An Intelligent System
For
Bearing Condition Monitoring

by

Jie Liu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Mechanical Engineering

Waterloo, Ontario, Canada, 2008

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Rolling-element bearings are widely used in various mechanical and electrical applications. Accordingly, a reliable bearing health condition monitoring system is very useful in industries to detect incipient defects in bearings, so as to prevent machinery performance degradation and malfunction. Although several techniques have been reported in the literature for bearing fault detection and diagnosis, it is still challenging to implement a bearing condition monitoring system for real-world industrial applications because of the complexity of bearing structures and noisy operating conditions. The objective of this thesis is to develop a novel intelligent system for more reliable bearing fault diagnostics. This system involves two sequential processes: feature extraction and decision-making. The proposed strategy is to develop advanced and robust techniques at each processing stage so as to improve the reliability of bearing condition monitoring. First, a novel wavelet spectrum analysis technique is proposed for the representative feature extraction. This technique applies the wavelet transform to demodulate the resonance signatures that are related to bearing health conditions. A weighted Shannon function is proposed to synthesize the wavelet coefficient functions to enhance feature characteristics. The viability of this technique is verified by experimental tests corresponding to various bearing health conditions. Secondly, an enhanced diagnostic scheme is developed for automatic decision-making. This scheme consists of modules of classification and prediction: a novel neuro-fuzzy classifier is developed to effectively integrate the strengths of the selected fault detection techniques (i.e., the resulting representative features) for a more accurate assessment of bearing health conditions; a novel multi-step predictor is proposed to forecast the future states of bearing conditions, which will be used to further enhance the diagnostic reliability. The investigation results have demonstrated that the developed intelligent diagnostic system outperforms other related bearing fault diagnostic schemes.

Acknowledgements

I would like to express my sincere appreciation to my supervisors, Dr. Wilson Wang and Dr. Farid Golnaraghi, for their enthusiastic supervision and patient guidance. I can not imagine having any better advisors for my doctoral study regarding their knowledge, perceptiveness, and encouragement. In addition, I wish to acknowledge the help of Dr. Fathy Ismail in previous collaboration projects.

I am grateful to my PhD committee members, Dr. Ming J. Zuo (University of Alberta), Dr. Paul Fieguth, Dr. Hamid Jahed, and Dr. William W. Melek, for their instructive guidance and help.

Financial support from the Graduate Scholarship of Natural Sciences and Engineering Research Council of Canada (NSERC), University of Waterloo, and my supervisors, is greatly appreciated.

Many thanks to my friends: Shahab Ghafari, Zai Li, Sherry Li, Siamak Arzanpour, Nima Eslaminasab, Jerry Dou, Yuliang Liu, Yin He, and many others for their friendship and help.

I am indebted, forever, to my wife, Lily Yin, my parents, and parents-in-law, for their understanding and endless encouragement.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Overview

The development of reliable monitoring systems has been the focus of various undertakings in a wide array of industries involving rotary machinery to prevent machine performance degradation and malfunction, or even catastrophic failures [1, 2]. A reliable monitoring system can be applied to schedule preventive and predictive maintenance operations without the need of periodic shutdowns for routine inspections. It can reduce costly repairs by quickly identifying the faulty components without examining all of the related components.

Most machine failures are related to mechanical transmission systems such as gears and bearings [1]. To date, a series of effective investigations have been conducted for gear fault diagnosis and prognosis [3-9]. Thus, this thesis is focused on the condition monitoring of rolling element bearings. From a mechanical structure standpoint, a bearing is different from a gear in that a bearing is a complex system which consists of two rings and a set of elements running in the tracks between the rings. The rolling elements may be balls, cylindrical rollers, tapered rollers, needles, or barrel rollers, encased in a cage that provides equal spacing and prevents internal strikes. From the signal processing perspective, a gear signal is periodic with the shaft rotation, which can be differentiated, in theory, by the technique of time synchronous average filtering [10]. The bearing signal is non-periodic in nature because of the slippage occurring between the mating components. Each bearing rotary component generates vibratory signals, and each component can experience damage. The signals generated by a bearing are non-stationary, especially when slipage occurs among the rotating elements (e.g., the rolling elements and ring raceways) [2]. Therefore, compared with a gear system analysis, a bearing fault diagnosis is an even more challenging task in condition monitoring, especially when the machine is operating in a noisy environment.

Bearing defects can be categorized into distributed and localized faults. Typical distributed bearing defects include surface roughness, waviness, misaligned races, and off-size rolling elements, which are usually caused by design and manufacturing errors, improper mounting, wear, and corrosion [11]. Localized bearing defects include cracks, pits, and spalls on the rolling surfaces, which are usually caused by plastic deformation, brinelling, and material fatigue [12]. Both distributed and localized bearing faults increase the noise and vibration levels, and can cause

machinery malfunction. From the standpoint of health condition monitoring, localized defect diagnostics are more important, because the spalling of races or rolling elements is the dominant style of the failure of rolling element bearings in real-world applications whereas many distributed faults originate from a localized spalling [13]. Consequently, the research in this thesis is focused on bearing systems with localized bearing faults.

1.2 Literature Review

This section reviews the state of the art related to bearing fault feature extraction, diagnostic decision-making, and machine fault state forecasting techniques.

1.2.1 Bearing Fault Detection Techniques

Several techniques have been reported in the literature for bearing fault detection. Based on the type of signals, they can be classified into acoustic signal analysis, temperature measurement, lubricant analysis, electrical current analysis, and vibration measurement [13].

The most effective acoustic-based bearing health monitoring is acoustic emission [18]. It is a transient impulse generated by the rapid release of strain energy in solid material under mechanical or thermal stress. The detection of cracks is the prime application of acoustic emission; therefore, this technique can be used as a tool for the detection of bearing faults and shaft cracks. Typically, the accuracy of these methods depends on the sound pressure and sound intensity data [143].

Bearing distributed defects generate excessive heat in the rotating components [20, 144]. Monitoring the temperature of a bearing housing or lubricant is the simplest method for fault detection in rotary machines. In wear debris analysis, the presence of metallic particles in the lubricant is detected by sensitive sensors [19, 145]. Furthermore, the analysis of the different metallic elements in the lubricant can facilitate the location of the fault.

The operating conditions of a machine can also be monitored by analyzing the spectrum of the motor current [21]. The changes in the electric signals are associated with the changes in the mechanical components of the machine; therefore, bearing faults can be detected by using the motor current related signal processing techniques [146].

Since the abnormal vibration of rotary machines is the first sensory effect of rotary component failure, vibration analysis is widely employed in various industries [14-16]. The vibration signal is generated by the interaction between the bearing defect and its mating component. Consequently, a vibration analysis can be employed for the diagnosis of all types of faults, either localized or

distributed. In this work, vibration measurement is applied for the detection of incipient bearing faults due to the ease of measurement and analysis [17].

A vibration-based signal analysis can be performed in the time domain [15, 18], the frequency domain [22, 23], or the time-frequency domain [24, 25]. In the time-domain analysis, a bearing fault is detected by monitoring the variation of some statistical indexes such as the root mean square (RMS) value, the crest factor, or the kurtosis. A bearing is believed to be damaged when a monitoring index exceeds thresholds; however, it is usually difficult to determine the appropriate thresholds because they may vary in different applications.

Frequency-domain methods may be the most commonly used approach in bearing fault detection, by which the bearing defects are detected based on the analysis of spectral information. The main advantage of the frequency-domain analysis over the time-domain analysis is that it is relatively easier to identify and isolate certain frequency components of interest [13]. The bearing health conditions are assessed by examining the fault-related characteristic frequency components in the spectra [26] or in some extended spectral expressions such as bispectrum or cepstrum maps [27, 28]. Frequency-based techniques, however, are not suitable for the analysis of non-stationary signals that are generally related to machinery defects [2].

Non-stationary or transient signals can be analyzed by applying time-frequency domain techniques such as the short-time Fourier transform (FT) [29], the Wigner-Ville distribution [30], and the wavelet transform (WT) [4]. In fault detection, the WT is the most popular time-frequency domain technique because of its more flexible multi-resolution [25, 29]. According to the signal decomposition paradigms, the WT can be classified as the continuous WT [7], the discrete WT [31], the wavelet packet analysis [32], and those WTs with post-processing schemes such as the singularity analysis, the FT, and the energy density analysis. In [33], for example, the singularity analysis was performed across all scales of the continuous WT to identify the temporal location of defect-induced bursts in the vibration signals. An approach based on the combination of the wavelet and FT was suggested for feature extraction in [34, 35]. A time-energy density analysis approach based on the WT was presented in [36], which could extract the fault related features by analyzing the energy distribution over different frequency bands.

Although a number of WT-based approaches have been reported in the literature for feature extraction, no technique has been proposed to deal with the deployment of WTs over a designated frequency band, and also no method has been suggested to effectively integrate the contributions from

various WT processes. To date, effectively extracting representative features related to initial bearing defects still remains a challenging task, particularly in real-world industrial applications [2, 37].

1.2.2 Diagnostic Decision-Making Techniques

Fault detection and diagnosis is a sequential process involving two steps: feature extraction and decision-making (diagnosis). Feature extraction is performed by the use of appropriate signal processing and/or post-processing techniques; it is a mapping process from the signal space to the feature space. The decision-making is a process to classify the obtained features into different categories. The traditional approach of fault diagnostics relies on human expertise to relate the extracted features to the faults, which is usually time-consuming and unreliable, particularly when multiple features are applied for fault diagnostics and when the data are noise affected [38, 39]. The alternative is to use automatic diagnoses. Several automatic decision-making techniques have been reported in the literature, which can be broadly classified into mathematical model-based methods [40] and data-driven approaches [41]. The latter is employed in this research because an accurate mathematical model is usually difficult to derive for complex mechanical systems, especially as the machine operates in uncertain and noisy environments.

The commonly used data-driven methods include pattern recognition methods and inference-based classification approaches. The pattern recognition methods include statistical classifiers, geometric methods, and polynomial classifiers [42], which have been implemented in certain fault detection applications [43-45]. However, these approaches cannot be used for time-varying systems, because they rely on the statistic measurements (e.g., density and probability) of the vibration data. Data driven (or inference-based) diagnostic classification can be performed by intelligent tools such as neural networks (NNs) [16, 46], fuzzy logic [47, 48], and synergetic schemes [49-51].

Neural classifiers are less affected by noise, and have been extensively applied in machine fault diagnostic applications [52-55]. After being properly trained, the NNs can represent the complex relationship between features and faults, and the extracted features can be processed in parallel. In [46], for example, a radial basis function (RBF) network was developed based on features from both the time and frequency domains; it was stated that a combination of the mean of the envelope signal and the 14th spectral amplitude from the power spectrum can lead to a satisfactory fault classification. An expert system based on fuzzy classification and linear membership functions (MFs) was presented in [48] for bearing fault detection, in which the input features were extracted by using the kurtosis method, the crest factor method, and the high frequency resonance technique. An NN-based system

was studied in [57] to detect the bearing faults based on the acoustic signals in a wayside environment; eighty-eight statistical features were generated from discrete wavelet transform coefficients, whereas thirteen optimal features were chosen by a genetic algorithm for classification. However, the disadvantages of NNs are that their internal layers are opaque to the users, and the convergence of learning is usually slow and not guaranteed [58].

Fuzzy logic reasoning deals with system uncertainties and ambiguities in a way that mimics human reasoning [59]. Fuzzy classification starts from highly formalized insights about the structure of categories, and then formulates expert knowledge in a linguistic form. The fuzzy diagnosis is conducted through adequate fuzzy operations [60, 61]. Due to their concise form, fuzzy classifiers have been applied in some real-world applications [47, 48, 62-66]. However, fuzzy systems do not have much learning capability, and it is usually cumbersome to establish an optimal inference rule base when more input variables are utilized [58].

To reap the benefits of both NNs and fuzzy logic techniques, they can be combined into synergism schemes, in which the NNs provide fuzzy systems with learning abilities, and the fuzzy logic provides NNs with a structural framework having a high-level fuzzy IF-THEN rule thinking and reasoning. In [67], for example, a neural fuzzy (NF) scheme was suggested for the diagnosis of localized defects in a ball bearing; the features were extracted from the standard deviation ratio of the wavelet decomposed signals of the testing and reference conditions. The adaptive network-based fuzzy inference system (ANFIS) was applied for bearing fault diagnosis in [68], and the input features are derived from the kurtosis analysis and spectrum peak ratios. A genetic algorithm based feature optimization technique was proposed by Liu (the author of this thesis) *et al.* for bearing fault diagnostics, in which an NF system is utilized for the decision making process [69].

Each of these aforementioned techniques, however, has its own merits and limitations, and can be used for some specific bearing applications only [1, 2, 59]. When a condition monitoring system is employed in real-world industrial applications, the critical issue is its reliability. Unreasonably missed alarms (i.e., the monitoring system cannot pick up existing faults) and false alarms (i.e., the monitoring system triggers an alarm because of noise, not real faults) can seriously mitigate the system's validity [41]. To this end, a more advanced diagnostic system is in demand to provide the industries a more accurate assessment of bearing health conditions.

1.2.3 Machine Fault State Forecasting Techniques

Machinery condition prognosis entails the use of current and previous machinery conditions to predict the future states of a mechanical system (or a component) of interest. Currently, several techniques have been reported in the literature for time series predictions [70]. The classical approaches for nonlinear time series predictions are the use of stochastic models such as the autoregressive (AR) model [71], the threshold AR model [72], the bilinear model [73], the projection pursuit [74], the multivariate adaptive regression splines [75], and the Volterra series expansion [76]. These approaches have found some applications in machinery system prognosis [11, 77-79]. However, it is usually difficult to develop an accurate analytical model for a complex mechanical system [80].

Recent interest in time series prediction has been focused on the use of flexible models such as NNs and fuzzy systems. The NNs have two typical connection architectures: the feedforward and the recurrent networks, and both have been employed in time series forecasting applications [81-87]. From the modelling perspective, a feedforward network is a special case of nonlinear AR models, whereas a recurrent network is a nonlinear AR moving average model. Consequently, it can be inferred that the recurrent NN predictors have an advantage over the feedforward NN predictors in a similar way that the AR moving average models have over the AR models. This conclusion has been quantitatively verified in [88] by simulations and practical tests.

Fuzzy predictors employ linguistic rules for system behaviour forecasting [89]. They begin with a formulated insight into the dynamic behaviours of the system, and then the expert knowledge is formulated into IF-THEN rules, in a way that mimics humans to perform a forecasting task. However, as stated before, fuzzy systems lack a learning capability, and sometimes, it is difficult to properly establish the fuzzy forecasting inference rules.

A solution to overcome these limitations is to use synergetic systems such as NF schemes in which a forecasting operation is performed by fuzzy logic, whereas fuzzy system parameters are trained by NN-based algorithms. In [90], the ANFIS predictor was implemented for time series prediction; the simulation results revealed that the ANFIS predictor provided a higher forecasting accuracy than both the classical AR models and the feedforward NNs. The research in [6, 8] demonstrated that if an NF predictor is properly trained, it even outperforms those based on recurrent NNs.

Although the NF predictors have demonstrated some superior properties to other classical forecasting tools, advanced research still needs to be done in the following aspects before it can be

applied to real-world industrial applications [90, 91]: 1) improving the convergence properties, particularly for multi-step-ahead predictions; and 2) enhancing the adaptive capability to accommodate time-varying system conditions.

1.3 Objectives

To tackle these challenges as discussed in the previous section, an advanced intelligent system is developed in this thesis in order to provide a more accurate assessment of bearing health conditions. As schematically illustrated in Figure 1.1, this intelligent system consists of two modules: the representative feature extraction by using the proposed signal processing technique and the decision-making by applying the suggested diagnostic scheme. The strategy is to develop advanced and more robust techniques at each processing stage to improve the reliability of bearing condition monitoring. The specific approaches are summarized as follows:

1) To develop a more reliable signal processing technique for bearing fault detection. The emphasis is placed on feature extraction and the analysis of non-stationary signatures that are generated, for example, by the faults on the rolling elements and rotating rings of bearings.

2) To develop an enhanced diagnostic (ED) scheme for automatic diagnostic decision-making. The suggested ED scheme consists of two modules:
   - A novel NF classifier is proposed to effectively integrate the strengths of several signal processing techniques for a more accurate assessment of the health condition of bearings.
   - A new multi-step predictor is developed and integrated into the ED scheme to forecast the future states of the bearing health condition, and to further enhance the diagnostic reliability.

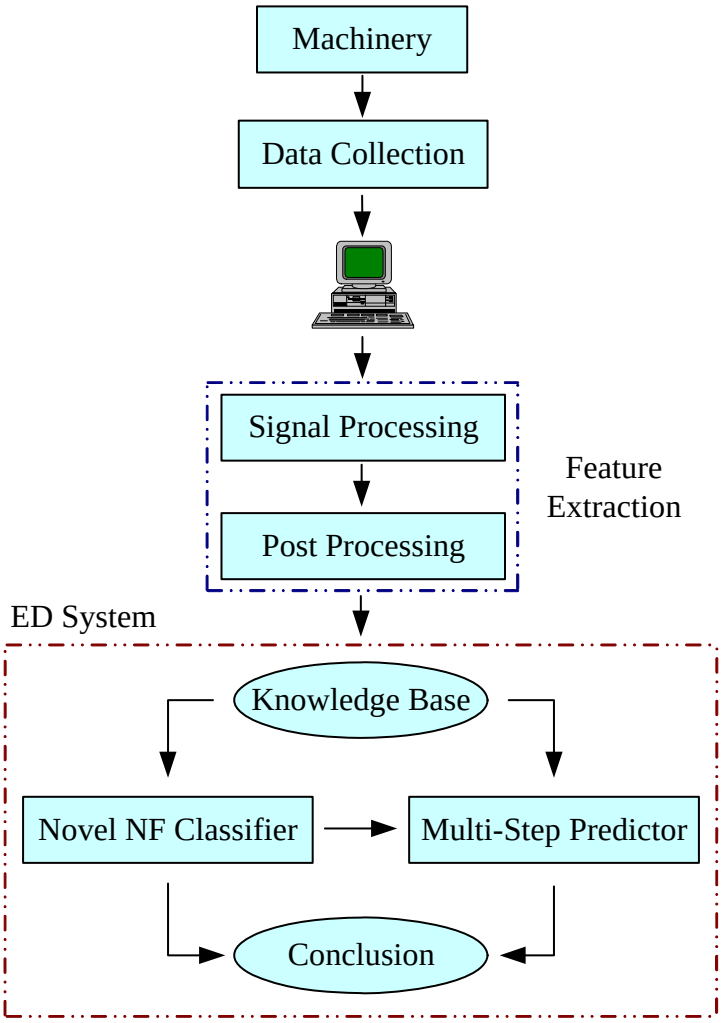

Figure 1.1. Schematic diagram of the developed intelligent system for bearing condition monitoring.

1.4 Experimental Setup

An experimental setup is employed in this research to collect the vibration signals for two purposes: to study the vibration signatures generated by incipient bearing faults, and to verify the techniques to be developed. This experimental setup is shown in Figure 1.2. The system is driven by a 3-hp induction motor, with the speed range between 20 and 4200 r/min. The shaft rotation speed can be controlled by a speed controller (Delta VFD-PU01). An optical encoder is used for shaft speed measurement. Thus, the shaft rotation speed is known directly from the reading of this speed controller, which can be further verified by applying the FT on the signals from the optical encoder. A flexible coupling is utilized to damp out the high-frequency vibration generated by the motor. Two ball bearings are fitted into the solid housings. Accelerometers (ICP-IMI, SN98697) are mounted on the housing of the tested bearing to measure the vibration signals along two directions. Considering the structure properties, the signal that is vertically measured is utilized for analysis, whereas the signal that is horizontally measured is used for verification. A static load is applied by two disks, and a variable load is applied by a magnetic brake system (Placid Industries, B150-24-H) through a bevel gearbox and a belt drive. A data acquisition board (NI PCI-4472) is employed for signal collection, which has built-in anti-aliasing filters with the cut-off frequency set at half of the sampling rate.

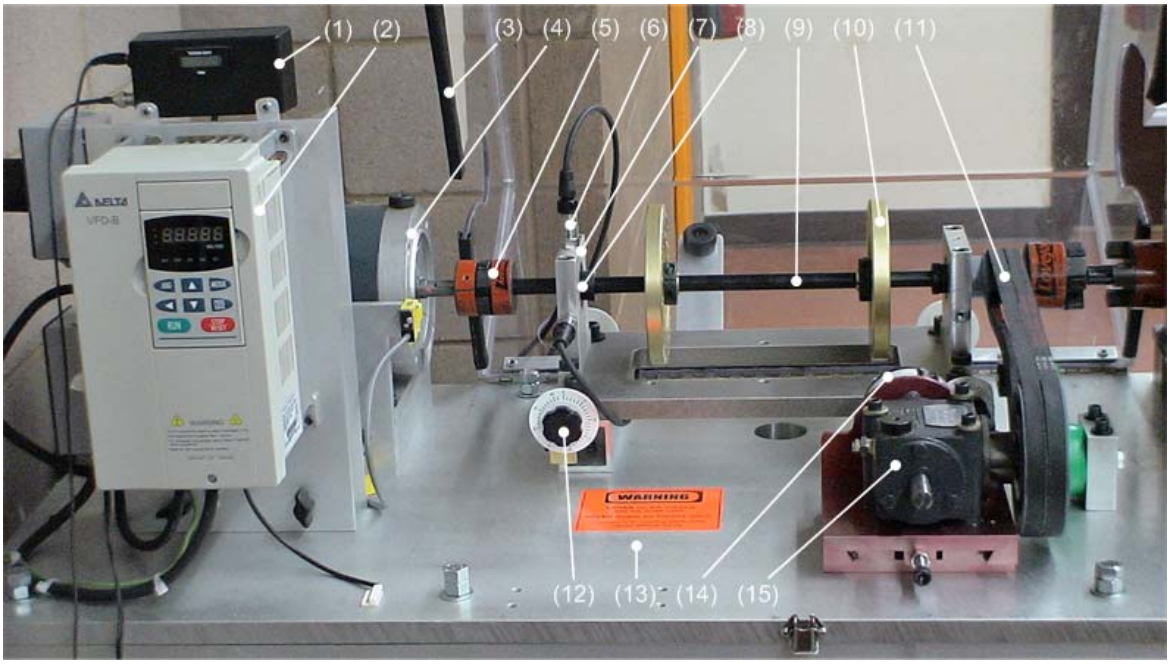

Figure 1.2. Experimental setup: (1)-digital encoder; (2)-variable speed control; (3)-enclosure; (4)-motor; (5)-flexible coupling; (6)-ICP accelerometer; (7)-bearing housing; (8)-tested bearing; (9)-hardened shaft; (10)-load disc; (11)-belt drive; (12)-alignment adjustor; (13)-base; (14)-magnetic load system; (15)-bevel gearbox.

The ball bearings of type MB ER-10K are utilized. Four bearing health conditions are considered: healthy bearings, bearings with outer race defects, bearings with inner race defects, and bearings with rolling element faults. Three sets of faulty bearings with different fault sizes are prepared and tested. The fault types and the corresponding fault dimension (in diameter) are listed in Table 1.1. Each bearing is tested under seven different shaft speeds (600, 900, 1200, 1500, 1800, 2100, and 2400 r/min) and two load levels (1.2 and 2.3 N · m), respectively. When a bearing is running at a specific shaft speed and load level, five segments of vibration signals are collected.

Table 1.1. Fault type and size

	Outer Race Defect	Inner Race Defect	Rolling Element Defect
SET 1 (mm)	0.35	0.41	0.29
SET 2 (mm)	0.39	0.47	0.35
SET 3 (mm)	0.44	0.56	0.38

1.5 Thesis Outline

Chapter 2 presents a new signal processing technique, the wavelet spectrum analysis, to extract the representative features that are related to the bearing health conditions. The investigation results demonstrate that this new signal processing technique is an effective bearing fault detection method, which is especially useful for non-stationary feature extraction and analysis.

In Chapter 3, a novel NF classifier is discussed. This classifier can integrate the strengths of the two suggested signal processing techniques, the wavelet spectrum analysis technique and the kurtosis ratio method, for a more reliable bearing fault diagnosis.

A multi-step NF predictor with a variable input pattern is proposed in Chapter 4 for dynamic system state forecasting. The performance of this new predictor is evaluated, in terms of the input pattern, structure, and the training algorithm, by simulations with both benchmark data sets and experimental data sets.

In Chapter 5, the newly-developed intelligent tools, the NF classifier and the NF predictor, are integrated to construct the ED scheme for comprehensive fault diagnostics in rolling element bearings.

Finally, the conclusion remarks and some future research topics are summarized in Chapter 6.

Chapter 2
A Wavelet Spectrum Analysis Technique

2.1 Problem Definition

Typically, the bearing defects arise during the operation. Therefore, the detection of these defects at their early stage without machine disassembly is pivotal in bearing condition monitoring. The magnitudes of the vibration signatures generated by these incipient bearing defects are related to the fault size, dynamic load, and the shaft speed [2]. To date, effectively extracting the representative features associated with incipient bearing defects still remains a challenging task, particularly in real-world industrial applications [37]. In this work, the bearing defects of three sizes as given in Table 1.1 in Chapter 1 will be analyzed, and a wavelet spectrum analysis (WSA) technique is proposed in Section 2.2 to extract the representative features for bearing fault detection.

2.2 Theory of Wavelet Spectrum Analysis

Whenever a fault occurs on a bearing component, stationary and/or non-stationary impacts are generated, which excite the bearing and its support structures. As a result, resonance signatures [26] are generated, which are usually buried in other high-amplitude vibration signals. In the proposed WSA technique, the first step is to find these resonance signatures; the second step is to synthesize the resulting wavelet coefficient functions by applying the proposed weighted Shannon function; the third step is to construct the autocorrelation spectrum of the synthesized function to highlight bearing characteristic frequencies [27]; finally, an averaging process is taken in the frequency domain to eliminate some random noises that are related to the variations in operating conditions.

2.2.1 Determination of Wavelet Functions

Given a rolling element bearing, the defect usually occurs on the fixed ring race first because the fixed ring material is subject to more dynamic load cycles [13]. Consider a bearing with a fixed outer ring, and suppose that a defect (e.g., a fatigue pit) has happened on the outer ring race. Each time a rolling element rolls over the pit, an impulse is generated due to the impact. This impulse excites the vibration resonance of the bearing and the around structures. In theory, the excited transient modes due to an outer race defect do not vary because the defect angular position remains the same as each

impact occurs. On the other hand, if a defect occurs on a rolling element or an inner ring race, the generated impulse transient modes will change in properties because the impact occurs at a different angular position as the bearing components rotate. Figure 2.1(a) shows a typical vibration signal generated by a bearing with an inner race defect. Because the defect is at its initial stage, it is difficult to directly recognize the bearing fault from the measured vibration signal without an appropriate signal processing technique. Figure 2.1(b) shows the residual signature that is obtained by high-pass filtering the original vibration data with a cut-off frequency 1000 Hz, in which the bearing fault related features are highlighted; these features can be further enhanced by using the proposed WSA technique as discussed below. It is also seen that the magnitudes of the impulse transients and the excited resonance modes vary over time, due to the variation of the angular positions of the impacts. Correspondingly, it is more suitable to apply the WT to analyze these non-stationary resonance signatures, which enables a higher time-resolution solution at high frequencies and *vice versa* [93]. Also for comparison, a typical vibration signal generated by a bearing with an outer-race defect, running at the same shaft speed under the same load level as what in Figure 2.1, is shown in Figure 2.2. It is seen that the signatures generated by outer-race defect bearing is stronger in magnitude than that by inner-race defect bearing, due to the difference in signal transmission pathway and the nature of resonance modes. Therefore, extracting the representative features for an inner-race defect bearing is more challenging than that for an outer-race defect bearing under the same operating condition. In the following context, the WT will be applied to analyze these non-stationary signatures generated by incipient bearing faults for feature extraction.

