term convergence indicator rather than a medium or long term convergence indicator. Slope can be calculated using

$$\gamma = a \tan \left(\frac{y_n - y_{n-m+1}}{x_n - x_{n-m+1}} \right). \quad (1)$$

The stopping criterion is satisfied when $\gamma \approx 0^\circ$ which means that the straight-line between two measurements properly spaced in time is nearly horizontal. In that case, it can be ensured that the variability of the EMA when new data is added to that specific bin is that low that the normal operation limits can be established reliably. Limits allow determining

whether or not the process is under control. The purpose is to provide a representative vibration level across a range of rotor speeds, for improving the reliability of wind turbine gearboxes and thereby help define appropriate vibration limits. If the limits are too far from the target value, small deviations from the target value may go undetected, but if the limits are too close to the target value there will be a large number of false alarms (meaning there will be a signal for action when the process mean is on target and no action is necessary). The baseline is generated using the mean (μ) of the values required according to the stopping criterion previously explained and their standard deviation (σ). Two kinds of limits are established: warning limits and action limits. Before constructing the control chart, the probability level (α) value for the warning and alarm limits has to be selected. According to the literature related to statistical process control (SPC) charts [56, 57], it has been convenient to set the warning limits so that, if the mean is on target, 95% of sample means lie within them. Limits are calculated as the mean plus/minus p times the standard deviation of the data stored in a specific bin. The action limits are set so that 99.8% of samples lie within them when the mean is on target. Given the value of α , the number of standard deviations (p) used for the calculation of the upper control limits (UCL) and lower control limits (LCL) is extracted from the t distribution table, using an input value equal to $1 - \alpha/2$. Therefore, for the warning limits, p is 1.96 which means that only 5% of the values should be outside the limits if the features under study are in the same condition as when the signature was calculated. For

the action limits, p is 3.09. The warning and action limits are defined as follows:

$$\begin{aligned} \text{Warning Limits} &= \mu \pm 1.96\sigma, \\ \text{Alarm Limits} &= \mu \pm 3.09\sigma. \end{aligned} \quad (2)$$

The standard deviation will be an estimation of the true value and 95% and 99.8% are figures which have been selected according to the literature. For these reasons and for simplicity the limits are often set at

$$\begin{aligned} \text{Warning Limits} &= \mu \pm 2\sigma, \\ \text{Alarm Limits} &= \mu \pm 3\sigma. \end{aligned} \quad (3)$$

In this specific case, the target value is not known and it will depend on the gearbox which is being monitored. When this happens, the data clustered within each bin is taken and the sample mean is used as a target value. One should clarify that this is accomplish just when the process is running satisfactory.

Finally, the new data gathered is plotted in front of the limits previously determined to evaluate the health of the gearbox. New RMS values falling out of the limits are a clear indication of a process out of target. Nevertheless, it does not necessarily mean that is due to a malfunction or defect within the turbine. The classification process to distinguish between faults and false alarms is carried out by the module described in the following section. Three areas can be distinguished in the baseline plot: Zone A is define as the area where the turbine operates free of fault; Zone B is the area between the warning and alarm limits and where minor faults will appear; and Zone C is the area above the alarm limit and where acute fault to danger will be localized.

On the other hand, if the termination criterion is not satisfied in a bin, more data is required and the whole process needs to continue.

2.2. Gearbox Fault Diagnosis for Wind Turbines Based on ICD Frequency Spectrum. The time series of abnormal events are subjected to a time-frequency signal processing technique which is suitable for nonlinear and nonstationary vibrations such as faulty gearbox signals. In a defective gear or rolling element the resonance frequency is excited and it is modulated by the transient impulse with defective frequency. The intrinsic characteristic-scale decomposition method (ICD) decomposes the nonstationary signal into a series of product components (PCs) so it is possible to diagnose gearbox and rolling elements defects. This method is needed in order to demodulate the feature and extract the fault frequency by carrying out envelope spectrum analysis of the first few PCs. Therefore, the output of this module will be the frequency spectrum of the first enveloped PCs which can identify the faulty frequencies in gearbox and rolling elements successfully. This method has been effectively applied by Li et al. [53] for faulty frequencies identification. The results demonstrate that ICD can get accurate monocomponents when the method is applied to broken tooth detection in gears or inner race defects detection in bearings.

TABLE 1: Specifications of the test's rig gearbox.

Ratio	1.5 : 1
Pinion teeth	18
Gear teeth	27
Forcing frequencies	Orders
Input frequency (f_{in})	1x
Output frequency (f_{out})	0.67x
Gear mesh frequency (GMF)	18x

Once the spectrum of the enveloped PCs is obtained, the remaining peaks in the FFT are compared with the forcing frequencies of the gearbox. If any of the frequency calculated correlates to any peak in the envelope spectrum coming from the ICD method, the false alarm hypothesis would be discarded and the source of defect would be identified. On the contrary, if there is a lack of correlation between the theoretical values and the envelope spectrum, the outlier being analysed is considered as a false alarm.

3. Laboratory Trials on a Test Rig

Experiments were conducted at laboratory scale to validate the methodology above explained. In this section, the layout and results obtained after applying the algorithm described in Figure 2 are presented.

3.1. Laboratory Experiments Set-Up. To verify the effectiveness of the proposed methodology, a couple of case studies were conducted at laboratory scale at the Machinery Fault Simulator from Spectraquest Inc. The aim was to recreate the stochastic behaviour of wind turbines by gathering data at different speeds so the results of this experiment can be useful for showcasing the applicability of the method to field trials in a wind turbine gearbox.

The test rig, utilized for the purpose of dataset collection, is shown in Figure 5(a). It consists of a 0.75 kW variable frequency drive (VFD) which controls the motor shown in Figure 5(a). In the first experiment, the motor is connected with the universal joint shaft to the single stage gearbox with spiral bevel gears. One of the two gears is shown in Figure 5(b). Figure 6 depicts a diagram of the gearbox. The specifications of the gearbox are shown in Table 1. Vibrations generated by this subsystem are measured in the radial direction, using a CTC AC150-1 accelerometer mounted vertically on the case. It is an industrial type IEPE accelerometer whose sensitivity is 100 mV/g and its frequency range goes from 1 Hz until 10 kHz. Input shaft speed is measured with a noncontacting proximity probe.

Two gearboxes were tested. The first gearbox was free of faults whereas the second one was in faulty conditions. Figure 5(b) shows the faulty gear from the second gearbox which is missing a tooth. Tests were performed in both gearboxes at variable speeds between 25 Hz and 45 Hz using a 5 Hz step.

For the second experiment, the motor is connected to a 91.4 cm shaft supported by one Rexnord ERI2K ball bearing which is mounted on a stand as shown in Figure 7. Damage

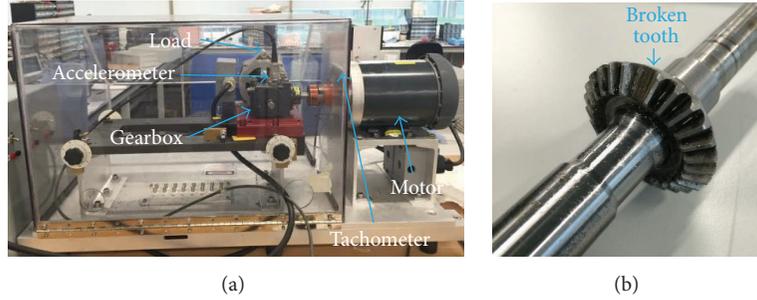


FIGURE 5: Layout of the test rig: (a) sensor attachment and (b) the faulty gear with a broken tooth.

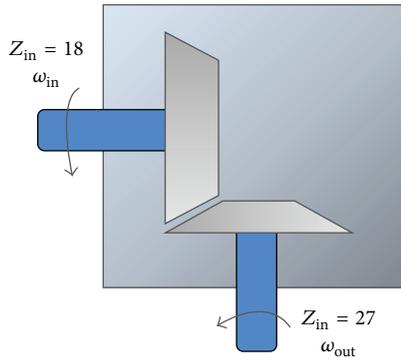


FIGURE 6: Diagram of the test rig's gearbox.

is introduced by replacing a healthy bearing by a predamaged specimen. The damage cases include ball spalling, inner race, and outer race. Li et al. [53] applied the ICD method to a bearing with inner race defect. In order to make a contribution to the current knowledge a bearing with an outer race defect was utilized for the experiments. In this way, the ICD method can be validated for another type of roller bearing fault. The forcing frequency (BPFO) can be calculated introducing the parameters defined in Table 2 into the following equation:

$$\text{BPFO} = \left(\frac{N_b}{2}\right) \left(1 - \frac{N_b}{d_p} \cos(\alpha)\right). \quad (4)$$

The value of the abovementioned forcing frequency expressed in orders is $3.044x$, where x is the rotating frequency at the gearbox input.

3.2. Results Analysis. The results from the laboratory experiments performed in a test rig gearbox are presented in Figure 8. First of all, healthy data was gathered to establish the normal operation condition limits. The system decides autonomously when the data gathered is enough for establishing these limits based on the convergence criteria previously described.

Convergence plots for each bin are shown in Figure 8. These plots reveal the effect of adding a new measurement to the current data. The convergence criteria marks with a light blue dotted line when the learning process is finished in each

TABLE 2: Specifications of the test's rig bearings.

MB ER12K bearing Rexnord	
Outer diameter D (mm)	47
Inner diameter d (mm)	19
Pitch diameter d_p (mm)	33.5
Ball diameter d_b (mm)	7.9
Outer ring width B (mm)	15.8
Number of ball N_b	8
Contact angle α ($^\circ$)	0

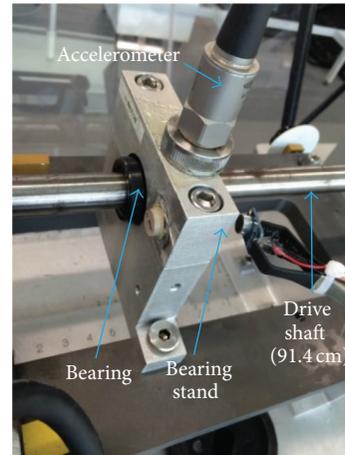


FIGURE 7: Experimental set-up with a faulty bearing.

bin, which does not generally happen simultaneously in all the bins. The number of files needed to determine the baseline limits reliably is depicted in Table 3. Following on, warning and alarm limits are calculated and plotted as dark and light blue lines, respectively (Figure 9).

A new set of 50 measurements was gathered for each bin when the gearbox was working in pristine conditions. This new set of data validates the capabilities of the normal operation limits calculated. The RMS values extracted from the vibration signals are plotted as dark blue points in Figure 9. Since the data was acquired in healthy conditions, features should fall within the limits, revealing that the gearbox status has not changed. The effectiveness of the baseline is evaluated by calculating the percentage of data that is within

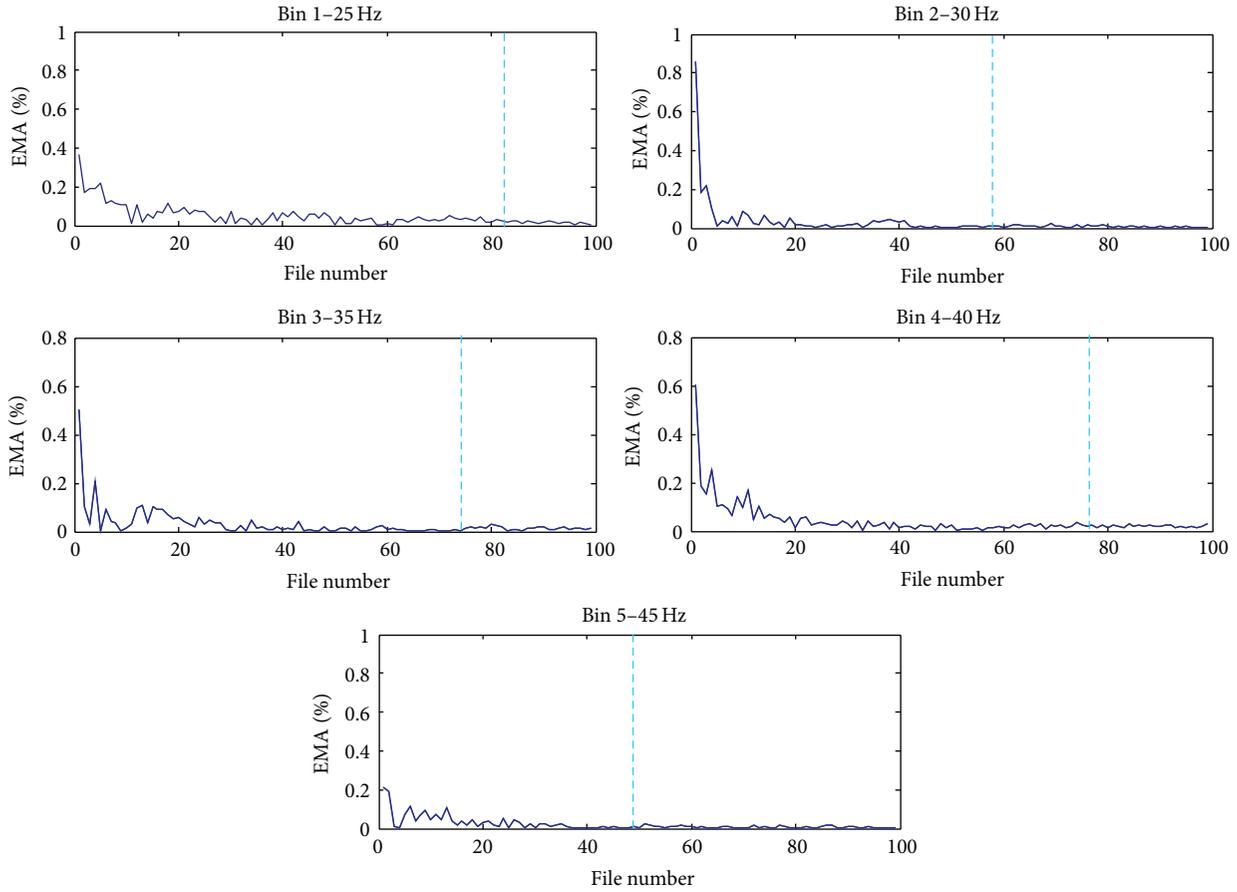


FIGURE 8: Vibration data convergence.

TABLE 3: Number of files for the baseline generation and effectiveness probed with an additional set of data.

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Rotation speed (Hz)	25	30	35	40	45
Number of files	84	58	72	77	48
Number of outliers	0	0	2	1	2
Data within the limits (%)	100	100	96	98	96

the limits. There are 2 outliers out of 50 in bins 3 and 5 whereas there is just one in bin 4. The algorithm assesses the gearbox as healthy for bins 1 and 2 in all cases. The outliers are located between the alarm and the warning limits. Therefore, the quality of the current results has to be assessed according to the 95% probability factor defined in the previous section of this paper. It can be concluded that the results are promising as the percentage of data lying within the limits is higher than 95%.

The healthy gearbox was replaced by the one with a broken tooth and new sets of 50 measurements were taken for each bin. The extracted RMS values are plotted as light blue

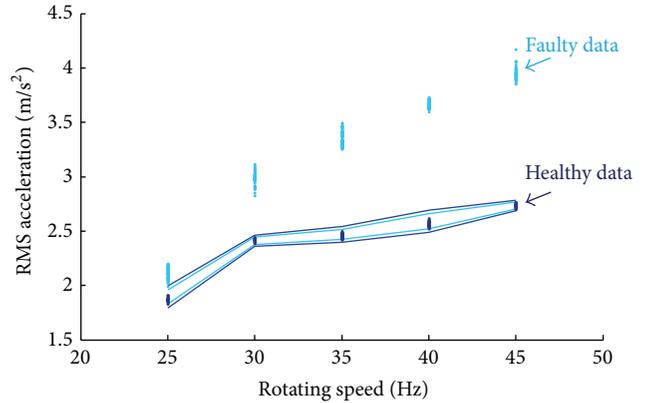


FIGURE 9: Baseline of the test rig operating at variable speed using the RMS value of the acceleration.

points in Figure 9. As can be seen, the new values are deviated from the targeted mean. The deviation in bin 1 is not big and it can even be seen that one of the points falls between the upper alarm and warning limits. However, for higher speeds all the points are outliers. In addition, positive correlation between the deviation from the targeted mean and the rotation speed has been noted. In conclusion, the baseline algorithm is experimentally demonstrated to be able

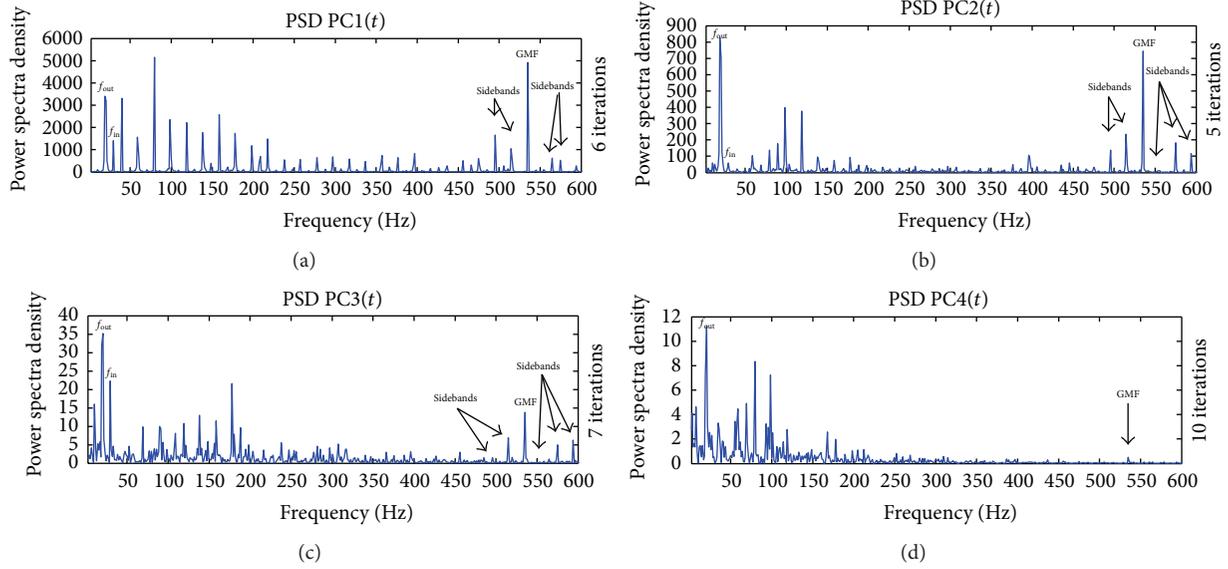


FIGURE 10: Envelope spectra of the obtained components from ICD (gearbox): (a) $PC1(t)$, (b) $PC2$, (c) $PC3(t)$, and (d) $PC4(t)$.

to distinguish between the healthy and the faulty condition of a gearbox successfully.

To verify the effectiveness of the ICD method revealing the source of defect, the algorithm is applied to one of the outliers falling in bin 2. That file was gathered when the variable frequency drive was running at 30 Hz in faulty conditions (Figure 5(b)). Given that the broken tooth was in the pinion, which is mounted on the output shaft, the frequency of interest is the rotational frequency of the output shaft ($f_{out} = 20.1$ Hz).

When the gear is operating with a localized fault, the vibration gathered can be classified as a multicomponent signal, amplitude-modulated (AM) and frequency-modulated (FM). It is not easy to find out which gear is faulty by simply performing the FFT. Therefore, further analysis in the time-frequency domain is required. The ICD algorithm is utilized to analyse the vibration signal coming from a gear in broken tooth fault condition. Firstly, the ICD method is applied to decompose the vibration signal in products. The decomposition process gives four PCs. Secondly, the PCs are enveloped and the frequency spectrum is performed to find the fault frequency. Results shown in Figure 10 are obtained after conducting the envelope spectrum analysis of the four PCs. As seen from Figures 10(a), 10(b), and 10(c) corresponding to the envelope spectrum of $PC1(t)$, $PC2(t)$, and $PC3(t)$, there are obvious spectrum lines of the faulty frequency ($f_{out} = 20.1$ Hz) and the gear mesh frequency (GMF = 540 Hz) and sidebands around it spaced a distance equal to the faulty frequency (499.8 Hz, 519.9 Hz, 560.1, and 580.2 Hz). The amplitude of the fault peak is lower than the amplitude of the peak related to the rotating speed. It can be confusing as someone could think that the defective gear is the one at the input of the gearbox when it is the opposite. As shown in Figure 10, the spectrum of the enveloped signal is cleaner with each new

iteration of the method but it stills keeping the frequency of interest (f_{out}). Figure 10(d) shows the spectrum of the last PC derived from the ICD method. This is the clearest PC spectrum since the number of peaks has been sufficiently reduced and there is a peak which is bigger than the rest. This peak, located at 20.1 Hz, brings to light the source of defect which is a broken tooth at the pinion. In the light of the results, the ICD method simplifies the defect identification process and consequently easing the automation of this type of processes.

Bearings are another important source of failure within gearboxes. Vibrations coming from bearings are AM-FM multicomponent signals which do not show the source of defect clearly. The features of the fault are modulated and its extraction is complicated. Therefore, the ICD method is applied to showcase its fault detection capabilities in other gearbox components. In this case, the aim is to detect an outer race defect in a roller ball bearing. As demodulation is needed, the signal is decomposed and the fault type identification is carried out by applying envelope spectra. Figure 11(a) shows the frequency spectrum of an outer race defect signal before any demodulation is applied. The rotation speed was 20 Hz at the acquisition moment. As can be seen, the governing frequency is the running speed set up in the VFD ($f_1 = 20$ Hz). The faulty feature can be identified in the spectra in spite of the fact that its amplitude is much lower (BPFO = 60.89 Hz). In addition, a clearest indication of the source of defect is needed due to the fact that the BPFO could be confused with the third harmonic of the rotating speed. ICD is applied to this fault signal and the decomposed products are shown in Figures 11(b) and 11(c). After two iterations, the BPFO becomes noticeable being the peak with the highest amplitude of the frequency spectrum (Figure 11(b)). The ICD has demonstrated to be effective identifying outer race defects in bearings.

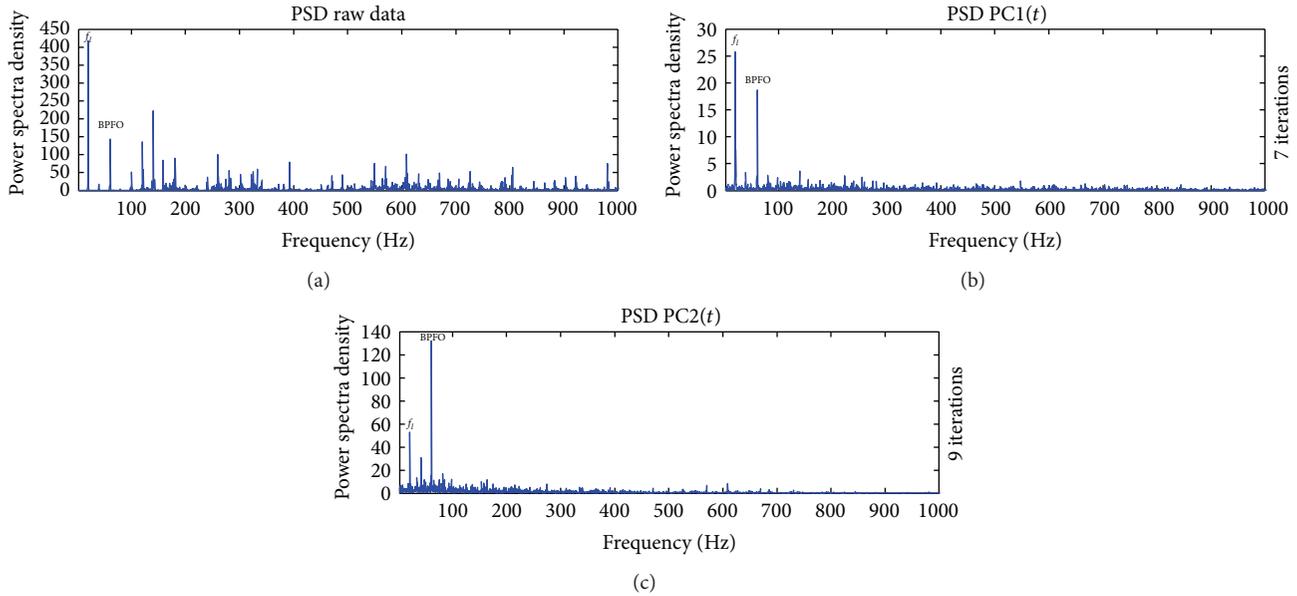


FIGURE 11: PSD of the raw signal and envelope spectra of the obtained components from ICD (bearing): (a) PSD raw data, (b) $PC1(t)$, and (c) $PC2(t)$.

4. Field Trials on a Vestas V90-3MW Wind Turbine

Experiments were conducted in a Vestas V90-3MW wind turbine to validate the methodology above explained on field. In this section, the set-up in the wind turbine gearbox and results obtained after applying the algorithm described in Figure 2 are presented. The aim of the experiments is to assess the health of the gearbox and if there is any sign of defect, identify the source.

4.1. Experimental Set-Up and Gearbox Analysis. The wind turbine selected for accomplishing the field trials was a Vestas V90-3MW which consist of a drive train (generator, main shaft, and gearbox), control system, three blades, yaw system, and the tower. Hereafter the wind turbine monitoring set-up is described. It consists of a set of three accelerometers, a data acquisition system (DAQ), referred to as CMS which was located in the inside of the nacelle and a data server (CMS server) located in the inside of the nacelle as well. The rotational speed signature (rpm) and power output (kW), which are recorded as an average of the last ten minutes, are supplied by the wind turbine operator.

The CMS utilizes vibration analysis for monitoring the gearbox of the wind turbine, which consists of two planetary gear stages and one helical gear stage, as shown in Figure 12. It is located in the middle of the wind turbine drive. Wind turbine gearboxes are used to increase the rotor speed to a speed suitable for the generation of electricity and operate under varying load conditions. Therefore, there is a low speed shaft (connected to the blades) and high speed shaft (connected to the generator) which are related through the gearbox (ratio 104.5582). Detailed information of this gearbox is provided in Table 4. The first stage is planetary, in which

TABLE 4: Vestas V90-3MW gear teeth information.

Gear stage	Number of gear teeth		
	Ring gear/gear	Planet	Sun/pinion
Planetary stage I	87	33	18
Planetary stage II	99	40	18
Helical stage III	91		33

TABLE 5: Rotational frequencies for each stage.

Gear stage	Frequency (orders)		
	Planet carrier/gear	Planet	Sun/pinion
Planetary stage I	1x	1.59x	5.8x
Planetary stage II	5.8x	8.5x	37.9x
Helical stage III	37.9x		104.5x

the planetary carrier (fixed by the ring gear) is the input, Table 4. The first stage is planetary, in which the planetary carrier (fixed by the ring gear) is the input and the sun gear is the output terminal. The second stage is equal to the first stage. The transmission ratio will not be the same as the number of teeth is different. The third stage consists of a gear which is connected to the output of the previous stage and a pinion. The pinion is at the output of the gearbox and determines the speed of the high speed shaft (HSS). The forcing frequencies to take into account when analysing gearbox data in the frequency domain are the rotational frequencies of the gears and the gear mesh frequency. These values are depicted in Tables 5 and 6, respectively.

For the application of vibration analysis to the wind turbine gearbox, three 3-axis piezoelectric accelerometers (Metra KS943 100 mV/g) with integrated built-in preamplifiers (IEPE standard) were used. Furthermore there is a need

TABLE 6: Forcing frequencies for each stage.

Gear mesh frequency (GMF)	Frequency (orders)
GMF stage I	104.91x
GMF stage II	681.98x
GMF stage III	3447.76x

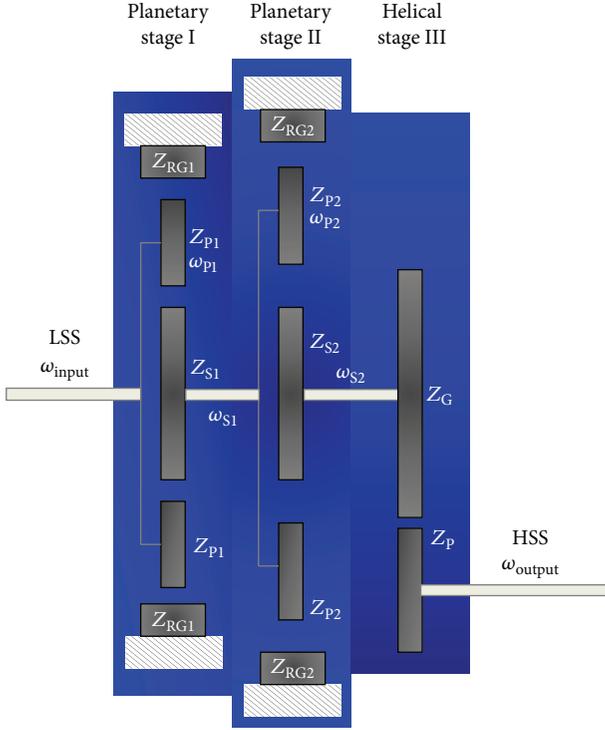


FIGURE 12: Sketch of the Vestas V90-3MW gearbox configuration.

for converters for each acceleration channel and a power supply of 24 V. Because the frequency range used is below 1 kHz, the sensors could be magnetically attached to the gearbox surface. Special magnets, which are available as accessories for this type of sensors, were used. The IEPE converters provide the data acquisition card with the voltage output of the 3-axis of each sensor which is proportional to the vibration measured by the transducers. The accelerometers were attached to the gearbox as shown in Figure 13: VIB1, first planetary stage; VIB2, second planetary stage, and VIB3, helical stage. Figure 13(b) shows the position of the sensor with respect to the global axis of the machine.

4.2. Results Analysis. The results from the field trials performed in a Vestas V90-3MW wind turbine between May and October 2015 are presented in Figure 14. A preliminary study based on the power curve for the Vestas V90 and the changes experienced by the features (RMS) extracted over time is presented. Then the baseline and ICD algorithms are applied to the field trials data for the health assessment of the monitored turbine.

The power curve for the Vestas V90-3MW, which shows the wind turbine's power output versus the rotor speed, is

shown in Figure 14. This is a performance indicator which can be used as first preventive maintenance indication that something is wrong. When a failure in the turbine occurs, power deviates from the normal power curve. In Figure 14, May and October power curves are superposed and no deviations are noticed. Therefore, this indicator designates no presence of malfunction at first sight.

Figure 15 depicts the clustering of the data features using a binning process based on the rotor speed, measured at the low speed shaft (16 rpm maximum). Figure 15 shows the number of files stored in each bin. Considering rotor speed as the binning parameter, the number of files increases from bin 18 to bin 30 which means that the rotor speed varied mostly between 10 rpm and 16 rpm during the months the system was installed in the wind turbine. The shape of the distribution makes sense bearing in mind that turbines operate at high speed for power generation purposes.

To get a better idea of how the data looks like and have a second performance indicator before the algorithm is applied, the RMS of the vibrations collected within the months of May and October are plotted in Figure 16. It gives a good overview of how the RMS changes versus time per bin. The RMS does not show large variations with time. However there are a couple of peaks which are of interest. These peaks are located in the first two bins and in the last one. The peaks recorded at low rotor speeds occurred at mid-August whereas the one captured at higher speeds happened at the beginning of September and its RMS value is around 8 m/s^2 . Further research revealed that those files which have higher RMS values were corrupted.

Following on, the baseline generation process and the health assessment based on its limits is presented. During the six months the field trials were taking place, the operator did not carry out any maintenance actions so it is assumed that the machinery is in a healthy status and the data collected is very valuable for the generation of a baseline. This baseline will mark the limits of normal operation conditions for this particular turbine. The baseline will be a useful tool for wind turbine operators so the maintenance actions are only carried out when deviations from the target value are noticed and verified as defect initiation, worsening or malfunction of the gearbox.

The convergence criterion is applied to all the bins. The results are depicted just for bin 30 though. Figure 17 shows how the EMA gets lower with every RMS value added to bin 30. The EMA in this bin starts dropping very fast from the beginning and converge after some measurement due to the fact that the EMA is close to 0%. The stopping criterion stipulates the minimum number of files as 34 (dotted line in Figure 17).

Figure 18 depicts the baseline obtained after calculating the minimum number of data necessary to establish the alarm and warning limits reliably in each bin. This baseline is unique for this turbine and is suitable for monitoring at low or high rotor speeds, covering the wind turbine through all its operational range. The number of files utilized for calculating the limits varies depending on the result obtained after evaluating the convergence criteria (see values in Table 7). The results

TABLE 7: Convergence per bin.