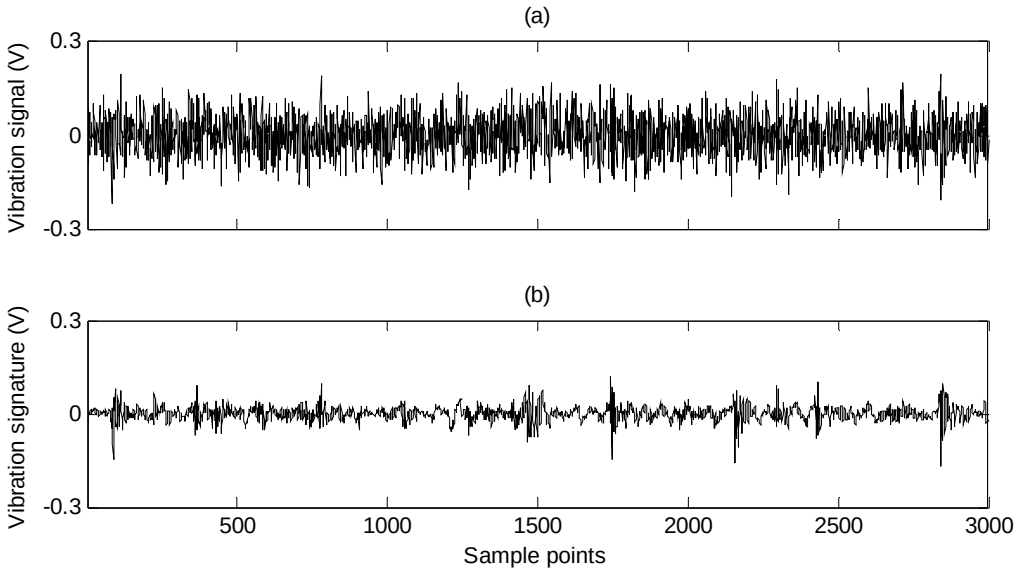

Figure 2.1. (a) vibration signal of a bearing with an inner race defect running at shaft speed 30 Hz and load level 2.3 N•m; (b) signature obtained by using a high-pass filter.

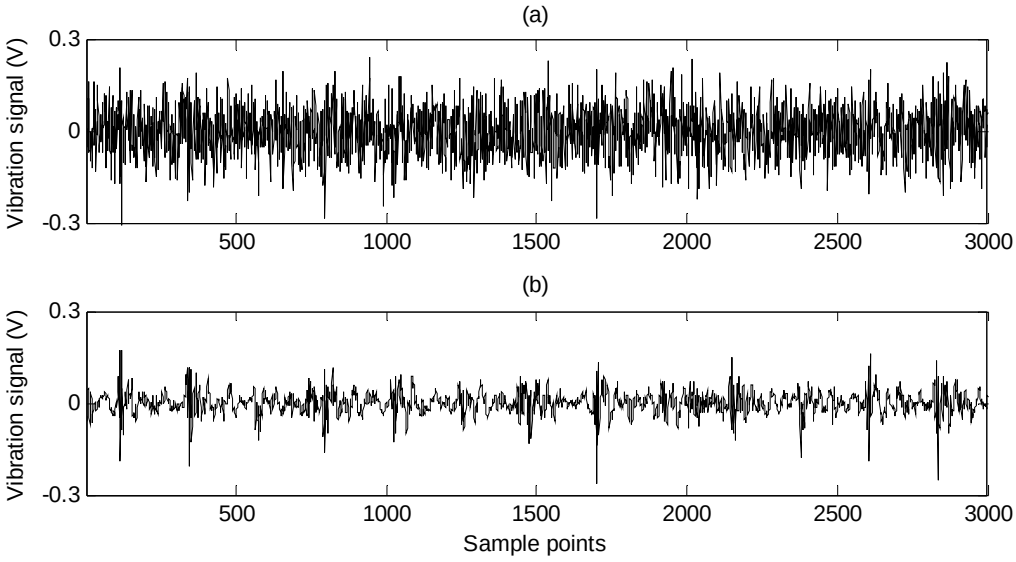

Figure 2.2. (a) vibration signal of a bearing with a outer race defect running at shaft speed 30 Hz and load level 2.3 N•m; (b) signature obtained by using a high-pass filter.

Given a time-domain signal $x(t)$, the wavelet coefficients are determined by

$$WT_x(t,s) = \int_{-\infty}^{+\infty} x(\tau)\sqrt{s}\, w^*\left(-s(t-\tau)\right) d\tau \ , \tag{2.1}$$

where $w^*(t)$ denotes the complex conjugation of mother wavelet function $w(t)$; *s* and *t* are the scale and time variables, respectively, which produce dilation and translation [92].

The choice of an appropriate mother wavelet depends on the signal properties and the purpose of the analysis. In bearing fault detection, the interest is to obtain the resonance features that are induced by a localized bearing fault. The selection of Morlet wavelet as the mother wavelet for the signal analysis is well accepted in the field of machinery fault detection [2, 4, 96]. The Morlet wavelet is a modulated Gaussian function [92]

$$w(t) = \exp\left(-\frac{t^2}{2b_0^2}\right)\exp\left(j2\pi f_0 t\right), \tag{2.2}$$

where b_0 is the spread of the Gaussian function and f_0 is the centre frequency of the pass-band of the mother wavelet. The real and imaginary parts of a complex-valued Morlet wavelet are shown in Figure 2.3.

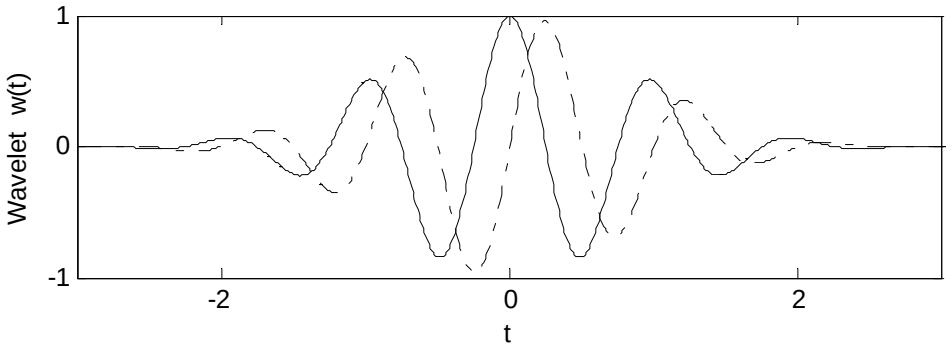

Figure 2.3. Real (solid line) and imaginary (dashed line) parts of the Morlet wavelet.

Given a mother wavelet, the following WT admissibility should be satisfied [92]:

$$0 < \int_{-\infty}^{+\infty} \frac{|W(f)|^2}{|f|}\, df < +\infty, \tag{2.3}$$

where $W(f)$ is the FT of $w(t)$. Because $w(t)$ decays over time, the admissibility (2.3) is equivalent to the requirement [93]:

$$\int_{-\infty}^{+\infty} w(t)\, dt = 0, \tag{2.4}$$

Strictly speaking, the Morlet wavelet does not satisfy the zero-mean requirement. However, the mean can become infinitely small if the term $b_0 f_0$ is sufficiently large. As $b_0 f_0$ increases, the duration of the wavelet expands, and the time resolution will decrease correspondingly. As a result, the obtained mother wavelet $w(t)$ may not be suitable to analyze fast-decaying transient signatures. Based on the above reasoning, the product of the spread and the scaled centre frequency is kept as a constant in this work, i.e.,

$$b_i f_i = \frac{b_0}{s_i}(f_0 s_i) = b_0 f_0 = \frac{2\pi b_0 f_0}{2\pi} = \frac{1}{\sqrt{2\ln 2}}, \tag{2.5}$$

where $2\pi b_0 f_0 = \pi\sqrt{2/\ln 2}$ was given in [92]; s_i represents the *i*th selected scale; b_i and f_i are the corresponding *i*th spread and centre frequency, respectively. Based on the relation between b_i and f_i as in (2.5), the mean of the obtained mother wavelet $w(t)$ remains at a level as low as 10^{-12} in this work, and the effective support will vary with the scaled centre frequency to accommodate the variation of the signatures of interest.

The FT of the mother wavelet $w(t)$ in (2.2) is given by

$$W(f) = b_0\sqrt{2\pi}\exp\left(-2b_0^2\pi^2(f - f_0)^2\right). \tag{2.6}$$

Thus, the FT of the dilated wavelet $w(st)$ becomes

$$W_s(f) = b_i\sqrt{2\pi}\exp\left(-2b_i^2\pi^2(f - f_i)^2\right). \tag{2.7}$$

For the chosen mother wavelet, the following equation holds, $w^*(-st) = w(st)$. Therefore, (2.1) can be implemented by using

$$WT_x(t,s) = \sqrt{s}F^{-1}\left[X(f)W_s(f)\right] \tag{2.8}$$

where $F^{-1}[\cdot]$ denotes the inverse FT, $X(f)$ is the FT of $x(t)$, and $W_s(f)$ is given by (2.7). In the implementation of (2.8), $X(f)$ is calculated only once, whereas $W_s(f)$ is calculated for all values of the scale variable *s*.

2.2.2 Strategy for Deployment of Wavelet Center Frequencies

High resonance signatures of the bearing structure are usually amplitude modulated by the bearing defect [26], thus the analysis of these resonance signatures plays an important role in bearing fault detection. The frequency band of interest is chosen as $[Nf_t, f_s/2.56]$ to reduce the interference effects from the low-frequency noisy components, where f_t is the shaft speed, *N* is the order of shaft harmonics, and Nf_t represents the lower bound frequency, the signal under which will not be considered for feature extraction; f_s is the sampling frequency, and the constant 2.56 is usually selected instead of 2 in industries [94] to avoid aliasing. The centre frequencies of the wavelet should be deployed to implement the WT over this designated frequency band, without the overlapping between the wavelet frequency bands. Based on (2.8), the 3-dB [147] bandwidth BW_i for the *i*th centre frequency f_i is derived as follows:

$$BW_i = \left[1-\lambda, 1+\lambda\right] f_i, \tag{2.9}$$

where $\lambda = \ln 2/\sqrt{2}\pi$. Beginning with the lower bound frequency Nf_t, the centre frequencies can be recursively calculated by

$$f_i = \frac{(1+\lambda)^{i-1}}{(1-\lambda)^i} Nf_t, \qquad i = 1, 2, \ldots, m\text{-}1 \tag{2.10}$$

$$f_i = \frac{1}{2}\left[\frac{f_s}{2.56} + f_{i-1}(1+\lambda)\right], \qquad i = m \tag{2.11}$$

where m is the number of the selected scales. The bandwidth between f_{m-1} and the upper bound frequency $f_s/2.56$ may be smaller than the 3-dB bandwidth of f_m, thus the middle point is utilized in (2.11).

2.2.3 A Weighted Shannon Function for Synthesizing Wavelet Coefficient Functions

Based on (2.10) and (2.11), the wavelet coefficients $|WT_x(t, s_i)|$ over seven bandwidths are derived and partly shown in Figures 2.4(c) – (i), and the corresponding time-domain vibration signatures are reproduced in Figures 2.4(a), (b). By comparing these figures, it is seen that the fault related features are usually reflected over several wavelet bands due to the variation of the transient modes. In the literature [34], one-scaled WT like those shown in Figure 2.4(e) is usually applied to demodulate the resonance vibration signatures for feature extraction. Consequently, the useful information from other frequency bands [e.g., Figures 2.4(c), (d), (f) – (i)] is missed. Accordingly, in this work, a weighted Shannon function is proposed next to integrate the contribution from the vibration signature of each frequency band for a more effective feature extraction.

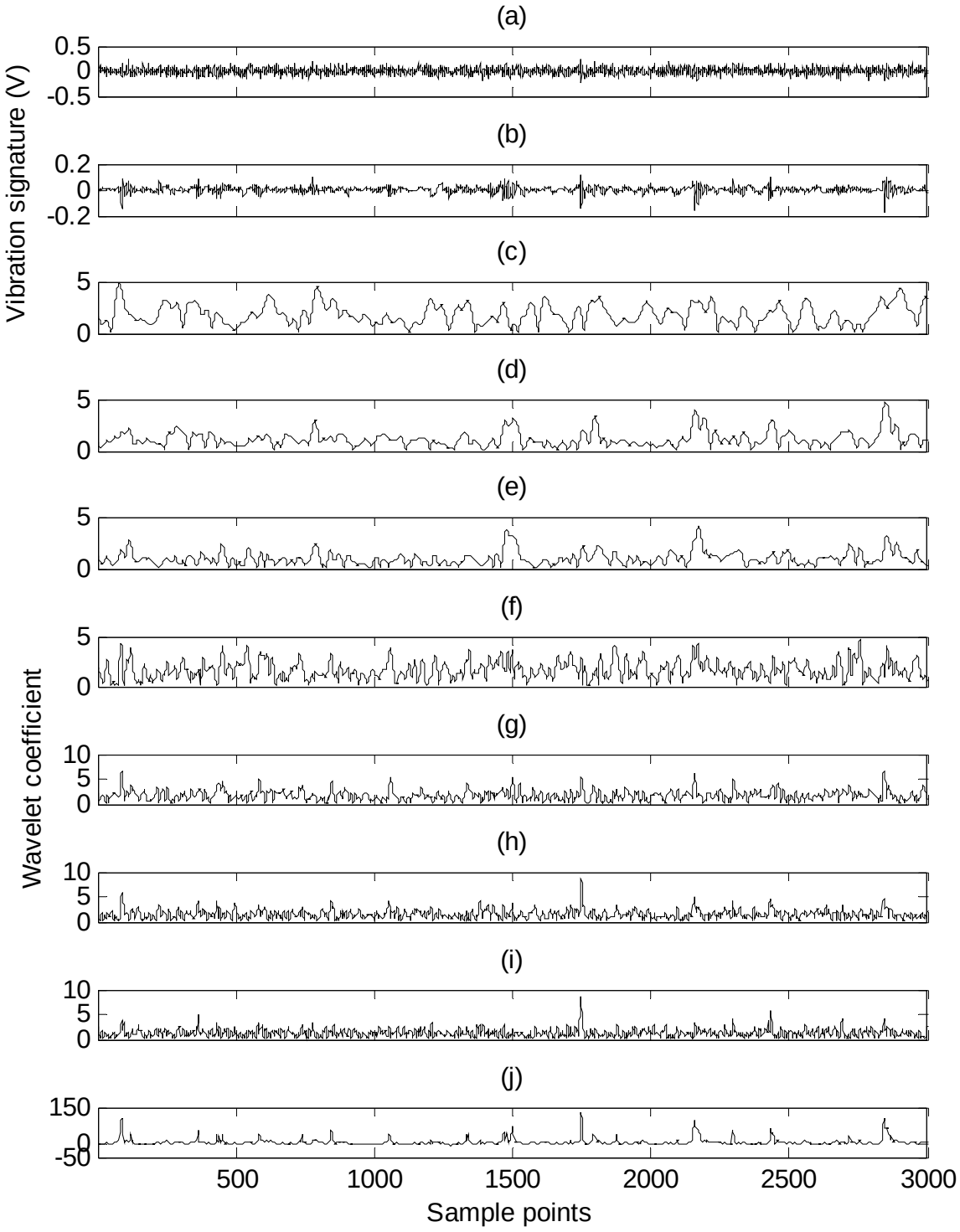

Figure 2.4. Vibration signatures (a) – (b) and wavelet coefficients: (c) – (i) obtained from seven frequency bands; (j) integrated using the proposed Shannon function (2.12).

This technique is proposed, based on an extended Shannon entropy function. The Shannon entropy in information theory is a measure of the degree of uncertainty that is associated with a random variable [95, 96]. In this study, a modification is made to that function to highlight the fault-related features and to minimize the interference information,

$$h(t_k) = \sum_{i=1}^{m} \left[\overline{WT}_x(t_k, s_i) \log_\beta \overline{WT}_x(t_k, s_i) \right], \tag{2.12}$$

where i = 1, 2, . . ., m; k = 1, 2, . . ., n; n is the number of discrete points of signal $x(t)$; β is the selected logarithm base that may vary from 2 to 10 (β=2 in this study); and the normalized wavelet coefficient $\overline{WT}_x(t_k, s_i)$ is given by

$$\overline{WT}_x(t_k, s_i) = \frac{|WT_x(t_k, s_i)|}{\sigma_i}, \tag{2.13}$$

where σ_i represents the spread estimate of $|WT_x(t, s_i)|$.

For a wavelet scale s_i, the term $\overline{WT}_x(t_k, s_i) \log_\beta \overline{WT}_x(t_k, s_i)$ is used to highlight the coefficients $\overline{WT}_x(t_k, s_i)$ whose magnitudes are larger than $\beta\sigma_i$, and diminish the others. The fault related features will correspond to the coefficients of higher magnitudes, which are derived from the resonance features that closely 'match' the dilated (i.e., s_i) wavelets. The threshold for determining whether the derived wavelet coefficients are bearing condition related depends on the selection of the design parameter β, such that a wavelet coefficient with a magnitude larger than $\beta\sigma_i$ is assumed to be bearing condition related. In operations, some noise effects may be also taken into the process over a specific bandwidth, however, these effects can be diminished in the "additive" operation (2.12). Figure 2.4(j) shows the integrated wavelet coefficients by which the bearing condition-related features are highlighted.

It can also be observed from Figure 2.4 that the bearing condition-related features become more prominent over some bandwidths [e.g., Figures 2.4(d), (e), (h), and (i)] and less pronounced over others [e.g., Figures 2.4(c), (f), and (g)]; each frequency band corresponds to a wavelet scale as in (2.9). If these wavelet coefficients are integrated for a lower-dimensional (1-D) feature representation,

their contribution to the feature extraction will differ from one centre frequency to another. To find an indicator to represent such an effect, the magnitude of the wavelet coefficient function is treated as a discrete random variable in this case, whereas its probability distribution corresponding to different scales is examined and plotted in Figure 2.5. It is seen that the tail length of the probability density function varies with respect to the wavelet scale s_i. This is because the characteristic defect frequency of a bearing is usually carried by resonance signatures generated by the impact as a rolling element strikes a defect. However, similar types of features may also be generated by other vibratory sources such as inherently varying compliance, misalignment, and imbalance [13]. When applying the WT to process the overall vibration signal from a faulty bearing, the coefficients $\left|WT_x(t,s_i)\right|$ obtained from the non-defect-related resonance signatures (or noise) usually exhibit the characteristics of lower amplitude and longer duration (counting for the major part of the total energy). On the other hand, the WT coefficients corresponding to the defect-related resonance signatures usually exhibit the characteristics of higher amplitude but with a much shorter duration (counting for a minor part of the overall energy) [26]. Fault detection is to extract the representative features related to machinery defects from a collected signal. From signal property standpoint, if a bearing is damaged, the magnitude distribution of the wavelet coefficients should have a longer tail, which indicates that fault related features buried in the resonance signatures have been enhanced. This long tail is induced by some coefficients with high magnitudes, whereas the magnitudes of most of coefficients are located around the sample mean. Therefore, a higher degree of asymmetry in the probability density function (pdf) of the wavelet coefficients is desirable for a faulty bearing condition since the resonance features generated by the faulty bearing are enhanced and manifested by the presence of the peaks in the coefficient functions. The asymmetric property of the probability distribution of the wavelet coefficients can be characterized by the skewness measure of the pdf of the resulting wavelet coefficients, i.e.,

$$\gamma_i = \frac{E[\left(\left|WT_x(t,s_i)\right| - \mu_i\right)^3]}{\sigma_i^3}, \tag{2.14}$$

where $E[\cdot]$ denotes the expectation function, μ_i and σ_i represent the mean and the spread of $\left|WT_x(t,s_i)\right|$, respectively. Figure 2.6 plots the values of the index γ_i computed from the distributions as illustrated in Figure 2.5. It is clear that this proposed skewness measure takes higher

values at scales s_2, s_3, s_6, s_7, and the corresponding wavelet coefficients over these scales are partly shown in Figures 2.4(d), (e), (h), and (i), respectively.

The suggested index γ_i is employed to measure the capability of the WT in extracting fault-related features, which can be implemented into the extended Shannon function (2.12),

$$H(t_k) = \sum_{i=1}^{m} \left[\gamma_i \overline{WT}_x(t_k, s_i) \log_\beta \overline{WT}_x(t_k, s_i) \right]. \tag{2.15}$$

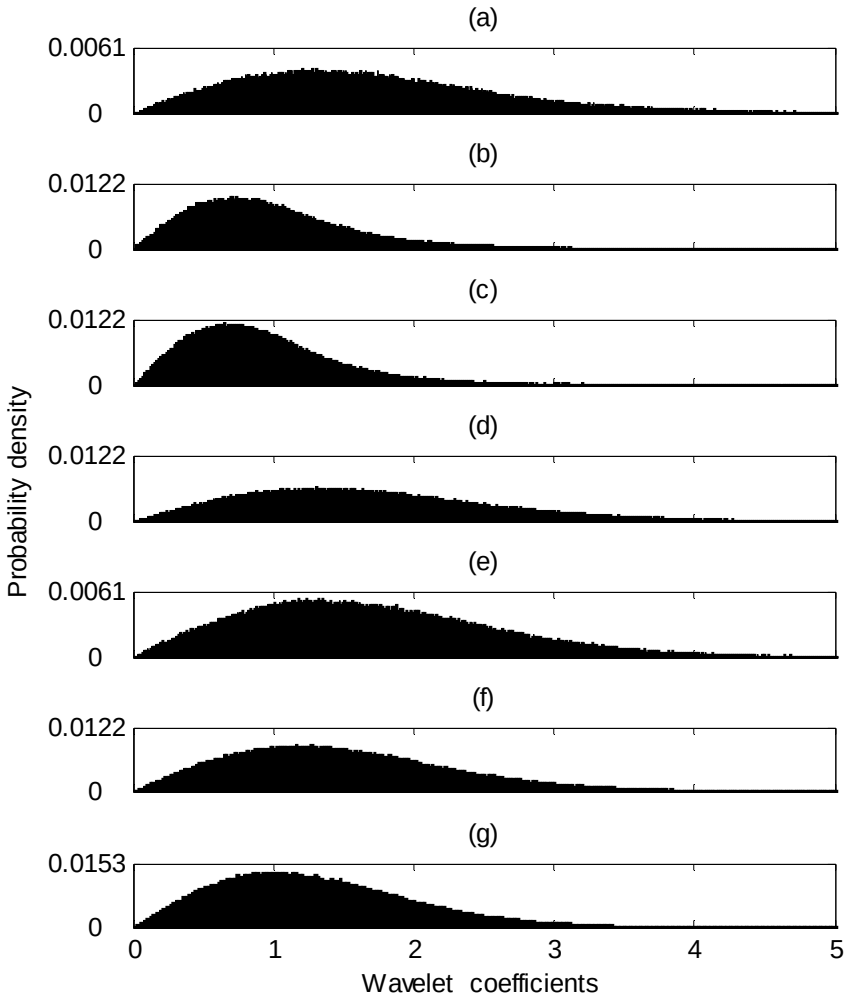

Figure 2.5. Probability distribution functions of magnitude of the wavelet coefficients corresponding to seven wavelet scales (s_1 to s_7); probability is evaluated by using its relative frequency of the occurrence with 1000 bins and 491520 sample points (partly shown in Figure 2.4).

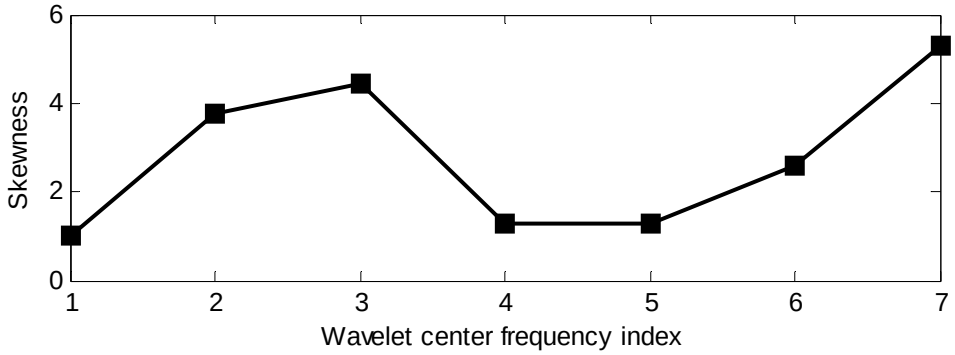

Figure 2.6. The trend of the proposed skewness measure γ_i corresponding to an inner race defect bearing (the signature is partly shown in Figure 2.1) when the wavelet scales varies from s_1 to s_7.

2.2.4 Averaged Autocorrelation Spectrum

Once the synthesized wavelet coefficients are obtained, the next step is to detect the characteristic defect frequency by calculating the averaged autocorrelation spectrum. The autocorrelation spectrum analysis takes two processes: do the autocorrelation on the synthesized wavelet coefficients $H(t)$ given by (2.15) to enhance the involved periodic features, and conduct the spectral analysis (FT) for periodic feature extraction. Specifically,

$$r_{xx}(l) = E[H(t)H^*(t+l)], \qquad l = 1, 2, \ldots, n\text{-}1 \tag{2.16}$$

$$R(f) = F[r_{xx}(l)] \tag{2.17}$$

where l is the lag index and $F[\cdot]$ denotes the FT. The autocorrelation spectrum is defined as:

$$\Phi(f) = R(f)R^*(f). \tag{2.18}$$

In implementation, the spectrums obtained by (2.18) from J segments of measured signals will be an averaged to reduce the effects of random noise,

$$\overline{\Phi}(f) = \frac{1}{j}\sum_{j=1}^{J}\Phi_j(f), \tag{2.19}$$

where j = 1, …, J. Bearing health conditions are estimated by analyzing the related characteristic frequency components on the resulting spectra, to be discussed in Section 2.3.

2.3 Performance Evaluation

The proposed WSA technique is applied in this section for bearing fault detection. Its performance will be compared with several related classical methods: the max-envelope approach [35], the envelope analysis [26], and the one-scaled WT [34]. In the max-envelope approach, only the maximum wavelet coefficient over the selected frequency bands given by (2.9) is selected as the resultant value. The envelope analysis [26] is applied to the signal bandpass-filtered around the system's resonant frequency [1500 2500] Hz, as shown in Figure 2.7. The one-scaled WT is performed with the centre frequency at 2000 Hz which is the middle of the system's resonant frequency band. Correspondingly, the vibration signatures within the resonance frequency band [1000 8000] Hz will be analyzed for bearing fault detection [26]; the lower bound frequency is chosen as 1000 Hz to eliminate the effects of shaft-related low-frequency components.

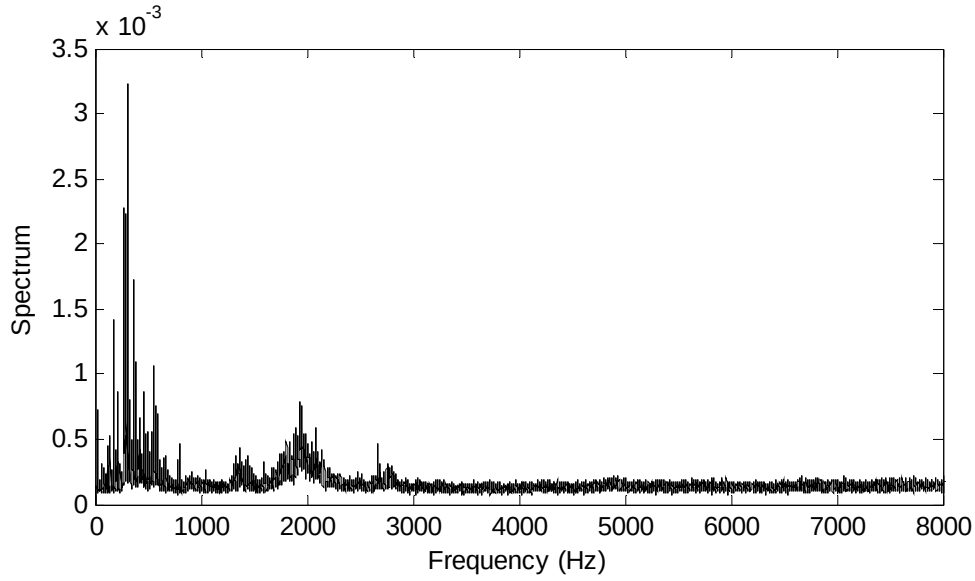

Figure 2.7. System's resonance frequency response.

The bearing fault is diagnosed by examining the related characteristic frequencies and their harmonics, whose theoretical values can be calculated based on the geometry of the bearing and the shaft speed [27]. For example, for a bearing with a fixed outer ring, the characteristic frequency for an inner race defect is

$$f_{id} = \frac{Zf_t}{2}\left(1+\frac{d}{D}\cos\alpha\right) \tag{2.20}$$

where Z is the number of rolling elements, d is the rolling element diameter, D is the bearing pitch diameter, a is the contact angle (a = 0 in this case), and f_t is the shaft speed.

The characteristic frequency for a rolling element defect is

$$f_{rd} = \frac{Df_t}{d}\left(1-\frac{d^2}{D^2}\cos^2\alpha\right). \tag{2.21}$$

The characteristic frequency for an outer race defect is

$$f_{od} = \frac{Zf_t}{2}\left(1-\frac{d}{D}\cos\alpha\right). \tag{2.22}$$

2.3.1 Experimental Data Preparation

To verify the effectiveness of the WSA technique in bearing fault detection, a large number of tests have been conducted and the experimental setup that is employed for these tests is described in Section 1.4 in Chapter 1. In the tests, four bearing health conditions are considered: healthy bearings, bearings with outer race defects, bearings with inner race defects, and bearings with rolling element faults. Three sets of faulty bearings with different fault sizes are prepared and tested, as listed in Table 1.1. Each bearing is tested under seven shaft speeds (600, 900, 1200, 1500, 1800, 2100, and 2400 r/min) and two load levels (1.2 and 2.3 N$\cdot$m), respectively. Thus, totally 42 different test cases have been conducted. In the following Section 2.3.2, the processing results from two cases (Case I and II) will be shown in figures to give the readers some institutive impressions, where Case I refers to bearings in SET 1 in Table 1.1 operating at shaft speed 32 Hz with a load 1.2 N$\cdot$m, and Case II represents bearings in SET 2 operating at shaft speed 40 Hz with a load 2.3 N$\cdot$m. Then in Section 2.3.3, the processing results from all the 42 cases will be summarized into a series of tables and

figures to show the effectiveness of the proposed WSA technique and the effects of the load, shaft speed, and the fault size on feature extraction.

2.3.2 Demonstrative Examples of Processing Results

1) Synthesized Wavelet Coefficients: Figure 2.8 shows parts of measured vibration signals at shaft speed 1920 r/min (32Hz) with the sampling frequency 20,480 Hz, from which the bearing imperfections cannot be directly recognized. Figure 2.9 plots some examples of the synthesized wavelet coefficients $H(t)$ calculated from (2.15) corresponding to different bearing health conditions. It can be seen that the bearing condition-related features are extracted from the weak vibration signatures. As shown in Figure 2.9(a), for a healthy bearing, the features related to the shaft rotation speed (power input) can be extracted clearly. The representative features are spaced by an interval of 640 sample data, corresponding to a repetition rate of 32 Hz and a sampling frequency 20,480 Hz. It is also seen that the features from outer race defects are stationary which is relatively easier to be extracted as shown in Figures 2.9(b) and (c). However, as seen in Figures 2.9(d) – (g), it is challenging to extract representative features from bearings with inner race defects or rolling element defects because the related resonance signatures are non-stationary.

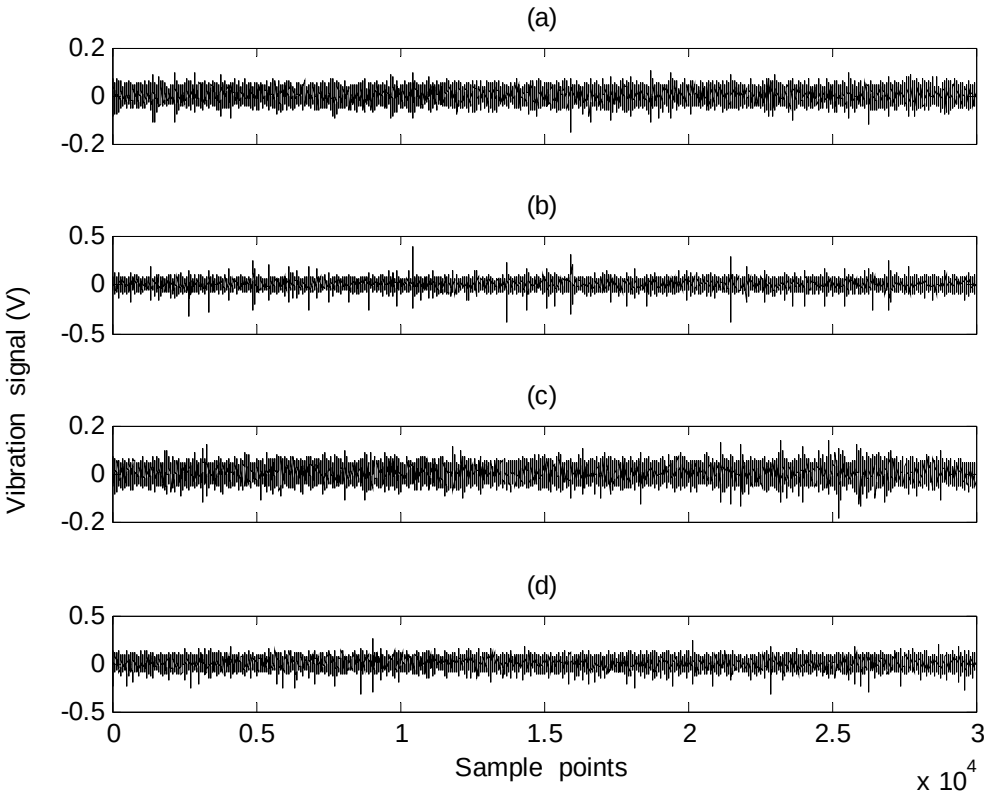

Figure 2.8. Vibration signals: (a) from a healthy bearing; (b) from an outer race defect bearing; (c) from an inner race defect bearing; (d) from a rolling element damaged bearing.

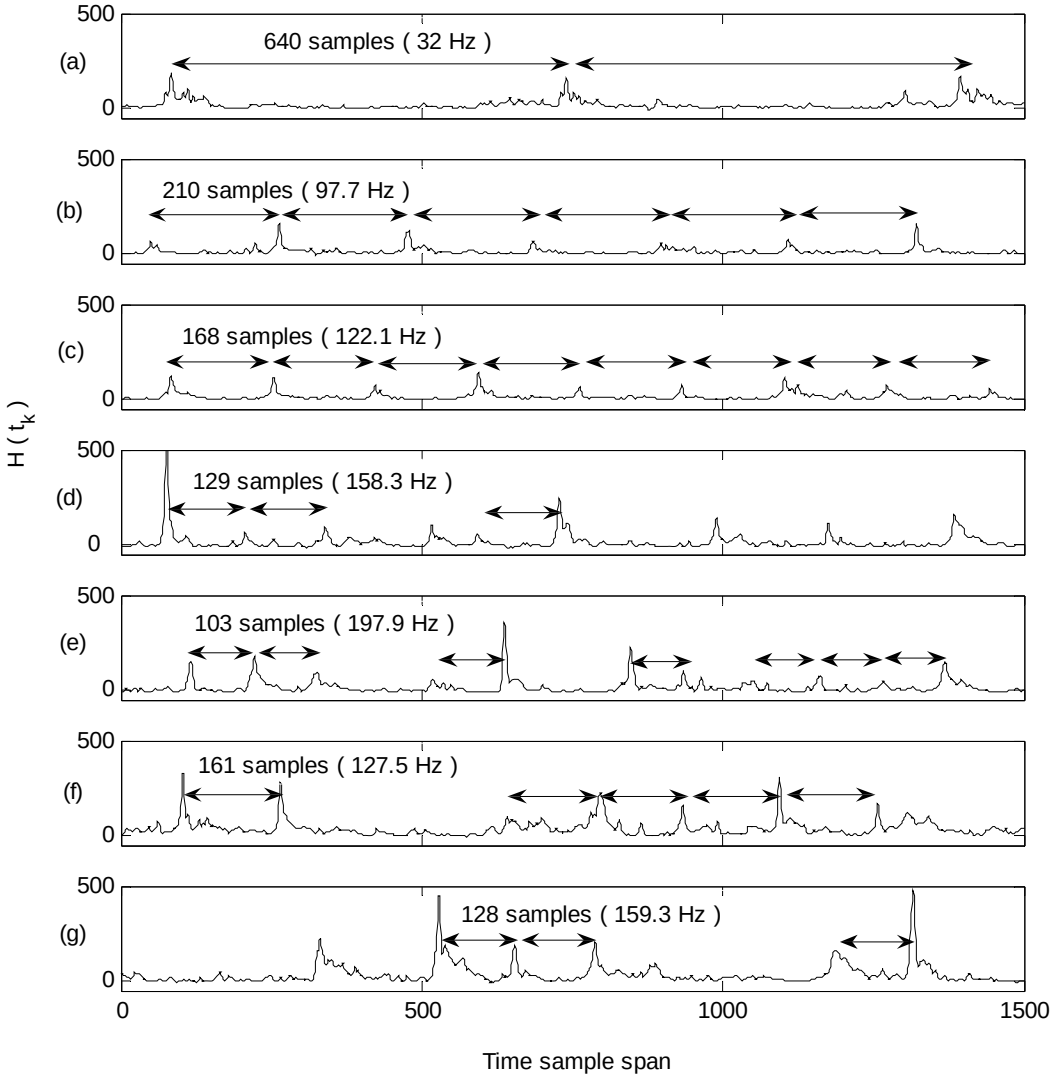

Figure 2.9. Synthesized coefficients where the repetition of the bearing condition related features is indicated using sampling intervals and the features are pointed by arrows: (a) from a healthy bearing; (b) from an outer race defect bearing in Case I; (c) from an outer race defect bearing in Case II; (d) from an inner race defect bearing in Case I; (e) from an inner race defect bearing in Case II; (f) from a rolling element damaged bearing in Case I; (g) from a rolling element damaged bearing in Case II.

2) Healthy Bearing: Bearing components generate vibratory signals that are related to the speed of shaft rotation (the power input). In theory, the shaft speed related resonance features can be recognized by the proposed technique. In industrial monitoring applications, the examination is mainly based on directly visual inspection of the resulting spectra. The following analyses are in terms of the recognition of the dominant frequency component in the spectra under an identical resolution. Figure 2.10 shows the processing results for a healthy bearing by using the WSA technique and three classical techniques. It is seen from Figure 2.10(a) that, based on the proposed WSA technique, the dominant spectral component is the shaft speed (f_t = 32 Hz). This high-resolution shaft speed information can be used not only as a healthy indicator of the bearing of interest, but also as a digital encoder in system condition monitoring and control applications.

This shaft speed can also be recognized in the spectra from the max-envelope approach [Figure 2.10(b)], although its resolution is lower than in Figure 2.10(a) due to the presence of some other spectral components. The results from the one-scaled WT [Figure 2.10(c)] and the envelope analysis [Figure 2.10(d)], however, may lead to an incorrect fault detection (i.e., a false alarm) because the third harmonic of shaft speed (96 Hz) is very close to the outer race defect characteristic frequency (f_{od} = 97.67 Hz).

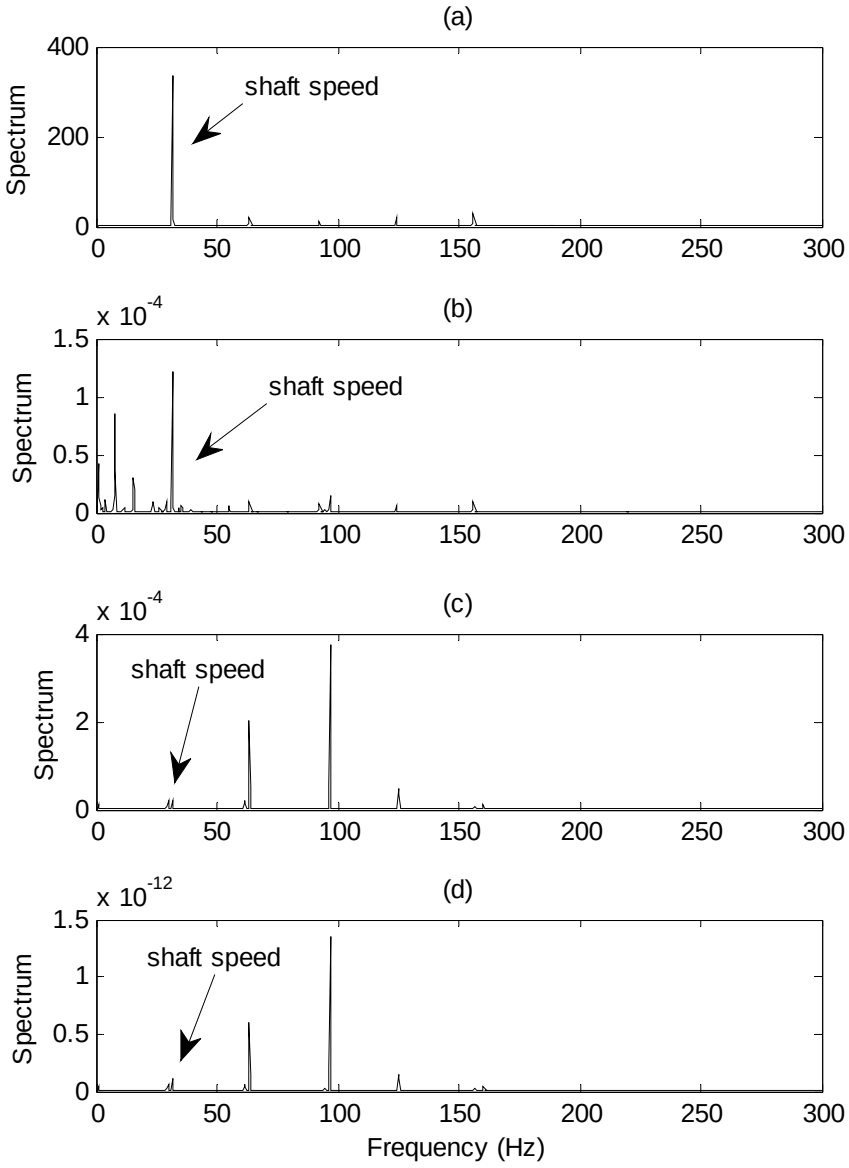

Figure 2.10. Healthy bearing processing results: (a) using the WSA method (2.19); (b) using the max-envelope approach [35]; (c) using the one-scaled WT [34]; (d) using the envelope analysis [26].

3) Outer Race Fault Detection: If the bearing is damaged, in theory, the corresponding characteristic defect frequency and/or its harmonics will appear in its spectra. As stated earlier, outer race fault detection is relatively easier because the defect resonance modes do not change dramatically. Figure 2.11 shows the processing results in test Case I by using the related techniques, and Figure 2.12 shows the processing results corresponding to test Case II, where the difference in Case I and Case II is given in Table 1.1 in Chapter 1. It is seen that each technique can recognize the presence of this bearing fault. It is also seen that the spectral component related to shaft speed disappears in this case because its spectral magnitude is much lower than these defect frequency components.

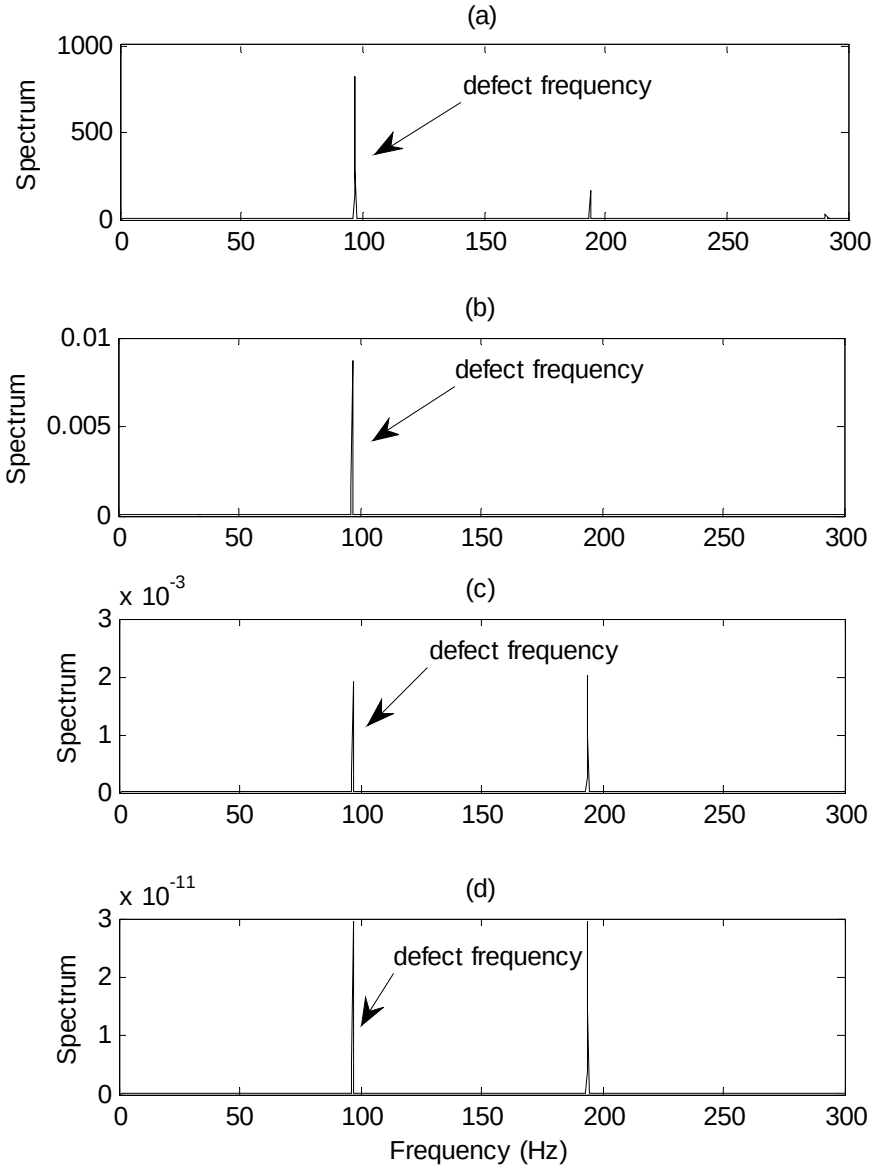

Figure 2.11. Outer race fault detection in Case I: (a) using the WSA method (2.19); (b) using the max-envelope approach [35]; (c) using the one-scaled WT [34]; (d) using the envelope analysis [26].

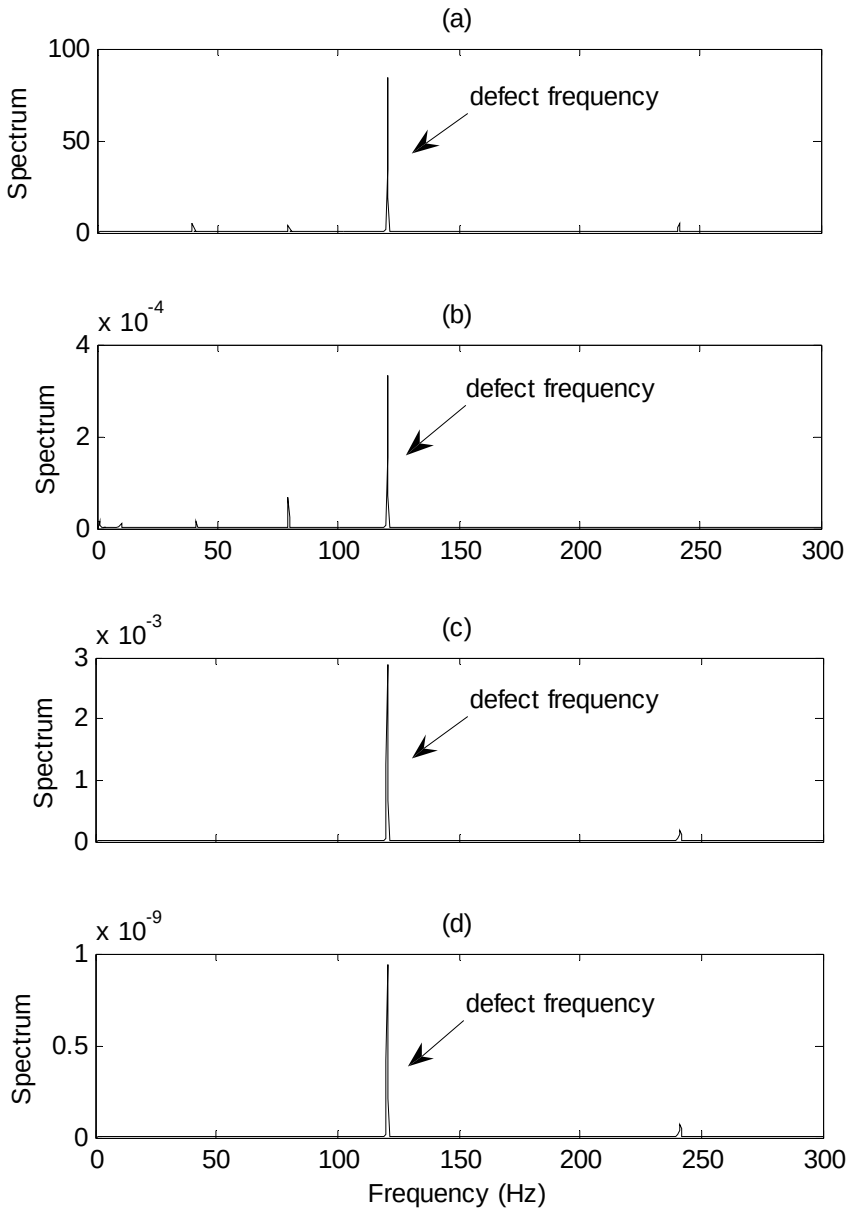

Figure 2.12. Outer race fault detection in Case II: (a) using the WSA method (2.19); (b) using the max-envelope approach [35]; (c) using the one-scaled WT [34]; (d) using the envelope analysis [26].

4) Inner Race Fault Detection: As discussed in Section 2.2, the detection of a fault on a rotating ring (bearing inner race in this work) and a rolling element is more challenging than on a fixed ring because the modes of the generated resonance signatures vary over time. Figures 2.13 and 2.14 show the processing results corresponding to the bearing signals with inner race defects. The characteristic frequencies f_{id} = 158.33 Hz in Case I and f_{id} = 197.91 Hz in Case II can be identified clearly in Figures 2.13(a) and 2.14(a), respectively, by using the WSA technique. However, when the max-envelope approach is applied, the defect frequency component becomes the third or fourth highest spectral component, as shown in Figures 2.13(b) and 2.14(b). This makes it difficult to detect this bearing defect based on the spectral analysis. Similar results can be seen in Figures 2.13(c), (d), and Figures 2.14(c), (d) by using the one-scaled WT and the envelope analysis, respectively. Unclear processing results may lead to false alarms or missed alarms in monitoring applications.