Bin number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Convergence	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Points used for baseline	41	31	41	45	45	37	7	5	12	5	10	9	9	9	7	10	9	19	98	143	47	69	55	56	39	35	31	39	27	30
Baseline finished									June	July	June							June	June	June	June	July	June	June	June	June	June	June	May	May

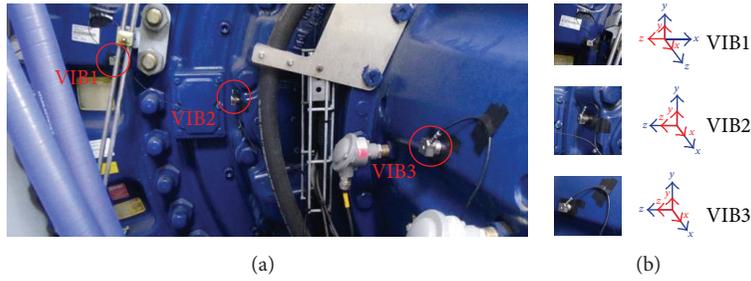
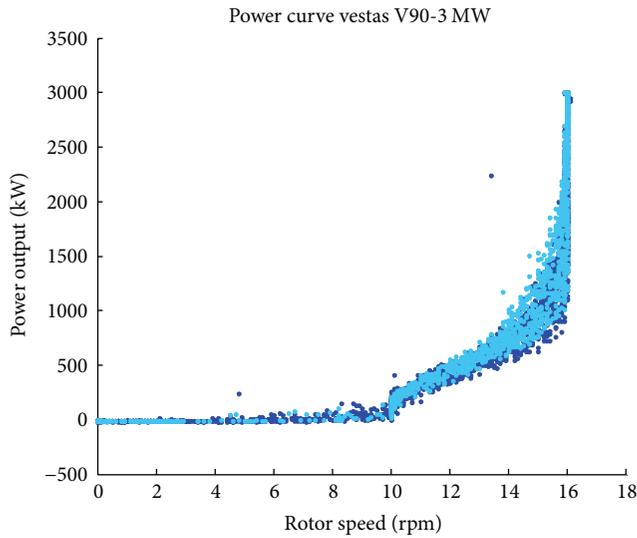


FIGURE 13: Sensors attachment on the gearbox: (a) set-up and (b) global (red) and local (blue) axis.



- May 2015
- October 2015

FIGURE 14: Scatter plot with power output as function of the rotor speed.

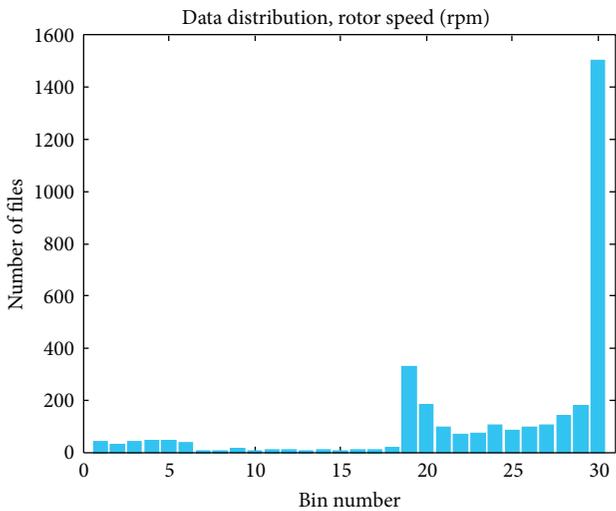


FIGURE 15: Distribution of the wind turbine data gathered using 30 bins.

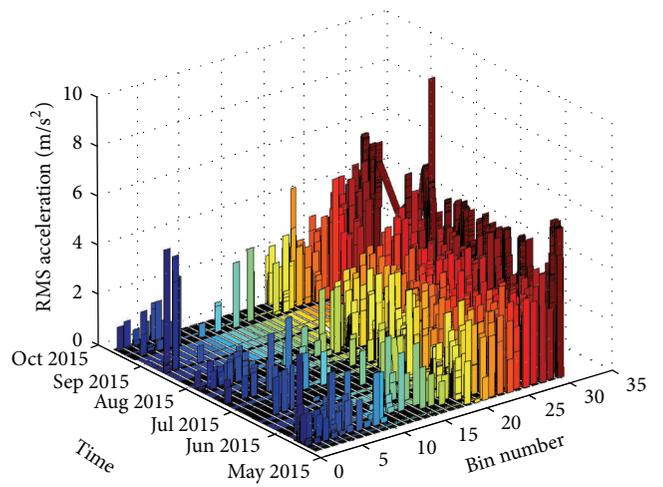


FIGURE 16: Variation on the RMS value of the acceleration per bin.

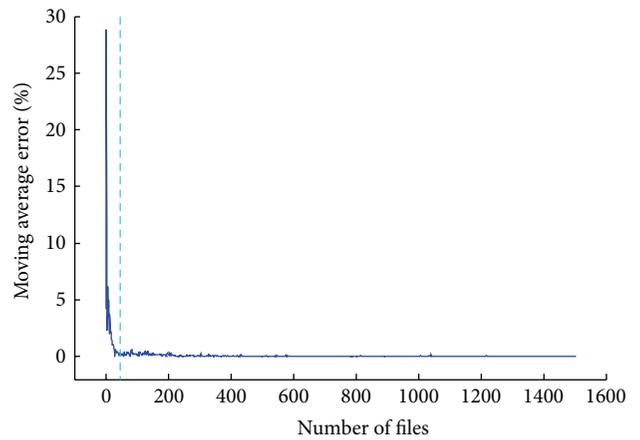


FIGURE 17: Convergence in bin 30.

point out that convergence was not reached in all the bins. According to the results, the fact of not reaching the convergence criterion in some bins is associated directly with the number of data clustered in them. As commented previously in the paper, the number of events recorded at low speed is very low. The method does not converge in the low speed bins, where there is not convergence of the method. It can be

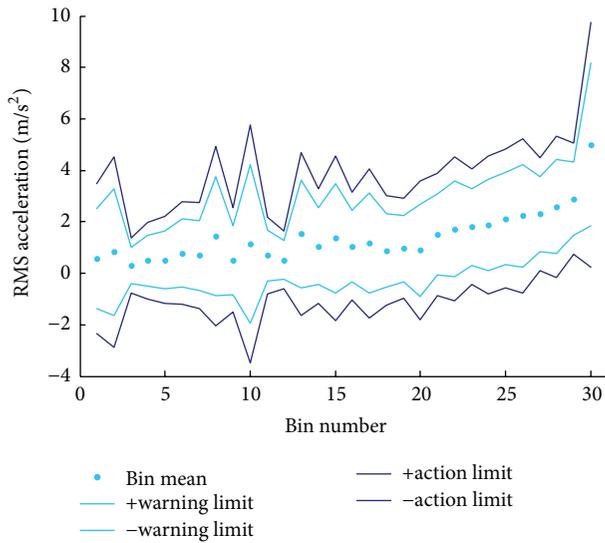


FIGURE 18: Baseline and targeted mean for a Vestas V90-3MW wind turbine.

noticed that the number of data clustered in them is less than 70 files. It does not mean that provisional values cannot be given to the limits. It means that more data is required in order to average the current value more and set the limits in a more confident way. At high rotor speeds the baseline has been properly defined, reaching the convergence in all the bins apart from bin 22 (which only contains 69 files). Most of the bins converged in June although some of them did it earlier (May) and others a bit later (July). Therefore, it can be concluded that a period of three months was necessary to define the baseline complying with the convergence criteria.

Figure 18 depicts the warning and action limits which define zones A, B, and C along with the targeted mean of the RMS values of the acceleration signals stored in each bin.

After defining the baseline, the machine goes into the monitoring stage. The data gathered after satisfying the convergence criterion is plotted along with the limits in Figure 19 to evaluate their deviation from the targeted mean. In this figure the limits still have the same values but there are two sets of data presented in different colours. The blue colour points out that convergence was not reached in those bins. On the other hand the green colour indicates the convergence within those bins and the points represent the data gathered after the baseline was defined. The results depicted in Figure 19 are very promising, specially bearing in mind that this data is totally independent of the data previously used for the limits establishment. The new data consist of more than 2000 RMS values which mostly fall within the limits. There are some outliers though. The percentage of outliers between the warning and alarm limits has been estimated: 0% in bin 20, 2% in bin 21, 1.6% in bin 26, and 2.5% in bin 30. Since it was chosen to add 2 times the standard deviation to the mean as a warning strategy a 5% of the data within each bin can fall outside the limits. The percentages are good as they are lower than a 5%. There are points above the alarm limits as well. For instance, that happens in bins 19, 29, and 30. The percentage of outliers

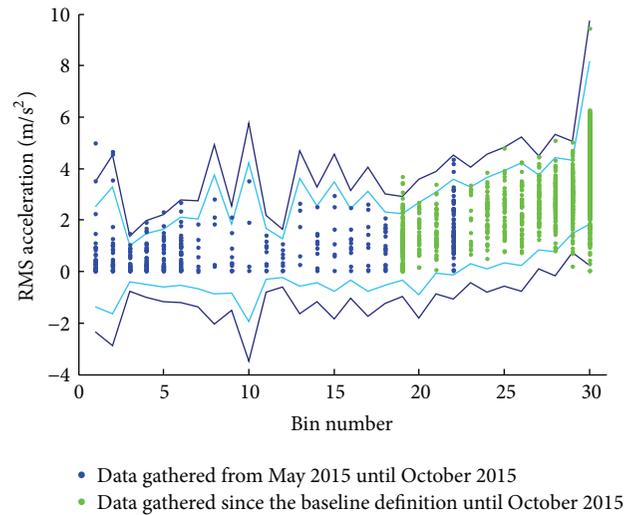


FIGURE 19: Monitoring using a 3-month established baseline.

in those bins is 0.4%, 0.3%, and 0.12%. Since it was chosen to add 3 times the standard deviation to the mean as a warning strategy a 0.2% of the data within each bin can fall outside the limits. The resulting percentage is good for bin 30 as it is lower than 0.2%. For the other two bins the percentages are acceptable as they are slightly bigger than 0.2%. Therefore the results shown in Figures 18 and 19 are a good tool for describing the acceptable levels of vibration of the wind turbine gearbox regarding the rotor speed of the wind turbine.

Outliers need to be further studied to find out the root cause of the deviations. One outlier per bin was analysed using the ICD method and the forcing frequencies were calculated according to the gearbox information provided in Table 6. The demodulation of the signal was carried out in six iterations which are shown in Figure 20. After $PC1(t)$ was obtained, the envelope spectrum of the signal shows the gear mesh frequency at the 3rd stage. This is the only peak that matches with the forcing frequencies theoretically calculated. In the following stages of the demodulation, the peak disappears which means that it was not a fault feature modulated by other frequencies. $PC2(t)$, $PC3(t)$, $PC4(t)$, $PC5(t)$, and $PC6(t)$ do not show explicit spectra lines of the fault frequencies of the gears. Therefore, the ICD does not show indication of gear defects within the gearbox.

5. Conclusions

This paper presents a novel approach for health condition monitoring and fault diagnosis in wind turbine gearboxes. The paper proposes baseline and ICD-based early fault diagnosis methodologies which are tested on laboratory and wind turbine vibration signals.

A baseline which represents the normal operation boundaries of the test rig and a Vestas V90-3MW have been produced. It is important to state that the resultant baseline from the laboratory trials has been validated and it has been accomplished with two sets of independent data (one healthy and another one from a gearbox with a broken tooth).

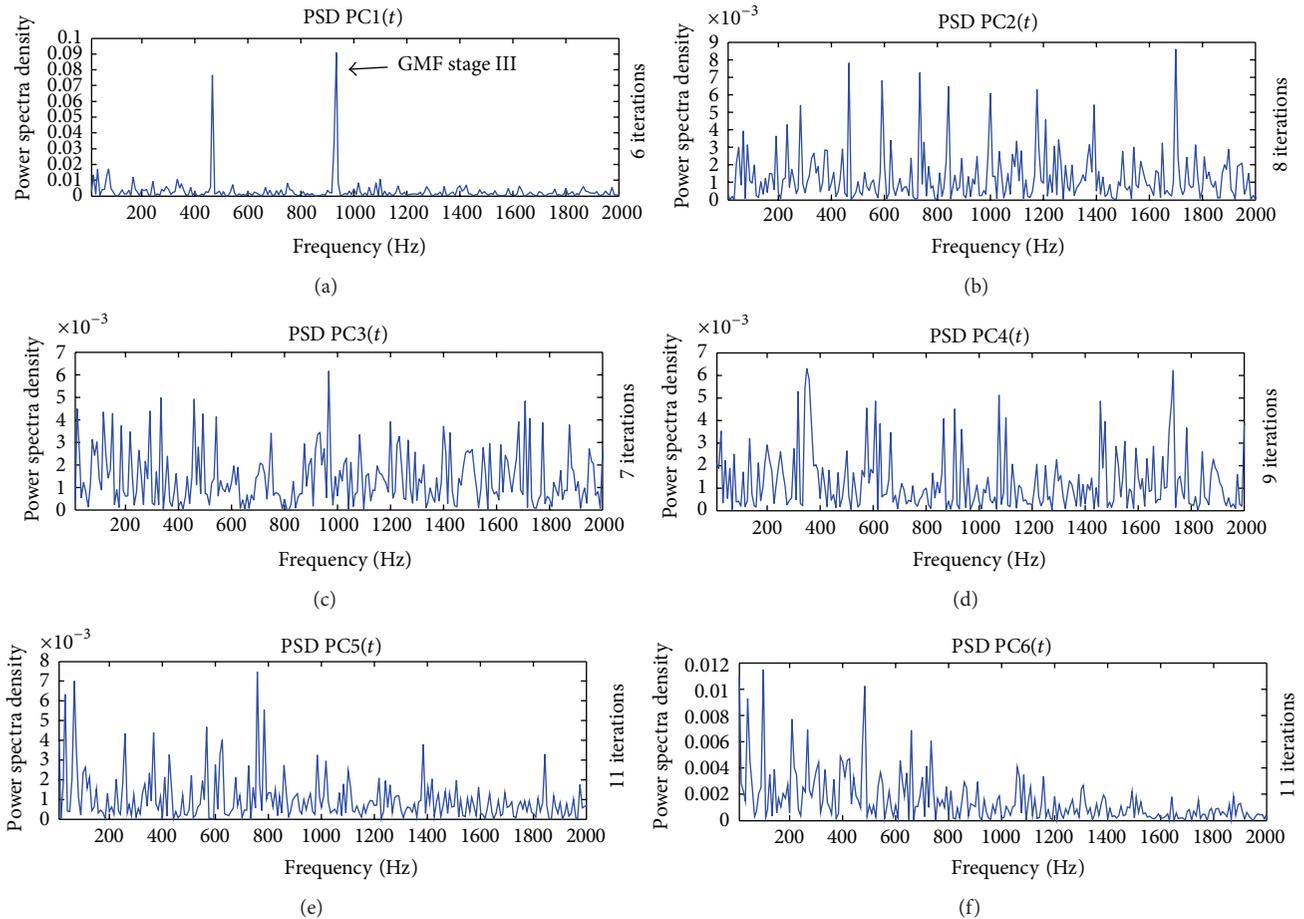


FIGURE 20: PSD of the envelope spectra of the obtained components from ICD (wind turbine data): (a) $PC1(t)$, (b) $PC2(t)$, (c) $PC3(t)$, (d) $PC4(t)$, (e) $PC5(t)$, and (f) $PC6(t)$.

Deviations from the targeted mean appear when the faulty set of data is utilized. The application of the baseline algorithm to the Vestas V90-3MW data establishes the range within the machine is running safely. Some outliers were identified at high running speeds, which is not very worrying as the outlier rate is lower or close to the value established by the probability factor. The validation expounds this methodology as a good practice for identifying those deviations which worries manufactures and wind turbine operators nowadays.

The process carried out for classifying the outliers at the laboratory demonstrates that the combination between the ICD method and the envelope spectrum analysis is superbly efficient for broken tooth and outer race fault diagnosis. The ICD method is experimentally demonstrated to be able to extract the early fault signature effectively. The algorithm was used to process the wind turbine outliers as well. However, the demodulated signal does not show any defect feature. This result, along with the output coming from the power curve indicator and the information provided by the operator regarding the health of the machinery, exhibits the lack of defect in the gears within the gearbox subsystem. Then, the outliers found at high speeds are considered false alarms.

The application of the ICD method simplifies the defect identification process due to the following two factors: the number of peaks in the spectrum is lower and the source of defect is shown as the peak with the biggest amplitude. Therefore, the ICD method eases the automation of the defect detection processes carried out by CMS which allows looking for the source of defect in real time.

Competing Interests

The authors declare that they have no competing interests.

References

- [1] S.-W. Choi, K. M. Farinholt, S. G. Taylor, A. Light-Marquez, and G. Park, "Damage identification of wind turbine blades using piezoelectric transducers," *Shock and Vibration*, vol. 2014, Article ID 430854, 9 pages, 2014.
- [2] S. Sheng and P. Veers, *Wind Turbine Drive Train Condition Monitoring—An Overview*, National Renewable Energy Laboratory, Golden, Colo, USA, 2011.
- [3] Y. Gui, Q. K. Han, Z. Li, and F. L. Chu, "Detection and localization of tooth breakage fault on wind turbine planetary gear

- system considering gear manufacturing errors,” *Shock and Vibration*, vol. 2014, Article ID 692347, 13 pages, 2014.
- [4] R. Moesgaard, H. Chandler, P. Barons, and G. Bakema, *Wind Energy—The Facts, The Economics of Wind Power*, Earthscan, 2009.
 - [5] S. Sheng, “Investigation of various wind turbine drivetrain condition monitoring techniques,” in *Proceedings of the Wind Turbine Reliability Workshop*, Albuquerque, NM, USA, August 2011.
 - [6] M. Strączkiewicz and T. Barszcz, “Application of artificial neural network for damage detection in planetary gearbox of wind turbine,” *Shock and Vibration*, vol. 2016, Article ID 4086324, 12 pages, 2016.
 - [7] Y. Feng, Y. Qiu, C. J. Crabtree, H. Long, and P. J. Tavner, “Monitoring wind turbine gearboxes,” *Wind Energy*, vol. 16, no. 5, pp. 728–740, 2013.
 - [8] S. Sheng, *Wind Turbine Drivetrain Condition Monitoring, Power Point Presentation*, NREL/NWTC, New York, NY, USA, 2011.
 - [9] F. P. García Márquez, A. M. Tobias, J. M. Pinar Pérez, and M. Papaalias, “Condition monitoring of wind turbines: techniques and methods,” *Renewable Energy*, vol. 46, pp. 169–178, 2012.
 - [10] F. P. Garcia, J. M. Pinar, A. Pliego, and M. Papaalias, “Identification of critical components of wind turbines using FTA over the time,” *Renewable Energy*, vol. 46, pp. 169–178, 2012.
 - [11] Z. Tian, T. Jin, B. Wu, and F. Ding, “Condition based maintenance optimization for wind power generation systems under continuous monitoring,” *Renewable Energy*, vol. 36, no. 5, pp. 1502–1509, 2011.
 - [12] C. C. Ciang, J.-R. Lee, and H.-J. Bang, “Structural health monitoring for a wind turbine system: a review of damage detection methods,” *Measurement Science and Technology*, vol. 19, no. 12, Article ID 122001, 20 pages, 2008.
 - [13] A. Kusiak and W. Li, “The prediction and diagnosis of wind turbine faults,” *Renewable Energy*, vol. 36, no. 1, pp. 16–23, 2011.
 - [14] D. Yang, H. Li, Y. Hu, J. Zhao, H. Xiao, and Y. Lan, “Vibration condition monitoring system for wind turbine bearings based on noise suppression with multi-point data fusion,” *Renewable Energy*, vol. 92, pp. 104–116, 2016.
 - [15] B. N. Madsen, *Condition monitoring of wind turbines by electric signature analysis [M.S. thesis]*, Technical University of Denmark, Copenhagen, Denmark, 2011.
 - [16] P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T. A. Tameghe, and G. Ekemb, “Wind turbine condition monitoring: state-of-the-art review, new trends, and future challenges,” *Energies*, vol. 7, no. 4, pp. 2595–2630, 2014.
 - [17] Z. Hameed, Y. S. Hong, Y. M. Cho, S. H. Ahn, and C. K. Song, “Condition monitoring and fault detection of wind turbines and related algorithms: a review,” *Renewable and Sustainable Energy Reviews*, vol. 13, no. 1, pp. 1–39, 2009.
 - [18] Noordzee Wind, “Operation Report 2009. Document No. OWEZ_R.000_20101112,” November 2010, http://www.noordzeewind.nl/wpcontent/uploads/2012/02/OWEZ_R.000_20101112_Operations_2009.pdf.
 - [19] C. J. Crabtree, D. Zappala, and P. J. Tavner, “Survey of commercially available condition monitoring systems for wind turbines,” Tech. Rep., Durham University, 2014.
 - [20] A. May, D. McMillan, and S. Thöns, “Economic analysis of condition monitoring systems for offshore wind turbine sub-systems,” *IET Renewable Power Generation*, vol. 9, no. 8, pp. 900–907, 2015.
 - [21] S. Soua, P. Van Lieshout, A. Perera, T.-H. Gan, and B. Bridge, “Determination of the combined vibrational and acoustic emission signature of a wind turbine gearbox and generator shaft in service as a pre-requisite for effective condition monitoring,” *Renewable Energy*, vol. 51, pp. 175–181, 2013.
 - [22] T. Toutountzakis and D. Mba, “Observations of acoustic emission activity during gear defect diagnosis,” *NDT & E International*, vol. 36, no. 7, pp. 471–477, 2003.
 - [23] C. Molina, “Contributions to the analysis of vibration and acoustic emissions for the condition monitoring of epicyclic gearboxes,” 2010.
 - [24] C. K. Tan, P. Irving, and D. Mba, “A comparative experimental study on the diagnostic and prognostic capabilities of acoustics emission, vibration and spectrometric oil analysis for spur gears,” *Mechanical Systems and Signal Processing*, vol. 21, no. 1, pp. 208–233, 2007.
 - [25] R. Ahmad and S. Kamaruddin, “An overview of time-based and condition-based maintenance in industrial application,” *Computers and Industrial Engineering*, vol. 63, no. 1, pp. 135–149, 2012.
 - [26] R. B. Randall, *Vibration-Based Condition Monitoring: Industrial, Aerospace and Automotive Applications*, John Wiley & Sons, New York, NY, USA, 2011.
 - [27] P. Vecer, M. Kreidl, and R. Smid, “Condition indicators for gearbox condition monitoring systems,” *Acta Polytechnica*, vol. 45, no. 6, pp. 3–17, 2005.
 - [28] K. Fischer, *Maintenance Management of Wind Power Systems by Means of Reliability-Centred Maintenance and Condition Monitoring Systems*, Chalmers University of Technology, Gothenburg, Sweden, 2012.
 - [29] S. Sheng, “Wind turbine gearbox condition monitoring round robin study—vibration analysis,” Tech. Rep., National Renewable Energy Laboratory, Denver, Colo, USA, 2012.
 - [30] D. F. Shi, L. S. Qu, and N. N. Gindy, “General interpolated fast fourier transform: a new tool for diagnosing large rotating machinery,” *Journal of Vibration and Acoustics*, vol. 127, no. 4, pp. 351–361, 2005.
 - [31] M. Inalpolat and A. Kahraman, “A dynamic model to predict modulation sidebands of a planetary gear set having manufacturing errors,” *Journal of Sound and Vibration*, vol. 329, no. 4, pp. 371–393, 2010.
 - [32] M. Inalpolat and A. Kahraman, “A theoretical and experimental investigation of modulation sidebands of planetary gear sets,” *Journal of Sound and Vibration*, vol. 323, no. 3–5, pp. 677–696, 2009.
 - [33] M. E. Badaoui, F. Guillet, and J. Danière, “New applications of the real cepstrum to gear signals, including definition of a robust fault indicator,” *Mechanical Systems and Signal Processing*, vol. 18, no. 5, pp. 1031–1046, 2004.
 - [34] Y. Itoh, T. Imazu, H. Nakamura, and T. Yamazaki, *Vibration Analysis Based on Time-Frequency Analysis with a Digital Filter: Application to Nonlinear System Identification*, Inter.noise, Melbourne, Australia, 2014.
 - [35] R. Bajric, N. Zuber, G. A. Skrimpas, and N. Mijatovic, “Feature extraction using discrete wavelet transform for gear fault diagnosis of wind turbine gearbox,” *Shock and Vibration*, vol. 2016, Article ID 6748469, 10 pages, 2016.
 - [36] L. Cohen, *Time Frequency Analysis*, vol. 778, Prentice Hall PTR, Englewood Cliffs, NJ, USA, 1995.

- [37] W. J. Staszewski, K. Worden, and G. R. Tomlinson, "Time-frequency analysis in gearbox fault detection using the Wigner-Ville distribution and pattern recognition," *Mechanical Systems and Signal Processing*, vol. 11, no. 5, pp. 673–692, 1997.
- [38] M. Timusk, M. Lipsett, and C. K. Mechefske, "Fault detection using transient machine signals," *Mechanical Systems and Signal Processing*, vol. 22, no. 7, pp. 1724–1749, 2008.
- [39] W. J. Staszewski and G. R. Tomlinson, "Application of the wavelet transform to fault detection in a spur gear," *Mechanical Systems and Signal Processing*, vol. 8, no. 3, pp. 289–307, 1994.
- [40] H.-J. Kou, H.-Q. Yuan, and X.-Y. Zhao, "Application of wavelet de-noising in fault diagnosis of wind turbine gearboxes," *Advanced Materials Research*, vol. 562-564, pp. 1091–1094, 2012.
- [41] F.-C. Zhou, "Research on online monitoring of gear fault of wind power gearbox based on undecimated wavelet transformation," *Advanced Materials Research*, vol. 383–390, pp. 4928–4931, 2012.
- [42] Y. Lu, J. Tang, and H. Luo, "Wind turbine gearbox fault detection using multiple sensors with features level data fusion," *Journal of Engineering for Gas Turbines and Power*, vol. 134, no. 4, Article ID 042501, 8 pages, 2012.
- [43] S. Yang, W. Li, and C. Wang, "The intelligent fault diagnosis of wind turbine gearbox based on artificial neural network," in *Proceedings of the International Conference on Condition Monitoring and Diagnosis (CMD '08)*, pp. 1327–1330, Beijing, China, April 2008.
- [44] I. Antoniadou, G. Manson, N. Dervilis, T. Barszcz, W. Staszewski, and K. Worden, "Condition monitoring of a wind turbine gearbox using the empirical mode decomposition method and outlier analysis," in *Proceedings of the 6th European Workshop on Structural Health Monitoring (EWSHM '12)*, pp. 1316–1323, July 2012.
- [45] W. Yang, R. Court, P. J. Tavner, and C. J. Crabtree, "Bivariate empirical mode decomposition and its contribution to wind turbine condition monitoring," *Journal of Sound and Vibration*, vol. 330, no. 15, pp. 3766–3782, 2011.
- [46] Y. Kopsinis and S. McLaughlin, "Investigation and performance enhancement of the empirical mode decomposition method based on a heuristic search optimization approach," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 1–13, 2008.
- [47] G. G. Pegram, M. C. Peel, and T. A. McMahon, "Empirical mode decomposition using rational splines: an application to rainfall time series," *Proceedings of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences*, vol. 464, no. 2094, pp. 1483–1501, 2008.
- [48] Y. Wang, Z. He, and Y. Zi, "A comparative study on the local mean decomposition and empirical mode decomposition and their applications to rotating machinery health diagnosis," *Journal of Vibration and Acoustics*, vol. 132, no. 2, Article ID 021010, 2010.
- [49] M. G. Frei and I. Osorio, "Intrinsic time-scale decomposition: time-frequency-energy analysis and real-time filtering of non-stationary signals," *Proceedings of the Royal Society of London A*, vol. 463, no. 2078, pp. 321–342, 2007.
- [50] G. T. Zheng and P. D. McFadden, "A time-frequency distribution for analysis of signals with transient components and its application to vibration analysis," *Journal of Vibration and Acoustics*, vol. 121, no. 3, pp. 328–333, 1999.
- [51] D. E. Newland, "Ridge and phase identification in the frequency analysis of transient signals by harmonic wavelets," *Journal of Vibration and Acoustics*, vol. 121, no. 2, pp. 149–155, 1999.
- [52] W.-X. Yang and P. W. Tse, "An advanced strategy for detecting impulses in mechanical signals," *Journal of Vibration and Acoustics*, vol. 127, no. 3, pp. 280–284, 2005.
- [53] Y. Li, M. Xu, Y. Wei, and W. Huang, "Rotating machine fault diagnosis based on intrinsic characteristic-scale decomposition," *Mechanism and Machine Theory*, vol. 94, pp. 9–27, 2015.
- [54] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, NY, USA, 2012.
- [55] C. R. Farrar and K. Worden, *Structural Health Monitoring: A Machine Learning Perspective*, John Wiley & Sons, New York, NY, USA, 2012.
- [56] A. Jessop, *statNotes: Some Statistics for Management Problems*, U. Durham University, World Scientific, 2016.
- [57] A. Grous, *Applied Quality Control: Fracture Mechanics*, vol. 3, John Wiley & Sons, New York, NY, USA, 2013.

Research Article

Shannon Entropy and K -Means Method for Automatic Diagnosis of Broken Rotor Bars in Induction Motors Using Vibration Signals

David Camarena-Martinez,¹ Martin Valtierra-Rodriguez,² Juan P. Amezcua-Sanchez,² David Granados-Lieberman,³ Rene J. Romero-Troncoso,¹ and Arturo Garcia-Perez¹

¹*División de Ingenierías, Universidad de Guanajuato, Campus Irapuato-Salamanca, Carretera Salamanca-Valle de Santiago Km 3.5 + 1.8, Comunidad de Palo Blanco, 36885 Salamanca, GTO, Mexico*

²*Facultad de Ingeniería, Universidad Autónoma de Querétaro, Campus San Juan del Río, Río Moctezuma 249, Col. San Cayetano, 76807 San Juan del Río, QRO, Mexico*

³*Departamento de Ingeniería Electromecánica, Instituto Tecnológico Superior de Irapuato, Carretera Irapuato-Silao Km 12.5, Colonia El Copal, 36821 Irapuato, GTO, Mexico*

Correspondence should be addressed to Arturo Garcia-Perez; agarcia@hspdigital.org

Received 2 June 2016; Revised 1 August 2016; Accepted 21 August 2016

Academic Editor: Lu Chen

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

For industry, the induction motors are essential elements in production chains. Despite the robustness of induction motors, they are susceptible to failures. The broken rotor bar (BRB) fault in induction motors has received special attention since one of its characteristics is that the motor can continue operating with apparent normality; however, at certain point the fault may cause severe damage to the motor. In this work, a methodology to detect BRBs using vibration signals is proposed. The methodology uses the Shannon entropy to quantify the amount of information provided by the vibration signals, which changes due to the presence of new frequency components associated with the fault. For automatic diagnosis, the K -means cluster algorithm and a decision-making unit that looks for the nearest cluster through the Euclidian distance are applied. Unlike other reported works, the proposal can diagnose the BRB condition during startup transient and steady state regimes of operation. Additionally, the proposal is also implemented into a field programmable gate array in order to offer a low-cost and low-complex online monitoring system. The obtained results demonstrate the proposal effectiveness to diagnose half, one, and two BRBs.

1. Introduction

Inductions motors are widely used in many applications because of their easy maintenance, ruggedness, low cost, versatility, and ease control [1]. During their service life, they are subject to unavoidable failures as a result of mechanical, environmental, thermal, and electrical stresses [2]. These faults such as bearing faults, air gap eccentricity, and broken rotor bars (BRBs) can yield a reduction on production, product quality, and an increase on costs, besides being a hazard for people and machinery [3]. Among the different faults that can occur in induction machines, BRB is a silent failure that allows operating the motor with apparent normality,

but it can cause an excessive vibration, a change in current consumption, and higher thermal stress with catastrophic consequences if the situation is not solved at early stages [4, 5]. In this regard, condition monitoring equipment has become an essential tool in many industrial areas. Yet, this task is very challenging because depending on the application the motor may be subject to transient and/or steady (nominal) regimes of operation, which changes its mechanical and electrical conditions by affecting and limiting the performance of equipment that only operates in a specific regime. From this point of view, an online and real-time monitoring system for an early detection of BRB in transient and steady regimes is a needed equipment in many industrial areas, since it will allow

scheduling maintenance operations in order to minimize its negative impact as well as saving time and money.

During the last decade, several vibration and current analysis-based processing techniques for BRB detection have been proposed. The conventional signal processing technique used to perform this task is the fast Fourier transform (FFT) [6–8]. However, it is limited in its capability for extracting features from signals that exhibit nonlinear and nonstationary characteristics, besides being susceptible to noise, making a correct identification of features related to the BRB fault difficult [9]. More recently, other powerful signal processing techniques, such as multiple signal classification (MUSIC) algorithm [10, 11], wavelet transform (WT) [9, 12–16], Empirical mode decomposition combined with Hilbert transform known as Hilbert-Huang transform (HHT) [16], and Wigner-Ville distribution (WVD) [17], have been used for BRB detection. Nevertheless, although prominent results have been obtained, the aforementioned signal processing techniques present some unresolved difficulties. For instance, MUSIC requires *a priori* knowledge of the interest frequencies and consumes significant computational resources [11]. The WT capabilities are significantly degraded in noisy signals, and the mother wavelet has to be appropriately chosen to obtain reliable results [4]. On the other hand, the WVD introduces cross-term interference in the estimated signal components, which inhibits the efficient estimation of the instantaneous frequencies, besides suffering aliasing problem [18]. The HHT suffers from the mode mixing effect, which means that waves with the same frequency are assigned to different intrinsic mode functions, affecting the accurate estimation of the instantaneous frequencies. In general, many advantages and disadvantages of the aforementioned techniques may be further discussed; yet, from a monitoring equipment viewpoint, two aspects become important. The first one is the performance capabilities; it means that the equipment does not degrade its performance when it analyzes transient or stationary signals. This desirable feature may be achieved either using a nonsusceptible signal processing technique or using different techniques for each scenario. The second one is the complexity since it may compromise the online analysis if low-end digital signal processors are used. In this regard, it would be desirable to have a signal processing algorithm with both the ability of identifying suitable and reliable features of signals for identifying BRB fault in different operating states and a low complexity for online analysis.