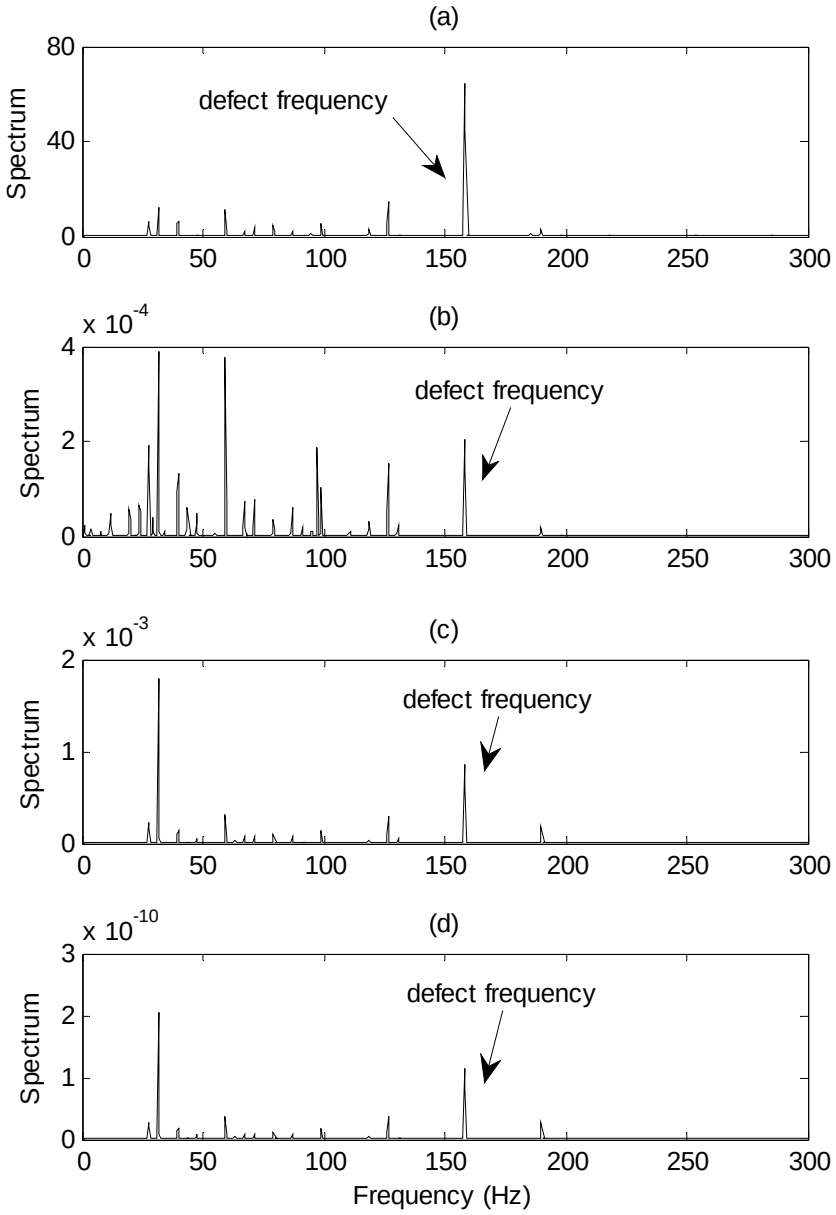

Figure 2.13. Inner race fault detection in Case I: (a) using the WSA method (2.19); (b) using the max-envelope approach [35]; (c) using the one-scaled WT [34]; (d) using the envelope analysis [26].

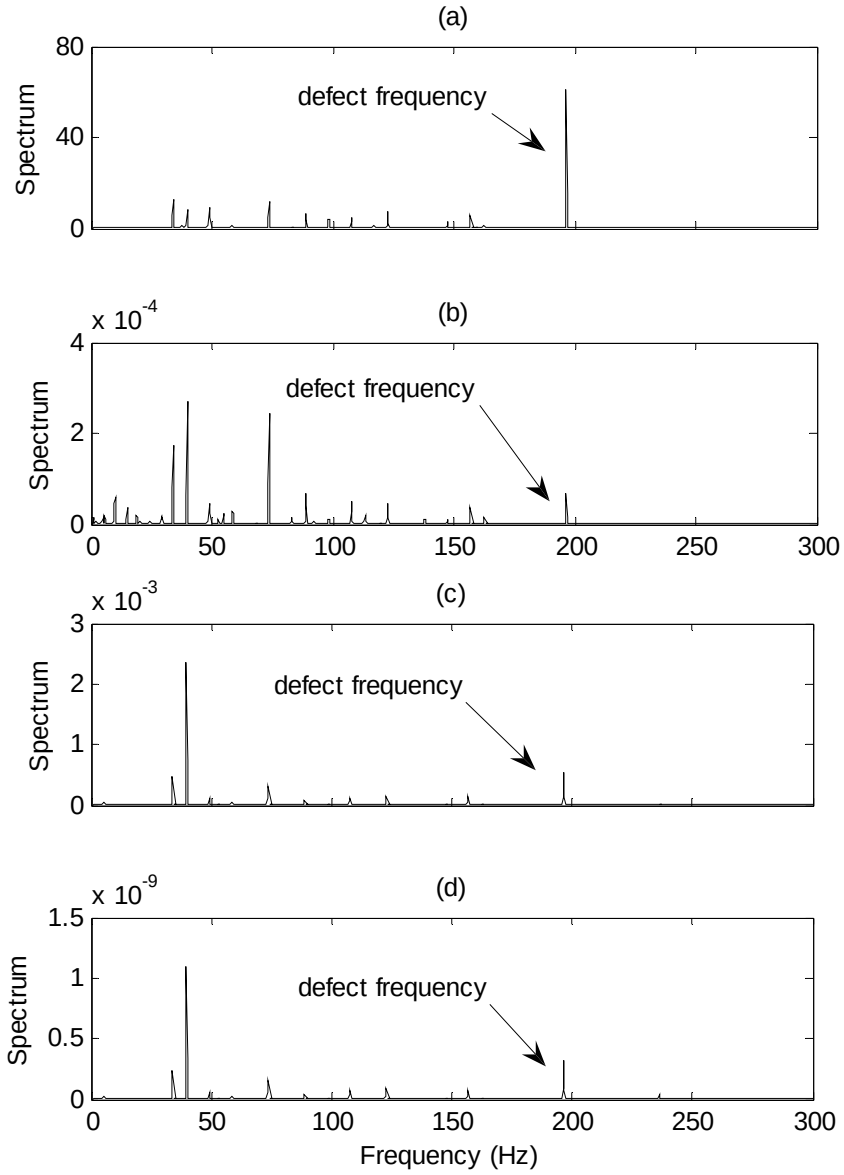

Figure 2.14. Inner race fault detection in Case II: (a) using the WSA method (2.19); (b) using the max-envelope approach [35]; (c) using the one-scaled WT [34]; (d) using the envelope analysis [26].

5) Rolling Element Fault Detection: Figures 2.15 and 2.16 illustrate two examples of the processing results corresponding to rolling element defects in test Case I and Case II, respectively. It is clear that the WSA technique is the only method that can recognize the existence of the bearing faults from weak resonance signatures (f_{rd} = 127.48 Hz in Case I; f_{rd} = 159.34 Hz in Case II). Other three methods have missed the bearing faults under these test conditions.

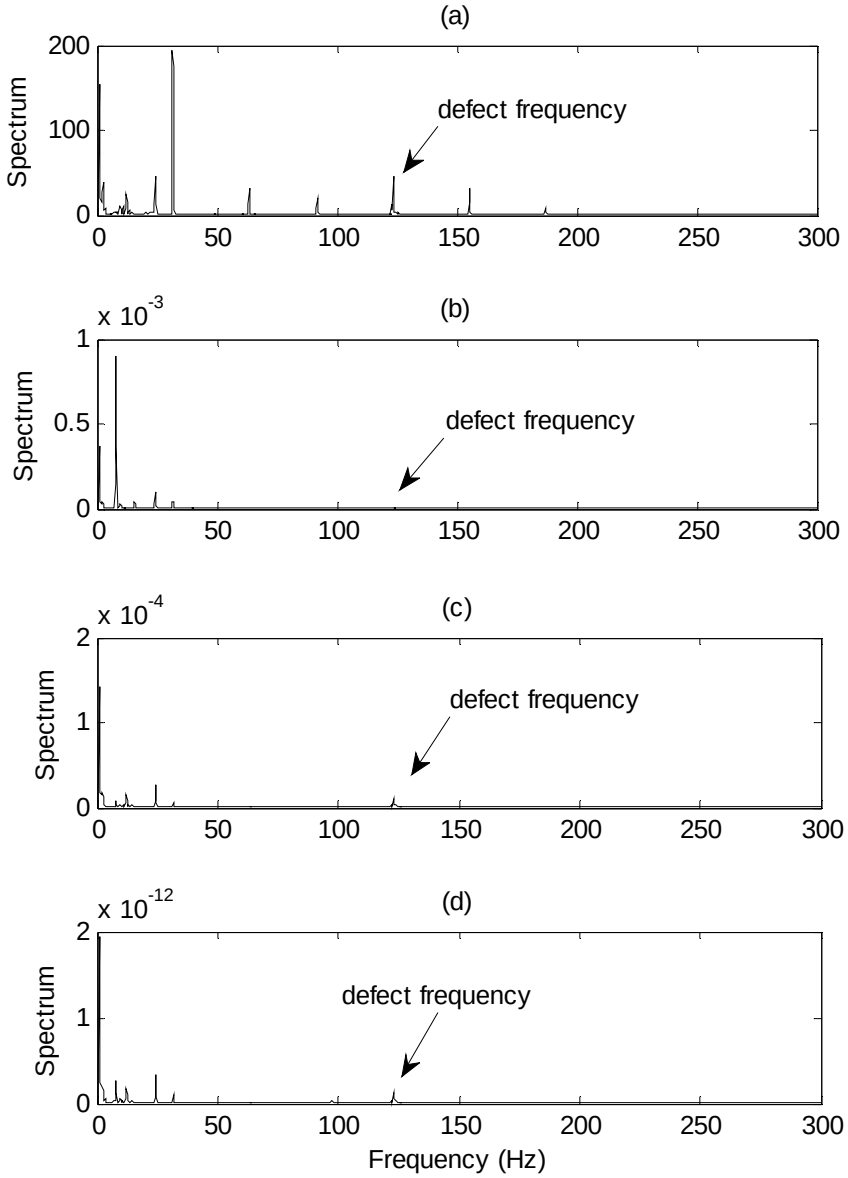

Figure 2.15. Rolling element fault detection in Case I: (a) using the WSA method (2.19); (b) using the max-envelope approach [35]; (c) using the one-scaled WT [34]; (d) using the envelope analysis [26].

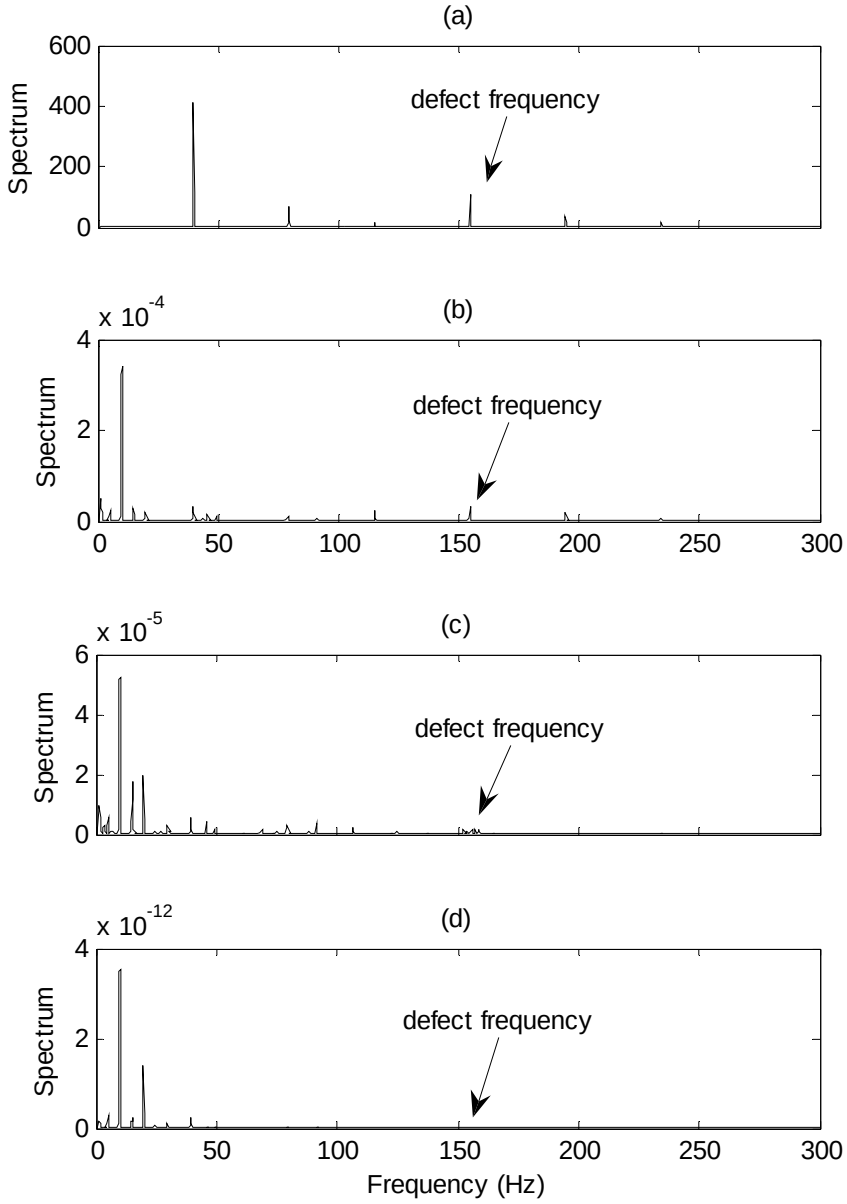

Figure 2.16. Rolling element fault detection in Case II: (a) using the WSA method (2.19); (b) using the max-envelope approach [35]; (c) using the one-scaled WT [34]; (d) using the envelope analysis [26].

2.3.3 Summary of Processing Results

What are shown in Section 2.3.2 are only some examples for illustration purpose. In this work, a total of 42 different test cases (three fault sizes, 7 shaft speeds, and 2 load levels, as given in Table 1.1) have been examined, to check the effectiveness of the proposed WSA technique in feature extraction and the effects of the load, shaft speed, and the fault size on its performance. In the test, the resulting spectrum magnitude $S^{c}_{m,d,l,s,j}(f)$ corresponding to each bearing characteristic frequency has been recorded, where *c* denotes the specific bearing health condition (*h*: healthy; *o*: outer-race defect; *i*: inner-race defect; and *r*: rolling element defect); *d* denotes the fault size (d_1: the smallest fault size in the tests; d_2: the medium fault size in the tests; d_3: the largest fault size in the tests); *m* denotes the *m*th method applied in this comparison study (m_1: the WSA technique by (2.19); m_2: the max-envelope approach [35]; m_3: the one-scaled WT [34]; and m_4: the envelope analysis [26]); *l* represents the load level (l_1: 1.2 N·m; l_2: 2.3 N·m); *s* represents the shaft speed (s_1: 600 r/min; s_2: 900 r/min; s_3: 1200 r/min; s_4:1500/min; s_5: 1800 r/min; s_6: 2100 r/min; s_7: 2400 r/min); *j* denotes the *j*th run under the same bearing operating condition (*j* = 1, 2, …, 5). Next, these processing results will be summarized into a series of tables and figures in the following way:

1) Finding the spectrum magnitude of the bearing characteristic frequencies. The spectrum corresponding to the bearing healthy (normal) condition ($S^{h}_{m,d,l,s,j}$), outer-race defect ($S^{o}_{m,d,l,s,j}$), inner-race defect ($S^{i}_{m,d,l,s,j}$) and rolling element defect ($S^{r}_{m,d,l,s,j}$) can be found from the resulting spectra,

$$S^{h}_{m,d,l,s,j}(f) = \max\left(\overline{\Phi}(f)\right), \quad f \in [f_t - f_{tol},\ f_t + f_{tol}] \cup [2f_t - f_{tol},\ 2f_t + f_{tol}] \tag{2.23}$$

$$S^{o}_{m,d,l,s,j}(f) = \max\left(\overline{\Phi}(f)\right), \quad f \in [f_{od} - f_{tol},\ f_{od} + f_{tol}] \cup [2f_{od} - f_{tol},\ 2f_{od} + f_{tol}] \tag{2.24}$$

$$S^{i}_{m,d,l,s,j}(f) = \max\left(\overline{\Phi}(f)\right), \quad f \in [f_{id} - f_{tol},\ f_{id} + f_{tol}] \cup [2f_{id} - f_{tol},\ 2f_{id} + f_{tol}] \tag{2.25}$$

$$S^{r}_{m,d,l,s,j}(f) = \max\left(\overline{\Phi}(f)\right), \quad f \in [f_{rd} - f_{tol},\ f_{rd} + f_{tol}] \cup [2f_{rd} - f_{tol},\ 2f_{rd} + f_{tol}] \tag{2.26}$$

where f_t is the shaft speed; f_{rd}, f_{id}, f_{od} are the theoretical characteristic frequencies of a bearing calculated by (2.20), (2.21), and (2.22), respectively; f_{tol} is the frequency tolerance

which depends on the spectral resolution Δf and shaft speed variation, $f_{tol} = f_t \times 4\% + \Delta f$ is used in this work.

2) Obtaining the normalized spectrum magnitude $\overline{S}^{c}_{m,d,l,s,j}$ corresponding to each bearing characteristic frequency,

$$\overline{S}^{h}_{m,d,l,s,j}(f) = \frac{S^{h}_{m,d,l,s,j}(f)}{\max(S^{h}_{m,d,l,s,j}(f),\, S^{o}_{m,d,l,s,j}(f),\, S^{i}_{m,d,l,s,j}(f),\, S^{r}_{m,d,l,s,j}(f))} \tag{2.27}$$

$$\overline{S}^{o}_{m,d,l,s,j}(f) = \frac{S^{o}_{m,d,l,s,j}(f)}{\max(S^{h}_{m,d,l,s,j}(f),\, S^{o}_{m,d,l,s,j}(f),\, S^{i}_{m,d,l,s,j}(f),\, S^{r}_{m,d,l,s,j}(f))} \tag{2.28}$$

$$\overline{S}^{i}_{m,d,l,s,j}(f) = \frac{S^{i}_{m,d,l,s,j}(f)}{\max(S^{h}_{m,d,l,s,j}(f),\, S^{o}_{m,d,l,s,j}(f),\, S^{i}_{m,d,l,s,j}(f),\, S^{r}_{m,d,l,s,j}(f))} \tag{2.29}$$

$$\overline{S}^{r}_{m,d,l,s,j}(f) = \frac{S^{r}_{m,d,l,s,j}(f)}{\max(S^{h}_{m,d,l,s,j}(f),\, S^{o}_{m,d,l,s,j}(f),\, S^{i}_{m,d,l,s,j}(f),\, S^{r}_{m,d,l,s,j}(f))} \tag{2.30}$$

3) Analyzing these normalized spectrum magnitudes $\overline{S}^{c}_{m,d,l,s,j}(f)$ by calculating the mean and standard deviation of each data group of interest. For example, to investigate the effectiveness of the proposed WSA technique with respect to other classical approaches, the following statistical indices can be examined,

$$\mu^{c}_{m}(f) = \underset{d,l,s,j}{\text{mean}}\, \overline{S}^{c}_{m,d,l,s,j}(f) \tag{2.31}$$

$$\sigma^{c}_{m}(f) = \underset{d,l,s,j}{\text{stddev}}\, \overline{S}^{c}_{m,d,l,s,j}(f) \tag{2.32}$$

To investigate the effect of the fault size on the performance of a specific method in feature extraction, the following statistical indices can be examined,

$$\mu^{c}_{m,d}(f) = \underset{l,s,j}{\text{mean}}\, \overline{S}^{c}_{m,d,l,s,j}(f) \tag{2.33}$$

$$\sigma^{c}_{m,d}(f) = \underset{l,s,j}{\text{stddev}}\, \overline{S}^{c}_{m,d,l,s,j}(f) \tag{2.34}$$

To investigate the effect of the shaft speed on the performance of a specific method in feature extraction, the following statistical indices can be examined,

$$\mu_{m,s}^{c}(f) = \underset{d,l,j}{\text{mean}}\, \bar{S}_{m,d,l,s,j}^{c}(f) \tag{2.35}$$

$$\sigma_{m,s}^{c}(f) = \underset{d,l,j}{\text{stddev}}\, \bar{S}_{m,d,l,s,j}^{c}(f) \tag{2.36}$$

To investigate the effect of the load on the performance of a specific method in feature extraction, the following statistical indices can be examined,

$$\mu_{m,l}^{c}(f) = \underset{d,s,j}{\text{mean}}\, \bar{S}_{m,d,l,s,j}^{c}(f) \tag{2.37}$$

$$\sigma_{m,l}^{c}(f) = \underset{d,s,j}{\text{stddev}}\, \bar{S}_{m,d,l,s,j}^{c}(f) \tag{2.38}$$

The processing results from all the test cases are summarized in Tables 2.1 – 2.9 and Figures 2.17 – 2.20. From Tables 2.1 – 2.3, it is observed that the proposed WSA technique (m_1) performs better than other three classical approaches (m_2: the max-envelope approach [35]; m_3: the one-scaled WT [34]; and m_4: the envelope analysis [26]) in feature extraction. Especially for inner-race fault diagnosis, the WSA technique can detect this bearing fault in all the tested cases. From Table 2.4, it can be seen that each fault detection technique can recognize the outer-race faults in the tested bearings. Correspondingly, Figure 2.17 plots the values of $\mu_m^c(f)$ given by (2.31) with respect to each bearing fault detection approach.

Table 2.1. Testing results of healthy bearings by (2.31) and (2.32) using four approaches

	m_1	m_2	m_3	m_4
$\mu_m^h(f)$	0.9605	0.7879	0.6123	0.6490
$\sigma_m^h(f)$	0.1670	0.3194	0.3836	0.3701
$\mu_m^h(f)/\sigma_m^h(f)$	5.7500	2.4666	1.5963	1.7538

Table 2.2. Testing results of rolling-element defect bearings by (2.31) and (2.32) using four approaches

	m_1	m_2	m_3	m_4
$\mu_m^r(f)$	0.8153	0.2321	0.7721	0.7126
$\sigma_m^r(f)$	0.1867	0.2206	0.3098	0.3490
$\mu_m^r(f)/\sigma_m^r(f)$	4.3676	1.0522	2.4923	2.0415

Table 2.3. Testing results of inner-race defect bearings by (2.31) and (2.32) using four approaches

	m_1	m_2	m_3	m_4
$\mu_m^i(f)$	1.0000	0.0597	0.6532	0.7052
$\sigma_m^i(f)$	0.0000	0.0700	0.2835	0.2682
$\mu_m^i(f)/\sigma_m^i(f)$	INF	0.8516	2.3037	2.6293

INF: infinite number

Table 2.4. Testing results of outer-race defect bearings by (2.31) and (2.32) using four approaches

	m_1	m_2	m_3	m_4
$\mu_m^o(f)$	1.0000	1.0000	1.0000	1.0000
$\sigma_m^o(f)$	0.0000	0.0000	0.0000	0.0000
$\mu_m^o(f)/\sigma_m^o(f)$	INF	INF	INF	INF

INF: infinite number

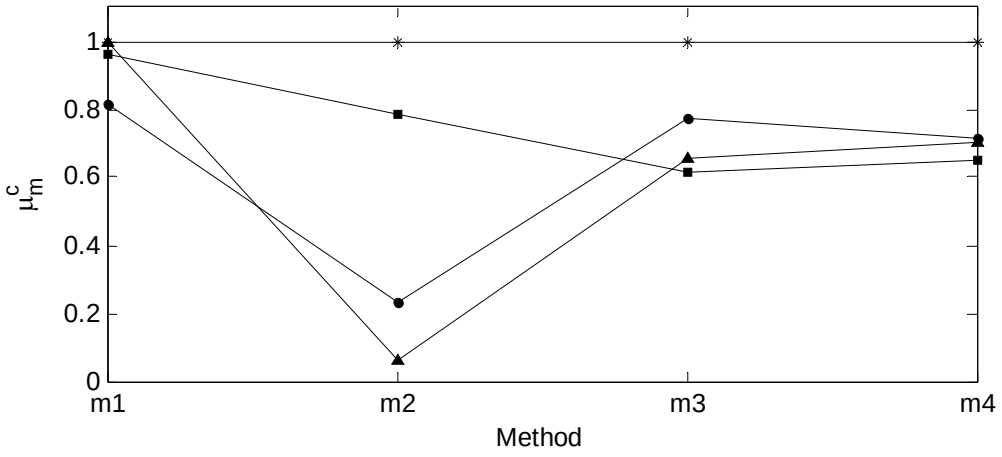

Figure 2.17. Effects of different fault detection approaches on the derived $\mu_m^c(f)$: square, healthy bearing; circle, rolling-element defect bearing; triangle, inner-race defect bearing; star, outer-race defect bearing.

The WSA technique can detect the inner-race bearing faults and outer-race bearing faults in all the tested cases; but can not achieve this high reliability when the bearing is healthy or rolling-element defective. To further investigate how the fault size, shaft speed, and the load affect the performance of the WSA technique in feature extraction, the processing results from these 42 test cases are summarized into several groups of interest, as given in Tables 2.5 – 2.9. Correspondingly, Figures 2.18 – 2.20 show the values of $\mu_d^c(f)$, $\mu_s^c(f)$, $\mu_l^c(f)$ for each interested group.

From the processing results in Tables 2.5 and 2.6, it is observed that the fault size affects the effectiveness of the WSA technique in the condition monitoring of healthy bearings and rolling-element defect bearings. As the fault size increases, it becomes easier for the WSA technique to recognize the current bearing health condition.

Table 2.5. Testing results of healthy bearings by (2.33) and (2.34) corresponding to three different fault sizes when the WSA technique is applied

	d_1	d_2	d_3
$\mu_d^h(f)$	0.9122	0.9782	0.9911
$\sigma_d^h(f)$	0.2606	0.0978	0.0615
$\mu_d^h(f)/\sigma_d^h(f)$	3.5010	9.9982	16.1181

Table 2.6. Testing results of rolling-element defect bearings by (2.33) and (2.34) corresponding to three different fault sizes when the WSA technique is applied

	d_1	d_2	d_3
$\mu_d^r(f)$	0.8089	0.8113	0.8257
$\sigma_d^r(f)$	0.1885	0.1828	0.1920
$\mu_d^r(f)/\sigma_d^r(f)$	4.2914	4.4362	4.2996

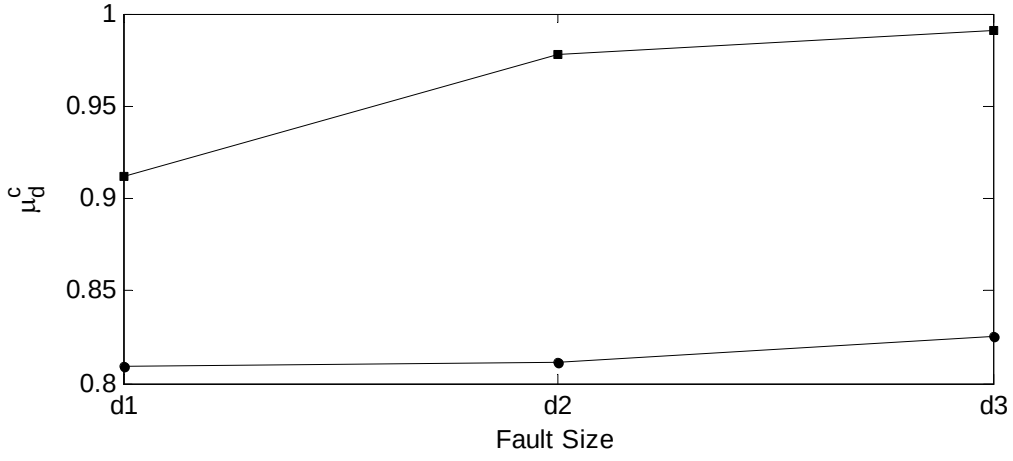

Figure 2.18. Effects of different fault sizes on the derived $\mu_d^c(f)$ when the proposed WSA technique (2.19) is applied: square, healthy bearing; circle, rolling-element defect bearing.

From the processing results in Table 2.7 and Figure 2.19, it can be seen that for a healthy bearing, the bearing condition can be recognized when the shaft speed is over 15 Hz (s_2); however, for the inner-race defect bearings, no uniform patterns can be observed from Table 2.8 and Figure 2.19. This is mainly because that for ball bearings, the rolling-element fault interacts with its mating components randomly.