Similar to signal processing techniques, the classification algorithms play an important role in the automatic diagnosis of faults [19]. Different classification techniques such as neural networks [20, 21] and fuzzy logic [22, 23] have been successfully applied for monitoring the condition of induction motors. Unfortunately, the neural networks and other conventional artificial intelligent techniques require enough samples and have limitations on generalization of results in models that can overfit the samples [13]. Therefore, having in mind that online monitoring equipment may require low-complexity procedures, a classification algorithm with a suitable accuracy without the need of complicated processing, that requires a small number of samples, and, mainly, that allows developing a methodology capable of identifying

several faults in different scenarios is a desirable tool. A promising classification technique is the K -means algorithm. It is a well-known signal classification technique that has been successfully utilized in many applications such as neuroscience [24], structural engineering [25], and mechanics [26]. This approach provides a high accuracy and good generalization for a small number of features; besides its computational cost during and after its design is relatively low.

In this work, a methodology to detect automatically the BRB fault in induction motors using vibration signals is presented. The proposal considers the analysis of both the startup transient and the steady state of operation, which is very important since the induction motor may be subject to both scenarios in real applications; besides, an implementation into a field programmable gate array (FPGA) is also presented as system-on-chip (SoC) solution. This allows offering a system for online and continuous monitoring. Regarding the BRB condition, half, one, and two bars are considered. For the analysis, the Shannon entropy is used as a measure of the information contained in the vibration signals. This information presents changes associated with the fault. Then, the obtained entropy values are classified for automatic diagnosis using the K -means algorithm. The results show that the proposal can be a low-complex and suitable tool for BRB detection in both the startup transient and the steady state of operation.

2. Theoretical Background

In this section, the two main topics of the proposed methodology are briefly described.

2.1. Entropy. In information theory, entropy describes how much information about the data randomness is provided by a signal or event [27]. It has been used in image processing [28], in gearbox fault detection [29], in structural health monitoring [30], for analysis of electroencephalogram signals to diagnose the patient's clinical condition [31], and in fault motor diagnosis [15, 19, 32] among others. In particular, Shannon entropy, named after Claude Shannon, of a random signal X with N possible outcomes $x_0, x_1, x_2, \dots, x_{N-1}$ and with a probability of $p(x_i)$ can be computed as follows:

$$H(X) = - \sum_{i=0}^{N-1} p(x_i) \log_2 [p(x_i)], \quad (1)$$

where it is bounded by $0 \leq H(X) \leq \log_2(N)$.

Due to the number of applications and to the requirements of processing time, a hardware processing unit based on FPGA for entropy estimation is presented by [33], where a simplified mathematical expression is given as follows:

$$E(X) = \log_2(N) - \left(\frac{1}{N}\right) \sum_{i=0}^{N-1} r_i \log_2(r_i), \quad (2)$$

where r_i is the incidence rate or histogram of a signal X ; therefore, $p(x_i)$ is r_i/N . In general, this expression follows the structure shown in Figure 1.

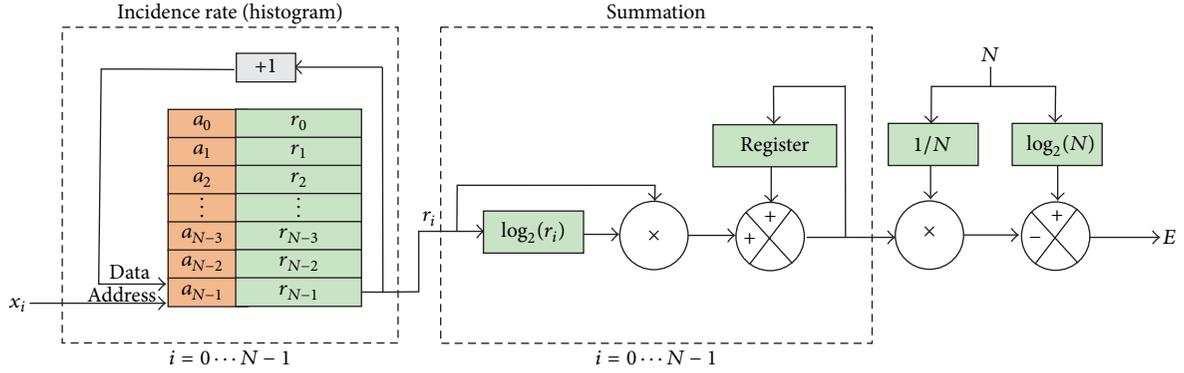


FIGURE 1: General structure for the entropy processor.

2.2. *K-Means*. In general, cluster analysis consists of creating groups of objects with similar features [34]. It implies that an object has to comply or has certain features for belonging to a specific group. In this regard, a classification task for unseen data can be carried out by looking for a group that fits better.

K-means is a simple and popular algorithm to solve clustering problems. The goal of the algorithm is to divide a data set $Y = \{y_1, y_2, \dots, y_N\}$ with N data into k clusters. The number of clusters is fixed *a priori*. The objective function based on squared Euclidian distances is calculated as follows [35]:

$$F(m_1, \dots, m_k) = \sum_{i=1}^k \sum_{j=1}^{M_i} \|y_{ij} - m_i\|^2, \quad (3)$$

where M_i is the number of objects of each i th cluster, y_{ij} is the j th object of the i th cluster, and m_i is the center of the i th cluster, which is defined as

$$m_i = \frac{1}{M_i} \sum_{j=1}^{M_i} y_{ij}, \quad i = 1, \dots, k. \quad (4)$$

The overall procedure is summarized as follows [34]: (1) select randomly the initial positions of the K -centroids; (2) assign the data set Y to the closest centroid; (3) relocate iteratively the K -centroids in order to minimize the objective function. It is worth noting that *K-means* algorithm is sensitive to the initial k clusters; however, it can be applied a number of times in order to find either the global objective function minimum or an acceptable error.

3. Proposed Methodology

The overall work is carried out in two stages, design and implementation, as shown in Figure 2.

3.1. *Design Stage*. In this stage, the extraction of the K -centroids for automatic diagnosis of BRBs in induction motors is carried out and shown in Figure 2(a). First, a set of vibration signals for each induction motor condition, healthy (HLT), half broken rotor bar (HBRB), one broken rotor bar (1BRB), and two broken rotor bars (2BRB), are acquired during the startup transient and the steady state. The vibration

signals are measured through a triaxial accelerometer (A_x , A_y , and A_z). Second, the Shannon entropy of each signal is computed using (2). Third, the *K-means* clustering algorithm is applied several times in order to look for the best clusters for each induction motor condition. The overall analysis is carried out using Matlab software. It is found, as will be discussed in the next sections, that the entropy values, E_x and E_z , present the best classification results; therefore, the K_x and K_z centroids for each cluster are selected as baseline for each condition.

3.2. *Implementation Stage*. In order to offer an online diagnosis tool, the proposed methodology is implemented into an FPGA as shown in Figure 2(b). First, the K -centroids for startup transient and steady state of each induction motor condition are stored into read-only memories, which are chosen by the user according to the actual motor operation. Second, the entropy values are estimated using the entropy processor, which follows the digital structure presented in Figure 1. Third, the Euclidian distances from entropy values to each cluster are computed using the coordinate rotation digital computer (CORDIC) algorithm [36]. Finally, a decision-making unit looking for the nearest cluster determines the induction motor condition.

4. Experimentation and Results

4.1. *Design Stage*. The vibration signals acquired for each induction motor condition during startup transient and steady state are presented in Figures 3 and 4, respectively. The total number of tests for each condition is 20, giving a total of 160 tests; their entropy values are shown in Figure 5. These values are bounded by the mean (μ) and ± 3 standard deviation (σ) in order to ensure an occurrence of 99.7% [33]; Table 1 presents the numerical values. The best *K-means* clusters are obtained using E_x and E_z since they offer more nonoverlapped patterns. The clusters are shown in Figure 6, where the Voronoi cells are used. The estimated values for the K_x and K_z centroids are summarized in Table 2.

4.2. *Experimental Setup*. The experimental setup used to test the proposal is shown in Figure 7(a). The induction

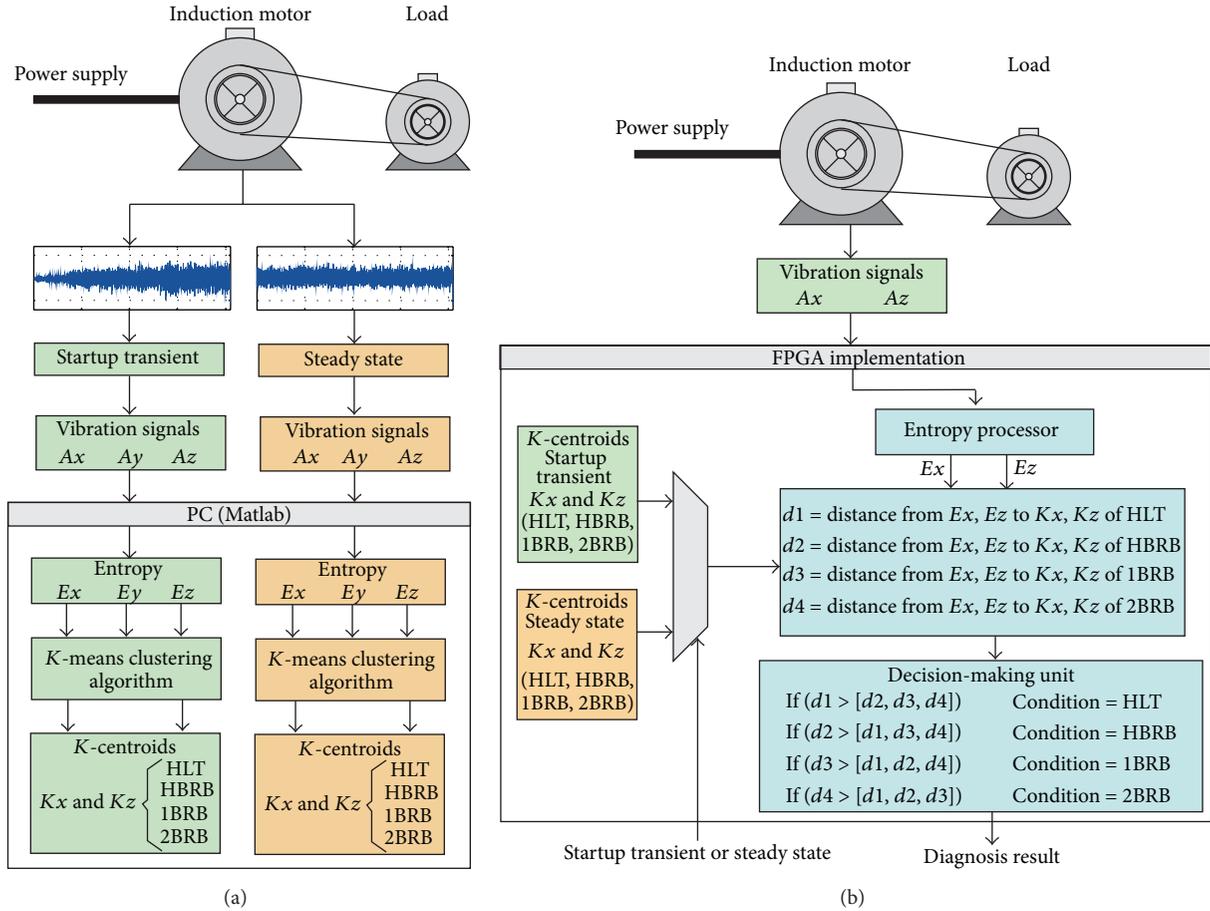


FIGURE 2: Proposed methodology: (a) design stage and (b) implementation stage.

TABLE 1: Entropy values.

Condition	Entropy (μ, σ)					
	Startup transient		Steady state			
	E_x	E_y	E_z	E_x	E_y	E_z
HLT	8.7530, 0.0416	9.1929, 0.0484	9.0489, 0.0644	8.6236, 0.715	9.4280, 0.0526	9.0940, 0.0941
HBRB	8.1024, 0.0445	8.9868, 0.0487	8.7272, 0.0566	7.8979, 0.0528	9.0849, 0.0544	8.6348, 0.0596
1BRB	9.0152, 0.0716	9.3918, 0.0345	9.7105, 0.0345	8.4532, 0.1763	8.9832, 0.0927	9.8297, 0.0927
2BRB	9.1787, 0.0511	9.4535, 0.0430	9.2498, 0.0589	8.8621, 0.0629	9.6350, 0.0647	9.2908, 0.0484

TABLE 2: Values of the K-centroids.

Induction motor condition	K-centroids			
	Startup transient		Steady state	
	K_x	K_z	K_x	K_z
HLT	8.7530	9.0489	8.6118	9.0811
HBRB	8.1024	8.7272	7.8979	8.6348
1BRB	9.0152	9.7105	8.4532	9.8297
2BRB	9.1787	9.2498	8.8614	9.2930

motor (model WEG 00136APE48T) has two poles and 28 bars. It is fed with 220 Vac at 60 Hz. The mechanical load is provided by an ordinary alternator. The 3-axis vibration

signal is measured by an accelerometer model LIS3L0AS4 placed on the top of the induction motor as shown in Figure 7(b). A 12-bit 4-channel analog to digital converter model ADS7841 is used for analog to digital conversion. The data acquisition system has a sampling frequency of 1500 Hz and captures 4096 samples (2.73 s), which is enough to capture the startup transient; the same time is used for the steady state. Figure 7(c) shows the four rotor conditions; the BRB condition is artificially created by drilling half, one, and two bars in the rotor without damaging the rotor shaft. The instrument is based on a proprietary Spartan 3E XC3S1600E FPGA with a frequency operation of 48 MHz; the FPGA resource utilization for the proposed methodology appears in Table 3. It is observed that it has a low consumption since a

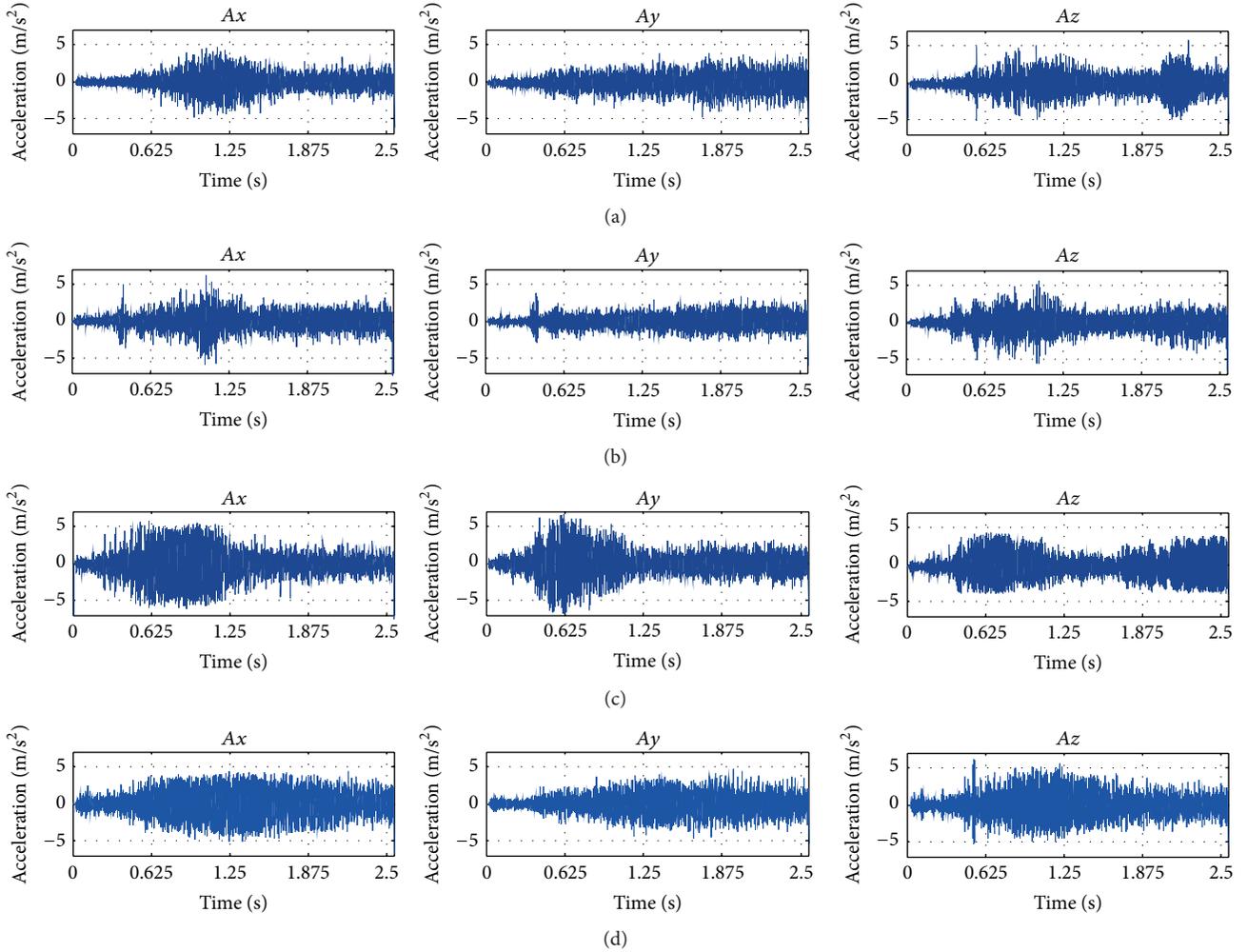


FIGURE 3: Vibration signals for startup transient: (a) HLT, (b) HBRB, (c) 1BRB, and (d) 2BRB.

TABLE 3: FPGA resources.

	Available	Used	Percentage (%)
Programmable logic	29504	1870	6.33
LUTs	29504	3445	11.67
Multipliers	36	2	5.55
Clock cycles		57632	

maximum of 11.67% is obtained; besides, the computational cost is also low since 57632 clock cycles (1.2 ms) are required to compute the final result. The overall system is tested 20 times for each motor condition. Tables 4 and 5 present the obtained classification results for startup transient and steady state, respectively. The correct results are located in the diagonals (highlighted in bold). Regarding the startup transient results, most the cases present an effectiveness of 100%, which means that, of the 20 actual HLT conditions, the system classifies 20 HLT conditions. A mistake for HBRB is obtained; yet, it is a positive false since it is classified as 1BRB. On the other hand, the classification results for steady state regimen have 100% of effectiveness in all cases.

TABLE 4: Classification results for startup transient tests (confusion matrix).

Induction motor condition	HLT	HBRB	1BRB	2BRB	Percentage of effectiveness (%)
HLT	20	0	0	0	100
HBRB	0	19	1	0	95
1BRB	0	0	20	0	100
2BRB	0	0	0	20	100

Finally, it is worth noting that different vibration levels generated by other faults, neighboring equipment, and uncontrolled conditions can be induced in the motor frame, which will modify the vibration patterns shown in Figure 5 and, as a consequence, the performance of the classification algorithm; in this regard, new calibrations in known conditions will be required.

4.3. Discussion. Table 6 summarizes a performance comparison of the proposed method against recently reported

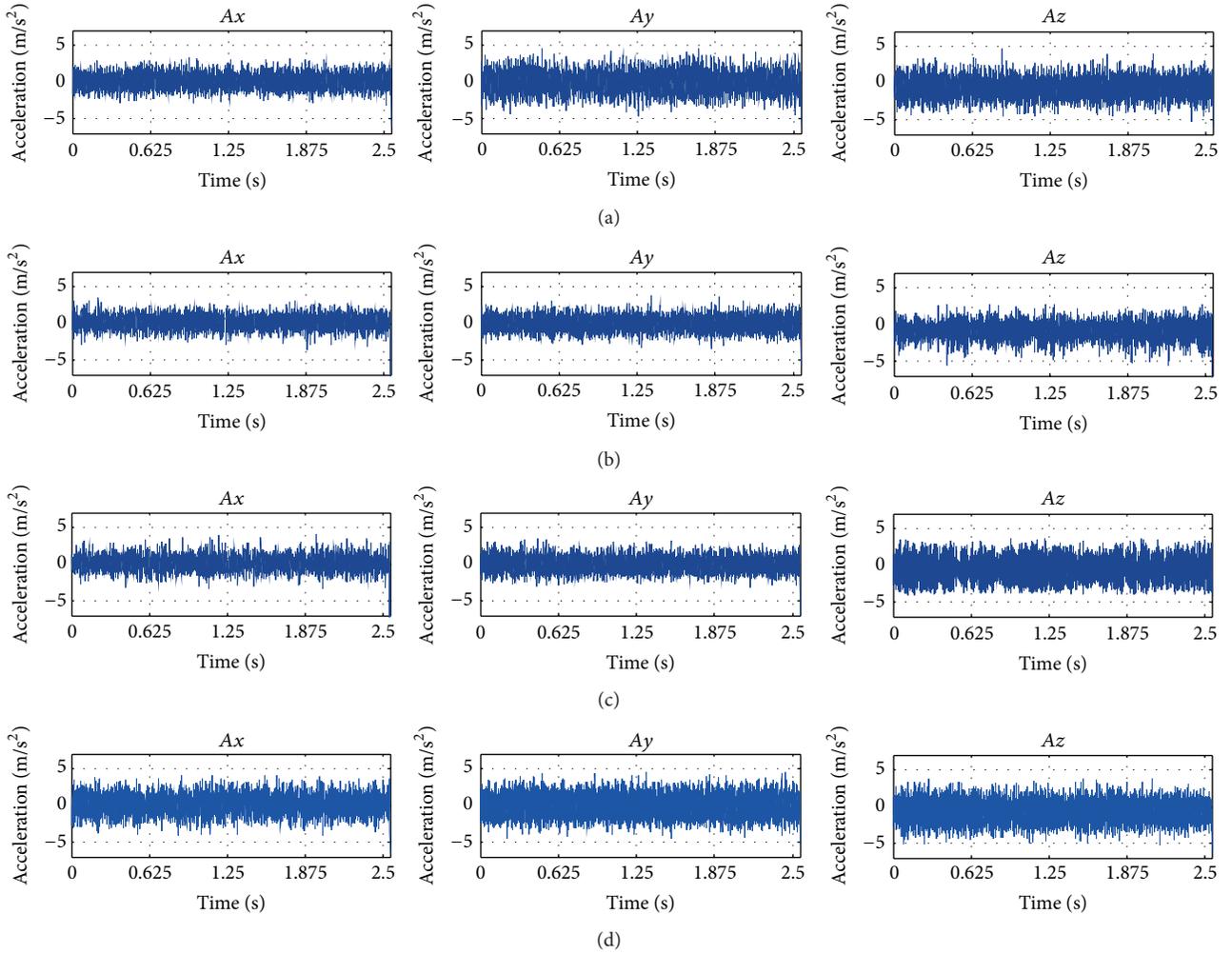


FIGURE 4: Vibration signals for steady state: (a) HLT, (b) HBRB, (c) 1BRB, and (d) 2BRB.

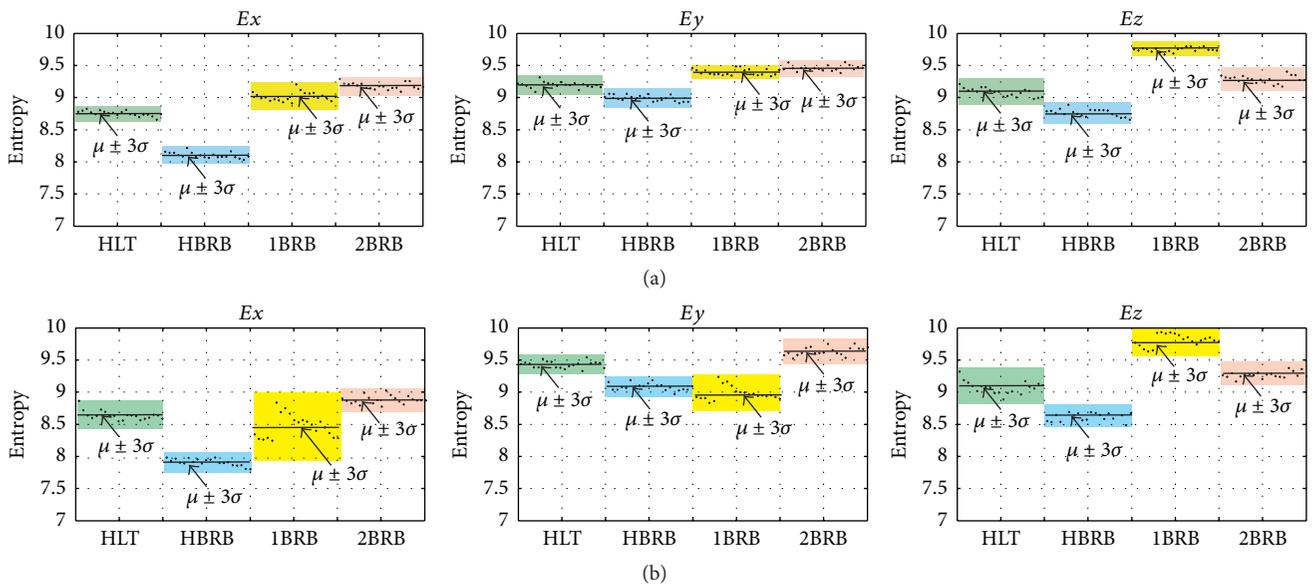


FIGURE 5: Entropy values for (a) startup transient tests and (b) steady state tests.

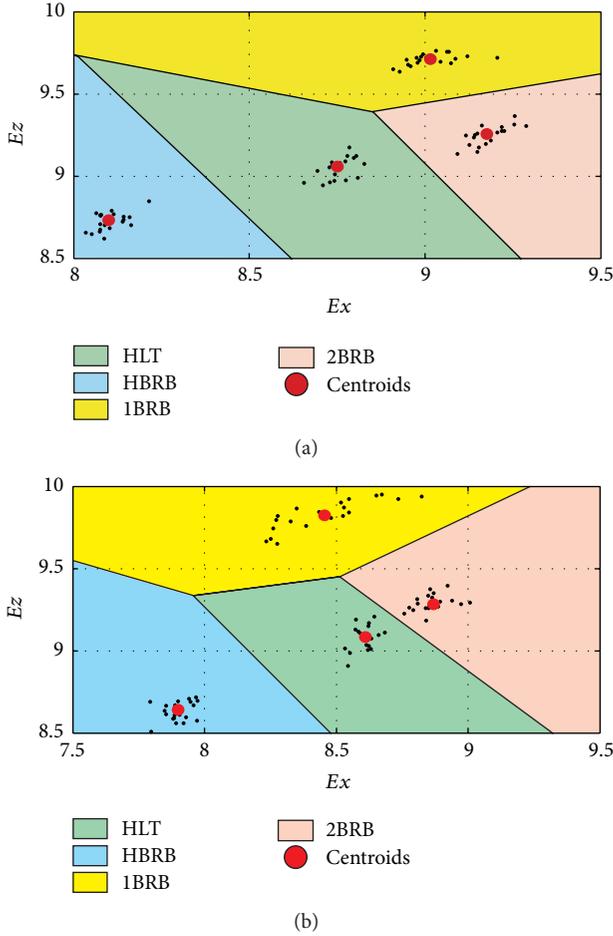
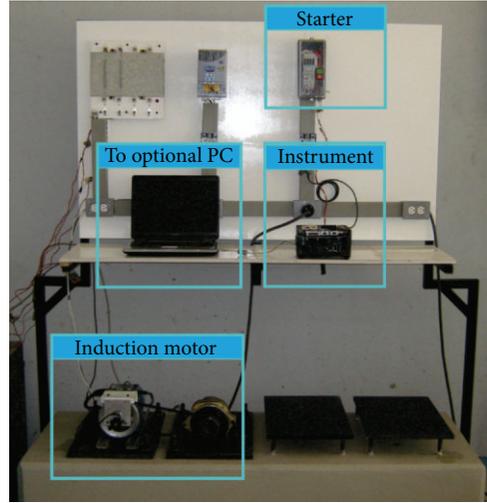


FIGURE 6: *K*-means clusters: (a) startup transient and (b) steady state.

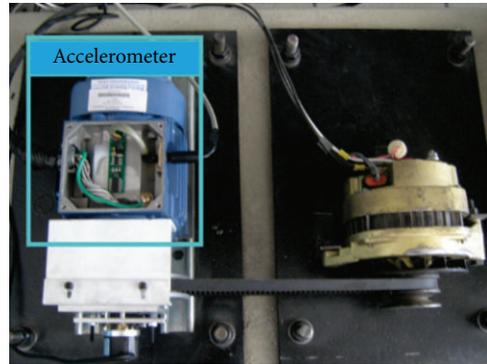
TABLE 5: Classification results for steady state tests (confusion matrix).

Induction motor condition	HLT	HBRB	1BRB	2BRB	Percentage of effectiveness (%)
HLT	20	0	0	0	100
HBRB	0	20	0	0	100
1BRB	0	0	20	0	100
2BRB	0	0	0	20	100

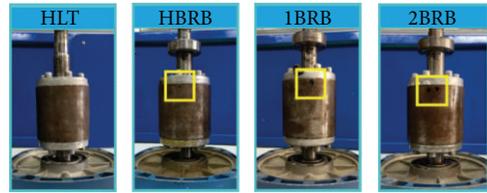
works in literature, where parameters such as methodology applied, analyzed state, and the percentage of effectiveness are considered. From Table 6, it is clear that the proposed methodology can detect the presence of BRB with high accuracy in both states of induction motor. Unlike other reported works [37–42], the proposal considers HBRB in both states. Although HBRB detection, in both states, has been considered in [5], the proposal outperforms in several cases the effectiveness for diagnosing the induction motor condition, which makes it more attractive for industrial applications. Further, it is important to mention that the proposal presents two essential characteristics: (1) the automatic



(a)



(b)



(c)

FIGURE 7: Experimentation: (a) experimental setup, (b) top view accelerometer, and (c) rotors.

identification of different grades of severity of broken bar faults and (2) the use of a single-parameter, which reduces the complexity and computational load. In addition, the designed FPGA implementation offers a low-cost and a real-time solution for online monitoring of induction motor condition.

5. Conclusions

In this work, a methodology based on Shannon entropy and *K*-means clustering algorithm for automatic diagnosis of BRBs during the startup transient and steady state is proposed. It allows diagnosing three severities of damage, HBRB, 1BRB, and 2BRB, as well as the HLT condition. Although a triaxial accelerometer to measure A_x , A_y , and

TABLE 6: Results and characteristics of previous works and of the proposed work.

Work	Applied methodology	Analyzed state	Effectiveness (%)			
			HLT	HBRB	IBRB	MBRB
Matić et al. [37]	(1) Feature selection through Hilbert transform. (2) Fault diagnosis classification through support vector machine.	Steady	100	—	100	—
Da Silva et al. [38]	(1) Feature selection through time-stepping finite-element simulations. (2) Fault diagnosis classification through a Bayesian classifier.	Steady	100	—	100	100 ^{3,5}
Georgoulas et al. [39]	(1) Extraction features by means of complex empirical mode decomposition. (2) Fault diagnosis classification through hidden Markov models.	Transient	100	—	100	100 ²
Keskes et al. [40]	(1) Stationary Wavelet Packet transform for statistical feature extraction. (2) Fault diagnosis classification through improved support vector machine.	Steady	100	—	99	99 ²
Rangel-Magdaleno et al. [41]	(1) Mathematical morphology. (2) Spectral analysis for identification of frequency components related to the fault using FFT. (3) Amplitude analysis from estimated spectrum for statistical feature extraction. (4) Fault diagnosis classification through a decision tree.	Steady	95	—	95	95 ²
Menacer et al. [42]	(1) Spectral analysis using discrete wavelet transform and Hilbert transform to obtain the envelope of the signal. (2) Fault diagnosis classification through measured features (Eigenvector changes).	Transient	100	—	100	100 ²
Valles-Novo et al. [5]	(1) Signal decomposition by means of empirical mode decomposition method.	Transient	98.2	99.7	98.2	—
	(2) Standard deviation of zero crossings and the time between successive zero crossings of the decompositions for fault diagnosis classification.	Steady	98.2	98.2	98.5	—
Proposal	(1) Shannon entropy for statistical feature extraction.	Transient	100	95	100	100 ²
	(2) Fault diagnosis classification through K -means method.	Steady	100	100	100	100 ²

HLT: healthy condition; HBRB: half broken rotor bar; IBRB: one broken rotor bar; MBRB: multiple broken rotor bars, where 2 means two broken bars, 3 means three broken bars, and 5 means five broken bars.

Az vibration signals is used, it is found that A_x and A_z provide enough information to diagnose the treated fault, which simplifies even more the computational cost of the proposal.