Table 2.7. Testing results of healthy bearings by (2.35) and (2.36) corresponding to seven different shaft speeds when the WSA technique is applied

	s_1	s_2	s_3	s_4	s_5	s_6	s_7
$\mu_s^h(f)$	0.6840	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\sigma_s^h(f)$	0.3771	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\mu_s^h(f)/\sigma_s^h(f)$	1.8138	INF	INF	INF	INF	INF	INF

INF: infinite number

Table 2.8. Testing results of rolling-element defect bearings by (2.35) and (2.36) corresponding to seven different shaft speeds when the WSA technique is applied

	s_1	s_2	s_3	s_4	s_5	s_6	s_7
$\mu_s^r(f)$	0.8563	0.7826	0.8816	0.8149	0.7470	0.7977	0.8221
$\sigma_s^r(f)$	0.3786	0.0327	0.0878	0.1165	0.0664	0.1362	0.2614
$\mu_s^r(f)/\sigma_s^r(f)$	2.2615	23.8994	10.0319	6.9934	11.2435	5.8540	3.1451

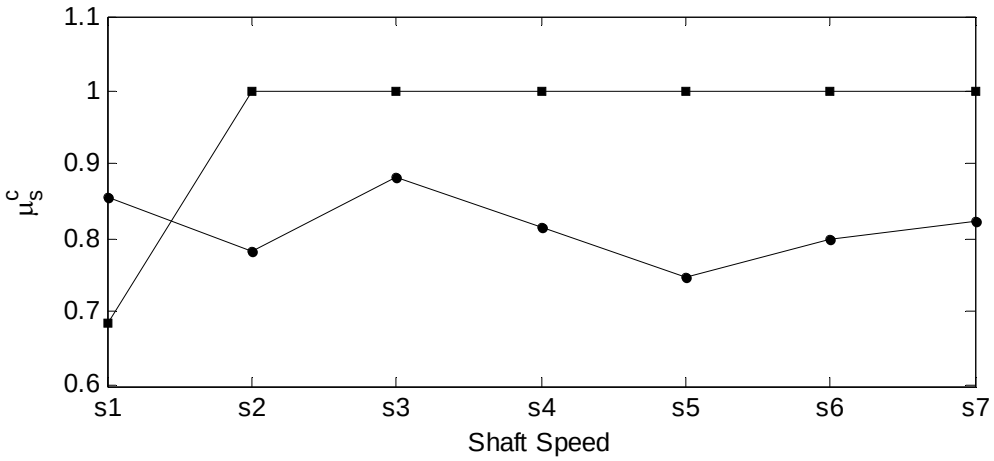

Figure 2.19. Effects of different shaft speeds on the derived $\mu_s^c(f)$ when the proposed WSA technique (2.19) is applied: square, healthy bearing; circle, rolling-element defect bearing.

In the tests, two levels of load are applied. Table 2.9 and Figure 2.20 show the processing results of $\mu_l^c(f)$ with respect to each load level. It can be seen that the load affects the performance of the WSA technique. A heavier load generates a vibration signature of high amplitude and the related spectral component will be more obvious.

Table 2.9. Testing results of healthy and rolling-element defect bearings by (2.37) and (2.38) corresponding to two different loads when the WSA technique is applied

	l_1	l_2		l_1	l_2
$\mu_l^h(f)$	0.9514	0.9696	$\mu_l^r(f)$	0.7869	0.8436
$\sigma_l^h(f)$	0.1812	0.1524	$\sigma_l^r(f)$	0.2065	0.1609
$\mu_l^h(f)/\sigma_l^h(f)$	5.2522	6.3629	$\mu_l^r(f)/\sigma_l^r(f)$	3.8105	5.2421

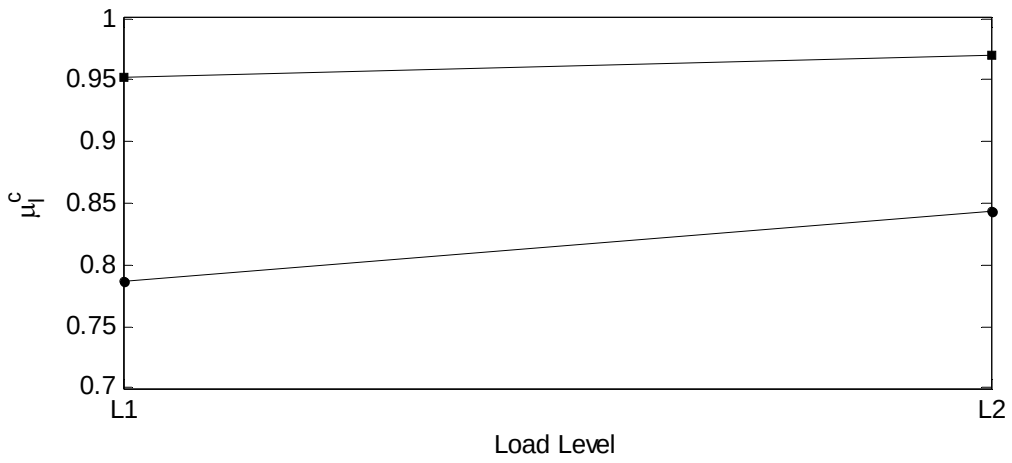

Figure 2.20. Effects of different loads on the derived $\mu_l^c(f)$ when the proposed WSA technique (2.19) is applied: square, healthy bearing; circle, rolling-element defect bearing.

2.4 Summary

A new signal processing technique, the wavelet spectrum analysis, is proposed in this chapter for bearing incipient fault-related feature extraction. This technique applies the WT to demodulate the resonance signatures over selected frequency bands to extract the representative features. A strategy is suggested for the deployment of the wavelet centre frequencies. A weighted Shannon function is proposed to synthesize the wavelet coefficient functions to enhance feature characteristics, whereas the applied weights are from a statistical index that quantifies the effect of different wavelet centre frequencies on feature extraction. An averaged autocorrelation spectrum is adopted to highlight the feature characteristics that are related to bearing health conditions. The performance of this proposed technique is examined by a series of experimental tests corresponding to different bearing conditions. The test results show that this new signal processing technique is an effective bearing fault detection method, which is particularly useful for non-stationary feature extraction and analysis.

The presented wavelet spectrum analysis technique in this chapter is to achieve the first research objective in Section 1.3: *1) to develop a more effective signal processing technique for feature extraction in bearing fault detection.*

Chapter 3
A Novel Neural Fuzzy Classifier

In this chapter, a novel neural fuzzy (NF) classifier is developed to effectively integrate the strengths of the selected signal processing techniques (or the resulting representative features) for a more accurate assessment of the bearing health conditions.

3.1 Overview

Several signal processing techniques have been reported in the literature for bearing fault detection [14, 17, 22-25]. Each technique, however, has its own advantages and limitations, and cannot be used independently for bearing applications [2, 37]. Although the proposed wavelet spectrum analysis (WSA) technique in Chapter 2 has demonstrated its superiority over other related classical techniques in enhancing feature characteristics, it is still challenging to reliably recognize some bearing health conditions such as healthy bearing condition and rolling-element defect bearing condition, based on dominant characteristic frequency analysis. If many techniques (or features) are applied for fault diagnosis, however, it is more possible to result in a conflicting result due to their limitations in robustness. Furthermore, most currently available fault diagnosis systems lack the adaptive capability to accommodate various bearing conditions in terms of load and speed variations. To approach these challenges, a NF classifier is developed in this chapter to integrate the strengths of the selected signal processing techniques for a more reliable bearing fault diagnosis.

3.2 Monitoring Indices

Two signal processing techniques are adopted in this bearing fault diagnosis: one is the WSA technique that has been presented in Chapter 2, and the other is the kurtosis ratio method [114] that will be discussed next.

3.2.1 Kurtosis Ratio Method

Statistical indicators have been widely investigated in machinery condition monitoring [15, 18, 103-106, 115, 116]. Recent studies have showed that the kurtosis ratio with a proper signal preprocessing can provide a compelling diagnosis of incipient bearing faults; this method is defined as the ratio of a standard kurtosis to a robust estimate of the kurtosis [114],

$$Kur(x) = \frac{E\{(x-\mu_x)^4\}}{\sigma_x^4}, \tag{3.1}$$

$$KR(x) = \frac{Kur(x)}{Kur(x_t)}, \tag{3.2}$$

where x is the high-pass filtered signal; x_t is the trimmed signal from x by removing the outlying samples [114]; *Kur* and *KR* denote the kurtosis and kurtosis ratio, respectively. Fault related signatures are usually modulated by the signals at high resonance frequencies of the bearing and surrounding structures [26, 97], thus a 1000Hz ($Nf_t = 35 f_t \approx 1000$ Hz) cut-off frequency is selected in this work to eliminate the effects of low-frequency noisy components. The monitoring indices, K_m and K_d, are employed in this application for bearing fault diagnosis, where K_m is the mean of the kurtosis ratios over five segments of the measured signals, and K_d is the corresponding standard derivation. Figure 3.1 shows an example of the processing results for both K_m and K_d corresponding to different bearing conditions. It is seen that the damaged bearings usually lead to a higher K_m than the healthy ones, and this phenomenon is even more obvious for the inner-race damaged bearings. It is also seen that the values of K_d from the healthy bearings and the bearings with inner-race or outer-race defects are usually small, whereas the bearings with rolling-element defects tend to have a higher K_d.

3.2.2 Monitoring Indices

The monitoring indices derived from the WSA technique are the normalized spectrum magnitudes $\bar{S}^c_{m,d,l,s,j}$ given by (2.27) – (2.30) in Section 2.33, whereas the monitoring indices derived from the kurtosis ratio method are K_m and K_d. These monitoring indices will be regarded as the inputs to the NF classifier to be developed,

$$x_1 = \bar{S}^h_{m,d,l,s,j}(f), \quad x_2 = \bar{S}^r_{m,d,l,s,j}(f), \quad x_3 = \bar{S}^i_{m,d,l,s,j}(f),$$

$$x_4 = \bar{S}^o_{m,d,l,s,j}(f), \quad x_5 = K_m, \qquad x_6 = K_d. \tag{3.3}$$

In bearing fault diagnoses, each of the first four indices, x_1 to x_4, will cooperate with the last two indices, x_5 to x_6, to examine a bearing health condition. The implementation procedure of this classifier is discussed in the following section.

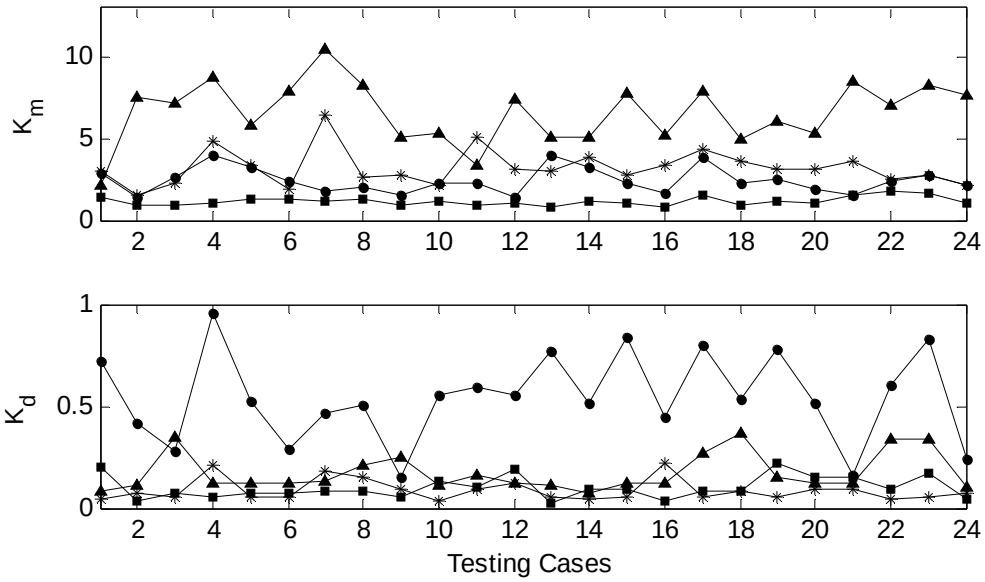

Figure 3.1. Reference functions K_m and K_d: square, healthy bearing; circle, bearing with rolling-element defects; triangle, bearing with inner-race defects; star, bearing with outer-race defects.

3.3 The Developed NF Classifier

3.3.1 Structure of NF Classifier

The proposed NF classier is to combine the features from the two selected techniques (with six monitoring indices as discussed in Section 3.2) to yield a more accurate assessment of the bearing health conditions. The diagnostic classification reasoning is conducted by fuzzy logic whereas the fuzzy system parameters are optimized by using the suggested training algorithm.

The conditions of a bearing are classified into two categories: *healthy* (C_1) or *damaged* (C_2). The damage may occur on a rolling element (C_{21}), inner race (C_{22}), or outer race (C_{23}). Taking the monitoring indices x_1 to x_6 as the input vector, the diagnostic classification, in terms of the diagnostic indicator *y*, can be formulated in the form of

$$\Re_j: \text{IF } (x_1 \text{ is } A_1^j) \text{ AND } (x_2 \text{ is } A_2^j) \text{ AND } \ldots \text{ AND } (x_6 \text{ is } A_6^j) \text{ THEN } (y \subset C_j \text{ with } w_j) \quad (3.4)$$

where A_i^j is the membership function (MF); i = 1, 2, …, 6, and j = 1, 2, . . ., *J*; *J* denotes the number of the rules; w_j is the weight factor to represent the contribution of the feature association in $\Re_j$ to the diagnostic operation; C_j is one of states C_1 and $C_2 = \{C_{21} \quad C_{22} \quad C_{23}\}$, depending on a specific pattern association. The number of the rules depends on classification operations and expert diagnostic knowledge, for example:

1) Based on the analysis in Chapter 2, it is found that outer-race fault detection is relatively easier than other types of bearing fault detection because the related resonance modes are stationary (e.g., Figures 2.11 and 2.12, Table 2.4). Accordingly, one MF is assigned to the input variable x_4.

2) The fault detection on a rotating ring or a rolling-element is more challenging because the generated resonance signatures are time-varying. However, the test results in Chapter 2 show that the proposed WSA technique performs better than other related techniques, and can also recognize the existence of an inner-ring defect due to its feature enhancement effect (e.g., Figures 2.13 – 2.16, Tables 2.2 and 2.3). Thus, two MFs are designated to the input variable x_2 whereas one MF is for x_3.

3) Whenever a defect occurs on a rolling element (e.g., a ball), the pitted position may or may not strike or be struck by a ring, depending on the ball orientation. The resulting impact resonance modes become random in nature [13]. Advanced studies [114] have revealed that a defect on a rolling element usually leads to a higher K_d, as illustrated in Figure 3.1. Accordingly, three MFs are assigned to the input variable x_6.

4) The most important task is to detect if the bearing of interest is healthy or faulty, no matter which component is damaged. The investigation results in Chapter 2 have showed that when the bearing is healthy, the WSA technique usually detects the shaft speed as the dominant spectral component whereas other spectral components are significantly suppressed; and also the kurtosis ratio K_m is at a low level (close to 1). However, when an incipient inner-race or outer-race fault occurs in bearings, the characteristic defect frequency will become dominant in the spectra (e.g., Figures 2.13 – 2.16, Tables 2.2 and 2.3), and K_m will correspondingly increase. To this end, two MFs are designated to the input variables x_1 whereas three MFs are for x_5.

Based on the above observations, the fuzzy IF-THEN rules are formulated and listed in Table 3.1, in which S_i, M_i and L_i represent, respectively, the *Small*, *Medium* and *Large* MFs for the *i*th input, $i = 1, 2, \ldots, 6$.

Table 3.1. Fuzzy IF-THEN rules in the developed NF classifier

$\Re_1$: IF (x_1 is S_1) AND (x_5 is S_5) AND (x_6 is S_6)	THEN ($y \subset C_1$ with w_1)
$\Re_2$: IF (x_1 is S_1) AND (x_5 is S_5) AND (x_6 is M_6)	THEN ($y \subset C_1$ with w_2)
$\Re_3$: IF (x_1 is S_1) AND (x_5 is M_5) AND (x_6 is S_6)	THEN ($y \subset C_1$ with w_3)
$\Re_4$: IF (x_1 is S_1) AND (x_5 is M_5) AND (x_6 is M_6)	THEN ($y \subset C_1$ with w_4)
$\Re_5$: IF (x_1 is L_1) AND (x_5 is S_5) AND (x_6 is S_6)	THEN ($y \subset C_1$ with w_5)
$\Re_6$: IF (x_1 is L_1) AND (x_5 is S_5) AND (x_6 is M_6)	THEN ($y \subset C_1$ with w_6)
$\Re_7$: IF (x_1 is L_1) AND (x_5 is M_5) AND (x_6 is S_6)	THEN ($y \subset C_1$ with w_7)
$\Re_8$: IF (x_1 is L_1) AND (x_5 is M_5) AND (x_6 is M_6)	THEN ($y \subset C_1$ with w_8)
$\Re_9$: IF (x_2 is S_2) AND (x_5 is M_5) AND (x_6 is M_6)	THEN ($y \subset C_{21}$ with w_9)
$\Re_{10}$: IF (x_2 is S_2) AND (x_5 is M_5) AND (x_6 is L_6)	THEN ($y \subset C_{21}$ with w_{10})
$\Re_{11}$: IF (x_2 is S_2) AND (x_5 is L_5) AND (x_6 is M_6)	THEN ($y \subset C_{21}$ with w_{11})
$\Re_{12}$: IF (x_2 is S_2) AND (x_5 is L_5) AND (x_6 is L_6)	THEN ($y \subset C_{21}$ with w_{12})
$\Re_{13}$: IF (x_2 is L_2) AND (x_5 is M_5) AND (x_6 is M_6)	THEN ($y \subset C_{21}$ with w_{13})
$\Re_{14}$: IF (x_2 is L_2) AND (x_5 is M_5) AND (x_6 is L_6)	THEN ($y \subset C_{21}$ with w_{14})
$\Re_{15}$: IF (x_2 is L_2) AND (x_5 is L_5) AND (x_6 is M_6)	THEN ($y \subset C_{21}$ with w_{15})
$\Re_{16}$: IF (x_2 is L_2) AND (x_5 is L_5) AND (x_6 is L_6)	THEN ($y \subset C_{21}$ with w_{16})
$\Re_{17}$: IF (x_3 is L_3) AND (x_5 is M_5) AND (x_6 is S_6)	THEN ($y \subset C_{22}$ with w_{17})
$\Re_{18}$: IF (x_3 is L_3) AND (x_5 is M_5) AND (x_6 is M_6)	THEN ($y \subset C_{22}$ with w_{18})
$\Re_{19}$: IF (x_3 is L_3) AND (x_5 is L_5) AND (x_6 is S_6)	THEN ($y \subset C_{22}$ with w_{19})
$\Re_{20}$: IF (x_3 is L_3) AND (x_5 is L_5) AND (x_6 is M_6)	THEN ($y \subset C_{22}$ with w_{20})
$\Re_{21}$: IF (x_3 is L_4) AND (x_5 is M_5) AND (x_6 is S_6)	THEN ($y \subset C_{23}$ with w_{21})
$\Re_{22}$: IF (x_3 is L_4) AND (x_5 is M_5) AND (x_6 is M_6)	THEN ($y \subset C_{23}$ with w_{22})
$\Re_{23}$: IF (x_3 is L_4) AND (x_5 is L_5) AND (x_6 is S_6)	THEN ($y \subset C_{23}$ with w_{23})
$\Re_{24}$: IF (x_3 is L_4) AND (x_5 is L_5) and (x_6 is M_6)	THEN ($y \subset C_{23}$ with w_{24})

The network architecture of this proposed NF classifier is schematically shown in Figure 3.2. Unless specified, all the network links have unity weights. The input nodes in layer 1 transmit the monitoring indices $\{x_1, \ldots, x_6\}$ to the next layer directly. Each node in layer 2 acts as an MF. The nodes in layer 3 perform the fuzzy *T*-norm operations. If a product operator is used, the firing strength of the rule $\Re_j$ is

$$\mu_j = \mu_{A_i^j}(x_i)\mu_{A_5^j}(x_5)\mu_{A_6^j}(x_6). \tag{3.5}$$

where $\mu_{A_i^j}(x_i)$, i = 1, …, 6, denotes the MF grade.

All the nodes in layer 3 form the rule base as listed in Table 3.1. The nodes in layer 4 and 5 perform the defuzzification operations: the belongingness grades to C_1, C_{21}, C_{22}, C_{23} are, respectively, given as

$$y_1 = \frac{\sum_{j=1}^{8}\mu_j w_j}{\sum_{j=1}^{8} w_j} = \sum_{j=1}^{8}\mu_j \overline{w}_j\ , \tag{3.6}$$

$$y_{21} = \frac{\sum_{j=9}^{16}\mu_j w_j}{\sum_{j=9}^{16} w_j} = \sum_{j=9}^{16}\mu_j \overline{w}_j\ , \tag{3.7}$$

$$y_{22} = \frac{\sum_{j=17}^{20}\mu_j w_j}{\sum_{j=17}^{20} w_j} = \sum_{j=17}^{20}\mu_j \overline{w}_j\ , \tag{3.8}$$

$$y_{23} = \frac{\sum_{j=21}^{24}\mu_j w_j}{\sum_{j=21}^{24} w_j} = \sum_{j=21}^{24}\mu_j \overline{w}_j\ , \tag{3.9}$$

The total belongingness grade to the damaged bearing condition C_2 is $y_2 = y_{21} + y_{22} + y_{23}$. Accordingly, the normalized classification indicator will be $y = y_1/(y_1 + y_2)$.

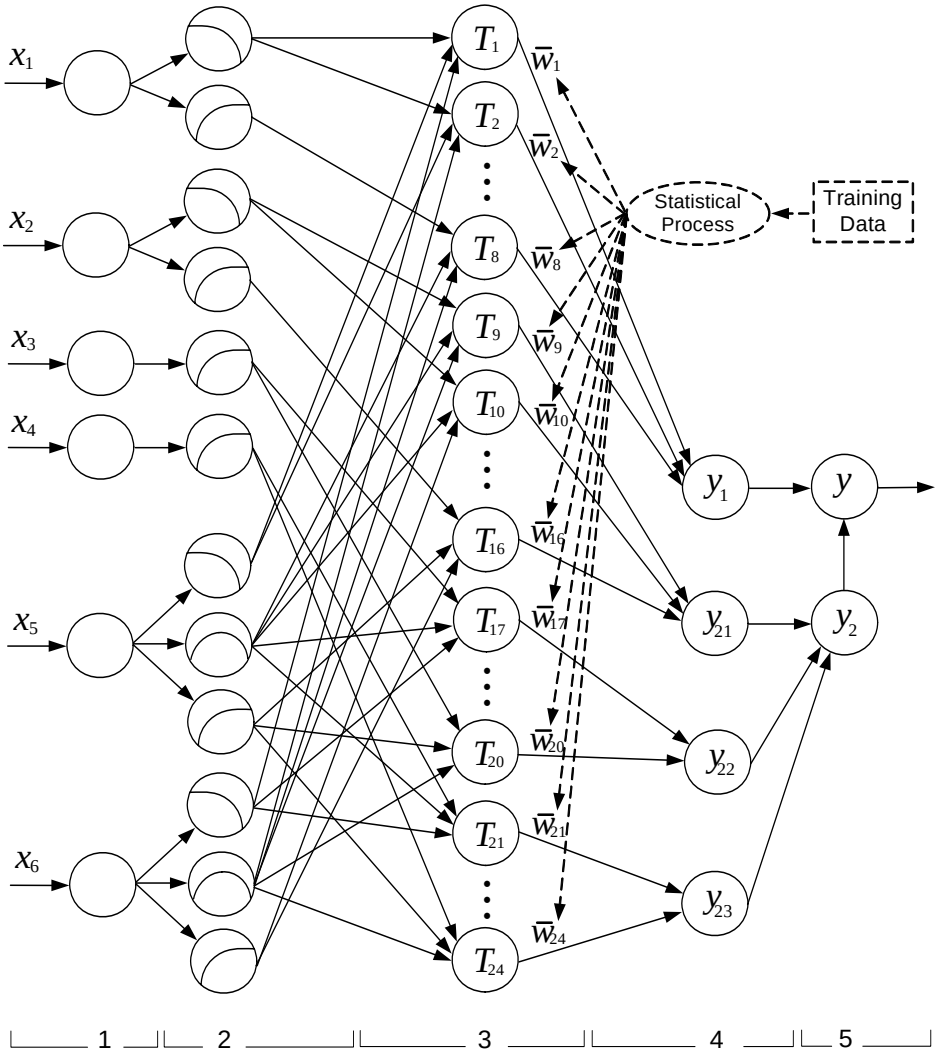

Figure 3.2. The network architecture of the developed NF classifier.

3.3.2 Training of NF Classifier

1) MF Parameter Training: The MF parameters of the NF classifier are optimized by using a recursive Levenberg Marquardt (LM) method. For a training data set, $\left\{x^{(m)} \quad y_d^{(m)}\right\}^T$, the input is $x^{(m)} = \left\{x_1^{(m)} \quad \cdots \quad x_6^{(m)}\right\}^T$; m = 1, 2, …, M; M is the total number of the training data sets; $y_d^{(m)}$ is the desired output in the form of

$$y_d^{(m)} = \begin{Bmatrix} y_{1d}^{(m)} & y_{21d}^{(m)} & y_{22d}^{(m)} & y_{23d}^{(m)} \end{Bmatrix}^T = \begin{cases} \{1 \quad 0 \quad 0 \quad 0\}^T, & if\ x^{(m)} \in C_1 \\ \{0 \quad 1 \quad 0 \quad 0\}^T, & if\ x^{(m)} \in C_{21} \\ \{0 \quad 0 \quad 1 \quad 0\}^T, & if\ x^{(m)} \in C_{22} \\ \{0 \quad 0 \quad 0 \quad 1\}^T, & if\ x^{(m)} \in C_{23} \end{cases}. \tag{3.10}$$

The objective function for all the *M* training data sets is defined as

$$E = \frac{1}{2}\sum_{m=1}^{M}\left[\left(y_1^{(m)} - y_{1d}^{(m)}\right)^2 + \left(y_{21}^{(m)} - y_{21d}^{(m)}\right)^2 + \left(y_{22}^{(m)} - y_{22d}^{(m)}\right)^2 + \left(y_{23}^{(m)} - y_{23d}^{(m)}\right)^2\right]. \tag{3.11}$$

where $y_1^{(m)}$, $y_{21}^{(m)}$, $y_{22}^{(m)}$, $y_{23}^{(m)}$ are the class belongingness indicators computed by (3.6) to (3.9), respectively. The sigmoid MFs, $u_{A_i^s}(x_i^{(m)}) = 1/(1+\exp[-a_i^s(x_i^{(m)} - b_i^s)])$ (i = 1, 2, …, 6; s = 1, 2), are utilized in this work, where $a_i^1 < 0$ and $a_i^2 > 0$ represent a *Small* and a *Large* MF, respectively. The Medium MFs for the inputs $x_5^{(m)}$ and $x_6^{(m)}$ are described by the Gaussian functions, $u_{A_i}(x_i^{(m)}) = \exp[(x_i^{(m)} - c_i)^2/(2\sigma_i^2)]$, where i = 5, 6. In training, the following constraints are applied to guarantee that the optimized NF classifier is linguistically interpretable, i.e., $b_i^1 \le c_i \le b_i^2$, where i = 5, 6.