The FPGA implementation as SoC solution for online and continuous monitoring is also developed. In general, a low consumption of FPGA resources with a maximum of 11.67% is obtained. This is possible since the K -centroids are computed offline and in the implementation stage only a search for the nearest cluster through the Euclidian distance is required. It is worth noting that the same methodology can diagnose the faults during both transient and steady regimens with a high accuracy; therefore, the processor cores can be reused with an only change in the selected K -centroids.

In a future work, different load conditions in both states will be explored in order to improve the usefulness and applicability of the proposed methodology. Besides that, other mechanical and electrical faults will be analyzed using the proposed methodology; in fact, other signals such as current or temperature will be also explored in order to obtain unique and reliable patterns for each condition.

Competing Interests

The authors declare that they have no competing interests.

References

- [1] P. Zhang, Y. Du, T. G. Habetler, and B. Lu, "A survey of condition monitoring and protection methods for medium-voltage induction motors," *IEEE Transactions on Industry Applications*, vol. 47, no. 1, pp. 34–46, 2011.
- [2] M. Hernandez-Vargas, E. Cabal-Yepez, and A. Garcia-Perez, "Real-time SVD-based detection of multiple combined faults in induction motors," *Computers and Electrical Engineering*, vol. 40, no. 7, pp. 2193–2203, 2014.
- [3] R. J. Romero-Troncoso, A. Garcia-Perez, D. Morinigo-Sotelo, O. Duque-Perez, R. A. Osornio-Rios, and M. A. Ibarra-Manzano, "Rotor unbalance and broken rotor bar detection in inverter-fed induction motors at start-up and steady-state regimes by high-resolution spectral analysis," *Electric Power Systems Research*, vol. 133, pp. 142–148, 2016.
- [4] J. A. Antonino-Daviu, M. Riera-Guasp, M. Pineda-Sanchez, and R. B. Pérez, "A critical comparison between DWT and

- Hilbert-Huang-based methods for the diagnosis of rotor bar failures in induction machines,” *IEEE Transactions on Industry Applications*, vol. 45, no. 5, pp. 1794–1803, 2009.
- [5] R. Valles-Novo, J. d. j. Rangel-Magdaleno, J. M. Ramirez-Cortes, H. Peregrina-Barreto, and R. Morales-Caporal, “Empirical mode decomposition analysis for broken-bar detection on squirrel cage induction motors,” *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 5, pp. 1118–1128, 2015.
- [6] Y.-H. Kim, Y.-W. Youn, D.-H. Hwang, J.-H. Sun, and D.-S. Kang, “High-resolution parameter estimation method to identify broken rotor bar faults in induction motors,” *IEEE Transactions on Industrial Electronics*, vol. 60, no. 9, pp. 4103–4117, 2013.
- [7] J. Kurek and S. Osowski, “Support vector machine for fault diagnosis of the broken rotor bars of squirrel-cage induction motor,” *Neural Computing and Applications*, vol. 19, no. 4, pp. 557–564, 2010.
- [8] S.-R. Huang, K.-H. Huang, K.-H. Chao, and W.-T. Chiang, “Fault analysis and diagnosis system for induction motors,” *Computers & Electrical Engineering*, 2016.
- [9] P. K. Kankar, S. C. Sharma, and S. P. Harsha, “Rolling element bearing fault diagnosis using wavelet transform,” *Neurocomputing*, vol. 74, no. 10, pp. 1638–1645, 2011.
- [10] A. Bellini, A. Yazidi, F. Filippetti, C. Rossi, and G.-A. Capolino, “High frequency resolution techniques for rotor fault detection of induction machines,” *IEEE Transactions on Industrial Electronics*, vol. 55, no. 12, pp. 4200–4209, 2008.
- [11] A. Garcia-Perez, J. P. Amezquita-Sanchez, A. Dominguez-Gonzalez, R. Sedaghati, R. Osornio-Rios, and R. J. Romero-Troncoso, “Fused empirical mode decomposition and wavelets for locating combined damage in a truss-type structure through vibration analysis,” *Journal of Zhejiang University: Science A*, vol. 14, no. 9, pp. 615–630, 2013.
- [12] M. Barakat, M. El Badaoui, and F. Guillet, “Hard competitive growing neural network for the diagnosis of small bearing faults,” *Mechanical Systems and Signal Processing*, vol. 37, no. 1-2, pp. 276–292, 2013.
- [13] Z. Liu, H. Cao, X. Chen, Z. He, and Z. Shen, “Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings,” *Neurocomputing*, vol. 99, pp. 399–410, 2013.
- [14] J. Rangel-Magdaleno, H. Peregrina-Barreto, J. Ramirez-Cortes, R. Morales-Caporal, and I. Cruz-Vega, “Vibration analysis of partially damaged rotor bar in induction motor under different load condition using DWT,” *Shock and Vibration*, vol. 2016, Article ID 3530464, 11 pages, 2016.
- [15] L. Zhang, L. Zhang, J. Hu, and G. Xiong, “Bearing fault diagnosis using a novel classifier ensemble based on lifting wavelet packet transforms and sample entropy,” *Shock and Vibration*, vol. 2016, Article ID 4805383, 13 pages, 2016.
- [16] R. Valles-Novo, J. D. J. Rangel-Magdaleno, J. M. Ramirez-Cortes, H. Peregrina-Barreto, and R. Morales-Caporal, “Empirical mode decomposition analysis for broken-bar detection on squirrel cage induction motors,” *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 5, pp. 1118–1128, 2015.
- [17] V. Climente-Alarcon, J. A. Antonino-Daviu, M. Riera-Guasp, and M. Vlcek, “Induction motor diagnosis by advanced notch FIR filters and the wigner-ville distribution,” *IEEE Transactions on Industrial Electronics*, vol. 61, no. 8, pp. 4217–4227, 2014.
- [18] A. L. Martinez-Herrera, L. M. Ledesma-Carrillo, M. Lopez-Ramirez, S. Salazar-Colores, E. Cabal-Yepez, and A. Garcia-Perez, “Gabor and the Wigner-ville transforms for broken rotor bars detection in induction motors,” in *Proceedings of the 24th International Conference on Electronics, Communications and Computers (CONIELECOMP '14)*, pp. 83–87, IEEE, Puebla, Mexico, February 2014.
- [19] S. Pan, T. Han, A. C. C. Tan, and T. R. Lin, “Fault diagnosis system of induction motors based on multiscale entropy and support vector machine with mutual information algorithm,” *Shock and Vibration*, vol. 2016, Article ID 5836717, 12 pages, 2016.
- [20] N. Saravanan and K. I. Ramachandran, “Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN),” *Expert Systems with Applications*, vol. 37, no. 6, pp. 4168–4181, 2010.
- [21] G. F. Bin, J. J. Gao, X. J. Li, and B. S. Dhillon, “Early fault diagnosis of rotating machinery based on wavelet packets—empirical mode decomposition feature extraction and neural network,” *Mechanical Systems and Signal Processing*, vol. 27, no. 1, pp. 696–711, 2012.
- [22] R. J. Romero-Troncoso, R. Saucedo-Gallaga, E. Cabal-Yepez et al., “FPGA-based online detection of multiple combined faults in induction motors through information entropy and fuzzy inference,” *IEEE Transactions on Industrial Electronics*, vol. 58, no. 11, pp. 5263–5270, 2011.
- [23] D. P. Winston and M. Saravanan, “Single parameter fault identification technique for DC motor through wavelet analysis and fuzzy logic,” *Journal of Electrical Engineering and Technology*, vol. 8, no. 5, pp. 1049–1055, 2013.
- [24] Y. Sharma and P. Kaur, “Detection and extraction of brain tumor from MRI images using K-means clustering and watershed algorithms,” *International Journal of Computer Science Trends and Technology*, vol. 3, no. 2, pp. 32–38, 2015.
- [25] H. Qarib and H. Adeli, “A new adaptive algorithm for automated feature extraction in exponentially damped signals for health monitoring of smart structures,” *Smart Materials and Structures*, vol. 24, no. 12, Article ID 125040, 2015.
- [26] H. Zhou, Y. C. Soh, and X. Wu, “Integrated analysis of CFD data with K-means clustering algorithm and extreme learning machine for localized HVAC control,” *Applied Thermal Engineering*, vol. 76, pp. 98–104, 2015.
- [27] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [28] Y. Wu, Y. Zhou, G. Saveriades, S. Agaian, J. P. Noonan, and P. Natarajan, “Local Shannon entropy measure with statistical tests for image randomness,” *Information Sciences*, vol. 222, pp. 323–342, 2013.
- [29] H. H. Bafroui and A. Ohadi, “Application of wavelet energy and Shannon entropy for feature extraction in gearbox fault detection under varying speed conditions,” *Neurocomputing*, vol. 133, pp. 437–445, 2014.
- [30] T.-K. Lin and J.-C. Liang, “Application of multi-scale (cross-) sample entropy for structural health monitoring,” *Smart Materials and Structures*, vol. 24, no. 8, Article ID 085003, 2015.
- [31] A. Lay-Ekuakille, P. Vergallo, G. Griffio et al., “Entropy index in quantitative EEG measurement for diagnosis accuracy,” *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 6, pp. 1440–1450, 2014.
- [32] E. Cabal-Yepez, R. J. Romero-Troncoso, A. Garcia-Perez, R. A. Osornio-Rios, and R. Alvarez-Salas, “Multiple fault detection

- through information entropy analysis in ASD-fed induction motors,” in *Proceedings of the 8th IEEE Symposium on Diagnostics for Electrical Machines, Power Electronics & Drives (SDEMPED '11)*, pp. 391–396, Bologna, Italy, September 2011.
- [33] E. Cabal-Yepetz, R. D. J. Romero-Troncoso, A. Garcia-Perez, and C. Rodriguez-Donate, “Novel hardware processing unit for dynamic on-line entropy estimation of discrete time information,” *Digital Signal Processing*, vol. 20, no. 2, pp. 337–346, 2010.
- [34] T. Velmurugan, “Performance based analysis between k-Means and Fuzzy C-Means clustering algorithms for connection oriented telecommunication data,” *Applied Soft Computing Journal*, vol. 19, pp. 134–146, 2014.
- [35] C. T. Yiakopoulos, K. C. Gryllias, and I. A. Antoniadis, “Rolling element bearing fault detection in industrial environments based on a K-means clustering approach,” *Expert Systems with Applications*, vol. 38, no. 3, pp. 2888–2911, 2011.
- [36] R. Andraka, “A survey of CORDIC algorithms for FPGA based computers,” in *Proceedings of the ACM SIGDA 6th International Symposium on Field Programmable Gate Arrays (FPGA '98)*, pp. 191–200, Monterey, Calif, USA, February 1998.
- [37] D. Matić, F. Kulić, M. Pineda-Sánchez, and I. Kamenko, “Support vector machine classifier for diagnosis in electrical machines: application to broken bar,” *Expert Systems with Applications*, vol. 39, no. 10, pp. 8681–8689, 2012.
- [38] A. M. Da Silva, R. J. Povinelli, and N. A. O. Demerdash, “Rotor bar fault monitoring method based on analysis of air-gap torques of induction motors,” *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2274–2283, 2013.
- [39] G. Georgoulas, I. P. Tsoumas, J. A. Antonino-Daviu et al., “Automatic pattern identification based on the complex empirical mode decomposition of the startup current for the diagnosis of rotor asymmetries in asynchronous machines,” *IEEE Transactions on Industrial Electronics*, vol. 61, no. 9, pp. 4937–4946, 2014.
- [40] H. Keskes, A. Braham, and Z. Lachiri, “Broken rotor bar diagnosis in induction machines through stationary wavelet packet transform and multiclass wavelet SVM,” *Electric Power Systems Research*, vol. 97, pp. 151–157, 2013.
- [41] J. J. Rangel-Magdaleno, H. Peregrina-Barreto, J. M. Ramirez-Cortes, P. Gomez-Gil, and R. Morales-Caporal, “FPGA-based broken bars detection on induction motors under different load using motor current signature analysis and mathematical morphology,” *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 5, pp. 1032–1040, 2014.
- [42] A. Menacer, M. Boumechraz, and H. Cherif, “DWT and Hilbert transform for broken rotor bar fault diagnosis in induction machine at low load,” *Energy Procedia*, vol. 74, pp. 1248–1257, 2015.

Research Article

Rolling Bearing Fault Diagnosis Based on STFT-Deep Learning and Sound Signals

Hongmei Liu,¹ Lianfeng Li,¹ and Jian Ma^{1,2}

¹*School of Reliability and Systems Engineering, Beihang University, Beijing, China*

²*Science & Technology on Reliability & Environmental Engineering Laboratory, Beijing, China*

Correspondence should be addressed to Jian Ma; 09977@buaa.edu.cn

Received 26 April 2016; Accepted 20 July 2016

Academic Editor: Fiorenzo A. Fazzolari

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

The main challenge of fault diagnosis lies in finding good fault features. A deep learning network has the ability to automatically learn good characteristics from input data in an unsupervised fashion, and its unique layer-wise pretraining and fine-tuning using the backpropagation strategy can solve the difficulties of training deep multilayer networks. Stacked sparse autoencoders or other deep architectures have shown excellent performance in speech recognition, face recognition, text classification, image recognition, and other application domains. Thus far, however, there have been very few research studies on deep learning in fault diagnosis. In this paper, a new rolling bearing fault diagnosis method that is based on short-time Fourier transform and stacked sparse autoencoder is first proposed; this method analyzes sound signals. After spectrograms are obtained by short-time Fourier transform, stacked sparse autoencoder is employed to automatically extract the fault features, and softmax regression is adopted as the method for classifying the fault modes. The proposed method, when applied to sound signals that are obtained from a rolling bearing test rig, is compared with empirical mode decomposition, Teager energy operator, and stacked sparse autoencoder when using vibration signals to verify the performance and effectiveness of the proposed method.

1. Introduction

As one of the most common components in rotating machinery, rolling bearings play a key role in maintaining the normal operation of entire machines. The faults of rolling bearings usually lead to a considerable decline in industrial productivity and can even cause enormous economic losses. To increase productivity and to reduce undesirable casualties, condition monitoring and fault diagnosis attract broad attention. In addition, the maintenance cost can be reduced, especially if the faults are identified before they become severe.

The features of sound signals can be used to detect faults in machines; for example, in the regular maintenance of a railway system, maintenance workers use clicking echoes to identify whether the train wheels are healthy. If the echo is dull, then a wheel could have internal cracks; otherwise, it is most likely normal. Similarly, experienced maintenance men in other engineering fields can judge whether a machine runs normally by recognizing the sound features. Sounds that are produced during running are characteristic of operating

under healthy conditions and differ across fault modes [1, 2]. Similarly, the sound signals change gradually while the components in the rolling bearings develop faults, and different faults produce different sounds. Because of these changes, the health status can be determined. At the same time, based on the differences between the fault modes, the various faults can be classified.

For fault diagnosis, high identification accuracy depends on having effective feature representations. However, noises and complex structures in the observed signal increase the difficulty of extracting valid characteristics. For this reason, a large amount of work regarding feature extraction and selection in fault diagnosis has been performed using different types of signals and algorithms.

In most of the existing diagnosis literatures based on vibration signals, the researchers either applied WT (wavelet transformation) to acquire time-frequency information of the signal and then extract features from the time-frequency spectra or employed EMD (empirical mode decomposition) [3], LMD (local mean decomposition) [4], and LCD (local

characteristic scale decomposition) [5] to adaptively decompose the original signal into a series of scales and then extract the energy or entropy, a complexity measure of the signal, to be the fault features. Usually, to cover sufficient fault information, it is inevitable that the dimension of the acquired features is sufficiently high so that visualization is difficult, while the classification performance can become poor. Therefore, a dimensionality reduction method, such as common PCA (principal component analysis) [6], KPCA (kernel principal component analysis) [7], ISOMAP (isometric feature mapping) [8], LLE (locally linear embedding) [9], or LTSA (local tangent space alignment) [10], is necessary to map high-dimensional data sequentially to low-dimensional space. Finally, the low-dimensional features are used for visualization analysis and to train a classifier such as SVM (support vector machine) [11] and KNN (k -nearest neighbor) [12] and neural network classifiers [13].

With regard to the foregoing analysis, in conventional fault diagnosis methods, scholars have spent a large amount of time on feature extraction, feature selection, and dimensionality reduction, which are also complicated and long-standing tasks. In 2006, Hinton and Salakhutdinov published a paper in *Science* [14], which proposes two core points. First, an artificial neural network with multihidden layers possesses excellent feature-learning ability, and the acquired features provide a more intrinsic and abstract representation of the raw data. Second, layer-wise pretraining can effectively overcome the training difficulties of the deep neural network. Since then, research on deep learning in academia and industry has raised a large amount of attention. Researchers on speech recognition at Microsoft Research and Google decreased the speech recognition error rate by 20%–30% when they adopted deep neural networks (DNNs). In 2012, amazing results emerged in image recognition where the error rate was largely reduced from 26% to 15% in the ImageNet evaluation. In the same year, DNN was also applied to the prognosis of drug activity in pharmaceutical companies and achieved the world's best accuracy, which was featured in the *New York Times*.

Despite its success in speech, image, and video recognition, the application of deep learning in mechanical fault diagnosis has received very little research attention. Deep learning is quite different from traditional diagnosis methods that require complicated and time-consuming feature extraction work, which only needs simple data preprocessing. STFT (short-time Fourier transform) is a simple, easy-to-apply signal transformation method that can transform time-domain signals into time-frequency space. In this paper, a combination of deep learning networks and STFT is proposed to solve fault diagnosis problems. SAE (stacked sparse autoencoder), a neural network that consists of multiple layers of basic autoencoders in which the outputs of each layer are wired to the inputs of the successive layer, can learn higher order feature representations of input signals. In the deep-layer networks of SAE, raw data can be represented in a much better form, enabling the classifier to provide more accurate results even with fewer training examples or less labeled training data.

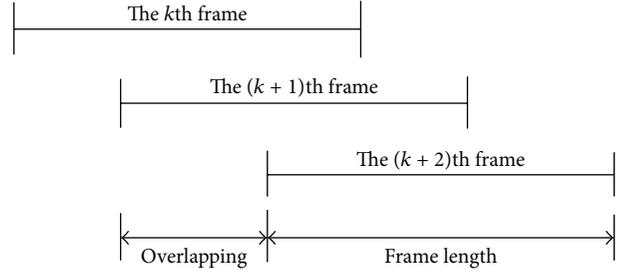


FIGURE 1: Data frames.

This paper is organized as follows. Section 2 introduces the basic principle of STFT. Section 3 proposes SAE based feature extraction. Section 4 describes softmax classifier-based pattern recognition. Section 5 outlines the implementation methodology of SAE with the softmax classifier and is followed by the conclusions in Section 6.

2. Time-Frequency Analysis of Sound Signals Using STFT

Fourier analysis decomposes a signal into its frequency components and determines their relative strengths. The Fourier transform is defined as

$$\begin{aligned} F(\omega) &= \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \longleftrightarrow \\ f(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega. \end{aligned} \quad (1)$$

This transform is applied to stationary signals whose properties do not evolve over time. When the signal is nonstationary, we can introduce a local frequency parameter in such a way that a local Fourier transform looks at the signal through a window over which the signal is approximately stationary. After multiplying it by a window, the signal is also truncated into short data frames, and to analyze the whole signal, the window is translated in time and then reapplied to the signal. The output of successive STFTs can provide a time-frequency representation of the signal [15].

Therefore, in this paper, a spectral analysis of sounds is performed by using STFT, in which the signal is divided into small sequential or overlapping data frames, as shown in Figure 1; then, FFT is applied to each data frame. The STFT positions a window function $\psi(t)$ at τ on the time axis and calculates the Fourier transform of the windowed signal as

$$F(\omega, \tau) = \int_{-\infty}^{\infty} f(t) \psi^*(t - \tau) e^{-j\omega t} dt. \quad (2)$$

The basic functions of this transform are generated by the modulation and transformation of the window function $\psi(t)$, where ω and τ are the modulation and translation parameters, respectively [16]. Commonly used windows are the rectangular window, Hamming window, Hanning window, and Blackman window. The first two windows are described as follows in (3) and (4).

For a rectangular window of size N ,

$$\omega(n) = \begin{cases} 1, & 0 \leq n \leq (N-1) \\ 0, & \text{others.} \end{cases} \quad (3)$$

For a Hamming window of size N ,

$$\omega(n) = \begin{cases} 0.5 \left(1 - \cos \frac{2\pi n}{N-1} \right), & 0 \leq n \leq (N-1) \\ 0, & \text{others.} \end{cases} \quad (4)$$

The rectangular window does not conform to the requirements for an excessively high side lobe, which leaks more energy. Therefore, the Hamming window is selected in this paper.

Given a signal $x(n)$, the discrete STFT for the frequency band k at time n is defined as

$$X_n(e^{j\omega_k}) = \sum_{m=-\infty}^{+\infty} x(m) \omega(n-m) e^{-j\omega_k m}, \quad (5)$$

where $\omega_k = 2\pi k/N$ is the frequency in radians; N is the number of frequency bands; $\omega(m)$ is the selected symmetric window of size L ; and $L \leq N$ if signal reconstruction is required.

It follows that (5) is equivalent to

$$X_n(e^{j\omega_k}) = e^{-j\omega_k n} \bar{X}_n(\omega_k), \quad (6)$$

where

$$\bar{X}_n(\omega_k) \sum_m x(n-m) \omega(m) e^{j\omega_k m} = x(n) h_k(n) \quad (7)$$

is the output of the k th complex band-pass filter, with impulse response $h_k(n)$ and center frequency f_k :

$$\begin{aligned} h_k(n) &= \omega(n) e^{j\omega_k n}, \\ f_k &= \frac{f_s k}{N} \text{ (Hz)}, \quad k = 0, 1, \dots, N-1. \end{aligned} \quad (8)$$

According to $\omega_k = 2\pi k/N$ as above, plugging into (5) yields

$$\begin{aligned} X(n, k) &= X_n(e^{j\omega_k}) \\ &= \sum_{m=-\infty}^{+\infty} x(m) \omega(n-m) e^{-j2\pi km/N}. \end{aligned} \quad (9)$$

Here, $|X(n, k)|$ is the short-time spectral amplitude estimate of $x(n)$. The power spectrum density (PSD) function is defined as

$$P(n, k) = |X(n, k)|^2 = x(n, k) x(\text{conj}(x(n, k))). \quad (10)$$

$P(n, k)$ is a two-dimensional, nonnegative, and real-valued function. It is easily proven that $P(n, k)$ is only a Fourier transform (FT) of the short-time autocorrelation function of the signal. The spectrogram algorithm [17] is an

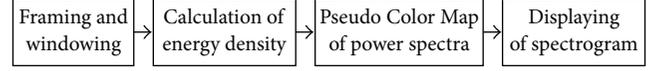


FIGURE 2: Generation of spectrogram.

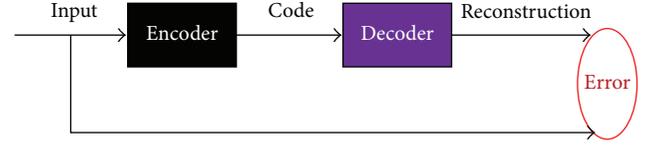


FIGURE 3: Schematic of an autoencoder.

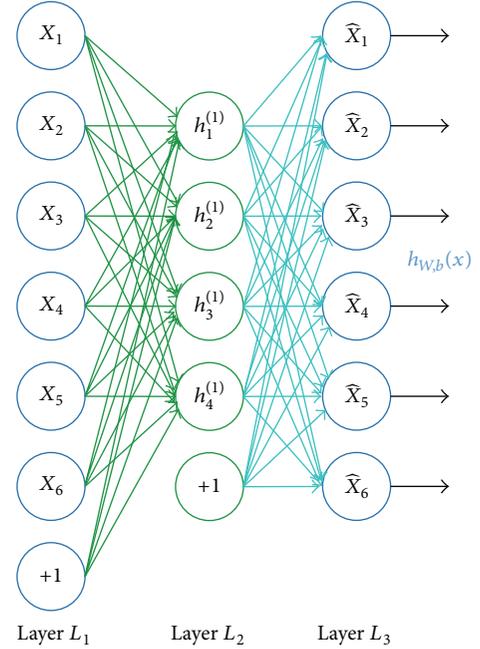


FIGURE 4: Structure of AE.

analysis algorithm that produces a two-dimensional image representation of sounds. PSD is expressed as the Pseudo Color Map (PCM), in other words, a spectrogram with a time axis and frequency axis. This time-frequency spectrum, which is sometimes called visual language, shows the dynamic characteristics of the sounds and enjoys significant practical worth. The spectrogram is acquired as shown in Figure 2.

3. Feature Extraction Using SAE

3.1. Autoencoder. As depicted in Figure 4, an autoencoder that was first introduced by Rumelhart et al. is a special neural network with three layers. A trained autoencoder can compute the input's representation from which the original data can be reconstructed with as much accuracy as possible [18], as shown in Figure 3. Recently, autoencoders were used in deep architectures as an unsupervised learning algorithm [19, 20].

An autoencoder takes an input vector $x^{(i)} \in [0, 1]$ that corresponds to the i th training example and first maps it to

the hidden layer $a \in [0, 1]$ (a is the activation vector of the first hidden layer), through deterministic mapping:

$$a = f_\theta(x^i) = \text{sigmoid}(W \cdot x^i + b) \quad (11)$$

parameterized by $\theta = \{W, b\}$. The resulting latent representation a is then mapped back to a reconstructed vector $h_{W,b}(x^{(i)}) \in [0, 1]$ in the input space [21, 22], as depicted in Figure 4:

$$h_{W,b}(x^{(i)}) = g_\theta(a) = \text{sigmoid}(W^T \cdot a + b^T). \quad (12)$$

(a) *Cost Function of an Autoencoder.* For a fixed training set $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$ of “ m ” training examples, the initial cost function is given by

$$\begin{aligned} J(W, b) &= \left[\frac{1}{m} \sum_{i=1}^m J(W, b; x^{(i)}, y^{(i)}) \right] \\ &\quad + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \\ &= \left[\frac{1}{m} \sum_{i=1}^m \frac{1}{2} \|h_{W,b}(x^{(i)}) - y^{(i)}\|^2 \right] \\ &\quad + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2, \end{aligned} \quad (13)$$

where the first term in $J(W, b)$ is an average sum-of-squares error term. Here, W and b are the same as mentioned in (11) and (12). The second term is a regularization term or weight decay term, which tends to decrease the magnitude of the weights and helps prevent overfitting [22]. Here, $h_{W,b}(x^{(i)})$ is the hypothesis and λ is a weight decay parameter.

(b) *Sparsity Constraint.* The network architecture should be designed such that each training sample can be properly represented by a unique code and, therefore, can be reconstructed from the code with a small reconstruction error. This goal can be effectively achieved by making the code a discrete variable with a small number of different values or by making the code have a lower dimension than the input; alternatively, the code could be forced to be a “sparse” vector in which most of the components are zero [23].

Sparse overcomplete representations have a number of theoretical and practical advantages. Overcomplete representations have a basis vector that is greater than the dimensionality of the input. In particular, they have good robustness to noise [24]. We want hidden units to be inactive most of the time; that is, the outputs of the neurons should be close to zero for the sigmoid activation function. Then, we will write $a_j^{(2)}(x)$ to denote the activation of this hidden unit when the network is given a specific input x . Furthermore, $\hat{\rho}_j = (1/m) \sum_{i=1}^m [a_j^{(2)}(x^{(i)})]$ denotes the average activation of hidden unit j (averaged over the training set). Then, the constraint $\hat{\rho}_j = \rho$ is imposed, where ρ is a sparsity parameter,

which is typically a small value close to zero; in our case, we used 0.1 [25].

To make the hidden unit’s activation values penalty term will give reasonable results. The close to zero, an extra penalty term that penalizes $\hat{\rho}_j$ deviating significantly from ρ is added in our optimization objective. Many choices of the penalty term will give reasonable results. The following is chosen [22]:

$$\sum_{j=1}^{s_2} \text{KL}(\rho \parallel \hat{\rho}_j) = \sum_{j=1}^{s_2} \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j}. \quad (14)$$

Here, s_2 is the number of neurons in the hidden layer, the index j sums up the hidden units in our network, and $\text{KL}(\rho \parallel \hat{\rho}_j)$ is the Kullback-Leibler (KL) divergence between a Bernoulli random variable with a mean of ρ and a Bernoulli random variable with a mean of $\hat{\rho}_j$.

On adding the penalty term, the overall cost function becomes

$$J(W, b)_{\text{sparse}} = J(W, b) + \beta \sum_{j=1}^{s_2} \text{KL}(\rho \parallel \hat{\rho}_j). \quad (15)$$

The term β controls the weight of the sparsity penalty term [22].

3.2. *Stacked Autoencoder.* An efficient way to learn a complicated map is to combine a set of simpler models that are trained sequentially. The combined model performs a nonlinear transformation on the input vectors and produces an output that will be used as an input for the next model in the sequence. As shown in Figure 5, each autoencoder produces a more abstract representation of its input from the former autoencoder, and therefore, some of the stacked autoencoders can be pretrained to produce a high-level representation of the input data. In addition, fine-tuning the network parameters based on the pretraining can prevent its solution from getting stuck at a poor local minimum [22].

3.3. *Unsupervised Feature Learning Using a Greedy Layer-Wise Approach.* To learn the high-level features of the input in an unsupervised fashion, a greedy layer-wise approach is applied to train each autoencoder in turn. Formally, for a stacked autoencoder with n layers, $W^{(k,1)}, W^{(k,2)}, b^{(k,1)}$, and $b^{(k,2)}$ denote the parameters $W^{(1)}, W^{(2)}, b^{(1)}$, and $b^{(2)}$ for the k th autoencoder. Then, the encoding step for the stacked autoencoder is given by running the encoding step of each layer in forward order [22]:

$$a^{(l)} = f(z^{(l)}). \quad (16)$$

The decoding step is in reverse order:

$$\begin{aligned} a^{(n+1)} &= f(z^{(n+1)}), \\ z^{(n+l+1)} &= W^{(n-l,2)} a^{(n+1)} + b^{(n-l,2)}, \end{aligned} \quad (17)$$

where $a^{(n)}$ is an activation value of the deepest layer of the hidden units.

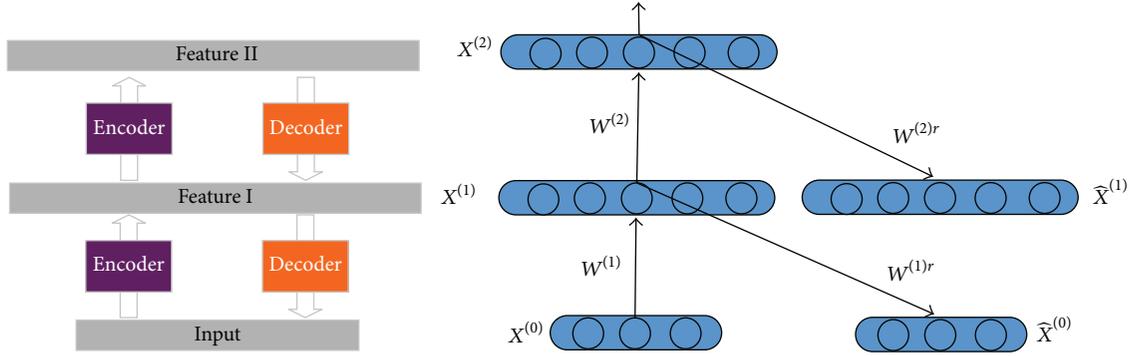


FIGURE 5: Schematic of SAE.

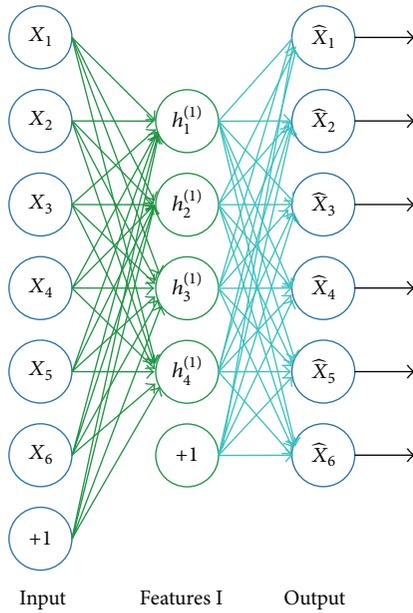


FIGURE 6: First-order representation.

First, we train the first layer on raw input to obtain the parameters $W^{(1,1)}$, $W^{(1,2)}$, $b^{(1,1)}$, and $b^{(1,2)}$. We use the first layer to transform the raw input into a vector that consists of the activation of the hidden units A . We train the second layer on this vector to obtain the parameters $W^{(2,1)}$, $W^{(2,2)}$, $b^{(2,1)}$, and $b^{(2,2)}$. We repeat this sequence of actions for subsequent layers, using the output of each layer as input for the subsequent layer.

In this paper, a stacked autoencoder of two hidden layers is trained for the rolling bearing fault identification. First, a sparse autoencoder is trained to learn the first-order features $h^{(1)(k)}$ of the inputs $x^{(k)}$ (as shown in Figure 6).

Next, we feed the raw input into this trained sparse autoencoder, obtaining the primary feature activation $h^{(1)(k)}$ for each of the inputs $x^{(k)}$. We then use these primary features as “raw input” for another sparse autoencoder to learn the secondary features $h^{(2)(k)}$ on these primary features (as shown in Figure 7).

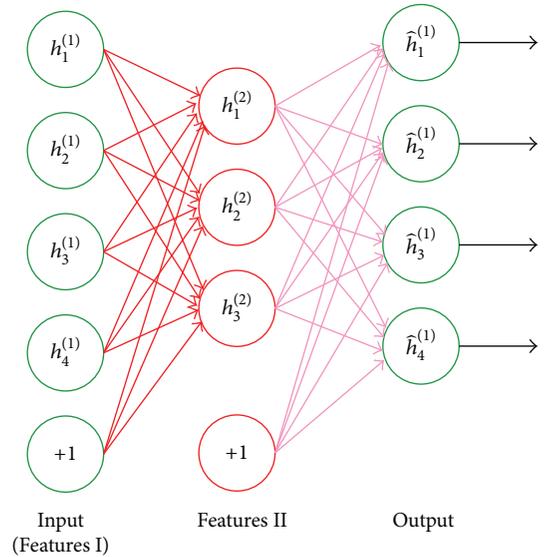


FIGURE 7: Second-order representation.

Next, we feed the primary features into the second sparse autoencoder to obtain the secondary feature activation $h^{(2)(k)}$ for each of the primary features $h^{(1)(k)}$ (which correspond to the primary features of the corresponding inputs $x^{(k)}$). The secondary features can then be treated as “raw input” to a softmax classifier, training it to map secondary features to the discrete digit labels (as shown in Figure 8).

Finally, two autoencoders and one classifier are wired together, building a stacked autoencoder with two hidden layers and a final softmax classifier layer that is capable of classifying the rolling bearing fault as desired (as shown in Figure 9).

3.4. Fine-Tuning Based on Back-Propagation. The greedy layer-wise approach pretrains the parameters of each layer individually while freezing the parameters for the remainder of the model. To produce better results, after this phase of training is completed, fine-tuning using backpropagation can be used to improve the results by tuning the parameters of all of the layers at the same time.

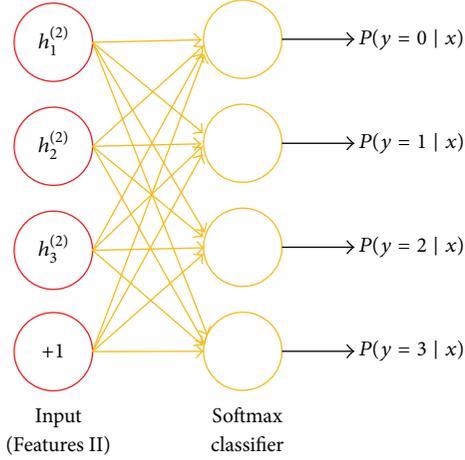


FIGURE 8: Softmax classifier.

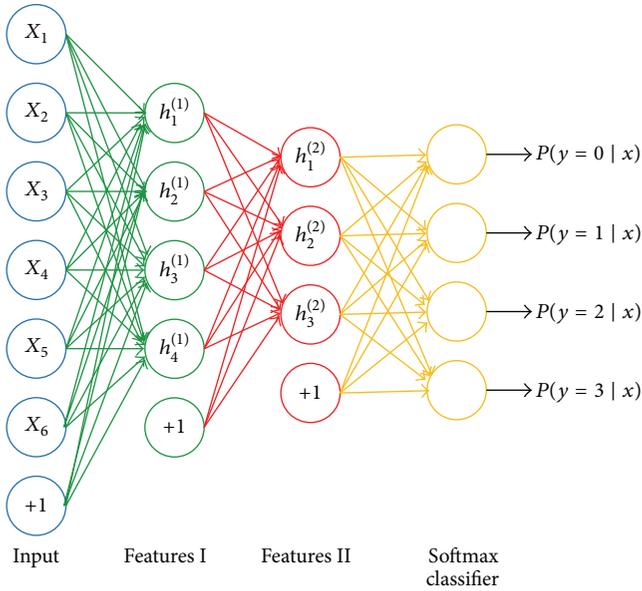


FIGURE 9: SAE with softmax classifier.

Fine-tuning of the weights of the network produces much better classification performance on the test data. It treats all of the layers of a stacked autoencoder as a single model, in such a way that in one iteration we can use the backpropagation algorithm to improve all of the weights in the stacked autoencoder. A summary of the fine-tuning with backpropagation using element-wise notation is given below [22]:

- (1) Perform a feedforward pass, computing the activation values for layers L_1 and L_2 , up to the output layer L_{n_l} , using the equations that define the forward propagation steps.
- (2) For the output layer (layer n_l), set

$$\delta^{(l)} = -(\nabla_{a^{n_l}} J) \cdot f'(z^{(n_l)}). \quad (18)$$

When using softmax regression, the softmax layer has $\nabla J = \theta^T(I - P)$, where I is the input labels and P is the vector of conditional probabilities.

- (3) For $l = n_{l-1}, n_{l-2}, \dots, 2$, we set

$$\delta^{(l)} = \left((W^{(l)})^T \delta^{(l+1)} \right) \cdot f'(z^{(l)}). \quad (19)$$

- (4) Compute the desired partial derivatives:

$$\begin{aligned} \nabla_{W^{(l)}} J(W, b; x, y) &= \delta^{(l+1)} (a^{(l)})^T, \\ \nabla_{b^{(l)}} J(W, b; x, y) &= \delta^{(l+1)}. \end{aligned} \quad (20)$$

In this paper, we could consider the softmax classifier as an additional layer, but its derivation is calculated in a different way. Specifically, we consider the “last layer” of the network to be the features that are input into the softmax classifier. Therefore, the derivatives (in Step (2)) are computed using $\delta^{(l)} = -(\nabla_{a^{n_l}} J) \cdot f'(z^{(n_l)})$, where $\nabla J = \theta^T(I - P)$.

All of the weights and biases of the network in Figure 9 have been improved in the above four steps. The pretrained and fine-tuned SAE possesses the basic characteristics and performances of biological neural systems, in which different hidden layers extract different abstract characteristics, and the more abstract high-level feature has obvious superiority for classification. For a complex morphological and topological structure, SAE can provide powerful capacity in nonlinear modeling or prognostics and has several obvious advantages in large-scale parallelism, distributed processing, and self-organizing or self-learning.

4. Pattern Classification Based on Softmax Regression

A softmax classifier is a generalized logistic regression where the class labels can take on multiple values [22, 26].

For the training set $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$, we have that $y^{(i)} \in \{1, 2, \dots, k\}$. For a given test input x , the hypothesis estimates the probability $p(y = j | x)$ or each value of $j = 1, \dots, k$, where k is the number of classes, that is, the estimate of the probability of the class label taking on each of the k different possible values. Thus, the hypothesis outputs a k -dimensional vector that gives the k estimated probabilities. Concretely, our hypothesis $h_\theta(x)$ takes the form [22]

$$h_\theta(x^{(i)}) = \begin{bmatrix} P(y^{(i)} = 1 | x^{(i)}; \theta) \\ P(y^{(i)} = 2 | x^{(i)}; \theta) \\ \vdots \\ P(y^{(i)} = k | x^{(i)}; \theta) \end{bmatrix} \quad (21)$$

$$= \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix},$$

where $\theta_1, \theta_1, \dots, \theta_k \in \mathfrak{R}^{n+1}$ are the model's parameters. Note that the term $\sum_{j=1}^k e^{\theta_j^T x^{(i)}}$ normalizes the distribution, in such a way that it sums to one. For convenience, we will also write θ to denote all of the parameters of our model. When softmax regression is implemented, it is usually convenient to represent θ as a k -by- $(n+1)$ matrix that is obtained by stacking up $\theta_1, \theta_1, \dots, \theta_k$ in rows, and thus, $\theta = \begin{bmatrix} \theta_1 & \theta_2 & \dots & \theta_k \end{bmatrix}^T$.

The cost function used by the softmax regression is given by

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k 1\{y^i = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \right] + \frac{\lambda}{2} \sum_{i=1}^k \sum_{j=0}^n \theta_{ij}^2. \quad (22)$$

In the equation above, $1\{\cdot\}$ is the indicator function, which means that $1\{\text{a true statement}\} = 1$ and $1\{\text{a false statement}\} = 0$. With this weight decay term (for any $\lambda > 0$), the cost function $J(\theta)$ is strictly convex and is guaranteed to have a unique solution. The Hessian is invertible, and because $J(\theta)$ is convex, algorithms such as gradient descent and L-BFGS (limited-memory Broyden-Fletcher-Goldfarb-Shanno) are guaranteed to converge to the global minimum [22].

One can show that the derivative of $J(\theta)$ is

$$\begin{aligned} \nabla_{\theta_j} J(\theta) &= -\frac{1}{m} \sum_{i=1}^m \left[x^{(i)} \left(1\{y^{(i)} = j\} - p(y^{(i)} = j | x^{(i)}; \theta) \right) \right] + \lambda \theta_j. \end{aligned} \quad (23)$$

By minimizing $J(\theta)$ with respect to θ , we will have a working implementation of softmax regression.

5. Rolling Bearing Fault Diagnosis Based on STFT and SAE

5.1. The Proposed Fault Diagnosis Scheme. In this section, a novel rolling bearing fault diagnosis method based on STFT and SAE is proposed, and Figure 10 briefly depicts the overall scheme for the fault identification.

(1) *Recording and Preprocessing.* Sound signals are acquired by a recording device, and each sample is approximately one minute in duration. Furthermore, the outliers in the data are removed or replaced.

(2) *The STFT Analysis of the Sound Signals.* In this step, the spectrogram algorithm is used to obtain the spectra and spectrum matrixes of the sounds, whose related parameter settings will be detailed in Section 5.2.4.

(3) *Data Normalization and Selection.* For convenient subsequent data processing, spectrum matrixes are normalized by column into gray-value matrixes. Min-max normalization,

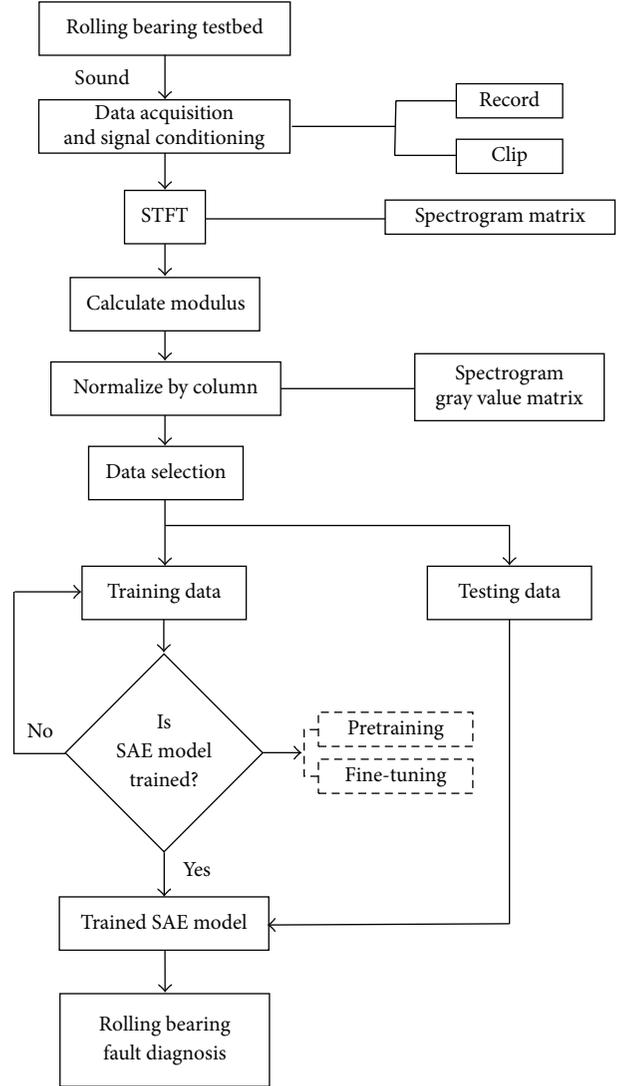


FIGURE 10: Flowchart of rolling bearing fault diagnosis.

called deviation normalization, is conducted in this paper, which maps each element of the matrixes to an integer value from 0 to 255. The transform function can be written as follows:

$$x^* = \frac{x - \min}{\max - \min} \times 255. \quad (24)$$

Here, \min is the minimum, while \max is the maximum in a column.

After the modulus of each spectrogram element is first determined, normalization is performed. Certain data from each column in the center of the matrixes is finally chosen to be inputs of the SAE network.

(4) *Fault Feature Extraction Based on SAE.* The SAE of two hidden layers can be trained by spectrogram data in an unsupervised way, which is a deep learning process. An eventual representation of the raw data is achieved by layer-by-layer learning, where the outputs of the first hidden layer become the inputs of the second hidden layer.

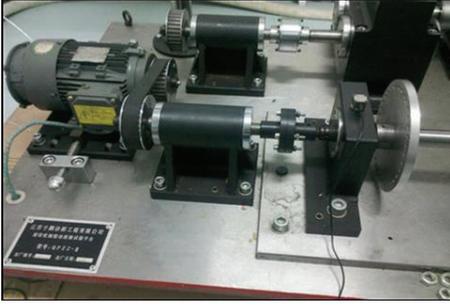


FIGURE 11: Rolling bearing test stand.

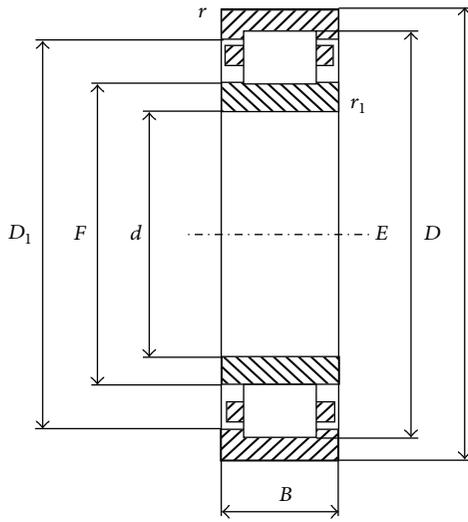


FIGURE 12: Profile of the rolling bearing.

(5) *Fault Modes Classification Based on a Softmax Classifier.* First, the eventual fault feature representation from the SAE is transformed into inputs of the softmax classifier. Through minimizing the cost function, the probability of each classification result will be calculated. Consequently, if one fault probability emerges as the maxima, then the input data can be identified as that fault.

5.2. Experimental Data Analysis

5.2.1. *Data Acquisition.* As shown in Figure 11, the test stand consists of a motor, a belt transmission, a coupling, and two bearing housings. The test bearings support a shaft with a turntable. In the test, four N205 bearings with different faults are installed and tested in turn, among which there are one normal bearing and three fault bearings of one inner-race fault, one outer-race fault, and one rolling parts fault. The structure and basic structural parameters of the tested rolling bearings are depicted, respectively, in Figure 12 and in Table 1. The sound data were acquired using a recorder, which was attached on a steel scaffold near the bearing block but without contacting the test stand, at 44,100 samples per second under the rotational speed of 1200 rpm.

5.2.2. *The STFT Analysis.* Spectrogram in the Matlab 8.1 function library is employed to extract the time-frequency

TABLE 1: Rolling bearing parameters.

d /mm	D /mm	D_0 /mm	d_0 /mm	α /deg	z /piece
25	52	38.5	7.5	0	12

The parameters in the table are as follows: d : inner diameter; D : outer diameter; D_0 : pitch diameter; d_0 : rolling parts diameter; α : contact angle; and z : number of rolling parts.

TABLE 2: Settings of STFT.

Window function	Window size	N_{overlap}	N_{fft}	f_s
Hamming window	44100	44000	44100	44100

The parameters in the table are as follows: N_{overlap} : the number of overlapping points; N_{fft} : the number of fast Fourier transform points; and f_s : the sampling frequency.

TABLE 3: Settings of SAE.

Layer 1	The number of input layer nodes	6500
	The number of hidden layer nodes	1000
	Sparsity	0.1
	Sparsity penalty factor of the loss function	3
Layer 2	Weight decay factor of the loss function	0.003
	The number of input layer nodes	1000
	The number of hidden layer nodes	100
	Sparsity	0.1
Layer 3	Sparsity penalty factor of the loss function	3
	Weight decay factor of the loss function	0.003
	The number of input layer nodes	100
	The number of hidden layer nodes	10
Softmax classifier	Sparsity	0.1
	Sparsity penalty factor of the loss function	3
Softmax classifier	Weight decay factor of the loss function	0.003
	The number of input layer nodes	10
	The number of output layer nodes	4

information in the sounds, and the spectrograms of the four fault modes are described in Figure 13.

5.2.3. *Data Normalization and Selection.* Min-max normalization is performed in this section to map each element of the spectrogram matrixes to an integer from 0 to 255. The acquired gray images are shown in Figure 14, and a method for selecting the SAE network's inputs is proposed in Figure 15.

5.2.4. *The Experiment on Fault Modes Classification.* According to the proposed diagnosis scheme, an SAE with a softmax classifier network is proposed to automatically identify the faults after simple data preprocessing. The experimental parameters are set up as follows.

(1) *Settings of STFT.* The parameter settings of the STFT are shown in Table 2.

(2) *Settings of SAE.* The parameter settings for SAE are detailed in Table 3.

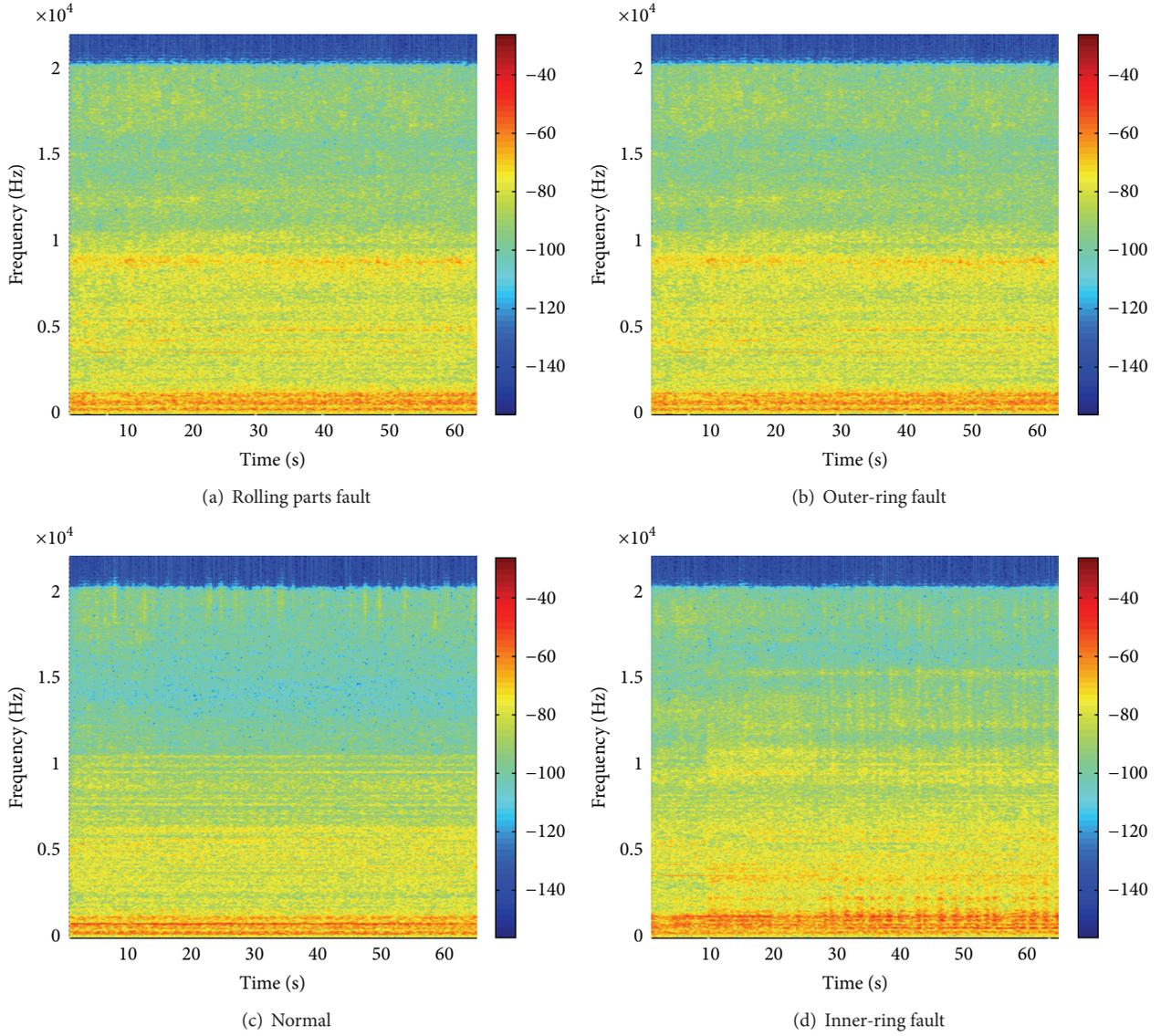


FIGURE 13: Spectrograms of the four fault modes.

TABLE 4: Data information.

Group index	Normal	Inner-race fault	Outer-race fault	Rolling parts fault	Total
G1	3000	3000	3000	3000	12000
G2	3000	3000	3000	3000	12000

5.2.5. Results of the Test Analysis. Under the above settings, SAE with a softmax classifier is trained and then used to recognize the faults of the rolling bearings by the sound signals. The proposed approach can be verified by a two-set cross-validation method, where the data are divided in half; one-half is selected to be the training data, and the other half is selected to be the testing data. An introduction to the data is shown in Table 4.

First, the proposed method is applied to identify whether a testing bearing is a failure or not. The experimental results are given in Table 5. From the chart, the classification

TABLE 5: Identification results on two fault modes.

Cross-validation	Classification accuracy after fine-tuning	Average
G1 for training and G2 for testing	98.35%	97.84%
G2 for training and G1 for testing	97.33%	

accuracy of each validation is higher than 97% after the networks are fine-tuned, and the average could reach up to

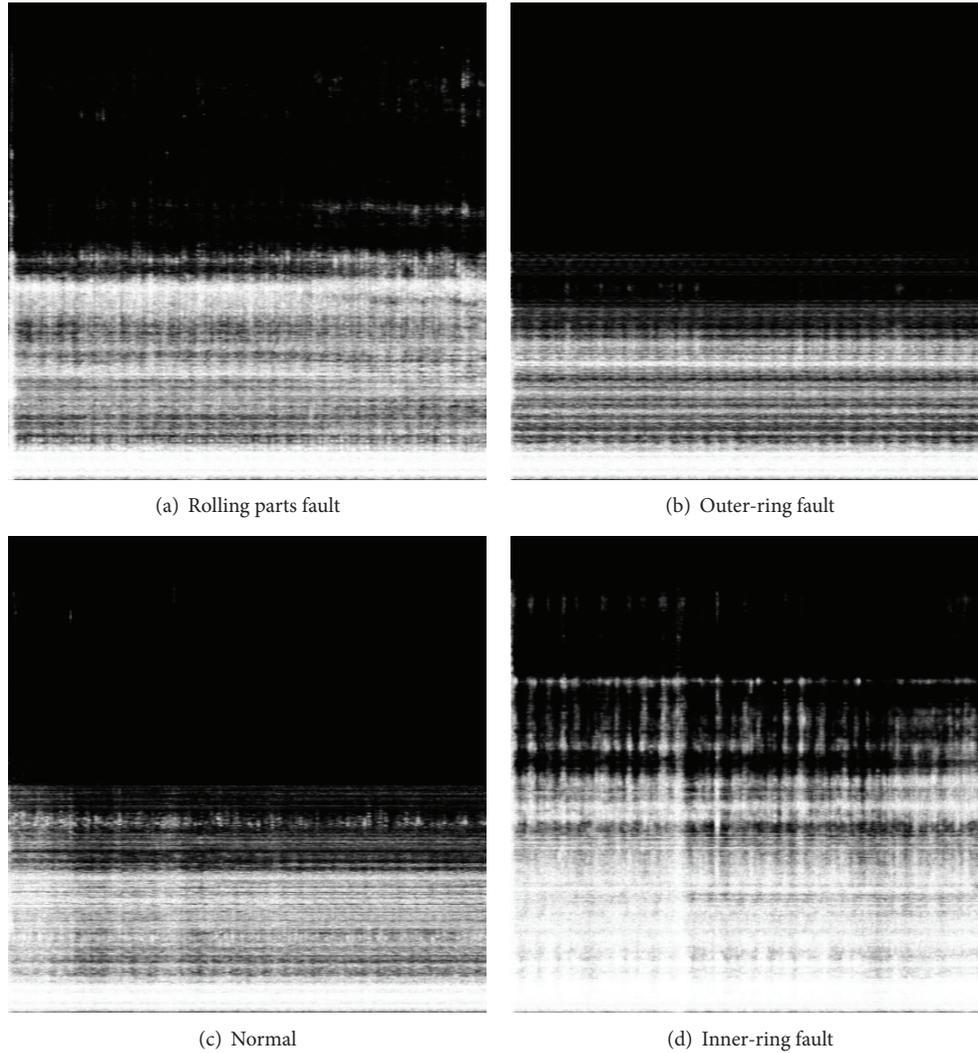


FIGURE 14: Gray images of four fault modes.

TABLE 6: Identification results on four fault modes.

Cross-validation	Normal	Inner-race fault	Outer-race fault	Rolling parts fault	Average
G1 for training and G2 for testing	100%	93.4%	100%	100%	95.68%
G2 for training and G1 for testing	100%	99.1%	90.24%	100%	

97.84%, which demonstrates that the method has excellent and powerful capability for use in health detection.

Next, this method is used to recognize the faults from the normal, inner-race fault, outer-race fault, and rolling bearing parts fault bearings. The diagnosis results are shown in Table 6. From the table, we can determine that the method has good recognition performance on four fault modes, and it increases the average recognition rate to 95.68%.

5.3. Comparisons of the Proposed Method with EMD-TEO and SAE Using Vibration Signals. In this subsection, based on vibration signals, EMD-TEO and SAE are also employed

to diagnose rolling bearing faults. The analysis results are illustrated in detail.

5.3.1. EMD-TEO Based on Vibration Signals. A fault diagnosis method based on empirical mode decomposition (EMD), Teager Energy Operator (TEO), and the softmax classifier is described as follows: First, vibration signals are decomposed into several Intrinsic Mode Components (IMFs) by using EMD. Second, TEO is used to extract the instantaneous amplitudes of the IMFs. Third, several amplitude ratios in the frequency spectra of demodulated IMFs are extracted as fault feature vectors, and then, Principal Components Analysis

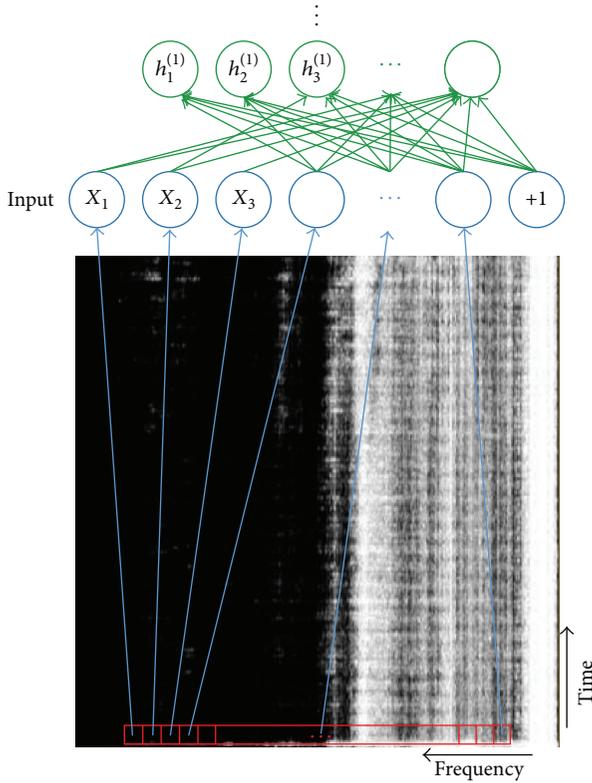


FIGURE 15: Data selection method.

TABLE 7: Fault identification results using EMD-TEO.

	Normal	Inner-race fault	Outer-race fault	Ball fault	Average
Accuracy	100%	90%	90%	100%	95%

(PCA) is applied for dimensionality reduction. Finally, these feature vectors are taken as inputs to train and test the softmax classifier. The diagnosis results are shown in Table 7.

5.3.2. SAE Based on Vibration Signals. In this part, SAE with a softmax classifier network is utilized to automatically identify faults based on vibration signals.

(1) *Settings of SAE.* The parameter settings of SAE are shown in Table 8.

(2) *Identification Results.* Under the above settings, SAE with a softmax classifier is trained and then used to recognize the faults of rolling bearings by vibration signals. The identification results are shown in Table 9.

5.3.3. Comparison Conclusion. From Tables 6, 7, and 9, SAE combined with STFT using sound signals can realize equal fault identification performance with traditional EMD-TEO and SAE based on vibration signals, but the EMD-TEO method spends too much time on artificially extracting the fault features, and specific instruments are required to acquire the vibration signals.

TABLE 8: Settings of SAE.

	The number of input neurons	128
	The number of hidden neurons	64
Layer 1	Sparsity	0.1
	Sparsity penalty factor of the loss function	3
	Weight decay factor of the loss function	0.003
	The number of input neurons	64
	The number of hidden layer neurons	32
Layer 2	Sparsity	0.1
	Sparsity penalty factor of the loss function	3
	Weight decay factor of the loss function	0.003
	The number of input neurons	32
	The number of hidden neurons	16
Layer 3	Sparsity	0.1
	Sparsity penalty factor of the loss function	3
	Weight decay factor of the loss function	0.003
Softmax classifier	The number of input neurons	16
	The number of output neurons	4

TABLE 9: Fault identification results using SAE.

	Normal	Inner-race fault	Outer-race fault	Ball fault	Average
Accuracy	100%	95.53%	91.17%	98.46%	96.29%

6. Conclusions

Because traditional feature extraction methods are time-consuming and require more experience, a novel rolling bearing fault diagnosis method based on STFT and a deep learning network is proposed. By STFT, the original sound signals are mapped into time-frequency space first. Then, SAE is proposed to automatically extract the intrinsic fault features of the rolling bearings. Last, softmax regression is utilized to recognize the fault modes of the feature vectors. Comparison results reveal that the proposed method outperforms traditional fault diagnosis method using vibration signals and realize equal fault identification performance with SAE based on vibration signals.

The proposed method is much easier to apply widely in a highly automated industry because it is data-driven without human interference. In particular, for large and nonstandard bearings, this method can be implemented to analyze fault locations and, thus, help operators and manufacturers to replace the faulty part. Due to the favorable robustness and diagnostic performance, this method can also be easily applied for fault diagnosis in a wide spectrum of machines.

Limited by the consumption of computer resources, to some extent, the proposed method might not be sufficiently satisfactory in “real time.” As expected, STFT and spectrogram functions quickly consume a vast amount of memory for their extensive matrix operations. Furthermore, the accuracy and efficiency of the proposed method would probably be influenced by changes in the working conditions, such as a changed rotation speed. Therefore, further study can be conducted on decreasing the consumption of computer

memory and increasing its adaptability to new working conditions in advance.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant nos. 51605014, 61074083, 51575021, and 51105019) as well as the Technology Foundation Program of National Defense (Grant no. Z132013B002).