In this work, the classical LM method is implemented recursively to improve the convergence properties of the NF classifier and to enhance its adaptive capability to accommodate the time-varying bearing conditions. The computation of the inverse Hessian matrix in the classical LM is time-consuming and impractical for real-time applications [90]. In this case, a remedy is to apply the matrix inversion lemma [117] to avoid the direct inversion of Hessian matrix. Instead of adding the $q \times q$ matrix $\mu(m)I$ at each step, only one diagonal element is added at a time. As a result, the nonlinear MF parameters of the NF classifier are recursively updated by

$$\hat{\Theta}(m) = \hat{\Theta}(m-1) - R(m-1)\varphi(m, \hat{\Theta}(m-1))^T \varepsilon(m, \hat{\Theta}(m-1)), \tag{3.12}$$

$$R(m)=\frac{1}{\lambda}\{R(m-1)-R(m-1)\varphi^{*}(m,\hat{\Theta}(m-1))^{T}[\lambda\Lambda^{*-1}(m) + \varphi^{*}(m,\hat{\Theta}(m-1))R(m-1)\varphi^{*}(m,\hat{\Theta}(m-1))^{T}]^{-1}\varphi^{*}(m,\hat{\Theta}(m-1))R(m-1)\} \tag{3.13}$$

where $0<\lambda\le 1$ is the forgetting factor; $\hat{\Theta}$ denote the set of the nonlinear parameters, $\hat{\Theta}\in R^{q\times 1}$; ε is the classification error vector, $\varepsilon\in R^{4\times 1}$; $\varphi=d\varepsilon/d\Theta$ is the Jacobian matrix; R is the inverse of an approximated Hessian matrix, and $R(0)$ is chosen as an identity matrix $\alpha_N I$ with a constant $\alpha_N\in[10^3 \quad 10^5]$; φ^{*} and Λ^{*} are, respectively, given by

$$\varphi^{*}(m,\hat{\Theta}(m-1))^{T}=\begin{bmatrix} & & 0\\ & & \vdots\\ & & 0\\ \varphi(m,\hat{\Theta}(m-1))^{T} & & 1\\ & & 0\\ & & \vdots\\ & & 0\end{bmatrix}\leftarrow \text{position } m \text{ (mod } q\text{)} + 1\,, \tag{3.14}$$

$$\Lambda^{*}(m)=\begin{bmatrix}1 & 0\\ 0 & q\mu(m)\end{bmatrix}. \tag{3.15}$$

The LM parameter $\mu(m)$ conversely affects the trust region radius [90], which should be adaptively modified based on the performance of the updated parameters. The following strategy is applied in this work: the initial value $\mu(1)$ is set to 50; if the objective function in (3.11) decreases as the parameters are updated, enlarge the $\mu(m)$ by $\kappa\mu(m)$; otherwise, if the objective function increases as the parameters are updated, reduce $\mu(m)$ by $\mu(m)/\kappa$, where κ is a design parameter (κ = 1.005 in this case). The detailed development of this recursive LM is found in Section 4.3.

2) Determination of Rule Weights: When multiple monitoring indices are employed for diagnostic classification, the contribution of each index association to the final decision varies, to a large degree, according to the situation under which the diagnostic decision is made. Such a

contribution can be represented by a weight factor for each rule [7]. Rule weights can be determined by several approaches such as choosing empirically [9], using information measures [62], or applying some training algorithms [119]. When the rule weights are updated by generally-used training methods, sometimes the results are unreasonable and difficult to interpret. For example, if a negative rule weight is obtained, this rule will never contribute to the classification output when a maximum inference operator is employed. In this work, the rule weight factor is determined by a statistical process, and then normalized within each classification category.

For a *Small* or a *Large* sigmoid MF, the critical point is selected as b_i^s (i = 1, 2, …, 6, and s = 1, 2) by which the membership grade is 0.5; whereas for a *Medium* Gaussian MF, the critical points are chosen as $c_i \pm \sqrt{2\ln 2}\sigma_i$. In order to determine if the current input vector satisfy a specific rule, each input is checked to see if it drops within the specified critical region, i.e., [0 b_i^1] for a *Small* MF, [b_i^2 $+\infty$] for a *Large* MF, and [$c_i - \sqrt{2\ln 2}\sigma_i$ $c_i + \sqrt{2\ln 2}\sigma_i$] for a *Medium* MF. The rule weight is determined from the conditional probability of the correct classification

$$w_i = \frac{\text{Prob}(\textit{correct classifications under this rule})}{\text{Prob}(\textit{all training cases that satisfy a specific rule})}. \tag{3.16}$$

For example, for the rule $\Re_{10}$ in Table 3.1, the diagnosis is based on the active monitoring indices, x_2, x_5 and x_6; thus, the rule weight, w_{10}, is calculated by

$$w_{10} = \frac{\text{Prob}[(x_2 < b_2^1) \cap (c_5 - \sqrt{2\ln 2}\sigma_5 < x_5 < c_5 + \sqrt{2\ln 2}\sigma_5) \cap (x_6 > b_6^2) \mid x \in C_1]}{\text{Prob}[(x_2 < b_2^1) \cap (c_5 - \sqrt{2\ln 2}\sigma_5 < x_5 < c_5 + \sqrt{2\ln 2}\sigma_5) \cap (x_6 > b_6^2) \mid x \in C_1 \cup C_2]} \tag{3.17}$$

The rule weights are then normalized within each classification category as in (3.6) to (3.9).

3.4 Performance Evaluation

The performance of the developed NF classifier is evaluated in a comparison with two related diagnostic schemes based on experimental tests.

3.4.1 Data Preparation

The experimental setup employed in this work is shown in Section 1.4. The tests are taken with the ball bearings of type MB ER-10K. Four bearing conditions are considered: healthy bearings (C_1), bearings with rolling element faults (C_{21}), bearings with inner race defects (C_{22}), and bearings with outer race defects (C_{21}). Three sets of faulty bearings with different fault sizes are prepared as in Table 1.1. Each bearing is tested under seven shaft speeds (600, 900, 1200, 1500, 1800, 2100, and 2400 r/min) and two load levels (1.2 and 2.3 $\mathrm{N}\cdot\mathrm{m}$). A total of 720 sets of data are collected under different operating conditions, which are listed in Table 3.2.

Table 3.2. Test conditions and collected data sets

Test Conditions	Total Data Sets	Training Data Sets	Checking Data Sets	Testing Data Sets
Healthy bearings	288	100	44	144
Bearings with rolling element defects	144	50	22	72
Bearings with inner-race defects	144	50	22	72
Bearing with outer-race defects	144	50	22	72

To properly train an NF scheme, sufficient representative training data sets, usually more than five times the number of the parameters to be updated, must be prepared for training [90]. The developed NF classifier has 48 parameters to be updated (24 fuzzy MF parameters and 24 rule weights). 250 data pairs are randomly chosen for training, whereas 110 pairs are randomly selected to check the validity of the updated models to prevent overfitting. The remaining data sets (360 pairs) are employed to test the resulting diagnostic scheme.

3.4.2 Performance Evaluation

The developed NF scheme (called *Scheme-1* in this case) is compared with two other related diagnostic classifiers to verify its effectiveness, as described below.

Scheme-2: A Pure Fuzzy Scheme with Unity Weights. The IF-THEN rules of this fuzzy scheme is the same as those in *Scheme-1*, which are listed in Table 3.1; the MF parameters in this scheme are valued based on the expertise; all the rule weights are set to be unity.

Scheme-3: An NF Scheme with Unity Weights. This scheme is the same as *Scheme-2* except that the MF parameters are initially valued based on the expertise and then trained by using the developed recursive LM method.

Scheme-1 and Scheme-3 are properly trained using the data sets as listed in Table 3.2. The diagnostic results are summarized in Table 3.3. It is seen that *Scheme-1* provides the best diagnostic results (with 93.2% reliability under these specific test conditions); this is because *Scheme-1* can optimize the fuzzy parameters and rule weights effectively, and possess the adaptive capability to accommodate different operating conditions. The worst performance is given by *Scheme-2*: with eleven false alarms and seven missed alarms, and the overall reliability is only 83.4%. This poor performance is mainly due to: 1) the lack of learning capability, such that the fuzzy MF parameters can not be optimized; and 2) a coarse fuzzy partition. Each input variable utilizes a few MFs compatible with *Scheme-1*. A further investigation shows that if a finer fuzzy partition is employed, the diagnostic performance of *Scheme-2* does improve to some extent; however, it still can not outperform other two schemes. The classification performance is improved from *Scheme-2* to *Scheme-3* when the MF parameters are optimized using the developed recursive LM method.

Table 3.3. Diagnostic testing results using different classification schemes

Diagnostic Scheme	False Alarms	Missed Alarms	Overall Accuracy
Scheme-1	4	3	93.2%
Scheme-2	11	7	83.4%
Scheme-3	5	6	88.5%

Considering the effects of different fuzzy partitions (rules) on classification, tuned rule weights,

instead of unity weights, can highlight the contribution from some specific rules. Table 3.4 illustrates the tuned rule weights in *Scheme-1* based on a statistical process in (3.20). It is seen that the rules $\mathfrak{R}_3$, $\mathfrak{R}_4$, $\mathfrak{R}_5$, $\mathfrak{R}_8$, $\mathfrak{R}_9$, $\mathfrak{R}_{11}$, $\mathfrak{R}_{13}$, $\mathfrak{R}_{15}$ contribute to the final decision less than the other rules in the rule base . Each signal processing technique (or the representative feature) has a limited capability in fault diagnostics. Even if the firing strengths of two fuzzy rules are identical, their diagnostic reliabilities may be different under different bearing test conditions. Therefore, rule weights play an important role in the diagnostic classification operations.

The diagnostic reliability of this proposed NF classifier can be further enhanced by incorporating the forecasting results from a multi-step predictor, which is discussed in Chapter 4 and 5.

Table 3.4. Rule weights before and after training

Rule Weights	w_1	w_2	w_3	w_4	w_5	w_6	w_5	w_8	w_9	w_{10}	w_{11}	w_{12}
Before Training	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
After Training	0.99	0.98	0.24	0.22	1.0	0.99	0.86	0.71	0.49	1.00	0.58	1.00
Rule Weights	w_{13}	w_{14}	w_{15}	w_{16}	w_{17}	w_{18}	w_{19}	w_{20}	w_{21}	w_{22}	w_{23}	w_{24}
Before Training	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
After Training	0.64	1.00	0.81	1.00	0.94	0.92	1.00	0.98	0.96	0.95	1.00	0.99

3.5 Summary

A neural fuzzy classifier is developed in this chapter to integrate the merits of two selected signal processing techniques (six resulting representative features) for a more reliable bearing condition monitoring. Expertise knowledge can be conveniently incorporated. To signify the effects of different fuzzy rules on classification, statistically-tuned rule weights are proposed. A training algorithm is suggested to fine-tune fuzzy system parameters and to improve the adaptive capability of this

classifier to accommodate different bearing conditions. The effectiveness of this classifier is verified by experimental tests corresponding to different bearing conditions. Test results show that the developed classifier is superior to other related diagnostic schemes in terms of prediction accuracy.

The work presented in this chapter is to achieve the research objective described in Section 1.3: *a NF classifier is proposed to effectively integrate the strengths of several signal processing techniques for a more accurate assessment of the health condition of bearings.*

Chapter 4
A Novel Multi-Step Predictor for Dynamic System Forecasting

A multi-step predictor with a variable input pattern is developed, based on a weighted recurrent NF paradigm, to forecast the future states of bearing health condition with the purpose to further enhance the diagnostic reliability. This chapter will first describe the structure and training of this predictor, and then evaluate its performance based on the benchmark data sets [132] and the gear fault state data sets [6, 7].

4.1 Overview

System state forecasting utilizes available observations to predict the future states of a dynamic system. For example, given a function

$$f(t+1) = F\big(\, f(t),\ \ f(t-1),\ \cdots,\ \ f(t-n+2),\ \ f(t-n+1)\big), \tag{4.1}$$

where $f(t+1)$ is the system output at a future time instant t+1, and $F(\cdot)$ means that the future state $f(t+1)$ is an unknown function of the current and previous observations, i.e., $f(t)$, $f(t-1)$, …, $f(t-n+1)$. The system state forecasting is to use the information of the current observation $f(t)$ and the previous observations $f(t-1)$, …, $f(t-n+1)$ to predict the future system state $f(t+1)$.

Several approaches have been reported in the literature for time series prediction [70]. The classical approaches are based on stochastic or related dynamic models [121-123]. However, an accurate analytical model is usually difficult to derive for a complex dynamic system, especially when the system operates under noisy and/or uncertain environment such as in real-world industrial applications. Since the last decade, more interests in time series forecasting have been focused on the use of data-driven flexible models such as various neural networks (NNs) [124, 125] and neural fuzzy paradigms [126, 127]. Jang [90, 128] proposed an ANFIS for time series forecasting; by simulation, it was found that the ANFIS performs better than both the stochastic models and the feedforward NNs. Wang *et al.* developed several flexible model-based predictors for machinery applications [6, 8]; their investigation results showed that if a NF predictor is properly trained, it performs better than both the feedforward and the recurrent network forecasting schemes.

Although an NF predictor has demonstrated some superior properties to other classical forecasting tools, advanced research still needs to be done in the following aspects before it can be applied to real-world industrial applications [91]: 1) improving the convergence properties, particularly for multi-step-ahead predictions; and 2) enhancing the adaptive capability to accommodate time-varying system conditions. In this chapter, an adaptive NF predictor is developed for multi-step forecasting applications, which is new in the following aspects: 1) a weighted recurrent NF paradigm is proposed for multi-step-ahead prediction; 2) a variable-step input pattern is suggested for multi-step forecasting operations; 3) a hybrid learning technique is adopted for the network training, which is based on a recursive LM algorithm and a recursive least squares estimate (RLSE).

4.2 The NF Multi-Step Predictor

Let *n* input states $f(t)$, $f(t-1)$, …, $f(t-n+1)$ be simply represented by $\{x_{r_0} \quad x_{r_1} \quad \cdots \quad x_{r_{n-1}}\}$, which are normalized into [0, 1]. Also, let the future state $f(t+1)$ be simply represented by x_{+r}, where *r* denotes the prediction step. *r* = 1 is for a single-step prediction whereas *r* > 1 corresponds to a multi-step prediction. The subscript r_i ($i = 0, 1, \cdots, n-1$) means that the input index is a function of the prediction step. In the literature, uniform input patterns are usually applied; that is, the input state vector is given as $\{x_0 \quad x_{-r} \quad x_{-2r} \quad \cdots \quad x_{-(n-1)r}\}$, which represents the current (x_0) and the previous states of a dynamic system [6, 8, 126-128]. Such a uniform input pattern is easy to implement; however, its disadvantage is that the information weights are paid equally to the previous states. To highlight the importance of recent state information, a variable-step input pattern is introduced in this work for the developed NF predictor. In implementation, 1) if *r* is an even number, the input variables are given as $\{x_0 \quad x_{-r/2} \quad x_{-3r/2} \quad x_{-3r} \quad \cdots\}$; 2) if *r* is an odd number, the input variables take $\{x_0 \quad x_{-F(r/2)} \quad x_{-F(3r/2)} \quad x_{-F(3r)} \quad \cdots\}$, where $F(\cdot)$ is a floor function to round the decimal towards an upper infinity.

To simplify illustration, two MFs, *small* and *large*, are assigned to each input state variable. The *r*-step-ahead state of the dynamic system x_{+r} can be formulated by

$$\Re_j: \text{ If } (x_{r_0} \text{ is } A_0^j) \text{ and } (x_{r_1} \text{ is } A_1^j) \text{ and } \ldots \text{ and } (x_{r_{n-1}} \text{ is } A_{n-1}^j) \text{ then } x_{+r} = f_j, \tag{4.2}$$

where $\Re_j$ denotes the *j*th fuzzy rule, $j = 1, 2, \ldots, 2^n$; A_i^j is the activated fuzzy set of x_{r_i} in the *j*th rule. The predictor (4.2) represents a forecasting scheme to estimate the future state of a single variable using the inputs in the form of a 1-D state vector. The consequent structure f_j can take different reasoning forms. To make it comparable with those reported in [6, 8, 70], a first-order TSK model is employed in this work, which has been proven to be more flexible in modeling,

$$f_j = c_0^j x_{r_0} + c_1^j x_{r_1} + \cdots + c_{n-1}^j x_{r_{n-1}} + c_n^j, \tag{4.3}$$

where c_i^j are constants.

The network architecture of the proposed NF predictor is schematically shown in Figure 4.1. It is a six-layer network in which each node performs a particular activation function on the incoming signals. The links have unity weights unless specified. The nodes in layer 1 transmit the input signals to the next layer. Each node in layer 2 acts as an MF. Different from the general predictors as reported in [6, 8, 128], this predictor has a weighted feedback link to each node in layer 2. These feedback links can deal with time explicitly as opposed to representing temporal information spatially. The context units copy the activations of the nodes from the previous time step, and allow the network to memorize the clues from the past, which forms a context for the current processing.

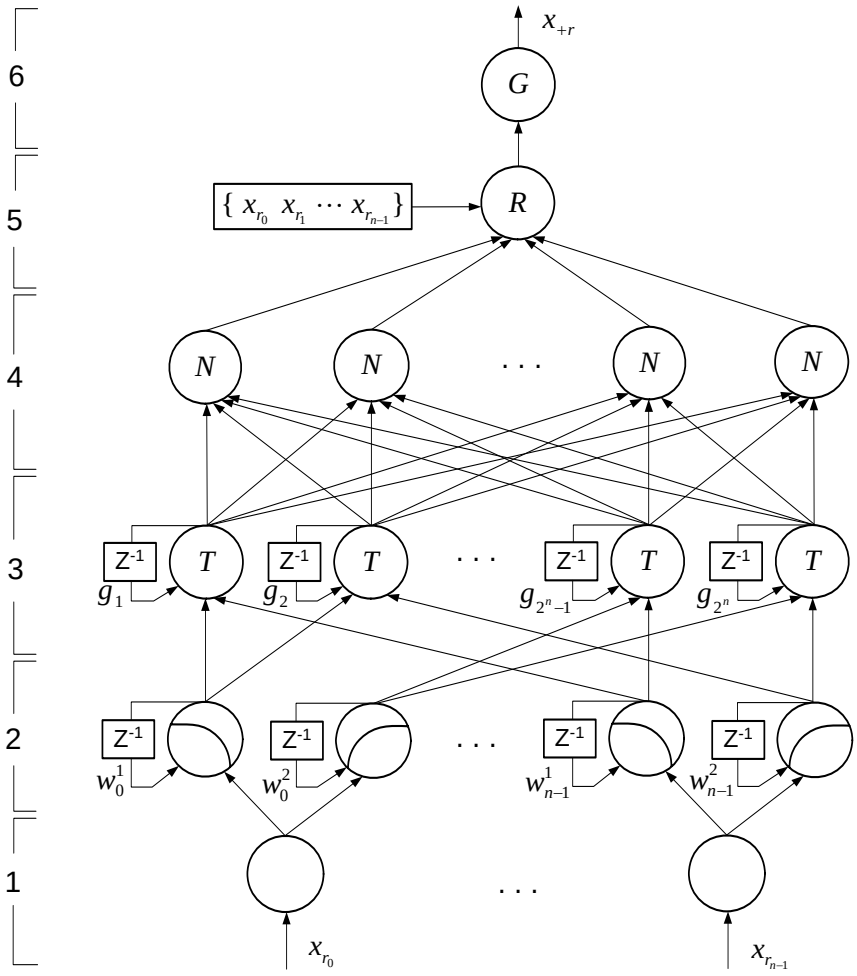

Figure 4.1. The network architecture of the proposed multi-step NF predictor.

The sigmoid MFs with parameters $\{a_i^s \; b_i^s\}$ are utilized in this case, in which $a_i^s > 0$ and $a_i^s < 0$, respectively, correspond to a *large* and a *small* MF. At time instant *t*, the node output is

$$u_{A_i^s}(X_{r_i}^{(t)}) = 1/(1+\exp[-a_i^s(X_{r_i}^{(t)} - b_i^s)]), \tag{4.4}$$

$$X_{r_i}^{(t)} = x_{r_i}^{(t)} + w_i^s u_{A_i^s}(X_{r_i}^{(t-1)}), \tag{4.5}$$

where i = 0, 1, …, n-1; s = 1, 2; $x_{r_i}^{(t)}$ is the input x_{r_i} at time t.

Each node in layer 3 performs a fuzzy T-norm operation. If a max-product operator is applied in this case, the rule firing strength is

$$u_j = \prod_{i=0}^{n-1} u_{A_i^s}(X_{r_i}), \quad j = 1, 2, \ldots, 2^n. \tag{4.6}$$

A weighted feedback link is applied to each node in this layer. The firing strength from the previous step is partially employed to provide more information to the network so as to improve the forecasting accuracy. The output of each node becomes

$$U_j = u_j^{(t)} + g_j u_j^{(t-1)}. \tag{4.7}$$

All the rule firing strengths are normalized in layer 4. After a linear combination of the input variables in layer 5, the predicted output x_{+r} is computed by using the centroid defuzzification:

$$x_{+r} = \sum_{j=1}^{2^n} \overline{U}_j \, f_j, \tag{4.8}$$

where $\overline{U}_j = U_j \Big/ \sum_{j=1}^{2^n} U_j$ is the normalized firing strength of the *j*th rule. The fuzzy system parameters are optimized by a hybrid adaptive training algorithm as discussed below.

4.3 Hybrid Adaptive Training Algorithm

The multi-step NF predictor as developed in Section 4.2 should be properly trained to generate an optimal input/output mapping. For offline training, the representative training data should cover all the possible application conditions [129-131]. Such a requirement is difficult to achieve in real-world applications because most machinery operates in noisy and/or uncertain environments. Usually, the classical forecasting schemes are employed for time-invariant systems or systems with slowly-varying model parameters. However, machinery dynamic characteristics may change suddenly, for example, just after repair or regular maintenance. In this section, a hybrid adaptive training technique is adopted to train this multi-step predictor to accommodate time-varying system conditions.

The data sets in the training database are represented as

$$T_d = \begin{bmatrix} & & \vdots & & \\ x_{r_0}^{(t-1)} & x_{r_1}^{(t-1)} & \cdots & x_{r_{n-1}}^{(t-1)} & d^{(t-1)} \\ x_{r_0}^{(t)} & x_{r_1}^{(t)} & \cdots & x_{r_{n-1}}^{(t)} & d^{(t)} \\ x_{r_0}^{(t+1)} & x_{r_1}^{(t+1)} & \cdots & x_{r_{n-1}}^{(t+1)} & d^{(t+1)} \\ & & \vdots & & \end{bmatrix}, \tag{4.9}$$

where $d^{(t)}$ denotes the actually observed output value at time step *t*. For an input data set $\{x_{r_0}^{(t)} \quad x_{r_1}^{(t)} \quad \cdots \quad x_{r_{n-1}}^{(t)}\}$, the *r*-step-ahead forecasted state $x_{+r}^{(t)}$ is computed by (4.8),

$$x_{+r}^{(t)} = \sum_{j=1}^{2^n} \overline{U}_j \, (c_0^j x_{r_0}^{(t)} + c_1^j x_{r_1}^{(t)} + \cdots + c_{n-1}^j x_{r_{n-1}}^{(t)} + c_n^j). \tag{4.10}$$

The optimization target is to minimize the following error function with respect to the adjustable parameters $\Theta = \{ a_i^s \quad b_i^s \quad w_i^s \quad g_j \quad c_i^j \}$,

$$V_t(\Theta) = \frac{1}{2} \sum_{\tau=1}^{t} \lambda^{t-\tau} (x_{+r}^{(\tau)} - d^{(\tau)})^2 = \frac{1}{2} \sum_{\tau=1}^{t} \lambda^{t-\tau} \varepsilon^2(\tau, \Theta), \tag{4.11}$$

where $\lambda \in [0, 1]$ is the forgetting factor, $0.9 < \lambda < 1$ is used in this work to avoid possible convergence instability. The point forecasting error $\varepsilon(t, \Theta)$ is a linear function of the consequent parameter c_i^j and a nonlinear function with respect to other parameters. Correspondingly, the recursive LM method is applied to optimize the nonlinear parameters $\Theta_N \in R^{q_N \times 1}$, and the RLSE method is employed to fine-tune the linear parameters $\Theta_L \in R^{q_L \times 1}$.

4.3.1 Optimization of Nonlinear Parameters

Let $\hat{\Theta}_N(t-1)$ be the optimal estimate at time step *t*–1, which minimizes $V_{t-1}(\Theta_N)$ and leads to $V'_{t-1}(\hat{\Theta}_N(t-1)) = 0$. It is desired to derive $\hat{\Theta}_N(t)$ to minimize $V_t(\Theta_N)$. This can be done by differentiating the Taylor-series expansion of $V_t(\Theta_N)$ around $\hat{\Theta}_N(t-1)$, that is,

$$\hat{\Theta}_N(t) = \hat{\Theta}_N(t-1) - \left[V_t''(\hat{\Theta}_N(t-1))\right]^{-1} V_t'(\hat{\Theta}_N(t-1))^T, \tag{4.12}$$

where $V_t'(\hat{\Theta}_N(t-1)) \in R^{1\times q_N}$. The first derivative of (4.10) with respect to Θ_N gives

$$V_t'(\Theta_N)^T = \lambda V_{t-1}'(\Theta_N)^T + \varphi(t,\Theta_N)^T \varepsilon(t,\Theta_N), \tag{4.13}$$

where $\varphi(t,\Theta_N) = d\varepsilon(t,\Theta_N)/d\Theta_N$ is the Jacobian matrix and $\varphi(t,\Theta_N) \in R^{1\times q_N}$. The second derivative of (4.10) with respect to Θ_N yields

$$V_t''(\Theta_N(t)) = \lambda V_{t-1}''(\Theta_N(t)) + \varphi(t,\Theta_N)^T \varphi(t,\Theta_N) + \varepsilon''(t,\Theta_N)^T \varepsilon(t,\Theta_N). \tag{4.14}$$

Substituting $\hat{\Theta}_N(t-1)$ for $\Theta_N(t)$ gives

$$\begin{aligned} V_t''(\hat{\Theta}_N(t-1)) = \lambda V_{t-1}''(\hat{\Theta}_N(t-1)) + \varphi(t,\hat{\Theta}_N(t-1))^T \varphi(t,\hat{\Theta}_N(t-1)) \\ + \varepsilon''(t,\hat{\Theta}_N(t-1))^T \varepsilon(t,\hat{\Theta}_N(t-1)) \end{aligned} \tag{4.15}$$

where $V_t''(\cdot) \in R^{q_N\times q_N}$ is the Hessian matrix. The term $\varepsilon''(t,\hat{\Theta}_N(t-1))^T \varepsilon(t,\hat{\Theta}_N(t-1))$ is approximated to zero as t increases. This is because when Θ_N approaches its true value, $\varepsilon(t,\Theta_N)$ will be white noise [117]; therefore, $\varepsilon(t,\Theta_N)$ may be considered to be of zero mean and independent of the states at time step t-1. In particular, it would be independent of $\varepsilon''(t,\Theta_N)$. Therefore, (4.15) becomes

$$V_t''(\hat{\Theta}_N(t-1)) = \lambda V_{t-1}''(\hat{\Theta}_N(t-1)) + \varphi(t,\hat{\Theta}_N(t-1))^T \varphi(t,\hat{\Theta}_N(t-1)) . \tag{4.16}$$

Since $V_{t-1}'(\hat{\Theta}_N(t-1)) = 0$, (4.13) becomes

$$V_t'(\hat{\Theta}_N(t-1))^T = \varphi(t,\hat{\Theta}_N(t-1))^T \varepsilon(t,\hat{\Theta}_N(t-1)). \tag{4.17}$$

Combining (4.12), (4.16) with (4.17) yields

$$\hat{\Theta}_N(t) = \hat{\Theta}_N(t-1) - H(t)^{-1}\varphi(t,\hat{\Theta}_N(t-1))^T \varepsilon(t,\hat{\Theta}_N(t-1)), \tag{4.18}$$

$$H(t) = \lambda H(t-1) + \varphi(t,\hat{\Theta}_N(t-1))^T \varphi(t,\hat{\Theta}_N(t-1)), \tag{4.19}$$

where $H(t)$ is the approximation of the Hessian matrix $V_t''(\hat{\Theta}_N(t-1))$ and $H(t) \in R^{q_N \times q_N}$. The LM method can properly process some ill-conditioned matrices, such as $\varphi(t,\hat{\Theta}_N(t-1))^T \varphi(t,\hat{\Theta}_N(t-1))$, by altering (4.19) to

$$H(t) = \lambda H(t-1) + [\varphi(t,\hat{\Theta}_N(t-1))^T \varphi(t,\hat{\Theta}_N(t-1)) + \mu(t)I], \tag{4.20}$$

where $I \in R^{q_N \times q_N}$ is the identity matrix. Equations (4.18) and (4.20) together constitute the recursive LM algorithm, whereas $\mu(t)$ is the LM parameter to be discussed later. The computation of the inverse Hessian matrix, however, is time-consuming and makes (4.17) impractical for real-time applications. A remedy is to apply the matrix inversion lemma [117] to avoid the direct inversion of Hessian matrix, which is given as

$$(A+UBV)^{-1} = A^{-1} - A^{-1}U(B^{-1}+VA^{-1}U)^{-1}VA^{-1}, \tag{4.21}$$

provided that *A* and $B^{-1}+VA^{-1}U$ are both invertible.

Instead of adding the $q_N \times q_N$ matrix $\mu(t)I$ at each step, only one diagonal element is added at a time. Let $J_{q_N}(t)$ be a matrix whose *t*(mod q_N)+1 diagonal element is 1 and the other elements are all zero. Over a period of q_N time samples, (4.20) can be rewritten as

$$H(t) = \lambda H(t-1) + [\varphi(t,\hat{\Theta}_N(t-1))^T \varphi(t,\hat{\Theta}_N(t-1)) + q_N \mu(t) J_{q_N}(t)], \tag{4.22}$$

which can be further expressed as

$$H(t) = \lambda H(t-1) + [\varphi^*(t,\hat{\Theta}_N(t-1))^T \Lambda^*(t)\varphi^*(t,\hat{\Theta}_N(t-1))], \tag{4.23}$$

where $\varphi^*(t,\hat{\Theta}_N(t-1))^T \in R^{q_N \times 2}$ and $\Lambda^*(t) \in R^{2\times 2}$ are, respectively, given as

$$\varphi^*(t,\hat{\Theta}_N(t-1))^T = \left[\begin{array}{c|c|c} & 0 & \\ & \vdots & \\ \varphi(t,\hat{\Theta}_N(t-1))^T & 0 & \leftarrow \text{position } t \ (\text{mod } q_N) + 1 \\ & 1 & \\ & 0 & \\ & \vdots & \\ & 0 & \end{array}\right], \tag{4.24}$$

$$\Lambda^*(t) = \begin{bmatrix} 1 & 0 \\ 0 & q_N \mu(t) \end{bmatrix}. \tag{4.25}$$

Let $A = \lambda H(t-1)$, $U = \varphi^*(t,\hat{\Theta}_N(t-1))^T$, $B = \Lambda^*(t)$, $V = \varphi^*(t,\hat{\Theta}_N(t-1)$, (4.18) and (4.23) can be expressed as

$$\hat{\Theta}_N(t) = \hat{\Theta}_N(t-1) - R(t-1)\varphi(t,\hat{\Theta}_N(t-1))^T \varepsilon(t,\hat{\Theta}_N(t-1)), \tag{4.26}$$

$$\begin{aligned} R(t) = \frac{1}{\lambda}\{ & R(t-1) - R(t-1)\varphi^*(t,\hat{\Theta}_N(t-1))^T [\lambda \Lambda^{*-1}(t) \\ & + \varphi^*(t,\hat{\Theta}_N(t-1)) R(t-1) \varphi^*(t,\hat{\Theta}_N(t-1))^T]^{-1} \varphi^*(t,\hat{\Theta}_N(t-1)) R(t-1)\} \end{aligned}, \tag{4.27}$$

where $R(t) = H^{-1}(t)$. (4.26) and (4.27) are employed to update the nonlinear parameters in this NF predictor. Some related implementation issues are listed as follows:

1) Initialization of $\hat{\Theta}_N(0)$ and *R*(0): A reasonable estimate can be applied to $\hat{\Theta}_N(0)$. In this work, $\{a_i^s \quad b_i^s\}$ take $\{-5 \quad 0.5\}$ and $\{5 \quad -0.5\}$ for initial *Small* and *Large* MFs, respectively. The initial feedback link weights are all set to zero. *R*(0) are chosen as an identity matrix $\alpha_N I$ with the large positive constant $\alpha_N \in [10^3 \quad 10^5]$.

2) Determination of $\mu(t)$: The LM parameter $\mu(t)$ conversely affects the radius of the trust region [90]. $\mu(t)$ can be adaptively modified based on the performance of the updated parameters. The following strategy is applied in this work: the initial value $\mu(0)$ is set to 100; if the objective function in (4.11) decreases as the parameters are updated, enlarge the $\mu(t)$ by $\kappa\mu(t)$.

Otherwise, if the objective function increases as the parameters are updated, reduce $\mu(t)$ by $\mu(t)/\kappa$, where κ is a design parameter and κ = 1.005 in this case.

3) The Jacobian matrix $\varphi(t,\Theta_N) \in R^{1\times q_N}$ for the developed recursive LM in (4.26) and (4.27) is defined as $\varphi(t,\Theta_N) = d\varepsilon(t,\Theta_N)/d\Theta_N = dx_{+r}^{(t)}/d\Theta_N$. The nonlinear parameters Θ_N include the premise parameters a_i^s , b_i^s , and the link weights w_i^s , g_j . Thus, this Jacobian matrix contains the partial derivative of the network output $x_{+r}^{(t)}$ with respect to each of these nonlinear parameters. Suppose at time *t* the sampled training data set is $\{x_{r_0}^{(t)} \quad x_{r_1}^{(t)} \quad \cdots \quad x_{r_{n-1}}^{(t)} \quad d^{(t)}\}$; if, for example, the sigmoid MFs with parameters $\{a_i^s \;\; b_i^s\}$ are applied, then

$$u_{A_i^s}(X_{r_i}^{(t)}) = 1/(1+\exp[-a_i^s(X_{r_i}^{(t)} - b_i^s)]), \quad i = 0, 1, \ldots, n\text{-}1;\ s = 1, 2. \tag{4.28}$$

$$X_{r_i}^{(t)} = x_{r_i}^{(t)} + w_i^s u_{A_i^s}(X_{r_i}^{(t-1)}). \tag{4.29}$$

The following equations can be derived:

$$\frac{\partial u_{A_i^s}}{\partial a_i^s} = \frac{\exp[-a_i^s(X_{r_i}^{(t)} - b_i^s)](X_{r_i}^{(t)} - b_i^s)}{(1+\exp[-a_i^s(X_{r_i}^{(t)} - b_i^s)])^2} = (X_{r_i}^{(t)} - b_i^s)(1-u_{A_i^s})u_{A_i^s}, \tag{4.30}$$

$$\frac{\partial u_{A_i^s}}{\partial b_i^s} = \frac{-\exp[-a_i^s(X_{r_i}^{(t)} - b_i^s)]a_i^s}{(1+\exp[-a_i^s(X_{r_i}^{(t)} - b_i^s)])^2} = a_i^s(u_{A_i^s} - 1)u_{A_i^s}, \tag{4.31}$$

$$\frac{\partial u_{A_i^s}}{\partial w_i^s} = \frac{\exp[-a_i^s(X_{r_i}^{(t)} - b_i^s)]a_i^s u_{A_i^s}(x_{r_i}^{(t-1)})}{(1+\exp[-a_i^s(X_{r_i}^{(t)} - b_i^s)])^2} = a_i^s u_{A_i^s}(X_{r_i}^{(t-1)})(1-u_{A_i^s})u_{A_i^s}. \tag{4.32}$$

From (4.6):

$$\frac{\partial u_j}{\partial u_{A_i^s}} = \frac{\partial \prod_{k=0}^{n-1} u_{A_k^s}}{\partial u_{A_i^s}} = \frac{\prod_{k=0}^{n-1} u_{A_k^s}}{u_{A_i^s}} = \frac{u_j}{u_{A_i^s}}. \tag{4.33}$$

Equations (4.7) and (4.8) yield

$$\frac{\partial x_{+r}^{(t)}}{\partial a_i^s} = \frac{\partial}{\partial a_i^s}\left(\frac{\sum_{k=1}^{2^n} U_k C_k^{(t)}}{\sum_{k=1}^{2^n} U_k}\right) = \sum_{j=1}^{2^{n-1}} \frac{C_j^{(t)} \sum_{k=1}^{2^n} U_k - \sum_{k=1}^{2^n} U_k C_k^{(t)}}{\left(\sum_{k=1}^{2^n} U_k\right)^2} \frac{\partial U_j}{\partial u_j} \frac{\partial u_j}{\partial a_i^s}$$

$$= \sum_{j=1}^{2^{n-1}} \frac{C_j^{(t)} \sum_{k=1}^{2^n} U_k - \sum_{k=1}^{2^n} U_k C_k^{(t)}}{\left(\sum_{k=1}^{2^n} U_k\right)^2} \frac{\partial u_j}{\partial u_{A_i^j}} \frac{\partial u_{A_i^s}}{\partial a_i^s} \quad , \tag{4.34}$$

$$= \left(\sum_{j=1}^{2^{n-1}} \frac{C_j^{(t)} \sum_{k=1}^{2^n} U_k - \sum_{k=1}^{2^n} U_k C_k^{(t)}}{\left(\sum_{k=1}^{2^n} U_k\right)^2} u_j^{(t)} \right) (X_{r_i}^{(t)} - b_i^s)(1 - u_{A_i^s})$$

where $C_j^{(t)} = c_0^j x_{r_0}^{(t)} + c_1^j x_{r_1}^{(t)} + \cdots + c_{n-1}^j x_{r_{n-1}}^{(t)} + c_n^j$, j denotes a specific rule in which the current parameter a_i^s is involved. A particular a_i^s should be involved in 2^{n-1} rules; correspondingly, the derivation should be undertaken on each of them before summation. Similarly,

$$\frac{\partial x_{+r}^{(t)}}{\partial b_i^s} = \left(\sum_{j=1}^{2^{n-1}} \frac{C_j^{(t)} \sum_{k=1}^{2^n} U_k - \sum_{k=1}^{2^n} U_k C_k^{(t)}}{\left(\sum_{k=1}^{2^n} U_k\right)^2} u_j^{(t)} \right) a_i^s (u_{A_i^s} - 1) \ , \tag{4.35}$$

$$\frac{\partial x_{+r}^{(t)}}{\partial w_i^s} = \left(\sum_{j=1}^{2^{n-1}} \frac{C_j^{(t)} \sum_{k=1}^{2^n} U_k - \sum_{k=1}^{2^n} U_k C_k^{(t)}}{\left(\sum_{k=1}^{2^n} U_k\right)^2} u_j^{(t)} \right) a_i^s u_{A_i^s} (X_{r_i}^{(t-1)})(1 - u_{A_i^s}) , \tag{4.36}$$

$$\frac{\partial x_{+r}^{(t)}}{\partial g_j} = \frac{C_j^{(t)} \sum_{k=1}^{2^n} U_k - \sum_{k=1}^{2^n} U_k C_k^{(t)}}{\left(\sum_{k=1}^{2^n} U_k \right)^2} u_j^{(t-1)}. \tag{4.37}$$

Then the Jacobian matrix should be in the form of

$$\left[\frac{\partial x_{+r}^{(t)}}{\partial a_i^s} \quad \frac{\partial x_{+r}^{(t)}}{\partial b_i^s} \quad \frac{\partial x_{+r}^{(t)}}{\partial w_i^s} \quad \frac{\partial x_{+r}^{(t)}}{\partial g_j} \right] \in R^{1\times(6n+2^n)}$$

4.3.2 Optimization of Linear Parameters

The linear parameters Θ_L in the predictor are optimized by applying RLSE method [6, 8, 90, 128]:

$$\hat{\Theta}_L(t) = \hat{\Theta}_L(t-1) + P(t) C_u [d^{(t)} - C_u^T \hat{\Theta}_L(t-1)], \tag{4.38}$$

$$P(t) = \frac{1}{\lambda}[P(t-1) - \frac{P(t-1) C_u C_u^T P(t-1)}{\lambda + C_u^T P(t-1) C_u}], \tag{4.39}$$

where $P(t) \in R^{q_L \times q_L}$ is the covariance matrix, $P(0) = \alpha_L I$, and $\alpha_L \in [10^2 \quad 10^6]$ is a constant; the vector $C_u \in R^{q_L \times 1}$ is given by

$$C_u = [\overline{U}_1^{(t)} x_{r_0}^{(t)} \quad \overline{U}_1^{(t)} x_{r_1}^{(t)} \quad \cdots \quad \overline{U}_1^{(t)} x_{r_{n-1}}^{(t)} \quad \overline{U}_1^{(t)} \quad \cdots \quad \overline{U}_{2^n}^{(t)} x_{r_{n-2}}^{(t)} \quad \overline{U}_{2^n}^{(t)} x_{r_{n-1}}^{(t)} \quad \overline{U}_{2^n}^{(t)}]^T. \tag{4.40}$$

In training, the predictor consequent linear parameters are optimized in the forward pass of each training epoch based on (4.38) and (4.39); whereas the premise nonlinear parameters are fine tuned in the backward pass based on (4.26) and (4.27). The training process is terminated as long as the training error is sufficiently small (e.g., less than 10^{-5}) or the number of training epochs reaches a predefined level.

4.4 Performance Evaluation

The performance of the proposed multi-step NF predictor is evaluated, in terms of the input pattern, structure and the training algorithm, by simulations with two benchmark data sets and gear fault-state data sets. The performance of this NF predictor is compared with those based on feedforward NNs [53], recurrent NNs [54] and the ANFIS scheme [128]. The ANFIS predictor has a similar architecture as shown in Figure 4.1 but without the weighted feedback links; the ANFIS is trained by a commonly-used gradient-LSE hybrid algorithm. To make the number of parameters compatible among different predictors, the feedforward NN predictor has two hidden layers with eight nodes for each layer, while the recurrent NN predictor has a similar structure to the feedforward NN except those weighted feedback links from the output layer to the second hidden layer. All the predictors have four input state variables.

4.4.1 Sunspot Activity Forecasting

A commonly used benchmark data set in forecasting research is the sunspot activity series, which has the natures of non-Gaussian and non-stationary. The available annual sunspot activity record is from years 1700 to 2007, which has a mean of 49.904 and a standard deviation of 40.430 (SIDC, http://sidc.oma.be/index.php3). The first 250 data sets are used for training, whereas the remaining data sets are for testing. Table 4.1 summarizes the processing results corresponding to several forecasting steps. It is seen that the predictors with a variable input pattern perform much better than those based on the uniform input pattern. This is because the predictors can pay more attention to recent system information for the establishment of input/output mapping. This property is more important for prediction over multiple steps. On the other hand, it is also observed that the proposed NF predictor outperforms other classical predictors due to its flexible reasoning structure and adaptive training capability.

Table 4.1. Prediction RMSE based on a sunspot activity record

	Feedforward NN		Recurrent NN		ANFIS		Proposed Predictor	
+ r	uniform	variable	uniform	variable	uniform	variable	uniform	variable
+ 2	45.50	36.74	73.83	69.78	66.63	47.67	19.83	17.65
+ 3	95.21	67.59	89.78	36.24	35.76	25.95	23.30	18.85
+ 4	112.36	73.42	54.55	38.43	53.12	58.73	20.01	12.66
+ 5	84.34	61.71	62.35	55.83	59.68	57.64	25.95	15.80

4.4.2 Mackey-Glass Data Forecasting

Another commonly used benchmark data set in forecasting research is Mackey-Glass data series given by (4.41) [132],

$$\frac{dx(t)}{dt} \approx \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t) \ . \tag{4.41}$$

Corresponding to the initial conditions $\tau = 30$, $dt = 1$, $x(0) = 1.2$, and $x(t) = 0$ for t < 0, 1200 data sets are selected, in which the first 600 data sets are used for training whereas the remaining data sets for testing the identified model. After 100 training epochs, the processing results from the related predictors are listed in Table 4.2. It is clear that all the predictors based on variable-step input patterns provide a better forecasting performance than those based on uniform-step input patterns. Furthermore, it is seen that the proposed NF predictor performs much better than other classical forecasting schemes, due to its adaptive training and specific reasoning architecture. The recurrent nodes in the NF predictor can store previous system information, which is important for an accurate forecasting operation, particularly for multi-step predictions based on only limited input state variables.

Table 4.2. Prediction RMSE based on Mackey-Glass differential equation

	Feedforward NN		Recurrent NN		ANFIS		Proposed Predictor	
+ r	uniform	variable	uniform	variable	uniform	variable	uniform	variable
+ 2	0.008	0.006	0.015	0.007	0.008	0.005	0.007	0.005
+ 3	0.026	0.014	0.044	0.021	0.024	0.013	0.020	0.014
+ 4	0.051	0.035	0.051	0.033	0.049	0.037	0.043	0.029
+ 5	0.097	0.056	0.069	0.048	0.076	0.061	0.066	0.045
+ 6	0.130	0.069	0.096	0.069	0.120	0.088	0.081	0.057
+ 7	0.082	0.068	0.097	0.067	0.11	0.076	0.070	0.054
+ 8	0.035	0.026	0.029	0.024	0.053	0.030	0.035	0.021

4.4.3 Gear Fault State Forecasting

A few examples related to gear condition monitoring from the previous works [6] are used as an instance to further examine the performance of the developed NF predictor for multi-step machinery forecasting applications. The experimental setup is described in [6] and the fault prognosis of the gear system is conducted gear by gear. The signal generated by each gear is first differentiated by applying a time synchronous filtering process [4, 10]. In machinery condition monitoring, a monitoring index should be sensitive to pattern modulation due to machinery faults but insensitive to noise. In this case, a wavelet amplitude-based index is applied and its derivation can be found in [6, 7]. Three types of gear faults are tested in that study, as represented in Figure 4.2: worn gears, chipped gears, and cracked gears. The analysis is for the gear with 16 teeth. The time step is chosen as 10 min. Table 4.3 summarizes the processing results based on the ANFIS and the proposed NF predictors.

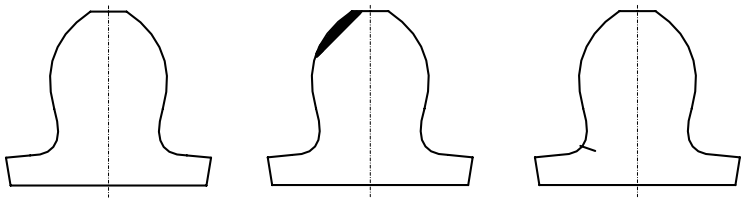

Figure 4.2. Gear conditions tested: (a) worn gear, (b) chipped gear, (c) cracked gear.

Table 4.3. Prediction RMSE in gear fault state forecasting

	Worn Gear		Chipped Gear		Cracked Gear	
+ r	ANFIS	Pro. Predictor	ANFIS	Pro. Predictor	ANFIS	Pro. Predictor
+ 2	0.08	0.06	0.41	0.28	2.39	0.44
+ 3	0.08	0.06	0.60	0.43	4.86	3.21
+ 4	0.15	0.08	0.68	0.65	1.26	0.73
+ 5	0.27	0.12	0.87	0.74	2.31	0.78
+ 6	0.28	0.21	0.69	0.53	4.12	0.88
+ 7	0.55	0.43	1.78	0.74	4.15	1.12
+ 8	0.72	0.64	1.33	0.92	6.06	1.87

1) Gear Wear Monitoring: Wear is a common fault type particularly in exposed gear systems. The proposed NF predictor is initially trained using data sets from the Mackey-Glass equation. Table 4.3 summarizes the processing results over several steps, and Figure 4.3 demonstrates an example of the performance comparison between the ANFIS and the proposed NF predictors in a five-step-ahead forecasting, where each time step is 10 min apart between two consecutive measurements. It is clear that the NF predictor can capture and track the system behavior more effectively than the ANFIS.

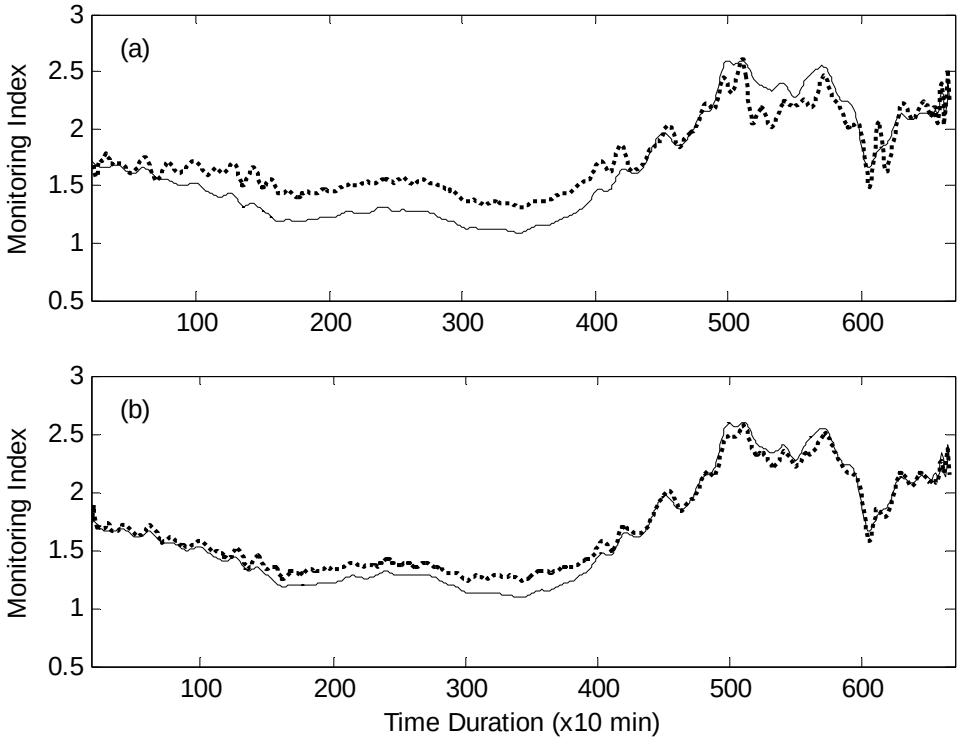

Figure 4.3. The five-step-ahead forecasting results in the worn gear test (solid curve): (a) by using the ANFIS predictor (dotted curve); (b) by using the proposed NF predictor (dotted curve).

2) Chipped Gear Monitoring: The predictors are trained using the previous worn-gear data sets, and then they are tested using the chipped-gear data sets. The processing results are listed in Table 4.3 whereas an example in two-step-ahead predictions is shown in Figure 4.4. It is seen that the output of each predictor match the actual state accurately in the period corresponding to the healthy gear condition (early section). After the chipped fault is introduced, the proposed NF predictor can catch up the new dynamic characteristics of the system quickly. However, the ANFIS predictor can not adapt itself effectively to the new system dynamics, and generates more errors in tracking the dynamic behavior of the system after the fault is introduced.

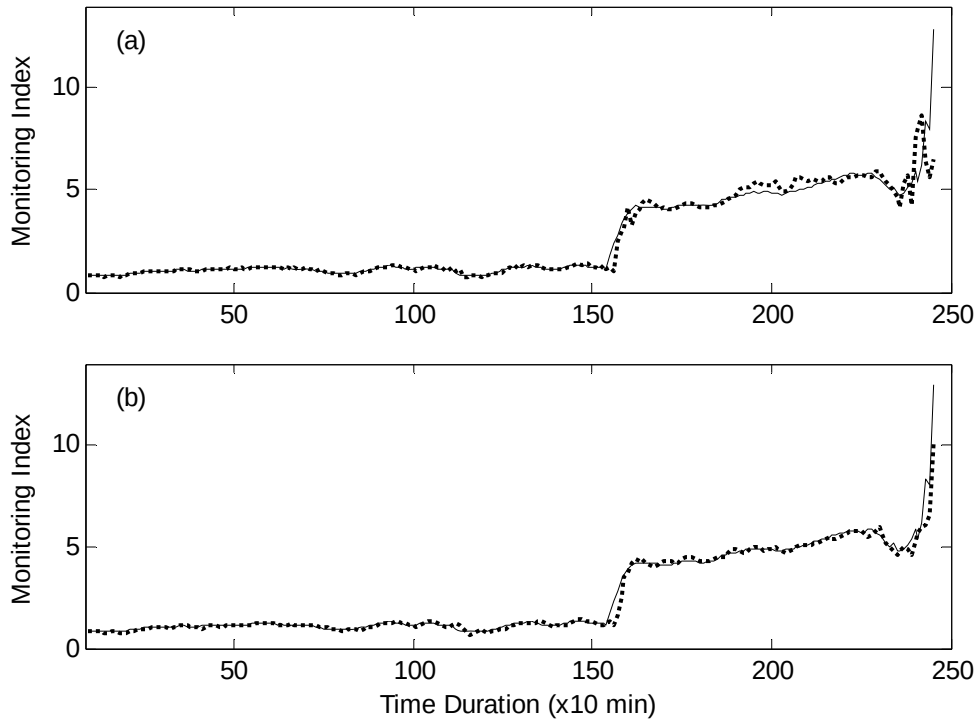

Figure 4.4. The two-step-ahead forecasting results in the chipped gear test (solid curve): (a) by using the ANFIS predictor (dotted curve), (b) by using the proposed NF predictor (dotted curve).