References

- [1] K. Shibata, A. Takahashi, and T. Shirai, "Fault diagnosis of rotating machinery through visualisation of sound signals," *Mechanical Systems and Signal Processing*, vol. 14, no. 2, pp. 229–241, 2000.
- [2] J. Lin, "Feature extraction of machine sound using wavelet and its application in fault diagnosis," *NDT & E International*, vol. 34, no. 1, pp. 25–30, 2001.
- [3] D. Yu, J. Cheng, and Y. Yang, "Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings," *Mechanical Systems and Signal Processing*, vol. 19, no. 2, pp. 259–270, 2005.
- [4] W. Y. Liu, W. H. Zhang, J. G. Han, and G. F. Wang, "A new wind turbine fault diagnosis method based on the local mean decomposition," *Renewable Energy*, vol. 48, pp. 411–415, 2012.
- [5] H. Liu, X. Wang, and C. Lu, "Rolling bearing fault diagnosis based on LCD-TEO and multifractal detrended fluctuation analysis," *Mechanical Systems and Signal Processing*, vol. 60, pp. 273–288, 2015.
- [6] W. Sun, J. Chen, and J. Li, "Decision tree and PCA-based fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 21, no. 3, pp. 1300–1317, 2007.
- [7] S. W. Choi, C. Lee, J.-M. Lee, J. H. Park, and I.-B. Lee, "Fault detection and identification of nonlinear processes based on kernel PCA," *Chemometrics and Intelligent Laboratory Systems*, vol. 75, no. 1, pp. 55–67, 2005.
- [8] Z. Li, X. Yan, C. Yuan, J. Zhao, and Z. Peng, "The fault diagnosis approach for gears using multidimensional features and intelligent classifier," *Noise & Vibration Worldwide*, vol. 41, no. 10, pp. 76–86, 2010.
- [9] Z. Wei, Z. Weijia, and L. Bin, "Fault diagnosis approach based on fractal dimension LLE and Fisher discriminant," *Chinese Journal of Scientific Instrument*, vol. 31, no. 2, pp. 325–333, 2010.
- [10] G. B. Wang, X. Q. Zhao, and Y. H. He, "Fault diagnosis method based on supervised incremental local tangent space alignment and SVM," *Applied Mechanics and Materials*, vol. 34–35, pp. 1233–1237, 2010.
- [11] L. Shuang and L. Meng, "Bearing fault diagnosis based on PCA and SVM," in *Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA '07)*, pp. 3503–3507, IEEE, Harbin, China, August 2007.
- [12] B. Bagheri, H. Ahmadi, and R. Labbafi, "Application of data mining and feature extraction on intelligent fault diagnosis by artificial neural network and k-nearest neighbor," in *Proceedings of the 19th International Conference on Electrical Machines (ICEM '10)*, pp. 1–7, IEEE, Rome, Italy, September 2010.
- [13] C. Chen and C. Mo, "A method for intelligent fault diagnosis of rotating machinery," *Digital Signal Processing*, vol. 14, no. 3, pp. 203–217, 2004.
- [14] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [15] M. K. Kıymık, I. Güler, A. Dizibüyük, and M. Akin, "Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application," *Computers in Biology and Medicine*, vol. 35, no. 7, pp. 603–616, 2005.
- [16] F. Jurado and J. R. Saenz, "Comparison between discrete STFT and wavelets for the analysis of power quality events," *Electric Power Systems Research*, vol. 62, no. 3, pp. 183–190, 2002.
- [17] L. Deng and I. Kheirallah, "Dynamic formant tracking of noisy speech using temporal analysis on outputs from a nonlinear cochlear model," *IEEE Transactions on Biomedical Engineering*, vol. 40, no. 5, pp. 456–467, 1993.
- [18] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [19] Y. Bengio and Y. LeCun, "Scaling learning algorithms towards AI," *Large-Scale Kernel Machines*, vol. 34, no. 5, pp. 1–41, 2007.
- [20] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, "Exploring strategies for training deep neural networks," *The Journal of Machine Learning Research*, vol. 10, pp. 1–40, 2009.
- [21] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th International Conference on Machine Learning*, pp. 1096–1103, ACM, Helsinki, Finland, July 2008.
- [22] Unsupervised Feature Learning and Deep Learning (UFLDL), 2011, http://deeplearning.stanford.edu/wiki/index.php/UFLDL_Tutorial.
- [23] Y. L. Boureau, S. Chopra, and Y. Lecun, "A unified energy-based framework for unsupervised learning," in *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pp. 371–379, 2007.
- [24] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pp. 315–323, Ft. Lauderdale, Fla, USA, April 2011.
- [25] N. K. Verma, V. K. Gupta, M. Sharma, and R. K. Sevakula, "Intelligent condition based monitoring of rotating machines using sparse auto-encoders," in *Proceedings of the IEEE Conference on Prognostics and Health Management (PHM '13)*, pp. 1–7, June 2013.
- [26] K. Duan, S. S. Keerthi, W. Chu, S. K. Shevade, and A. N. Poo, "Multi-category classification by soft-max combination of binary classifiers," in *Multiple Classifier Systems*, vol. 2709 of *Lecture Notes in Computer Science*, pp. 125–134, Springer, Berlin, Germany, 2003.

Research Article

Bearing Fault Diagnosis Based on Deep Belief Network and Multisensor Information Fusion

Jie Tao,^{1,2} Yilun Liu,^{1,3} and Dalian Yang^{1,4}

¹School of Mechanical and Electrical Engineering, Central South University, Changsha 410083, China

²Key Laboratory of Knowledge Processing and Networked Manufacturing, Hunan University of Science and Technology, Xiangtan 411201, China

³Light Alloy Research Institute, Central South University, Changsha 410083, China

⁴Hunan Provincial Key Laboratory of Health Maintenance for Mechanical Equipment, Hunan University of Science and Technology, Xiangtan 411201, China

Correspondence should be addressed to Jie Tao; caroltaojie@126.com

Received 20 April 2016; Revised 3 August 2016; Accepted 11 August 2016

Academic Editor: Ganging Song

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

In the rolling bearing fault diagnosis, the vibration signal of single sensor is usually nonstationary and noisy, which contains very little useful information, and impacts the accuracy of fault diagnosis. In order to solve the problem, this paper presents a novel fault diagnosis method using multivibration signals and deep belief network (DBN). By utilizing the DBN's learning ability, the proposed method can adaptively fuse multifeature data and identify various bearing faults. Firstly, multiple vibration signals are acquainted from various fault bearings. Secondly, some time-domain characteristics are extracted from original signals of each individual sensor. Finally, the features data of all sensors are put into the DBN and generate an appropriate classifier to complete fault diagnosis. In order to demonstrate the effectiveness of multivibration signals, experiments are carried out on the individual sensor with the same conditions and procedure. At the same time, the method is compared with SVM, KNN, and BPNN methods. The results show that the DBN-based method is able to not only adaptively fuse multisensor data, but also obtain higher identification accuracy than other methods.

1. Introduction

Bearing is one of the critical components which has a broad range of application in mechanical equipment. Due to the overload, fatigue, wear, corrosion, and other reasons, bearing is easily damaged in the process of machine operation. As a matter of fact, more than 50% of rotating machine malfunctions are related to bearing faults [1, 2]. Actually, a rolling bearing fault may lead to equipment intense shaking, apparatus shutdown, stopping producing, and even casualties. In general, the early weak fault of bearing is complicated and hard to detect [3, 4]. Therefore, bearing state monitoring and analysis is very important, in which it can discover early weak fault of the bearing and control the fault damage situation in time.

Recently, fault detection and diagnosis of bearing has been attracting considerable attention. Among all the kinds

of bearing fault diagnosis methods, vibration signal analysis is one of the most principal and useful tools [2]. In vibration-based bearing fault diagnosis, there are two kinds of approaches that have been proven effective to fault diagnosis: signal processing and pattern recognition [1, 3]. Conventional signal processing techniques such as fast Fourier transform (FFT), wavelet transforms (WT), and empirical mode decomposition (EMD) have been applied to bearing fault diagnosis and achieved some effectiveness [5, 6]. For pattern recognition approaches, artificial intelligence and machine learning are extensively used and studied, for example, fuzzy logic, support vector machine (SVM), and artificial neural network (NN) [7, 8]. However, most research only focused on single vibration analysis in bearing fault diagnosis. In fact, when using a single sensor vibration, the fault characteristics are very weak and useful information is limited. So, it requires intricate signal processing and feature

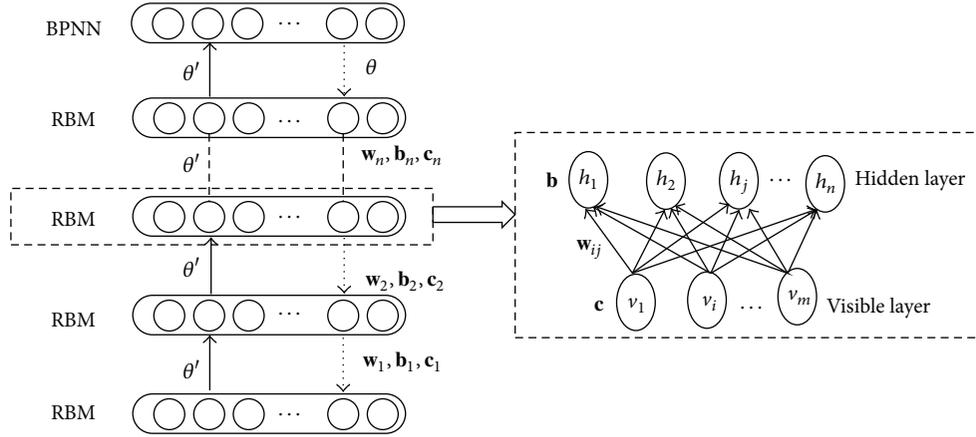


FIGURE 1: Basic structure of DBN.

extraction. Sometimes the accuracy of fault diagnosis is not stable.

To improve the diagnosis accuracy of bearing, some researches put forward the multisignals. At present, fusion of multisource signals mainly focused on three aspects: data level, feature level, and decision level. Among them, the data level fusion primarily mixed the diagnosis objects such as temperature, pressure, and vibration signals [9, 10]. This needs various kinds of sensors and instruments in the process of data gathering. The monitoring cost is expensive and the manipulation is complicated. However, in the convergence of feature level with the same kinds of signals, it needed complex signal analysis and weighted calculation [11, 12]. These methods had some shortcomings such as poor real-time property and weak generalization ability. In the decision level, the intelligent approaches introduced by the literature are, for example, expert systems, decision tree, and SVM [13, 14]. However, these methods all belong to the shallow learning method; the learning ability is lower.

Recently, deep learning became popular in artificial intelligence and machine learning [15]. As a key framework of deep learning, deep belief network (DBN) is primarily constituted by stacked restricted Boltzmann machines (RBM) which is a generative stochastic neural network that can learn probability distribution over abundant data [16]. In 2006, Hinton and colleagues utilized contrastive divergence to advance the RBM training process that greatly improved the learning efficiency of the DBN. The essence of DBN is the capability to automatically extract features through a successive learning process; it can mine the features from different aspects of the data in lower levels as input for the next layer [17, 18]. In addition, DBN accomplishes the learning process with an unsupervised pretraining and supervised fine-tuning. So, DBN has more mapping capability and extensive adaptability by a hierarchical structure. Due to the great advantages of DBN, it has obtained good effect in areas such as natural language understanding, image processing, speech recognition, and document recognition [18–20].

Lately, DBN gets the preliminary application in the field of fault diagnosis. Shao et al. [21] developed particle swarm to optimize the structure of the DBN and applied it

to analyze the simulation signals and experimental signals of a rolling bearing, which obtained more accurate and robust results than other intelligent methods. Tamilselvan et al. [22] originally presented a novel multisensor diagnosis methodology which used the DBN in system health diagnosis such as aircraft engine and electric power transformer. Gan et al. [23] constructed a two-layer DBN of rolling-element bearing fault diagnosis, and experiments showed that DBN got highly reliable results compared to those obtained by SVM and BPNN; Lei et al. [24] proposed a method for multistage gear fault diagnosis with deep learning, which can adaptively extract available fault characteristics from the original data and acquire higher diagnostic accuracy than subsistent methods. Tran et al. [25] presented an approach to implement DBN and multi-information for fault diagnosis of reciprocating compressors.

This paper focuses on the early weak fault of rolling bearing and applies the DBN to integrate the time-domain features of multivibration. The remainder of this paper is organized as follows. In Section 2, the methodologies of deep belief network are introduced. In Section 3, the process of multivibration signal fusion is described. In Section 4, a bearing test rig is explained and experiments are conducted for the proposed method. In Section 5, implementation of classifier based on the DBN model is presented. The obtained results and their evaluation are described. Finally, conclusions and future work are given in Section 6.

2. Deep Belief Network

2.1. Deep Belief Network Architecture. DBN is a model based on probability of energy generation, which comprises multiple layers of restricted Boltzmann machines (RBM) and a backpropagation neural network (BPNN) [16]. Figure 1 is the fundamental structure of DBN; the multilayered architecture makes sure that DBN can be trained through bottom-up learning in a sequence of RBMs and top-down fine-tuning by BPNN [17].

Restricted Boltzmann machine, the key prototype of DBN, is structured by a layer of visible (or input) units and a layer of hidden (or output) units. As every unit is binary,

it is trained by the activation probabilities. The units in the same layer are not connected to each other but have directed symmetrical connections to the units in the next layer. In DBN, the hidden layer of the RBM becomes the visible layer of the next RBM, so they set up a successive hierarchy by stacked RBMs.

In RBM, the visible node is denoted by v_i and the hidden node is represented by h_j . The weights between v_i and h_j are directed and denoted by \mathbf{w}_{ij} . The visible and hidden nodes have their biases represented by vectors \mathbf{c} and \mathbf{b} , respectively. \mathbf{b} , \mathbf{c} , and \mathbf{w}_{ij} of all RBMs make up the parameter set θ in DBN. As the values of θ define a probability distribution over the joint states of the visible and hidden nodes by an energy function,

$$E(\mathbf{v}, \mathbf{h}) = -\sum_{i=1}^m v_i \mathbf{c}_i - \sum_{j=1}^n h_j \mathbf{b}_j - \sum_{i=1}^m \sum_{j=1}^n v_i h_j \mathbf{w}_{ij}. \quad (1)$$

The ultimate purpose of DBN training is to find the best θ , which can minimize the model energy error and make the model at an equilibrium state. So, the energy function is utilized to define the joint probability distribution between \mathbf{v} and \mathbf{h} as follows:

$$P(\mathbf{v}, \mathbf{h} | \theta) = \frac{1}{Z(\theta)} e^{-E(\mathbf{v}, \mathbf{h} | \theta)}, \quad (2)$$

$$Z(\theta) = \sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h} | \theta)}.$$

Since DBN has no intralayer connections, the conditional probability distributions of visible and hidden nodes can be calculated by

$$P(v_i = 1 | \mathbf{h}) = \frac{1}{1 + \exp(-\mathbf{b}_i - \sum_j h_j \mathbf{w}_{ij})}, \quad (3)$$

$$P(h_j = 1 | \mathbf{v}) = \frac{1}{1 + \exp(-\mathbf{c}_j - \sum_i v_i \mathbf{w}_{ij})}. \quad (4)$$

2.2. The DBN Training Process. Generally, the DBN training procedure includes two parts: pretraining and fine-tuning. The pretraining is an unsupervised learning procedure which used the unlabeled data to train the individual RBM. The fine-tuning is a supervised learning process which utilized the backpropagation algorithm to further adjust the parameters.

In the pretraining, each layer is trained by the RBM rules. Since the RBM model is with binary units, it can be learned by stochastic gradient descent on the negative log-likelihood probability of the training data. The functions are as follows:

$$\frac{\partial \ln p(\mathbf{v}; \theta)}{\partial \mathbf{w}_{ij}} = \langle v_i h_j \rangle_d - \langle v_i h_j \rangle_m,$$

$$\frac{\partial \ln p(\mathbf{v}; \theta)}{\partial \mathbf{b}} = \langle h_j \rangle_d - \langle h_j \rangle_m, \quad (5)$$

$$\frac{\partial \ln p(\mathbf{v}; \theta)}{\partial \mathbf{c}} = \langle v_i \rangle_d - \langle v_i \rangle_m,$$

where $\langle \cdot \rangle_d$ denotes an expectation of the data distribution and $\langle \cdot \rangle_m$ is an expectation of the distribution defined by the model.

With the RBM property, it is easy to compute an unbiased sample of $\langle \cdot \rangle_d$ to the data distribution. However, obtaining an unbiased sample of $\langle \cdot \rangle_m$ is quite difficult [23]. Actually, the RBM learning method closely approximates the gradient objective function called contrastive divergence (CD) [17], in which $\langle \cdot \rangle_m$ is substituted by k iterations of Gibbs sampling as expressed in (6), where an iteration of alternating Gibbs sampling includes updating all parallel visible nodes by using (3), subsequently updating all parallel hidden nodes by (4).

$$\frac{\partial \ln p(\mathbf{v}; \theta)}{\partial \mathbf{w}_{ij}} \approx \langle v_i h_j \rangle^0 - \langle v_i h_j \rangle^k,$$

$$\frac{\partial \ln p(\mathbf{v}; \theta)}{\partial \mathbf{b}} \approx \langle h_j \rangle^0 - \langle h_j \rangle^k, \quad (6)$$

$$\frac{\partial \ln p(\mathbf{v}; \theta)}{\partial \mathbf{c}} \approx \langle v_i \rangle^0 - \langle v_i \rangle^k.$$

Actually, one-step Gibbs sampling has been shown to perform surprisingly well [17]. Based on (6), the updated methods for all parameters are given by the following equation, where σ represents learning rate whose value is between 0 and 1:

$$\Delta \mathbf{w} \leftarrow \sigma \left(\langle v_i h_j \rangle^0 - \langle v_i h_j \rangle^1 \right),$$

$$\Delta \mathbf{b} \leftarrow \sigma \left(\langle h_j \rangle^0 - \langle h_j \rangle^1 \right), \quad (7)$$

$$\Delta \mathbf{c} \leftarrow \sigma \left(\langle v_i \rangle^0 - \langle v_i \rangle^1 \right).$$

In the training process, dataset is usually divided into minibatches with a small number of data vectors and the values of θ are updated after handling each minibatch. To stabilize the RBM learning procedure, a momentum (λ) is often utilized in updating the synaptic weights and biases. With momentum (λ), the θ update, at the current epoch, can be associated with the θ update in the preceding epoch and calculated as

$$\mathbf{w}_n \leftarrow \lambda \mathbf{w}_{n-1} + \sigma \left(\langle v_i h_j \rangle^0 - \langle v_i h_j \rangle^1 \right),$$

$$\mathbf{b}_n \leftarrow \lambda \mathbf{b}_{n-1} + \sigma \left(\langle h_j \rangle^0 - \langle h_j \rangle^1 \right), \quad (8)$$

$$\mathbf{c}_n \leftarrow \lambda \mathbf{c}_{n-1} + \sigma \left(\langle v_i \rangle^0 - \langle v_i \rangle^1 \right).$$

After the bottom-up successive learning, the following step of the DBN training is top-down fine-tuning. Fine-tuning is a supervised learning process which used the backpropagation (BPNN) to further decrease the training error and advance the classification accuracy of the DBN. As the BPNN is supervised learning, fine-tuning uses labeled data for the DBN training. Unlike the unsupervised training in DBN that only deals one RBM at a time, the BPNN simultaneously trains all layers in DBN. The training error of BPNN is calculated with model outputs and the target label data. And the backpropagation learning is continued until the model output attains the maximum number of epochs.

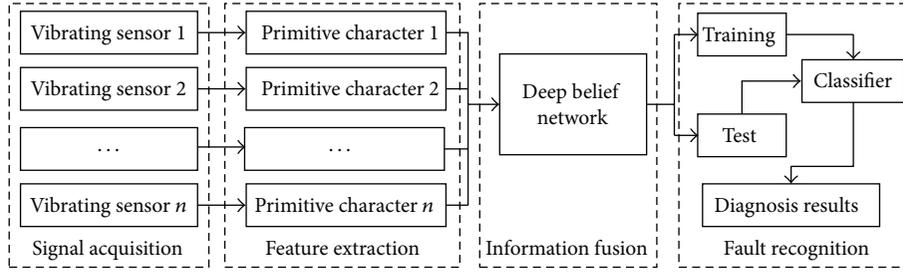


FIGURE 2: The flow diagram of multisensor information fusion.

3. Multisignal Fusion with DBN

Multisensor information fusion technology can obtain more accurate, rich fault features from vibration signals [12]. However, in the conventional information integration, signal processing needs to master a lot of signal processing technologies and to be combined with rich experience in engineering practice to extract fault features. Meanwhile, in the pattern recognition, traditional machine learning only contains single nonlinear transform structure; it cannot adaptively integrate the multi-information [20, 26].

In this paper, we apply the deep belief network (DBN) to adaptively fuse multivibrations. There are four main processes in the proposed bearing: multichannel signal acquisition, feature extraction, information fusion, and fault recognition.

As shown in Figure 2, firstly, the vibration signals are acquainted by each sensor. Secondly, some time-domain characteristics are extracted from original signal of every individual sensor. Thirdly, without any artificial selection, features data of all signal sensors are put into the DBN and generate appropriate DBN classifier. Finally, the integrated information is used to train or test the classifier, and then the classifier puts out the diagnosis results and completes fault diagnosis.

Since the DBN has a hierarchical structure which can extract the features from various aspects of the data by a layer-by-layer successive learning procedure [17], the multi-information fusion, based on deep belief network, can get rid of complex signal processing and complicated experience [24]. It takes the unsupervised learning with RBM and directly extracts feature from the multivibrations and then uses the best parameters to design DBN and completes the multi-information integration.

However, the structure of DBN is closely related to the number of hidden nodes and hidden layers; if the DBN structure is too simple, learning ability is so poor that it cannot effectively integrate the multi-information. Meanwhile, if the DBN structure is too complicated, it not only wastes running time but also produces problems such as overfitting, local extremism, and training failure [26]. Therefore, a method based on data reconstruction error is used to determine the structure of information fusion in DBN.

Figure 3 introduces the optimization process for the DBN structure. The reconstruct error is computed with the model outputs and the objective label data. At the beginning of the procedure, multichannel signal information is put into the

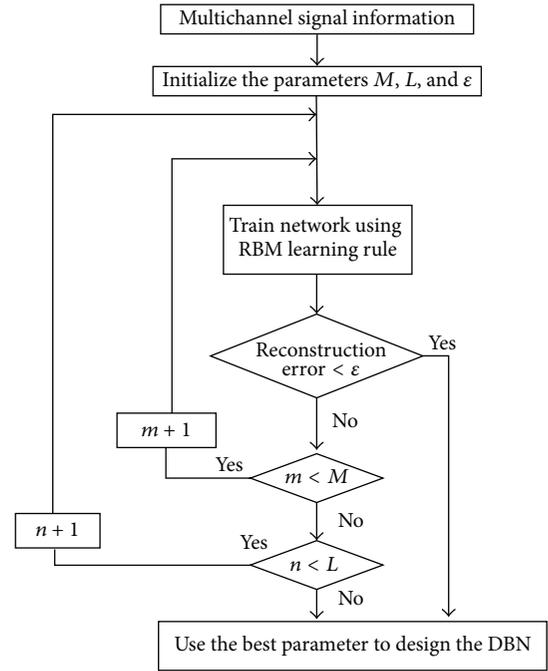


FIGURE 3: The flow chart of the optimization DBN in signal fusion.

DBN and the parameters of M , L , and ϵ are initialized, where the M , L , and ϵ are the max values of the hidden nodes, hidden layers, and reconstruction error, respectively. Then, DBN calculates the reconstruction error of training dataset by RBM learning rules. If the reconstruction error is less than ϵ , it finishes the optimization and puts out the parameters (θ) of DBN. Otherwise, it increases the number of hidden nodes or hidden layers. If the numbers overflow M or L , the procedure finds the best reconstruction error from history and builds the DBN for multi-information fusion.

Table 1 summarizes the procedure of bearing fault diagnosis using multi-information fusion with DBN. As shown in the table, the first step is gating the vibration signals from multichannels and collecting vibration data from each sensor. As the raw samples are nonlinear and unstable, it is necessary to extract some features from each sample. Then, the preprocessed vibration data are divided into training and testing datasets. The DBN structure is optimized by reconstruction error of training dataset and obtains the suitable DBN to accomplish the multi-information fusion.

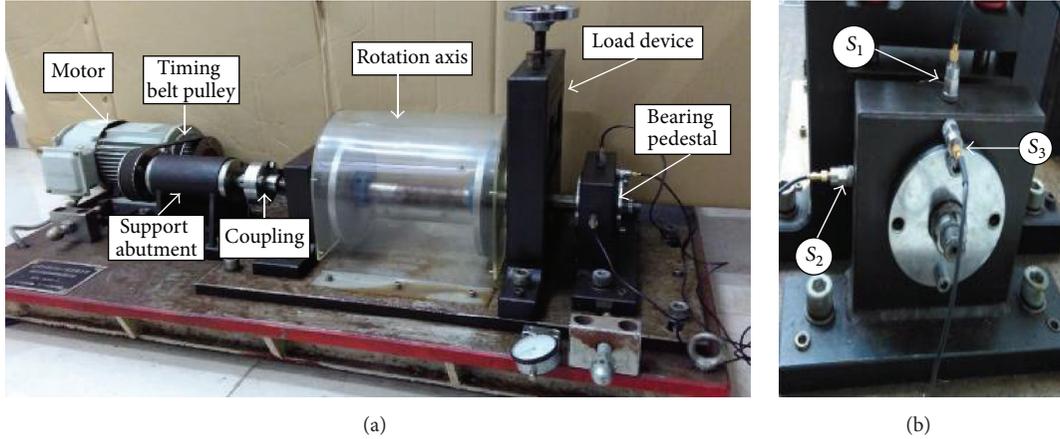


FIGURE 4: Rotating machinery fault simulation platform of QPZ-II: (a) experiment platform; (b) sensors locations.

TABLE 1: Procedure for bearing fault diagnosis using multi-information with DBN.

Step	Description
Step 1	Gather multichannel vibration signals
Step 2	Extract features of each channel sample
Step 3	Input all features of training samples and initial parameters of the DBN
Step 4	Optimize the DBN structure using reconstruction error of multi-information fusion
	(1) Each layer of the DBN is trained using RBM learning rule
	(2) Fine-tune the DBN using backpropagation learning
	(3) Calculate the reconstruction error using model outputs and the target label data
	(4) If the reconstruction error is smaller than ϵ output the DBN structure; otherwise, $m \leftarrow m + 1$ and return to step (1)
(5) If the reconstruction error is not smaller than ϵ and $m = M$, $l \leftarrow l + 1$ and return to step (1) until l is more than n	
Step 5	Develop the DBN using the structure with the best reconstruction error
Step 6	Perform diagnosis using the training DBN classifier model

4. Experimental Setup

In order to measure the validity of the suggested method, a bearing experimental platform is set up as shown in Figure 4. The bearing fault simulation platform was produced by Qian Peng Company with QPZ-II in China.

As shown in Figure 4, the experimental table is mainly constituted with motor, belt coupling, bearing pedestal, and so on. The bearing is installed in the pedestal, and three magnet acceleration sensors are installed in the pedestal, labeled by S_1 , S_2 , and S_3 , respectively. The position of S_1 is located on the vertical side of the bearing pedestal; S_2

and S_3 are, respectively, located on the lateral and front of the bearing pedestal. In the experiments, the variety of typical fault bearings can be installed and dismantled for multivibration collection.

The test bearings are produced by Harbin Bearing Manufacturing Company, China, with the bearing designation being NU205, which have 13 cylindrical rollers. The inner diameter is 25 mm, the outer diameter is 52 mm, and the thickness of the bearing is 15 mm. As shown in Figure 5, four experiments are carried out under each of the following bearing health conditions: the inner race fault, outer race fault, ball fault, and normal. All the faults are linear cutting with electrical discharge machining, and the cutting diameter is 0.5 mm; the cutting depth is 0.3 mm.

In the process of testing, a variable velocity motor directly drives a shaft. The belt on the right of the shaft brings along the coupling which runs with the same speed of motor. In the experiment, the sampling frequency is 10000 hz, the bearing speed is 1200 rpm, and the sampling time is 5 seconds.

According to the steps shown in Table 1, each experiment continuously acquainted 50000 signal points. Meanwhile, the bearing rotated 100 cycles. We select the signal points of a rotation cycle to construct a sample. So, 500 signal points constitute a data sample. There are four conditions defined for classification and 400 (100×4) training samples in dataset. Then, we randomly selected 200 samples constituting the test dataset. The dataset description is shown in Table 2. When the rolling bearing has local damage, it will cause the vibration signal mutation. The local damage position is different and the change of the vibration signal usually is not the same. Figure 6 is the amplitudes waveform of rolling bearing in different conditions.

It is seen from Figure 6 that the vibration signals waveforms are similar, and it is difficult to distinguish the various fault types of rolling bearings. So, some time-domain features are extracted from the original signals; the method is as follows:

- (1) X_i ($i = 1, 2, \dots, n$) is the discrete-time series of the i th sensor, and the vibration signals of bearing rotating a cycle are $X_i = [x_1, x_2, \dots, x_m]$, $m = (f \times 60) \div h$,

TABLE 2: Sample distribution of normal and different faults.

Rolling bearing condition	Fault diameter/mm	Fault depth/mm	Training samples	Test samples
Normal	—	—	100	
Inner race fault	0.5	0.3	100	200
Outer race fault	0.5	0.3	100	
Ball fault	0.5	0.3	100	



FIGURE 5: Pictures of the testing bearings: (a) inner race fault; (b) ball fault; (c) outer race fault; (d) normal.

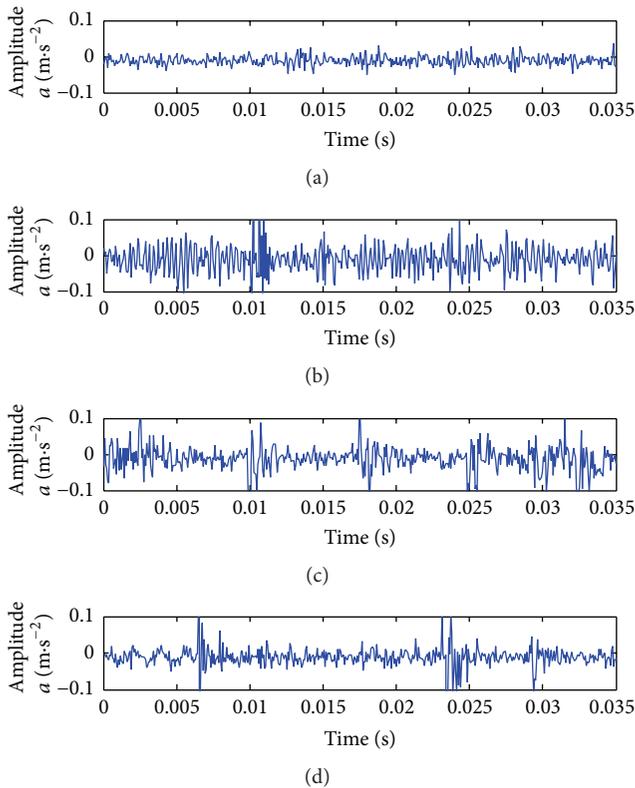


FIGURE 6: Instantaneous amplitudes waveform of rolling bearing in different conditions: (a) normal; (b) inner race fault; (c) outer race fault; (d) ball fault.

where f is the sampling frequency (Hz) and h is the rotational velocity (rpm).

- (2) According to Table 3, compute the time-domain statistics of the vibration signals X_i and get feature vector of the i th sensor as $A_i = [a_1, a_2, \dots, a_{14}]$.
- (3) Three sensors constitute the multisensor feature vector $T = [A_1, A_2, A_3]$.
- (4) Normalize the feature vector $T_i = [t_1, t_2, \dots, t_m]$, $m = n \times 14$.

5. Results and Discussions

5.1. Multisignals and Individual Signals. As the signals pre-processing is clearly explained in Section 4, 14 classical time-domain features are computed from the raw signals.

To illustrate the property of multivibrations fusion, the method based on each single sensor is also measured with the same conditions. These methods are represented as sensors 1, 2, and 3, corresponding to the three individual sensors. The input vectors of DBN in single sensor experiment have only 14 features extracted from vibration signals. The input vectors of DBN in multisensors have 42 features. The DBNs' structures are shown in Table 4.

The DBN structure of multisensors is 42-12-12-4. That is to say, the input layer contained 42 nodes and the output layer included 4 nodes, which depended on the dimensions of the input and output data. There are two hidden layers in the architecture; both hidden layers contained 12 hidden neurons, respectively. However, in sensors 1, 2, and 3 the input nodes are 14 and the hidden nodes are 8 in every hidden layer. The learning rate and momentum are used to adjust the model error and training efficiency. The learning rate in the experiment is selected as 0.01, and the momentum is 0.02 [20]. In the sequential training of every individual RBM, the pretraining of each RBM is accomplished with 20 iterations.

TABLE 3: Statistics features in time domain.

	Parameter	Definition
(1)	Average value \bar{x}	$\frac{1}{n} \sum x_i$
(2)	Maximum value x_{\max}	$\max(x_i)$
(3)	Minimum value x_{\min}	$\min(x_i)$
(4)	Peak-to-peak x_p	$\max(x_i) - \min(x_i)$
(5)	Mean absolute x_{abs}	$\frac{1}{n} \sum x_i $
(6)	Variance x_{var}	$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$
(7)	Standard deviation x_{std}	$\left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2}$
(8)	Kurtosis x_{kur}	$\frac{\sum_{i=1}^n (x_i - \bar{x})^4}{(n-1) x_{\text{std}}^4}$
(9)	Root mean square x_{rms}	$\left(\frac{1}{n} \sum_{i=1}^n x_i ^{1/2} \right)^2$
(10)	Shape factor x_{sf}	$\frac{x_{\text{rms}}}{\bar{x}}$
(11)	Peak factor x_{pf}	$\frac{x_p}{x_{\text{rms}}}$
(12)	Skewness x_{ske}	$\frac{\sum_{i=1}^n (x_i - \bar{x})^3}{(n-1) x_{\text{std}}^3}$
(13)	Impulse factor x_{if}	$\frac{x_p}{\bar{x}}$
(14)	Margin factor x_{mf}	$\frac{x_p}{x_{\text{rms}}}$

In the fine-tuning of model parameters, stochastic gradient descent (SGD) is used to further reduce the training error and improve the information fusion. However, in this research, the SGD takes the minibatch to globally adjust the parameters in DBN. Since there are 400 samples in the training dataset, the number of minibatch is 10 in the experiments.

We use the training dataset to train the DBN model and use the testing dataset to test the model identification accuracy. The classification process is repeated for 25 times and the classification results are averaged as shown in Figure 7.

The average accuracy of the training samples in multisensors is 97.5%, and the number of correct classification samples is 390. This is much higher than those using other methods, which are 91.5%, 85%, and 87.5% with 366, 340, and 350, respectively. The average accuracy of testing samples in multisensors fusion is 95.5%, and the number of correct classification samples is 191. This is much higher than those using other methods, which are 89%, 78.5%, and 75% with 178, 157, and 150, respectively.

Compared to those individual sensors, the training and testing accuracies of multisensors information fusion are obviously higher than other methods. In the three individual sensors, the classified accuracy of S_1 is better than S_2 and S_3 . It is indicated that the sensor put on the vertical location of the testing pedestal is most sensitive to the bearing faults. The results prove that it is more effective to integrate the signals from multisensors than to use the vibration from individual ones.

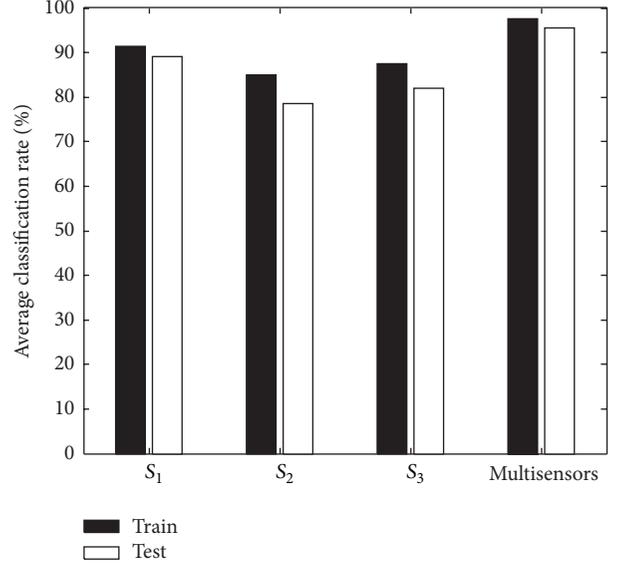


FIGURE 7: Classification rate of individual sensor and multisensors.

5.2. Deep Belief Network and Traditional Methods. To introduce the effectiveness of the suggested method, the comparative methods are also tested with the same data such as SVM, BPNN, and KNN. Without manual feature selection, the 42-dimension feature vector is used directly for rolling bearing fault diagnosis. The architecture of BPNN is 42-12-4; the decision function of KNN is Euclidean distance; the SVM used RBF kernel, which is the most popular and suitable kernel function applied. The penalty factor and the radius of the SVM kernel are 0.216 and 0.3027, each of which is determined through genetic algorithm. All the experiments are finished in the computer with Core i5 Intel central processing unit and 4G memory. In the DBN, SVM, BPNN, and KNN, the experiments are, respectively, repeated 15 times. The average classification accuracies are shown in Tables 5 and 6.

In the training experiments, DBN achieved 95.72% identification accuracy which is better than that of SVM (92.28%), KNN (90.06%), and BPNN (83.63%) for multivibration signals. For the testing experiments, the average accuracy of DBN is 93.17%; meanwhile SVM is 90.13%, KNN is 85.23%, and BPNN is 78.13%. In brief, the experiments results explain that the suggested methods have higher reliability and better accuracy than SVM, KNN, and BPNN in rolling bearing fault diagnosis.

The algorithms of DBN, SVM, KNN, and BPNN all trained the model by the same dataset and generated the classifier to carry out the bearing fault diagnosis. However, the stability and generalization ability are different in these methods. As shown in Tables 5 and 6, the training results of DBN approximately agree with the testing results in the 15 experiments. The testing accuracies in SVM and KNN are lower than training accuracies by 3–5%. The classification accuracy of BPNN is decreased obviously in test experiments.

In the experiments, the training datasets are selected from sequence data samples of all kinds of fault condition,

TABLE 4: The parameters of DBN in sensors 1–3 and multisensors.

Parameters	Sensor 1	Sensor 2	Sensor 3	Multisensors
Input	14	14	14	42
Output	4	4	4	4
Number of hidden layers	2	2	2	2
Number of neurons in a layer	8	8	8	12
Iteration	20	20	20	20
Minibatch size	10	10	10	10
Momentum	0.02	0.02	0.02	0.02
Learning rate	0.01	0.01	0.01	0.01

TABLE 5: Results of training datasets.

Method	Experiment results of each group (%)															Average (%)
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
DBN	95.25	95	95.5	96	95	97.25	95	95.5	95.25	96	97	96.25	95.5	96	95.25	95.72
SVM	94.5	90	91	94	92.5	91	92.5	94	93.5	91	89.5	91	92	94.5	93.25	92.28
KNN	90	88	87.25	87	93.75	89.75	90	89	92	92	92.5	90.25	90	89	90.5	90.06
BPNN	84.5	82.5	80	80.5	82.5	84	86	80	82.5	85	86.5	84	83	88.5	85	83.63

TABLE 6: Results of testing datasets.

Method	Experiment results of each group (%)															Average (%)
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
DBN	93.5	95	93.5	95.5	94.5	96.5	94.5	94.5	94	93.5	96	94.5	95	96	94.5	94.73
SVM	91	89	90	90.5	91	89.5	90	90	91	90	89	90	88.5	92	90.5	90.13
KNN	85	86	82.5	84.5	87.5	85.5	88	84.5	86	84.5	85	84	85.5	84.5	85.5	85.23
BPNN	82.5	80	78.5	77.5	75.5	75	74.5	78	75.5	78	76.5	78	80	80.5	82	78.13

and they orderly composite the training sample datasets. As the testing samples are randomly selected from the various states datasets, both sample category and sample order are random. SVM, to classify data identification, mainly depends on the kernel functions with the training set, which is closely related to the quantity and quality dataset. When using the KNN to classify the data identification, the results mainly are determined by the distance function, and once the distance function is selected, it will not be able to transform. So, the testing accuracies are much less than in training set of SVM and KNN. BPNN is a typically shallow learning model, which involved no more than one nonlinear feature transformation and has difficulty in representing complex functions with poor performance and generalization ability.

Compared with the traditional machine learning and signal processing technology, DBN has the merit to get rid of the dependence on signal processing technology. On the other hand, DBN can adaptively extract the fault feature without restrictive assumptions or complex parameter adjustment. Consequently, it is nothing strange that the DBN as a promising method has been effectively applied in multivibrations fusion.

6. Conclusions

Multiple sensors installed on various locations of bearing pedestal can supply abundant information for fault diagnosis

and detection. Based on this observation, a novel technique using deep belief network for the multivibrations fusion is put forward in this paper. Some conventional time-domain features are extracted from three accelerometer vibration sensors. Without manual feature selection, the features are used directly as the input vectors of the DBN. The obtained accuracy of multisensors is 97.5% which is about 10% higher than single sensor. At the same time, the mean accuracy of DBN, SVM, KNN, and BPNN is, respectively, 93.17%, 90.13%, 85.23%, and 78.13%. It suggests that DBN is more effective and stable for the identification of rolling bearing fault diagnosis than other methods. From the results, it can be realized that DBN is able to adaptively integrate available fault features from multisensors and it obtained higher identification accuracy than traditional methods.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This paper was supported by Nature Science Fund of China, under Grants nos. 61402167 and 51375500, and the National Science and Technology Support Project of China, under Grant no. 2015BAF32B01.

References

- [1] Y. Yang, D. Yu, and J. Cheng, "A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM," *Measurement*, vol. 40, no. 9-10, pp. 943-950, 2007.
- [2] L. Zhang, G. Xiong, H. Liu, H. Zou, and W. Guo, "Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference," *Expert Systems with Applications*, vol. 37, no. 8, pp. 6077-6085, 2010.
- [3] I. El-Thalji and E. Jantunen, "A summary of fault modelling and predictive health monitoring of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 60-61, pp. 252-272, 2015.
- [4] G. Wang, Z. He, X. Chen, and Y. Lai, "Basic research on machinery fault diagnosis-what is the prescription," *Journal of Mechanical Engineering*, vol. 49, no. 1, pp. 63-72, 2013.
- [5] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: a review with applications," *Signal Processing*, vol. 96, pp. 1-15, 2014.
- [6] J. Ben Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Applied Acoustics*, vol. 89, no. 3, pp. 16-27, 2015.
- [7] D. Yang, Y. Liu, S. Li, X. Li, and L. Ma, "Gear fault diagnosis based on support vector machine optimized by artificial bee colony algorithm," *Mechanism and Machine Theory*, vol. 90, pp. 219-229, 2015.
- [8] A. Moosavian, H. Ahmadi, A. Tabatabaefar, and M. Khazaei, "Comparison of two classifiers; K-nearest neighbor and artificial neural network, for fault diagnosis on a main engine journal-bearing," *Shock and Vibration*, vol. 20, no. 2, pp. 263-272, 2013.
- [9] P. Shakya, A. K. Darpe, and M. S. Kulkarni, "Bearing diagnosis using proximity probe and accelerometer," *Measurement*, vol. 80, pp. 190-200, 2016.
- [10] M. S. Safizadeh and S. K. Latifi, "Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell," *Information Fusion*, vol. 18, no. 1, pp. 1-8, 2014.
- [11] X. Li, D. Yang, D. Guo, and L. Jiang, "Fault diagnosis method based on multi-sensors installed on the base and KPCA," *Chinese Journal of Scientific Instrument*, vol. 32, no. 7, pp. 1551-1557, 2011.
- [12] M. Dong and D. He, "Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis," *European Journal of Operational Research*, vol. 178, no. 3, pp. 858-878, 2007.
- [13] T. P. Banerjee and S. Das, "Multi-sensor data fusion using support vector machine for motor fault detection," *Information Sciences*, vol. 217, no. 24, pp. 96-107, 2012.
- [14] L.-L. Jiang, Y.-L. Liu, X.-J. Li, and A.-H. Chen, "Gear fault diagnosis based on SVM and multi-sensor information fusion," *Journal of Central South University*, vol. 41, no. 6, pp. 2184-2188, 2010.
- [15] J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, vol. 61, pp. 85-117, 2015.
- [16] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504-507, 2006.
- [17] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527-1554, 2006.
- [18] D. Yu, L. Deng, I. Jang, P. Kudumakis, M. Sandler, and K. Kang, "Deep learning and its applications to signal and information processing," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 145-154, 2011.
- [19] R. Sarikaya, G. E. Hinton, and A. Deoras, "Application of deep belief networks for natural language understanding," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 4, pp. 778-784, 2014.
- [20] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, "Why does unsupervised pre-training help deep learning?" *Journal of Machine Learning Research*, vol. 11, pp. 625-660, 2010.
- [21] H. Shao, H. Jiang, X. Zhang, and M. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Measurement Science and Technology*, vol. 26, no. 11, Article ID 115002, 2015.
- [22] P. Tamilselvan, Y. Wang, and P. Wang, "Deep belief network based state classification for structural health diagnosis," *Reliability Engineering and System Safety*, vol. 115, no. 3, pp. 124-135, 2013.
- [23] M. Gan, C. Wang, and C. Zhu, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 72-73, pp. 92-104, 2016.
- [24] Y. Lei, F. Jia, X. Zhou, and J. Lin, "A deep learning-based method for machinery health monitoring with big data," *Journal of Mechanical Engineering*, vol. 51, no. 21, pp. 49-56, 2015.
- [25] V. T. Tran, F. Althobiani, and A. Ball, "An approach to fault diagnosis of reciprocating compressor valves using Teager-Kaiser energy operator and deep belief networks," *Expert Systems with Applications*, vol. 41, no. 9, pp. 4113-4122, 2014.
- [26] G.-Y. Pan, W. Chai, and J.-F. Qiao, "Calculation for depth of deep belief network," *Control and Decision*, vol. 30, no. 2, pp. 256-260, 2015.

Research Article

Fault Diagnosis for Rolling Bearing under Variable Conditions Based on Image Recognition

Bo Zhou^{1,2} and Yujie Cheng^{1,2}

¹School of Reliability and Systems Engineering, Beihang University, No. 37, Xueyuan Road, Haidian District, Beijing 100191, China

²Science & Technology on Reliability and Environmental Engineering Laboratory, Beijing 100191, China

Correspondence should be addressed to Yujie Cheng; chengyujie@buaa.edu.cn

Received 25 May 2016; Accepted 14 July 2016

Academic Editor: Minvydas Ragulskis

Copyright © 2016 B. Zhou and Y. Cheng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rolling bearing faults often lead to electromechanical system failure due to its high speed and complex working conditions. Recently, a large amount of fault diagnosis studies for rolling bearing based on vibration data has been reported. However, few studies have focused on fault diagnosis for rolling bearings under variable conditions. This paper proposes a fault diagnosis method based on image recognition for rolling bearings to realize fault classification under variable working conditions. The proposed method includes the following steps. First, the vibration signal data are transformed into a two-dimensional image based on recurrence plot (RP) technique. Next, a popular feature extraction method which has been widely used in the image field, scale invariant feature transform (SIFT), is employed to extract fault features from the two-dimensional RP and subsequently generate a 128-dimensional feature vector. Third, due to the redundancy of the high-dimensional feature, kernel principal component analysis is utilized to reduce the feature dimensionality. Finally, a neural network classifier trained by probabilistic neural network is used to perform fault diagnosis. Verification experiment results demonstrate the effectiveness of the proposed fault diagnosis method for rolling bearings under variable conditions, thereby providing a promising approach to fault diagnosis for rolling bearings.

1. Introduction

Rolling bearings are considered to be a critical mechanical component in industrial applications. The bearings' defects usually lead to a considerable decline in plant productivity and may even cause huge economic losses [1, 2]. Thus, it is important to diagnose rolling bearing fault to keep the bearings in good technical state.

Vibration-based methods have garnered particular attention due to their noninvasive nature and their high reactivity to incipient fault. Therefore, vibration signal analysis is vital for rolling bearings fault diagnosis due to its connection to fault feature extraction accuracy [3]. Aiming to extract the fault features, many feature extraction methods, including Wigner-Ville distribution (WVD) [4], wavelet packet decomposition (WPT) [5, 6], and empirical mode decomposition (EMD) [7–9], have been proposed and have been demonstrated to be powerful. Additionally, many fault diagnosis methods also have been proposed, such as fast spectral

kurtosis based on genetic algorithms [10], multiscale entropy and adaptive neurofuzzy inference system [11], and time varying singular value decomposition [12]. However, most of these methods are proposed based on the assumption that the rolling bearings operate under fixed conditions when performing fault diagnosis. Moreover, the application of these methods is limited because of the tough, complex, and particularly variable working environment of rolling bearings [13, 14]. Therefore, it is important to investigate the fault diagnosis method suitable for varying conditions.

Many studies have researched rolling bearings fault diagnosis. However, thus far, few researchers have studied fault diagnosis under variable conditions. In 1990, Potter [15] proposed a constant angular sampling method (i.e., order tracking) that utilized the electronic impulse angular encoder and solved the frequency smearing phenomenon of the spectrum caused by fluctuating rotating speeds and realized the fault diagnosis for rotating machines. Considering the special analog hardware whose function is sampling data increases

the cost of equipment; Fyfe and Munck [16] developed the computed order tracking (COT) technique based on order tracking to realize fault diagnosis for rotating machines. However, the COT may make the carrier frequencies of the transient responses, which are caused by the faults at various speeds, expand to a wider scope because the natural characteristic of the bearing system hardly changes, which is not beneficial for extracting the fault characteristic. In addition, [13] has proposed a new method for rolling bearing fault diagnosis under variable conditions. This new method utilizes LMD-SVD to extract features, but LMD also has the problem of iterative calculation capacity, frequency aliasing, end effect and other issues. Because of problems associated with the above methods, we need to research a new method for bearing vibration signal feature extraction, a method based on the nonstationary and nonlinear bearing vibration signals, thereby achieving fault diagnosis under variable conditions.

Scale invariant feature transform (SIFT), an image invariant feature extraction method, can recognize the same image when it is rotated, scaled, translational, and affine transformed. By extracting the 128-dimensional feature containing scale, orientation, and location information, SIFT can perform image recognition and matching under translation, rotation, scaling, and brightness changes [17]. Many studies have used SIFT to recognize images. For example, Montazer and Giveki [18] have utilized SIFT to extract image features and match them to a database (i.e., a content-based image retrieval system). Li et al. [19] have employed Robust SIFT to match remote sensing images, and a number of studies have also applied SIFT to such methods as facial expression recognition [20], ear recognition for a new biometric technology [21], and wheat grain classification [22]. Inspired by SIFT, the vibration signals of rolling bearings are considered to be transformed into images. The recurrence plot (RP) is a kind of method to describe the recursive behavior of dynamic orbit in the phase-space reconstruction; it is an important method to analyze the instability of time series. RPs of rolling bearing vibration signals under different conditions reveal translation and scaling characteristics, so RP is employed to transform the vibration signals under different conditions into images and SIFT is utilized to extract the features of transformed RPs, which is without interference of working conditions.

After the 128-dimensional invariant features are extracted, to reduce the data redundancy between the extracted features and the occupation of computer resources, a dimensionality reduction method is utilized to identify the low-dimensional structure hidden in high-dimensional data. Principal component analysis (PCA) is a widely utilized dimension reduction technique performed by linearly transforming a high-dimensional input space onto a lower dimensional one in which the components are uncorrelated. However, PCA will not perform well when the process exhibits nonlinearity. Hence, kernel principal component analysis (KPCA) was developed to overcome the limitations of PCA in dealing with the nonlinear system [23].

This paper is structured as follows. Section 2 first introduces the image transformation method, which generates images for the following recognition. Then, SIFT, the core

of this paper, is described, which is utilized to extract the stable fault features under variable working conditions. Subsequently, KPCA is introduced for the dimensionality reduction. At last, probabilistic neural network (PNN) is described for the final fault classification. Section 3 describes the entire fault diagnosis method for rolling bearing under variable conditions, including descriptions of the experimental data, image transformation, feature extraction, and fault classification. Section 4 includes the results and discussion, and the conclusions are presented in Section 5.

2. Related Theories

2.1. Recurrence Plot. To achieve fault diagnosis under variable conditions, image transformation for SIFT is important to ensure success. Therefore, choosing a good image transformation method is particularly important. On account of the nonlinear and nonstationary characteristics of rolling bearing signals, detecting dynamical changes in complex systems is one of the most difficult problems. Recursiveness is one of the basic characteristics of a dynamic system, and the recurrence plot (RP) based on this characteristic is a good dynamic mainstream shape-description method. Through the black and white dots in the two-dimensional space, the recursive state can be visualized in the phase space [24]. This approach can uncover hidden periodicities in a signal in the recurrence domain. These periodicities are not easily noticeable, and it is an important method that analyzes the periodic, chaotic, and nonstationary of time series. The following theories are related.

The RP analysis is based on the phase-space reconstruction theory, which is described as follows.

- (1) For a time series, u_k ($k = 1, 2, \dots, N$), whose sample interval is Δt , we chose the mutual information method and CAO algorithm to calculate the suitable embedding dimension m and delay time τ , which could reconstitute the time series. The reconstructed time series is

$$x_i = (u_i, u_{i+\tau}, \dots, u_{i+(m-1)\tau}), \quad (1)$$

$$i = 1, 2, \dots, N - (m - 1)\tau.$$

- (2) Calculate the Euclidean norm of x_i and x_j in the reconstructed phase space [25]:

$$S_{ij} = \|x_i - x_j\|, \quad (2)$$

$$i = 1, 2, \dots, N - (m - 1)\tau; \quad j = 1, 2, \dots, N - (m - 1)\tau.$$

- (3) Calculate the recurrence value [26]:

$$R(i, j) = H(\varepsilon_i - S_{ij}), \quad (3)$$

$$i = 1, 2, \dots, N; \quad j = 1, 2, \dots, N,$$

where ε is the threshold value and $H(r)$ is the Heaviside function:

$$H(r) = \begin{cases} 1 & r \geq 0 \\ 0 & r < 0. \end{cases} \quad (4)$$

- (4) Utilize a coordinate graph whose abscissa is i and whose ordinate is j to draw $R(i, j)$, where i and j are the time series labels and the image is RP.

2.1.1. Mutual Information Method. The mutual information method estimating the delay time has been proposed by Fraser and Swinney, based on the Shannon information theory, which is widely used in phase-space reconstruction [27].

The Shannon theory shows that we can obtain the information content of a_i from the event b_j :

$$I_{AB}(a_i, b_j) = \log_2 \left[\frac{P_{AB}(a_i, b_j)}{P_A(a_i) P(b_j)} \right]. \quad (5)$$

The relationship between a_i and b_j could be expressed with comentropy I_{AB} :

$$I_{AB} = \sum_{ij} P_{AB}(a_i, b_j) \log_2 \left[\frac{P_{AB}(a_i, b_j)}{P_A(a_i) P(b_j)} \right]. \quad (6)$$

Apply the theory of the mutual information, and set A is

$$\{A : a_i = x_i = x(t_0 + i\tau_x)\} \quad (7)$$

and set B is

$$\{B : b_i = x_i = x(t_0 + i\tau_x + \tau)\}. \quad (8)$$

Then, (6) translates into

$$I_{AB}(\tau) = \sum_i P[x(t_0 + i\tau_x), x(t_0 + i\tau_x + \tau)] \cdot \log_2 \left\{ \frac{P[x(t_0 + i\tau_x), x(t_0 + i\tau_x + \tau)]}{P[x(t_0 + i\tau_x)] P[x(t_0 + i\tau_x + \tau)]} \right\}. \quad (9)$$

Usually at the beginning, $I_{AB}(\tau)$ is large; therefore, we can obtain an infinite amount of information in $x(t) = x(t + 0)$. $x(t_0 + i\tau_x)$ and $x(t_0 + i\tau_x + \tau)$ are completely independent for chaotic signals when τ is large; when $\tau \rightarrow \infty$, $I(\tau) \rightarrow 0$. Generally the first minimum of $I_{AB}(\tau)$ is selected as the delay time.

2.1.2. CAO Algorithm. The CAO algorithm was proposed by CAO in 1997, and it has excellent properties to clearly distinguish real signal and noise, as well as high computational efficiency [28]. First, we calculated the distance of the points under the embedded dimensionality:

$$a(i, m) = \frac{\|u_i(m+1) - u_{n(i,m)}(m+1)\|}{\|u_i(m) - u_{n(i,m)}(m)\|}, \quad (10)$$

$$i = 1, 2, \dots, N - m,$$

where $\|\cdot\|$ is ∞ the norm of the vector; $u_i(m+1)$ is i th vector after phase-space rebuilding, and the embedded dimension is $m+1$; $u_{n(i,m)}(m+1)$ is the nearest vector from $u_i(m+1)$.

Next, we calculated the average value of the distance change under the same dimension:

$$E(m) = \frac{1}{N-m} \sum_{i=1}^{N-m} a(i, m), \quad (11)$$

where $E(m)$ is the average value of all $a(i, m)$.

Finally, according to the discriminant equation

$$E_1(m) = \frac{E(m+1)}{E(m)} \quad (12)$$

when $m > m_0$, $E_1(m)$ stops changing or changes slowly, and $m_0 + 1$ is the minimum embedding dimension.

2.2. SIFT Theory. Recognizing the images that are rotating, scaling, and translating refers to finding the stable points of the images. These points, such as the corners, blobs, T-junctions, and light spots in dark regions, do not disappear with the rotation, scale, translation, and brightness changes. SIFT was developed by Nurhaida et al. to extract distinctive invariant features from images that can be used to perform reliable matching between different views of an object or scene [29]. SIFT has five basic steps: constructing scale space, extreme points detection, precise location of key points, orientation assignment, and descriptor calculation [30].

2.2.1. Gaussian Blur. SIFT finds key points in the different scale spaces, and the acquisition of scale space needs to be realized using Gaussian blur. Lindeberg has proved that Gaussian convolution kernel is the only kernel to achieve scale transformation, and it is the only linear kernel [31].

Gaussian blur is an image filter that utilizes normal distribution to calculate the fuzzy template, and the template is used to perform convolution operations with the original image to achieve the transition of fuzzy images.

The normal distribution equation of N dimensional space is

$$G(r) = \frac{1}{\sqrt{2\pi\sigma^2}^N} e^{-r^2/(2\sigma^2)}, \quad (13)$$

where σ is the standard deviation of the normal distribution; the larger σ is, the fuzzier image is. r is the fuzzy radius that refers to the distance between the template element and the center of the template. If the two-dimensional template size is $m \times n$, then (x, y) on the template corresponding to the Gauss equation is

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-((x-m/2)^2 + (y-n/2)^2)/2\sigma^2}. \quad (14)$$

According to the value of σ , the size of the Gaussian blur template matrix is $(6\sigma + 1) \times (6\sigma + 1)$. Equation (14) is used to calculate the value of the Gaussian template matrix; convolution is calculated with the original image, and the Gaussian blur image of the original image is obtained.

2.2.2. Scale Space Construction

(1) *Scale Space Theory.* Scale space theory was first proposed by Iijima in 1962, and it was widely used in the field of computer vision after being promoted by Duits et al. [32].

The basic concept of scale space is as follows. A scale parameter is introduced in the image model, and the scale space sequence at multiple scales is obtained through the continuous change of scale parameter. The principal contours are extracted from the scale space of these sequences, and the principal contours are utilized as a feature vector to realize edge detection, corner detection, and feature extraction at different resolutions.

(2) *Representation of Scale Space.* The scale space $L(x, y, \sigma)$ of an image is defined as the convolution calculation between the Gauss function $G(x, y, \sigma)$ and the original image $I(x, y)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad (15)$$

where $*$ represents convolution:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-((x-m/2)^2 + (y-n/2)^2)/2\sigma^2}, \quad (16)$$

where m and n are the dimensionality of the Gaussian template determined by $(6\sigma + 1) \times (6\sigma + 1)$. (x, y) is the pixel location in the image. σ is the scale space factor, the smaller value of which is the least amount of the smoothed image, and the corresponding scale is smaller.

(3) *Gaussian Pyramid.* The pyramid model of an image is as follows: the original image is constantly downsampling, and it generates a series of different sizes of images, ranging from large to small and from the bottom to the top, thereby constructing a tower-shaped model. The original image is the first stratum of the pyramid, and the new image obtained through downsampling is a stratum of the pyramid. The number of strata in the pyramid is jointly decided through the size of the original and top images. The equation is as follows:

$$n = \log_2 \{ \min(M, N) \} - t, \quad (17)$$

$$t \in [0, \log_2 \{ \min(M, N) \}],$$

where M and N are the sizes of the original image and t is logarithm of the minimum dimensionality of the top image.

To reflect the continuity of scale, the Gaussian pyramid introduces the Gaussian filter on the simple downsampling, as shown in Figure 1. The image in each stratum calculates the Gaussian blur using different parameters; thus, each stratum of the pyramid contains multiple Gaussian blur images. The images in each stratum are named octaves. The initial image (bottom image) of an octave in the Gaussian Pyramid is obtained by sampling from the last third image of the previous octave of images.

(4) *DOG Pyramid.* In 2002, Mikolajczyk found that the scale normalization of the Laplacian Gaussian function can produce the most stable image features compared to other feature extraction functions. The difference of Gaussian (DOG)

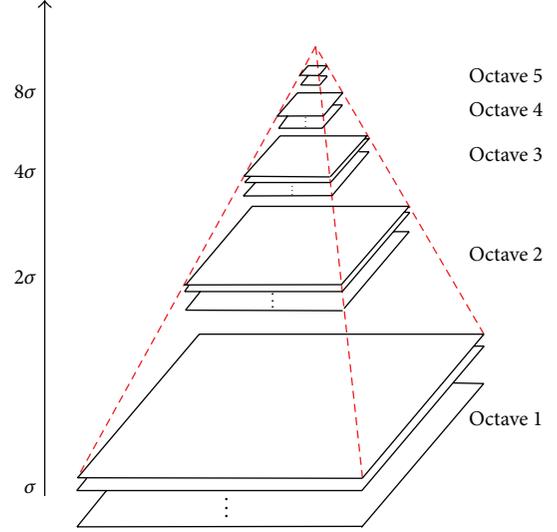


FIGURE 1: Gaussian pyramid.

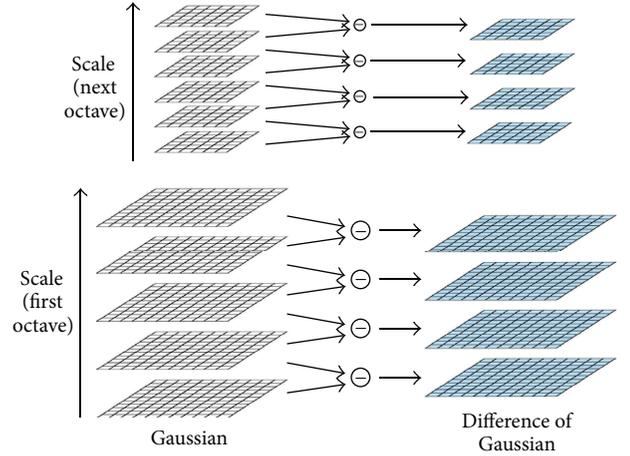


FIGURE 2: Gaussian pyramid generation.

function is similar to the scale normalization of the Laplacian Gaussian function [17]. Therefore, the DOG filter is applied to the input image. The image is gradually downsampled, and the filtering is performed at several scales. Figure 2 demonstrates the creation process for the DOG filters at different scales.

2.2.3. *Extreme Point Detection.* The key points are the local extreme points (in the DOG space) whose initial exploration is accomplished by comparing the two adjacent images of each DOG in the same group. To determine the key points, a $3 \times 3 \times 3$ neighborhood comparison is used, as shown in Figure 3. Each pixel processed by the DOG pyramid is compared with 26 points of its 3-dimensional neighborhood to obtain the maximum or minimum, as the preliminary feature points.

2.2.4. Precise Location of Key Points

(1) *Location of Interpolation.* The extreme points detected by the above methods are the extreme points of the discrete

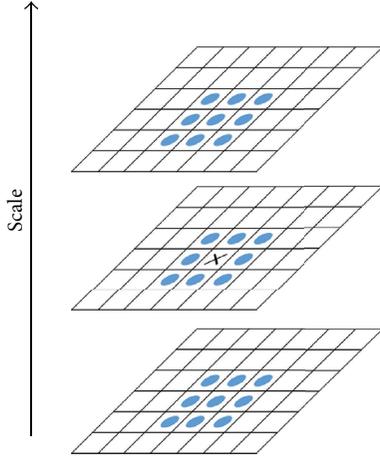


FIGURE 3: Key point localization.

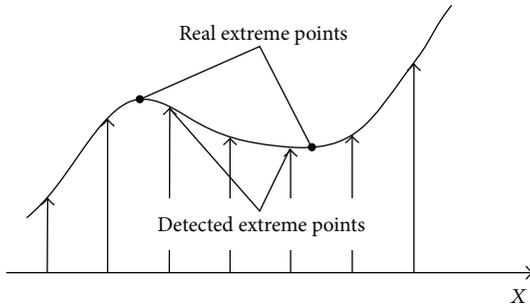


FIGURE 4: Difference between discrete space and continuous space.

space, which are not the real extreme points. Figure 4 shows the difference of the extreme point of a two-dimensional function in discrete and continuous space. SIFT utilizes the linear interpolation method to obtain accurate key points.

(2) *Remove Edge Response*. The detected key points are further examined to choose the “best” candidates. The stability of the resulting set of key points is determined. Locations with low contrast and unstable locations along edges are discarded by calculating the ratio of the square of the matrix trace and determining the Hessian matrix.

2.2.5. *Orientation Assignment*. To determine the descriptor with rotation invariance, the local features of images are needed to assign a reference orientation for each key point. By calculating the gradient orientations of the neighborhood pixels of key points, the orientation parameter is specified for each feature point. The gradient values and orientations in (x, y) are

$$m(x, y) = \sqrt{[L(x+1, y) - L(x-1, y)]^2 + [L(x, y+1) - L(x, y-1)]^2} \quad (18)$$

$$\theta(x, y) = \tan^{-1} \left\{ \frac{[L(x, y+1) - L(x, y-1)]}{[L(x+1, y) - L(x-1, y)]} \right\}.$$

The gradient histogram statistical method is employed to further ascertain the orientation of the key point. The gradient values of the key points in the neighborhood window are calculated with the key point as the center and 1.5σ as the radius. The 360° of a circle are divided into 36 bins by drawing the gradient histogram, the contribution of each neighborhood point to the orientation of the gradient decreases with the increase of the distance between the neighborhood and the key point. The peak of the histogram is the main orientation of the key points.

After selecting the main orientation, there may also be one or more peaks whose values are more than 80% of the main peak. To enhance the robustness of the match, some key points whose locations and scales are the same as the original key point will be employed.

2.2.6. *Descriptor Calculation*. After performing the above steps, the key points have the location, scale, and orientation. The next step is to create a descriptor for the key points. First, the coordinate axes are rotated as the key points to ensure the rotation invariance. Then, a 16×16 window is taken; the center of this window has the key points shown in Figure 5(a). Each grid represents a pixel in the scale space of the neighborhood of the key point; the orientation of the arrow represents the gradient orientation of the pixel, and the length of the arrow represents the gradient mode. In the figure, the circle represents the Gauss range weighted. Next, the gradient orientation histogram of 8 orientations is calculated in every 4×4 image, and the seed point is formed by drawing the cumulative value of each gradient orientation, as shown in Figure 5(b). In this figure, a key point is composed of 16 (4×4) seed points, each of which has eight orientation vectors. The key point can generate 128 ($4 \times 4 \times 8$) data sets and then form a 128-dimensional feature vector. The concept of the neighborhood orientation information alliance enhances the antinoise ability and also provides good fault tolerance for the feature matching with the localization error.

2.3. *Kernel Principal Component Analysis*. KPCA projects the m dimensional observed data matrix ($\mathbf{X} \in \mathbf{R}^m$, input space) onto a high-dimensional feature space \mathbf{F} , which can be expressed as

$$\Phi : \mathbf{R}^m \longrightarrow \mathbf{F}. \quad (19)$$

Similar to PCA, KPCA aims to project a high-dimensional feature space onto a lower space, in which the principal components are linear, uncorrelated combinations of the feature space [33]. The covariance matrix in the feature space can be formulated as

$$\mathbf{C} = \frac{1}{N} \sum_{i=1}^N \Phi(x_i) \Phi^T(x_i). \quad (20)$$

The characteristic equation is

$$\mathbf{C}\mathbf{V} = \lambda\mathbf{V}, \quad (21)$$

where $\Phi(x_i)$ is i th sample in the feature space with zero mean, N denotes the sample size, and T is the transpose operation.

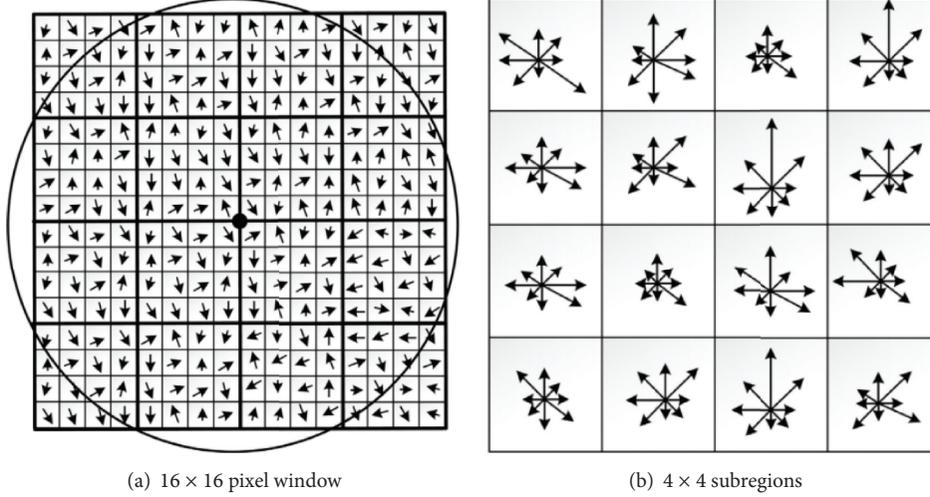


FIGURE 5: Image gradient and key point descriptor.

Let $\theta = [\Phi(x_1), \dots, \Phi(x_N)]$ be the data matrix in the feature space. Hence, C can be expressed as $C = \theta\theta^T/N$. Due to the difficulty of obtaining Φ , a Gram kernel matrix K is determined as follows to directly avoid eigen-decomposition C :

$$K(x_i, x_j) = \Phi^T(x_i), \Phi(x_j), \quad (22)$$

where $K = \theta^T\theta$; therefore, the inner product in the feature space (see (20)) can be obtained by introducing a kernel function to the input space.

Let

$$V = \sum_{i=1}^N \alpha_i \Phi(x_i). \quad (23)$$

Simultaneously, (20), (21), and (22) obtain the following equation:

$$\sum_{i=1}^N \sum_{j=1}^N \Phi(x_i) \alpha_j K_{ij} = N\lambda \sum_{i=1}^N \alpha_i \Phi(x_i). \quad (24)$$

The vector of (23) is

$$K\alpha = N\lambda\alpha = \lambda'\alpha. \quad (25)$$

To extract the principal components, the projection of feature vector V_j in the feature space is calculated:

$$(V_j)^T \Phi(x) = \sum_{i=1}^N \alpha_i \Phi^T(x_i) \Phi(x) = \sum_{i=1}^N \alpha_i K(x_i, x). \quad (26)$$

2.4. Probabilistic Neural Network. The probabilistic neural network was proposed by Specht in 1990 [34]. It is a feed-forward neural network that was developed from the radial basis function, and its theoretical basis is the Bayes minimum

risk rule (Bayes decision theory). As one of the radial basis networks, the PNN is suited to pattern classification, it places the Bayes decision analysis (with the Parzen window function) into the framework of a neural network, and the Bayes classification is produced by combining the Bayes decision and nonparametric estimation of probability density function. It can be described in the following manner. Assuming that there are two fault modes (θ_A and θ_B) for a fault feature sample $X = (x_1, x_2, x_3, \dots, x_n)$,

$$\text{if } h_A l_A f_A(X) > h_B l_B f_B(X), \quad \text{then } X \in \theta_A, \quad (27)$$

$$\text{if } h_A l_A f_A(X) < h_B l_B f_A(X), \quad \text{then } X \in \theta_B,$$

where h_A and h_B denote the prior probability of fault modes θ_A and θ_B , generally, $h_A = N_A/N$, $h_B = N_B/N$, and N_A and N_B are the number of training samples of θ_A and θ_B , respectively, and N is the total number of training samples. l_A is the cost factor used to classify the feature X sample belonging to θ_A into mode θ_B (falsely), l_B is the cost factor used to classify the feature sample X belonging to θ_B into mode θ_A (falsely). f_A and f_B are the probability density functions of the fault modes θ_A and θ_B , respectively.

Figure 6 shows the PNN structure that demonstrates that the input mode X is divided into 2 types. As shown in Figure 6, the PNN is a feed-forward neural network with a 4-layer structure: the input layer, pattern layer, summation layer, and output layer. The input layer transmits input samples to each node of the pattern layer. The node of the pattern layer calculates the weighted sum of the data passed by the input node, following the operation of a nonlinear operator, which transmits the results to the summation layer. The nonlinear operator is

$$g(z_j) = \exp \left[\frac{(z_j - 1)}{\sigma^2} \right]. \quad (28)$$

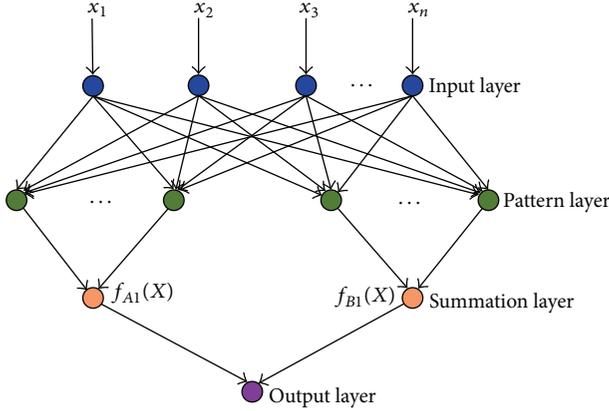


FIGURE 6: The basic PNN structure.

Assuming that X and W are standardized into unit lengths, (1) is equivalent to

$$g(z_j) = \exp \left[-\frac{(W_j - X)^t (W_j - X)}{2\sigma^2} \right], \quad (29)$$

where W is the weight vector.

The summation layer sums up the input from the pattern layer and then obtains the estimated probability densities. The classification result selected by the output layer is the maximum output of the summation layer.

The PNN is equal to the Bayes pattern classification method, which utilizes the Gauss kernel multivariate probability density function. The density function can be estimated as follows:

$$F_A(X) = \frac{1}{(2\pi)^{p/2}} \frac{1}{\delta^p m} \sum \exp \left[-\frac{(X - X_{Aj})^T (X - X_{Aj})}{2\delta^2} \right], \quad (30)$$

where X is input sample vector, s is the number of sample vectors' variables, X_{Aj} (the weight in the PNN) is j th training vector of fault mode A , and m is the number of training samples belonging to mode A ; δ is the smoothing parameter.

3. Method for Fault Diagnosis of Rolling Bearing under Variable Working Conditions

Inspired by SIFT, this study proposes a novel fault diagnosis for rolling bearings under variable conditions. The diagnostic procedure is shown in Figure 7.

The diagnostic process generally consists of four steps. First, the vibration signals in different fault modes under different conditions are transformed into RPs that are regarded as the objects of SIFT. To more accurately reconstruct the phase space, the delay time and embedded dimensionality are calculated using the mutual information algorithm and CAO algorithm is used to calculate the RPs. Second, the SIFT

extracts the invariable features of the RPs (as described in Section 2), and the constructing scale space, extreme point detection, precise locations of the key points, orientation assignment, and descriptor calculation are determined to achieve the salient invariable features. Third, due to the high-dimensional vector of the extraction features, KPCA is employed by introducing a kernel function to reduce the dimensionality. Finally, the PNN is used as a classifier to diagnose the fault classification using data from one of the conditions to train the neural network and using data from the other conditions to test the proposed method.

4. Results and Discussion

In this section, vibration data of rolling bearings collected from the Case Western Reserve University Bearing Data Center under different working conditions and fault modes were utilized to validate the effectiveness of the proposed method.

4.1. Description of the Experimental Data. The experimental data used to test and verify the proposed method were obtained from the Bearing Data Center of Case Western Reserve University, Cleveland, OH, USA. The experimental setup used a Reliance Electric 2HP motor connected to a dynamometer, which was used as the prime mover to drive a shaft coupled with a bearing housing. Faults (i.e., size 7 mils, 14 mils, 21 mils, and 28 mils) were introduced into the drive-end bearing (6205-2RS JEM SKF) and fan-end (NTN equivalent bearing) of a motor using the electric discharge machining (EDM) method, with the motor speed varied at 1730, 1750, 1772, and 1797 rpm, respectively. These faults were introduced separately at the inner raceway, rolling element (ball), and outer raceway [35]. To quantify the stationary effect of the outer raceway faults, experiments were conducted for the FE and DE bearings, with outer raceway faults located at 3 o'clock, 6 o'clock, and 12 o'clock. An impulsive force was applied to the motor shaft, and the resulting vibration was measured using two accelerometers, one mounted on the motor housing and the other placed at the 12 o'clock position of the outer race of the drive-end bearing. Digital data were collected at 12,000 samples per second, and data were also collected at 48,000 samples per second for the drive-end bearing faults.

In this study, the DE bearing data for the normal, inner race fault, outer race fault, and rolling element fault with the speed varied between the 4 conditions were acquired for the fault pattern classification, and the fault diameters were 21 mils. The fault information (21 mils and outer race fault at 6 o'clock with four speeds), in terms of the test bearings, is listed in Table 1.

4.2. Image Transformation of the Vibration Signals under Different Conditions. In this section, the vibration signals under different conditions are transformed into 2-dimensional images, which facilitate the extraction of invariable features for the fault classification. As previously mentioned, RP can uncover the hidden periodicities in a signal in the recurrence domain, and it is important that the method analyzes the periodic, chaotic, and nonstationary elements of the time

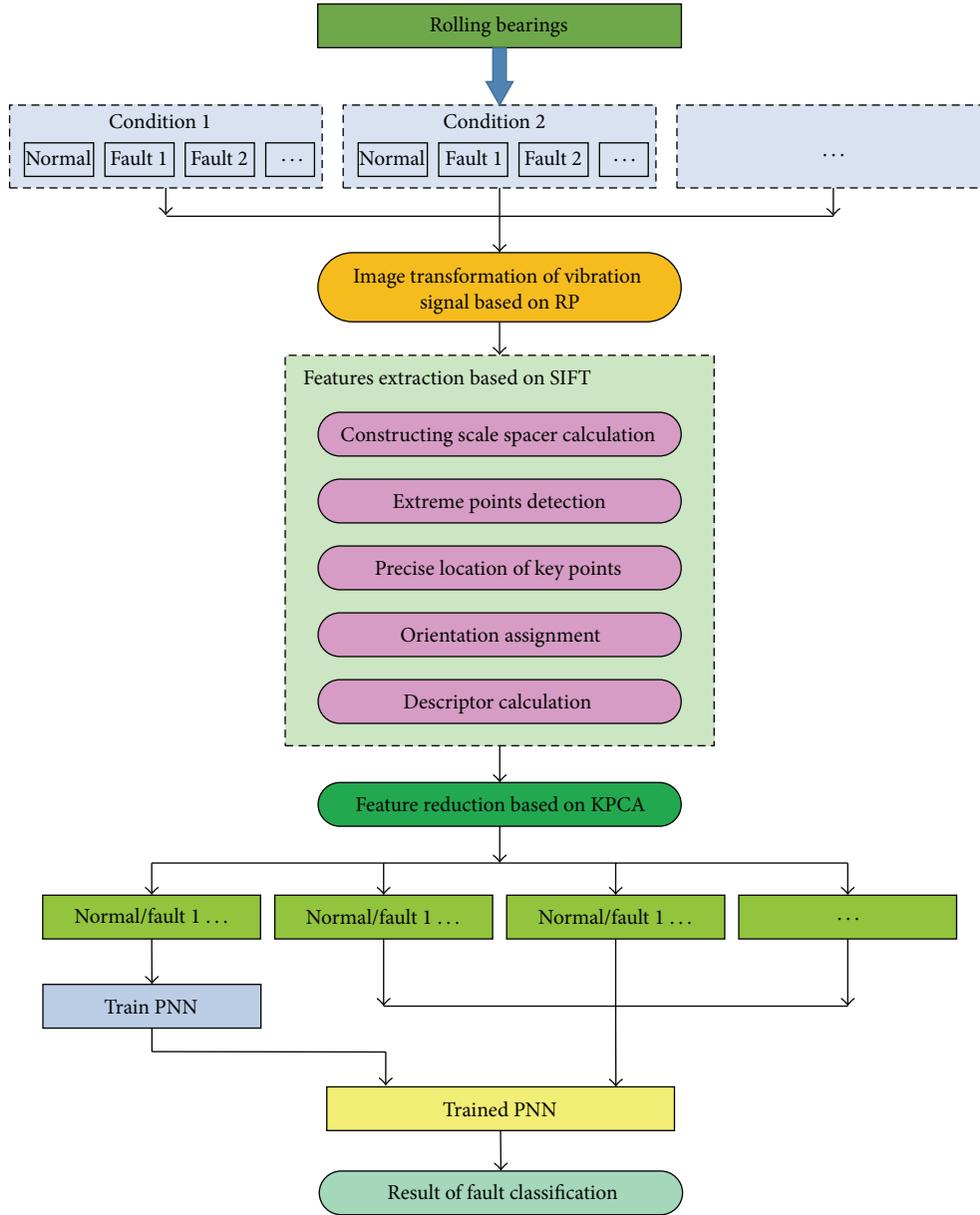


FIGURE 7: A flowchart of the proposed method.

TABLE 1: Data declaration.

Fault diameter (mils)	Motor speed (rpm)	Inner race	Ball	Outer race		
				3 o'clock	6 o'clock	12 o'clock
21	1797	213.mat	226.mat		238.mat	
	1772	214.mat	227.mat		239.mat	
	1750	215.mat	228.mat		240.mat	
	1730	217.mat	229.mat		241.mat	

series. Thus, RPs are particularly suitable for the image transformation of vibration signals without loss of signal information.

To accurately confirm the suitable embedding dimension, m and delay time τ of each signal for the phase-space

reconstruction, mutual information algorithm, and CAO algorithm were used. The parameters m and τ for each condition are shown in Table 2. Unfortunately, to guarantee the calculation speed, for the data segments of each vibration signal, only 1,000 points were chosen to transform into an RP

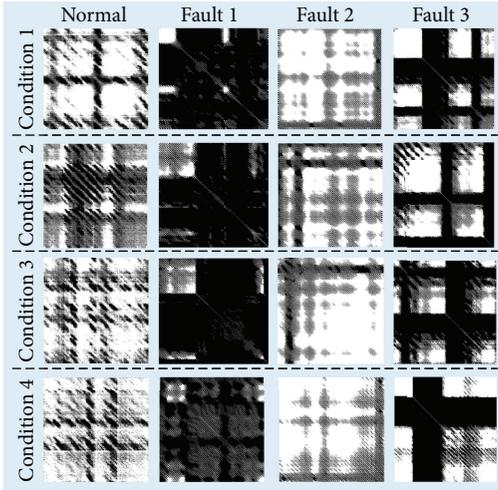


FIGURE 8: RPs transformed by fault mode vibration data under different conditions.

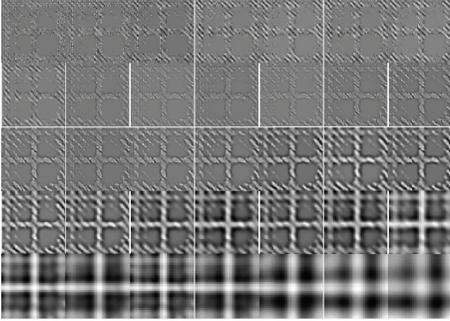


FIGURE 9: The normal DOG scale space under condition 1.

(with dimensionalities of $N \times N$) when reconstructing the phase-space due to calculating the ergodic Euclidean norms of x_i and x_j . This experiment selected a 20-set data segment for each signal. Figure 8 shows the RPs in each fault mode under 4 different conditions that were randomly selected in a 20-set data segment. From the results, we see that the size of the RPs in the different modes under different conditions reveals slight differences.

In Figure 8, the row represents the condition changes, and the column represents the different fault modes (i.e., the first column shows the RPs in the normal state, the second column shows the RPs in the inner fault state, the third column shows the RPs in the element fault state, and the fourth column shows the RPs in the outer fault state). In Figure 8, we can see that the RPs of the different fault modes under different conditions have different structural characteristics, while the same fault modes under different conditions are notably similar. Affected by the condition changes, the RPs under the different conditions show the translation variation, scale variation, and combination of these changes.

4.3. Feature Extraction Based SIFT and the Dimensionality Reduction. In this section, the invariable features in each fault mode under the different conditions are extracted from

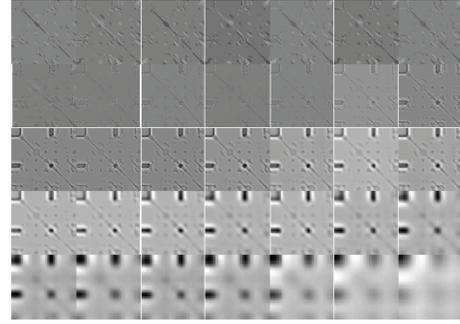


FIGURE 10: The DOG scale space of the inner race fault under condition 1.

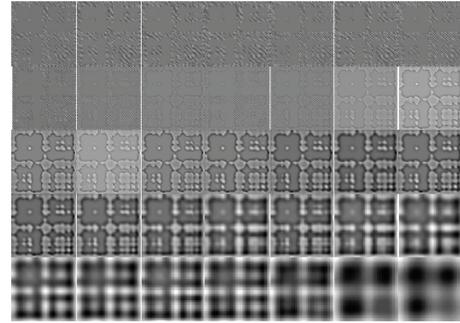


FIGURE 11: The DOG scale space of the element fault under condition 1.

the transformed RPs, based on SIFT. Using SIFT, the scales, orientation, and locations of the key points are calculated. The scale information is obtained by establishing the difference of the Gaussian pyramid, which has 7 octaves. Each octave has 5 strata, and the scale factor σ is utilized to make the image fuzzy between the different strata. Because of the length of the paper, the DOG scale space of four different fault modes (only under condition 1) are shown in Figures 9–12; the locations of the key points are calculated by locating the extreme value points and through further interpolation to determine the exact extreme points on a continuous space. The detected key points of the RPs are shown in Figure 13, and the orientations of the key points are obtained by calculating the gradient orientation of the neighborhood pixels of the key points. The orientation parameters are specified for each key point through the gradient histogram statistics. After performing the above steps, the descriptor of the key point is established through a 128-dimensional vector.

Due to the essential features of RPs hidden in the high-dimensional space, which makes the calculation difficult, the aforementioned KPCA method was used to reduce dimensionality. First, the input space is mapped onto a high-dimensional feature space using the kernel function, and then the PCA is used to reduce the dimensionality. However, the features of the low-dimensional space are also too large and complex to be taken as feature vectors. To solve this problem and to improve the robustness of the feature vectors, singular value decomposition (SVD) was utilized in this

TABLE 2: The experiment parameters of each fault mode under different conditions.

Conditions	Parameters	Normal	Inner race fault	Element fault	Outer race fault
Condition 1	m	15	12	12	12
	τ	4	5	5	5
Condition 2	m	15	12	13	12
	τ	4	5	5	5
Condition 3	m	15	20	13	12
	τ	4	2	5	5
Condition 4	m	16	11	13	12
	τ	4	5	5	5

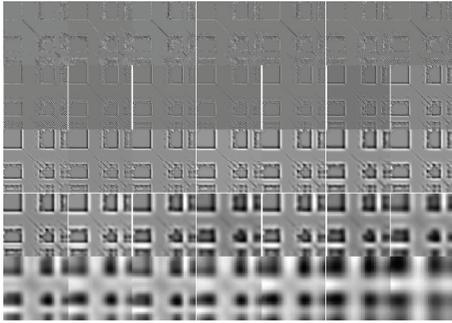


FIGURE 12: The DOG scale space of the outer race fault under condition 1.

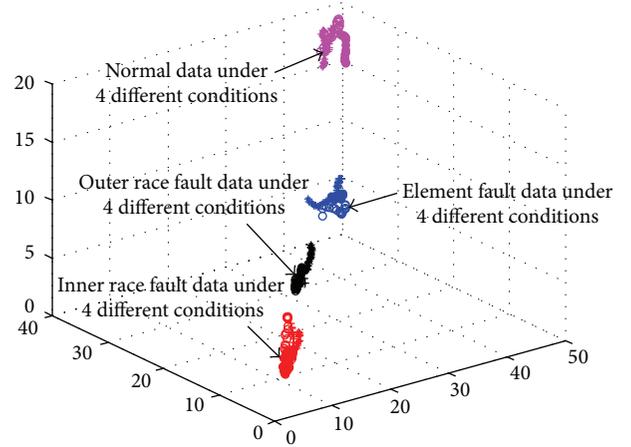


FIGURE 14: The feature scatter diagram in the 3-dimensional space.

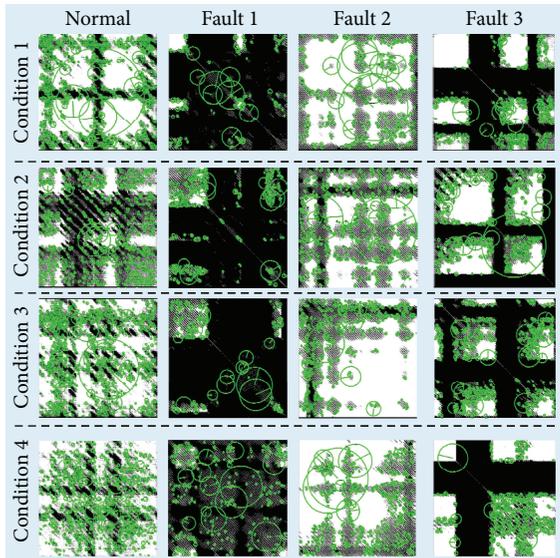


FIGURE 13: The detected interest points in RPs.

paper to compress the scale of the fault feature vectors and to obtain more stable feature vectors [13]. Figure 14 shows the 3-dimensional visual feature points reduced by KPCA and SVD.

4.4. Fault Classification Based on PNN. In this paper, PNN is employed as the classifier to classify the extracted features from vibration signals under different conditions, which are processed by SIFT and KPCA. To verify that the training

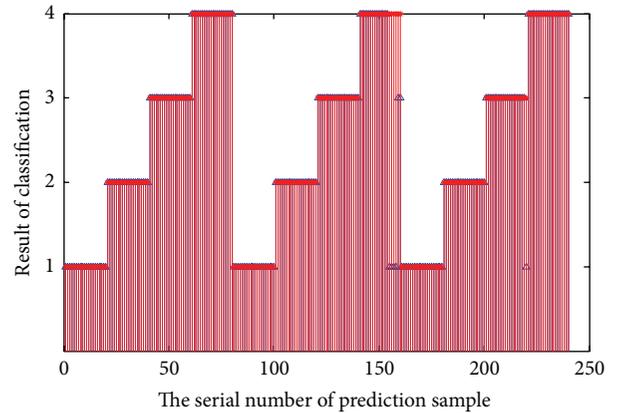


FIGURE 15: The classification result of the first cross-validation group.

data for the different conditions are effective, cross-validation is also necessary; the vibration data collected under each condition are orderly selected as training data, and the data collected under the other three conditions are used as test data, as shown in Table 3.

In each cross-validation, the training data and test data are composed as follows.

TABLE 4: The detailed error samples statistics of cross-validation.

The work conditions of training data	The work conditions of test data	Normal	Inner race fault	Element fault	Outer race fault	Mean value/sum		
1	2	Number of data	20	20	20	20	80	
		Number of wrong samples	0	0	0	0	0	
		Error rate	0	0	0	0	0	
	3	Number of data	20	20	20	20	80	
		Number of wrong samples	0	0	0	6	6	
		Error rate	0	0	0	0.3	0.075	
	4	Number of data	20	20	20	20	20	
		Number of wrong samples	0	0	1	0	1	
		Error rate	0	0	0.05	0	0.0125	
	2	1	Number of data	20	20	20	20	80
			Number of wrong samples	0	0	0	0	0
			Error rate	0	0	0	0	0
3		Number of data	20	20	20	20	80	
		Number of wrong samples	0	1	0	5	6	
		Error rate	0	0.05	0	0.25	0.075	
4		Number of data	20	20	20	20	80	
		Number of wrong samples	0	1	0	0	1	
		Error rate	0	0.05	0	0	0.125	
3		1	Number of data	20	20	20	20	80
			Number of wrong samples	0	0	1	5	6
			Error rate	0	0	0.05	0.25	0.075
	2	Number of data	20	20	20	20	80	
		Number of wrong samples	0	0	0	0	0	
		Error rate	0	0	0	0	0	
	4	Number of data	20	20	20	20	80	
		Number of wrong samples	0	0	0	0	0	
		Error rate	0	0	0	0	0	
	4	1	Number of data	20	20	20	20	80
			Number of wrong samples	0	0	6	0	6
			Error rate	0	0	0.3	0	0.075
2		Number of data	20	20	20	20	80	
		Number of wrong samples	0	0	0	0	0	
		Error rate	0	0	0	0	0	
3		Number of data	20	20	20	20	80	
		Number of wrong samples	0	0	0	0	0	
		Error rate	0	0	0	0	0	

TABLE 5: The result of fault diagnosis for bearing under variable conditions.

Groups of cross validation	1	2	3	4	Mean value/sum
Number of wrong samples	7	7	6	6	26
Number of test data	240	240	240	240	960
Classification accuracy	97.08%	97.08%	97.5%	97.5%	97.29%

Competing Interests

The authors declare that there is not any potential conflict of interests in the research.

Acknowledgments

This study was supported by the Fundamental Research Funds for the Central Universities (Grant no. YWF-16-BJ-J-18) and the National Natural Science Foundation of China (Grant no. 51575021), and the Technology Foundation Program of National Defense (Grant no. Z132013B002).

References

- [1] G. Wang, X. Feng, and C. Liu, "Bearing fault classification based on conditional random field," *Shock and Vibration*, vol. 20, no. 4, pp. 591–600, 2013.
- [2] H. Liu, X. Wang, and C. Lu, "Rolling bearing fault diagnosis based on LCD-TEO and multifractal detrended fluctuation analysis," *Mechanical Systems and Signal Processing*, vol. 60-61, pp. 273–288, 2015.
- [3] Y. Lv, R. Yuan, and G. Song, "Multivariate empirical mode decomposition and its application to fault diagnosis of rolling bearing," *Mechanical Systems and Signal Processing*, vol. 81, pp. 219–234, 2016.
- [4] G. Chen, J. Chen, and G. M. Dong, "Chirplet Wigner-Ville distribution for time-frequency representation and its application," *Mechanical Systems and Signal Processing*, vol. 41, no. 1-2, pp. 1–13, 2013.
- [5] Z. Zhang, Y. Wang, and K. Wang, "Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network," *Journal of Intelligent Manufacturing*, vol. 24, no. 6, pp. 1213–1227, 2013.
- [6] L.-S. Law, J. H. Kim, W. Y. H. Liew, and S.-K. Lee, "An approach based on wavelet packet decomposition and Hilbert-Huang transform (WPD-HHT) for spindle bearings condition monitoring," *Mechanical Systems and Signal Processing*, vol. 33, pp. 197–211, 2012.
- [7] F. Wu and L. Qu, "Diagnosis of subharmonic faults of large rotating machinery based on EMD," *Mechanical Systems and Signal Processing*, vol. 23, no. 2, pp. 467–475, 2009.
- [8] T. Han, D. Jiang, and N. Wang, "The fault feature extraction of rolling bearing based on EMD and difference spectrum of singular value," *Shock and Vibration*, vol. 2016, Article ID 5957179, 14 pages, 2016.
- [9] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 108–126, 2013.
- [10] Y. Zhang and R. B. Randall, "Rolling element bearing fault diagnosis based on the combination of genetic algorithms and fast kurtogram," *Mechanical Systems and Signal Processing*, vol. 23, no. 5, pp. 1509–1517, 2009.
- [11] L. Zhang, G. Xiong, H. Liu, H. Zou, and W. Guo, "Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference," *Expert Systems with Applications*, vol. 37, no. 8, pp. 6077–6085, 2010.
- [12] S. Zhang, S. Lu, Q. He, and F. Kong, "Time-varying singular value decomposition for periodic transient identification in bearing fault diagnosis," *Journal of Sound and Vibration*, vol. 379, pp. 213–231, 2016.
- [13] Y. Tian, J. Ma, C. Lu, and Z. Wang, "Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine," *Mechanism and Machine Theory*, vol. 90, pp. 175–186, 2015.
- [14] A. B. Ming, W. Zhang, Z. Qin, and F. Chu, "Fault feature extraction and enhancement of rolling element bearing in varying speed condition," *Mechanical Systems and Signal Processing*, vol. 76-77, pp. 367–379, 2016.
- [15] M. C. W. Potter, "Tracking and resampling method and apparatus for monitoring the performance of rotating machines," patents, 1990.
- [16] K. R. Fyfe and E. D. S. Munck, "Analysis of computed order tracking," *Mechanical Systems and Signal Processing*, vol. 11, no. 2, pp. 187–202, 1997.
- [17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [18] G. A. Montazer and D. Giveki, "Content based image retrieval system using clustered scale invariant feature transforms," *Optik*, vol. 126, no. 18, pp. 1695–1699, 2015.
- [19] Q. Li, G. Wang, J. Liu, and S. Chen, "Robust scale-invariant feature matching for remote sensing image registration," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 2, pp. 287–291, 2009.
- [20] H. Soyel and H. Demirel, "Localized discriminative scale invariant feature transform based facial expression recognition," *Computers and Electrical Engineering*, vol. 38, no. 5, pp. 1299–1309, 2012.
- [21] L. Ghoulmi, A. Draa, and S. Chikhi, "An ear biometric system based on artificial bees and the scale invariant feature transform," *Expert Systems with Applications*, vol. 57, pp. 49–61, 2016.
- [22] M. Olgun, A. O. Onarcan, K. Özkan et al., "Wheat grain classification by using dense SIFT features with SVM classifier," *Computers and Electronics in Agriculture*, vol. 122, pp. 185–190, 2016.
- [23] C. K. Yoo and I.-B. Lee, "Nonlinear multivariate filtering and bioprocess monitoring for supervising nonlinear biological processes," *Process Biochemistry*, vol. 41, no. 8, pp. 1854–1863, 2006.
- [24] J. Yan, Y. Wang, G. Ouyang, T. Yu, and X. Li, "Using max entropy ratio of recurrence plot to measure electrocorticogram changes in epilepsy patients," *Physica A: Statistical Mechanics and Its Applications*, vol. 443, pp. 109–116, 2016.

- [25] N. Marwan, J. Kurths, and P. Saparin, "Generalised recurrence plot analysis for spatial data," *Physics Letters A*, vol. 360, no. 4-5, pp. 545–551, 2007.
- [26] D. Yang, W.-X. Ren, Y.-D. Hu, and D. Li, "Selection of optimal threshold to construct recurrence plot for structural operational vibration measurements," *Journal of Sound and Vibration*, vol. 349, pp. 361–374, 2015.
- [27] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, no. 2, pp. 1134–1140, 1986.
- [28] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena*, vol. 110, no. 1-2, pp. 43–50, 1997.
- [29] I. Nurhaida, A. Noviyanto, R. Manurung, and A. M. Arymurthy, "Automatic Indonesian's batik pattern recognition using SIFT approach," *Procedia Computer Science*, vol. 59, pp. 567–576, 2015.
- [30] L. Lenc and P. Král, "Automatic face recognition system based on the SIFT features," *Computers and Electrical Engineering*, vol. 46, pp. 256–272, 2013.
- [31] T. Lindeberg, "Scale-space for discrete signals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 3, pp. 234–254, 1990.
- [32] R. Duits, L. Florack, J. de Graaf, and B. ter Haar Romeny, "On the axioms of scale space theory," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 3, pp. 267–298, 2004.
- [33] C.-Y. Cheng, C.-C. Hsu, and M.-C. Chen, "Adaptive kernel principal component analysis (KPCA) for monitoring small disturbances of nonlinear processes," *Industrial & Engineering Chemistry Research*, vol. 49, no. 5, pp. 2254–2262, 2010.
- [34] D. F. Specht, "Applications of probabilistic neural networks," *Neural Networks*, vol. 3, no. 1, pp. 109–118, 1990.
- [35] J. Yu, "Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models," *Mechanical Systems and Signal Processing*, vol. 25, no. 7, pp. 2573–2588, 2011.