3) Cracked Gear Monitoring: After the predictors are trained by using the new data sets from the above chipped gear testing, they are applied to forecast the data sets corresponding to cracked gear testing. From the processing results as summarized in Table 4.3, apparently the NF predictor can outperform the ANFIS predictor in forecasting the future states of the tested gear system. Figure 4.5 demonstrates another example of six-step-ahead prediction performances from both predictors. It is clear that the NF predictor works accurately in this case; it can capture the system's dynamic behavior effectively. Unfortunately, the ANFIS predictor fails in tracking the system properties in this test due to its lack of the adaptive capability to accommodate time-varying system conditions.

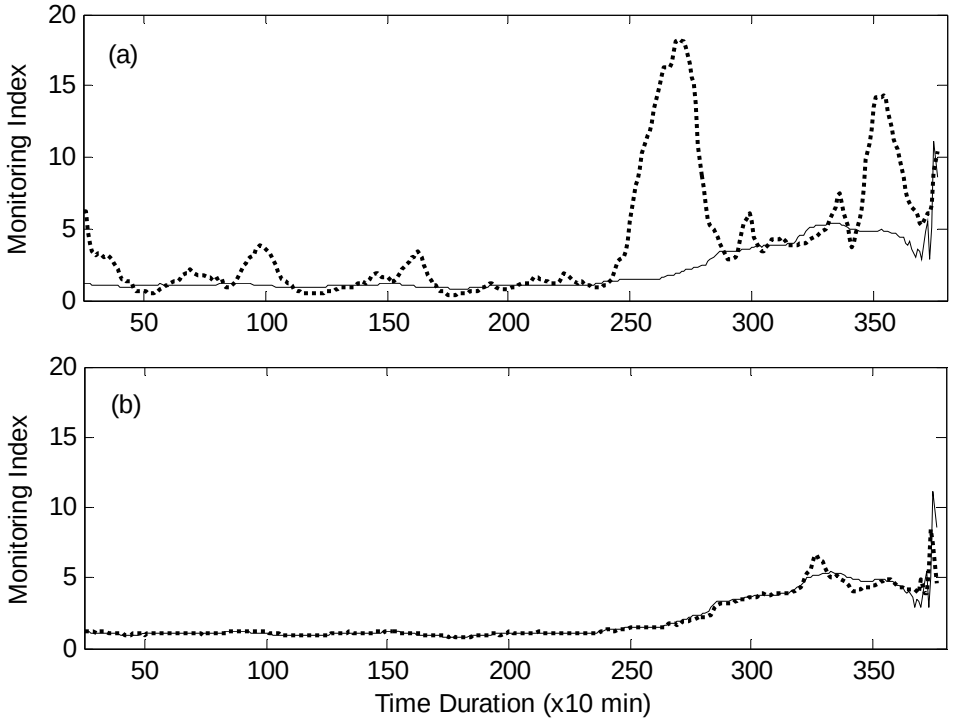

Figure 4.5. The six-step-ahead forecasting results in the cracked gear test (solid curve): (a) by using the ANFIS predictor (dotted curve), (b) by using the proposed NF predictor (dotted curve).

4.5 Summary

A neural fuzzy predictor has been developed in this chapter for multi-step-ahead forecasting of dynamic system properties. The forecasting reasoning is performed based on a novel weighted recurrent neural fuzzy paradigm, while the system parameters are adaptively updated by a hybrid training technique. This training technique combines the recursive Levenberg-Marquart Method and the recursive least square estimate algorithm. A variable-step input pattern is applied to highlight the importance of recent state information in multi-step input/output data mapping. Comparison studies using two benchmark data sets have shown that the proposed multi-step predictor has a better convergence and a higher forecasting accuracy than the classical predictors. This predictor is also implemented for gear system monitoring. Test results have demonstrated that this predictor is a reliable forecasting tool. It can capture the system's dynamic behavior quickly and track the system's characteristics accurately.

The work presented in this chapter is to achieve the research objective described in Section 1.3: *a multi-step predictor is developed and integrated into the ED scheme to forecast the future states of the bearing health condition, and to further enhance the diagnostic reliability*.

Chapter 5
An Enhanced Diagnostic Scheme for Decision-Making

The proposed NF classifier in Chapter 3 can effectively integrate six representative features from two selected signal processing techniques for a more accurate assessment of bearing health condition. The proposed NF predictor in Chapter 4 has been proven to be a reliable forecasting tool and can provide multi-step-ahead forecasting of dynamic system properties. In this chapter, these two intelligent tools are integrated to construct an enhanced diagnostic (ED) scheme so as to further improve the bearing fault diagnostic reliability.

5.1 Final Decision-Making Procedures

The developed intelligent system is reproduced in Figure 5.1, in which the proposed ED scheme is employed for the final decision making. This ED scheme consists of both the classification and prediction modules: the NF classifier is applied to estimate the current bearing health indicator *y* based on the derived monitoring indexes; the multi-step predictor is then employed to forecast the future states of the bearing health condition *y'* based on the preceding diagnostic indicator states obtained in previous processing steps. The final decision regarding the health condition of the bearing is made by using the following inference:

$\Re_1^{'}$: IF ($y \subset C_1$) AND ($y' \subset C_1$) THEN (Bearing is *healthy*) (5.1)

$\Re_2^{'}$: IF ($y \subset C_2$) AND ($y' \subset C_2$) THEN (Bearing is *damaged*) (5.2)

$\Re_3^{'}$: OTHERWISE, (Bearing is *possibly damaged*) (5.3)

If the bearing is *damaged* ($\Re_2^{'}$), an alarm signal is triggered. If the bearing is *possibly damaged* ($\Re_3^{'}$), care should be taken to this bearing during the following monitoring operations, and the final decision should be made according to the next-step diagnostic results.

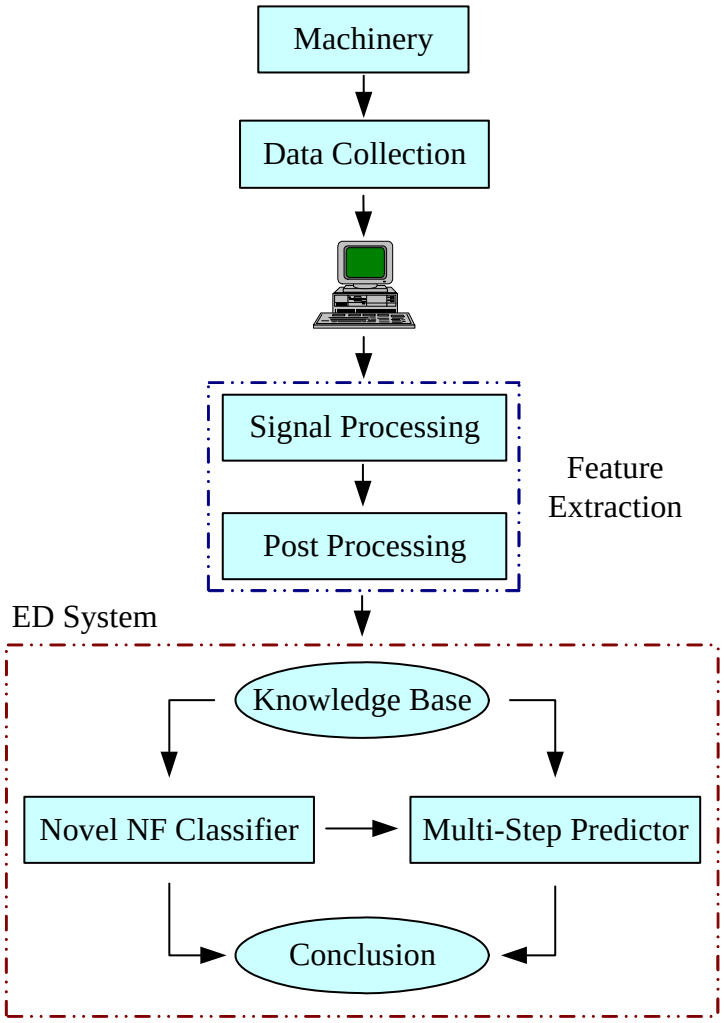

Figure 5.1. Schematic diagram of the developed intelligent system for bearing condition monitoring.

5.2 Performance Evaluation

A series of tests, corresponding to various bearing health and operating conditions, have been conducted; the details of the data preparation process can be seen in Section 3.4.1. The developed ED scheme is compared with other three related diagnostic schemes in terms of diagnostic reliability. *System-1* refers to the developed NF classifier in Chapter 3; *System-2* is a pure fuzzy scheme with unity weights; and *System-3* is an NF scheme with unity weights. The detailed descriptions of these related diagnostic schemes can be found in Section 3.4.2.

The processing results are summarized in Table 5.1. It is seen that although the developed *System-1* in Chapter 3 can provide a more reliable bearing health condition monitoring than *System-2* and *System-3*, four false alarms and three missed alarms are still resulted in these tests with the overall reliability 93.2%. It also shows that the ED scheme can accurately diagnose all the related bearing health conditions (with 100% reliability under these specific test conditions). The proposed ED scheme outperforms all the other related schemes. This is because, on one hand, the ED scheme can optimize the fuzzy parameters and rule weights, and possess the adaptive capability to accommodate different operating conditions; on the other hand, the integrated multi-step NF predictor can effectively enhance the diagnostic reliability by employing more information related to the future states of the bearing health conditions.

Table 5.1. Diagnostic testing results using different classification schemes

Diagnostic Scheme	False Alarms	Missed Alarms	Overall Accuracy
Scheme-1	4	3	93.2%
Scheme-2	11	7	83.4%
Scheme-3	5	6	88.5%
The Proposed ED Scheme	0	0	100%

5.3 Summary

An enhanced diagnostic system is developed in this chapter for bearing fault diagnostics. This system comprises of two modules: 1) A novel neural fuzzy classifier, as described in Chapter 3, is to effectively integrate the merits of several signal processing techniques for a more positive assessment of bearing health condition; 2) A multi-step predictor, as described in Chapter 4, is integrated into the enhanced diagnostic system to forecast the future state of the bearing health condition, and this result is employed to further enhance the diagnostic reliability. Experimental tests corresponding to different bearing health conditions have demonstrated the superior fault diagnostic capability of the proposed technique over other related diagnostic schemes.

The work presented in Chapters 3, 4, and 5 is to achieve the research objective described in Section 1.3: *to develop an enhanced diagnostic scheme for automatic diagnostic decision-making*.

Chapter 6
Conclusions and Future Works

6.1 Conclusions

Although several techniques have been reported in the literature for bearing fault detection and diagnosis, it is still challenging to implement a reliable condition monitoring system for real-world industrial applications because of complex bearing structures and noisy operating conditions. The theme of this thesis is to develop a novel intelligent system to tackle these related challenges. The strategy is to develop more robust techniques at each processing stage to improve the condition monitoring reliability.

First, a new signal processing technique, the wavelet spectrum analysis, is proposed to extract the representative features that are related to the incipient bearing faults. A strategy is suggested for the deployment of wavelet centre frequencies. To enhance the feature characteristics, a weighted Shannon function is proposed to synthesize the wavelet coefficient functions; the applied weights are determined by a statistical index to quantify the effect of the wavelet bandwidth on feature extraction. An averaged autocorrelation power spectrum is adopted to highlight the bearing characteristic features in the spectra. A series of experimental tests, corresponding to various bearing health conditions, demonstrate the superior capability of the proposed wavelet spectrum technique to the related classical bearing fault detection methods in non-stationary feature extraction and analysis.

Secondly, an enhanced diagnostic (ED) scheme is proposed for the final decision making on the bearing health conditions. This scheme consists of two modules: a classifier and a multi-step predictor. The novel neuro-fuzzy classifier is developed to integrate the merits of the two selected signal processing techniques (the developed wavelet spectrum analysis technique and the kurtosis ratio method) for a more positive assessment of the bearing health condition. The diagnostic reliability of this classifier is further enhanced by integrating the information from the developed multi-step predictor. In addition, a hybrid training technique is proposed to fine-tune the fuzzy system parameters and to improve the adaptive capability of the ED scheme to accommodate different bearing conditions. Simulations, based on both the benchmark data sets and gear fault-state data sets, demonstrate that the developed multi-step predictor is an accurate and robust forecasting tool. The effectiveness of the proposed ED scheme is verified by experimental tests corresponding to different

bearing conditions. The test results show that the developed ED scheme is a reliable diagnostic tool, and it outperforms other related diagnostic schemes.

6.2 Future Works

The future research works should be conducted in the following subjects:

1) The proposed bearing fault detection techniques and decision-making schemes will be applied to other mechanical systems such as gearboxes and engines.
2) More investigation related to the diagnoses of advanced bearing faults and distributed bearing defects will be conducted.
3) A structure-evolving neural fuzzy system will be explored for machinery condition monitoring. The rule base will be established using a fuzzy clustering algorithm, by which the noise-affected rules (clusters) can be properly removed whereas the structure of the rules can be adaptively updated according to different machinery conditions.
4) Implement the developed monitoring tools for real-world industrial monitoring applications, so as to improve production quality and to reduce costs.
5) A novel nonlinear mapping technique will be developed to map any separable nonlinear function into a fuzzy logic framework. Some primary investigation has been undertaken and summarized in Appendix A. Such a mapping frame will provide an alternative approach to designing a fuzzy system. It can guarantee that the synthesized fuzzy system has an identical performance to its crisp-domain counterpart. It will provide insight into relationship between the input vector and the output functions. Once the primary fuzzy frame is established, its performance can be optimized by offline expert knowledge and online system training.
6) A genetic programming based technique has been developed in Appendix B for feature reconstruction, which would be useful to specific fault diagnostic applications. However, advanced research needs to be conducted to investigate the analytical relationships between the classical features and the formulated features such that these constructed features can be physically interpretable.

Appendix A
Mapping Separable Nonlinear Functions into a Fuzzy Framework

A.1 Overview

Designing a fuzzy logic system (FLS) is usually time-consuming and there is no guarantee that the resulting system can provide a desired performance [59]. Therefore, to facilitate the FLS design process, a lot of research has been conducted to investigate the transformation relations between conventional system and FLS. Early works in [134, 135] demonstrated that an FLS can be constructed roughly from a conventional linear function. Subsequent development showed that the output from an FLS can be simply represented by a linear parametric function of the inputs [136, 137], from which the synthesis of the FLS could be more straightforward. This method was further improved in [138], and the fuzzy partition of the universe of discourse of the inputs was taken arbitrarily depending on the design requirements. A more general strategy was proposed by Kubica *et al.* [139], whereby virtually any linear system with finite dimension can be transformed to an equivalent FLS.

In contrast to the synthesis strategies for linear systems, the research in constructing FLS from nonlinear functions remains a challenge. Although a few methods were reported in the literature [140-142], the FLS designed using these methods is only an approximated version of the existing nonlinear system, thus the performance of the initially-synthesized FLS cannot be accurately predicted. In this work, a novel method is proposed to effectively design FLS which can generate the prescribed performance of the system so as to provide a well-defined prototype for the FLS [143, 144]. The developed technique extends the related previous work in [139] to include the nonlinear systems.

A.2 A Novel Fuzzy Framework for FLS Design

The FLS design process requires establishing the rule-base, determining appropriate shapes and spans of the input and output fuzzy sets, and selecting appropriate inference and defuzzification methods. The technique proposed in this section prescribes an automated and efficient procedure for the FLS design, which in turn ensures that the performance of the FLS is exactly identical to that of the given nonlinear system. The class of nonlinear systems considered in this study is mathematically described in the form of

$$u(x_1,\ x_2,\ \cdots,\ x_n) = F\big(\ f_1(x_1),\ f_2(x_2),\ \cdots,\ f_n(x_n)\ \big) \tag{A.1}$$

where u is the system output, $f_i(x_i)$ is a sub-function of the state variable x_i (input) within the universe of discourse, and $F(\cdot)$ is a parametric function of $f_i(x_i)$ that is formed by conventional arithmetic operators. It is seen that (A.1) represents a class of nonlinear systems. The procedure for constructing an FLS using the proposed methodology is described in the following subsections.

A.2.1. Fuzzy Input Sets and Fuzzification

The proposed FLS design technique applies singleton fuzzification, whereas the input MFs are constructed such that the sum of the membership grades for a particular input is unity, i.e.,

$$\sum_{s=0}^{z_i-1} \mu_{A_i^s}(x_i) = 1, \qquad 1 \le i \le n\,. \tag{A.2}$$

where $\mu_{A_i^s}(x_i)$ are the MFs, i designates the ith input, s denotes the sth fuzzy set, z_i is the number of fuzzy sets spanning the universe of discourse of input x_i, and n represents the number of the inputs.

To take the nonlinear aspects of the given nonlinear system into the FLS and to realize the exact mapping, semi-prescribed MFs are employed. The MFs for an arbitrary input x_i are specified based on the property of sub-function $f_i(x_i)$, i.e.,

$$\mu_{A_i^{s,1}}(x_i) = \frac{f_i\left(\overline{A}_i^{\,s+1}\right) - f_i(x_i)}{f_i\left(\overline{A}_i^{\,s+1}\right) - f_i\left(\overline{A}_i^{\,s}\right)}, \tag{A.3}$$

$$\mu_{A_i^{s+1,0}}(x_i) = \frac{f_i(x_i) - f_i\left(\overline{A}_i^{\,s}\right)}{f_i\left(\overline{A}_i^{\,s+1}\right) - f_i\left(\overline{A}_i^{\,s}\right)}, \tag{A.4}$$

where $\mu_{A_i^{s,1}}(x_i)$ represents the right-side of the MF of input fuzzy set A_i^s and $\mu_{A_i^{s+1,0}}(x_i)$ is the left-side of the MF of fuzzy set A_i^{s+1}. In addition, $f_i(\overline{A}_i^{\,s})$ and $f_i(\overline{A}_i^{\,s+1})$ are the sub-functions evaluated at the input points $\overline{A}_i^{\,s}$ and $\overline{A}_i^{\,s+1}$ (i.e., modal points) that correspond to the input values by which the membership grade is unity, as depicted in Figure A.1.

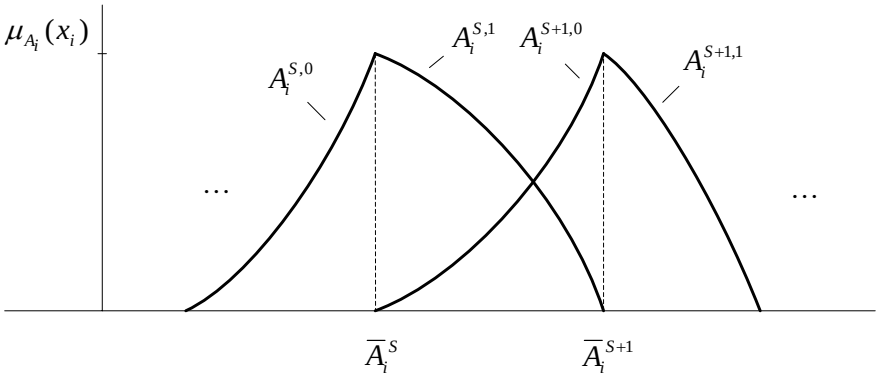

Figure A.1. Input membership functions.

A few issues need to be addressed: 1) particular attention is given to the modal points in the input space whose membership grades are unity; 2) the fuzzy sets defined here are usually asymmetric, thus two independent sides of the MFs are utilized to describe the shape of a fuzzy set; c) for each real-valued crisp input, at most two fuzzy sets have nonzero membership grades; d) if the sub-function $f_i(x_i)$ is not monotonic, the resulting MF may exhibit unusual characteristics, i.e., the membership grades at those points farther away from the modal point are greater than the grades at closer points; however, it can be realized that if the modal points are selected at the positions where the slope changes polarity, the resulting MFs will be monotonic.

In [139], it was shown that a linear system can be cast into a fuzzy framework which can achieve an identical functionality as the original system. The proposed technique in this work extends the approach in [139] to include nonlinear systems. For example, if $f_i(x_i)$ in (A.1) is in the form of

$$f_i(x_i) = kx_i^a, \tag{A.5}$$

the MFs of input x_i can be simplified as

$$\mu_{A_i^{s,1}}(x_i) = \frac{\left(\overline{A}_i^{s+1}\right)^a - (x_i)^a}{\left(\overline{A}_i^{s+1}\right)^a - \left(\overline{A}_i^{s}\right)^a}, \tag{A.6}$$

$$\mu_{A_i^{s+1,0}}(x_i) = \frac{(x_i)^a - (\overline{A}_i^s)^a}{(\overline{A}_i^{s+1})^a - (\overline{A}_i^s)^a}. \tag{A.7}$$

Specially, if a = 1, (A.1) and (A.5) represent a linear system and the resulting MFs are triangular in shape, which were systematically analyzed in [139].

The following analysis is valid with any number of fuzzy sets ($z_i \geq 2$). However, at a given time instant, only two fuzzy sets for each input will have nonzero membership grades under the aforementioned conditions. Therefore, without loss of generality, the following analysis will only consider two fuzzy sets with nonzero membership grades spanning over each input space such that the subsequent explanation can be represented in a more tractable manner.

A.2.2 Rule Base and Fuzzy Inference

The rule base contains the regionalized linguistic mapping from the input fuzzy space to the output fuzzy space. Using a T-norm, for example, a fuzzy rule can be represented as:

$$\mu_{A_1^l \times \cdots \times A_n^l \to B^l}(x,u) = \mu_{A_1^l \times \cdots \times A_n^l \to B^l}(x) * \mu_{B^l}(u). \tag{A.8}$$

Consider an arbitrary set of inputs defined on X, represented by the fuzzy set P_X . Each entry in the rule base generates an output fuzzy set $P_X \circ R^l$ based on the composition inference, i.e.,

$$\mu_{P_X \circ R^l}(u) = \max_{x \in X} [\mu_{P_X}(x) * \mu_{A_1^l \times \cdots \times A_n^l \to B^l}(x,u)]. \tag{A.9}$$

A product operator is applied in (A.9). This is because that the designed FLS is desired to be functionally identical to the original nonlinear system. To accomplish this, the input sets are modeled with respect to the original nonlinear system and are “scaled” by the output sets. This “scaling” requires a “multiplication” between the input and output sets.

The rule base should be fully populated, which contains all possible fuzzy associations. Hence, the total number of rules M is equal to the product of the number of fuzzy sets spanning each of the n inputs, i.e., $M = \prod_{i=1}^{n} z_i$. Since only two fuzzy sets have nonzero membership grades for a given input (let these sets be A_i^0 and A_i^1 , i.e., $A_i^0 = A_i^s$, and $A_i^1 = A_i^{s+1}$ as x_i drops between $\overline{A}_i^s$ and

$\overline{A}_i^{s+1}$), the total number of contributing fuzzy rules for a system with *n* inputs would be $M_c = 2^n$ at a particular time instant. For simplicity, a technique proposed in [139] is employed here to ensure that only the contributing fuzzy rules are considered in the analysis. Let a contributing fuzzy rule be represented by an (*n*+1)-tuple of the form $(A_1^{b_1}, A_2^{b_2}, \cdots, A_n^{b_n}; B^l)$, where $b_i \in \{0,1\}$ and $A_i^{b_i}$ is a fuzzy set in the antecedent of the *l*th rule, $0 \le l \le M_c - 1$. Consider the b_i terms as digits in an *n*-bit binary number, then the contributing rule number *l* in a base-10 number system can be constructed as

$$l = b_1 \times 2^{n-1} + b_2 \times 2^{n-2} + \cdots + b_n \times 2^0 = (b_1 b_2 \cdots b_n)_{base-2}. \tag{A.10}$$

where b_1 is the Most Significant Bit and b_n is the Least Significant Bit. Thus, a convenient way to ensure that only those rules with nonzero output membership grades are considered in the analysis is to list the binary equivalence of the rule numbers ($0 \le l \le 2^n - 1$). For example, consider a 3-input 1-output system, there should be $M_c = 2^3 = 8$ rules in the rule base as illustrated in Table A.1, where "$*$" denotes T-norm.

Table A.1. Bitwise rule base construction

Rule No.	(Rule No.)$_{\text{base-2}}$	Fuzzy Rule
0	0 0 0	$A_1^0 * A_1^0 * A_1^0 \rightarrow B^0(B^{000})$
1	0 0 1	$A_1^0 * A_1^0 * A_1^1 \rightarrow B^1(B^{001})$
2	0 1 0	$A_1^0 * A^1 * A_1^0 \rightarrow B^2(B^{010})$
3	0 1 1	$A_1^0 * A_1^1 * A_1^1 \rightarrow B^3(B^{011})$
4	1 0 0	$A_1^1 * A_1^0 * A_1^0 \rightarrow B^4(B^{100})$
5	1 0 1	$A_1^1 * A_1^0 * A_1^1 \rightarrow B^5(B^{101})$
6	1 1 0	$A_1^1 * A_1^1 * A_1^0 \rightarrow B^6(B^{110})$
7	1 1 1	$A_1^1 * A_1^1 * A_1^1 \rightarrow B^7(B^{111})$

A.2.3 Fuzzy Output Sets and Defuzzification

The output fuzzy sets are chosen to be symmetrical and unimodal with their centers located at $\overline{B}^l$, where $\overline{B}^l$ is defined as

$$\overline{B}^l = F\left(f_1(\overline{A}_1^l),\ f_2(\overline{A}_n^l),\ \cdots,\ f_n(\overline{A}_n^l) \right),\ \ 0 \le l \le M_c - 1. \tag{A.11}$$

The net result of these output sets is that the output sets are effectively singletons at $\overline{B}^l$ when a centroidal defuzzification method is applied. From (A.11), it is seen that special attention is given to those points in the input space that satisfy the rules exactly, i.e.,

$$\prod_{i=1}^{n} \mu_{A_i^l}\left(\overline{A}_i^l\right) = 1. \tag{A.12}$$

In the cases in which the rules are fully satisfied, the consequent of each fuzzy rule is assigned to a unique output set. The centroid of each output set is positioned so as to provide the same output as by the given nonlinear system with the same set of inputs. On the other hand, the input MFs are specified such that the membership grades vary nonlinearly between input sets; this nonlinearity remains the same characteristics as in the given nonlinear system, which in turn can guarantee that the FLS output is a nonlinear interpolation of the original nonlinear system when a rule in the rule base is not satisfied exactly. The output of the constructed FLS is determined by a weighted average of the output singletons $\overline{B}^l$ of the participating fuzzy rules:

$$u(x) = \frac{\sum_{l=1}^{M_c} \overline{B}^l \left(\mu_{P_X \circ R^l}\left(\overline{B}^l\right) \right)}{\sum_{l=1}^{M_c} \left(\mu_{P_X \circ R^l}\left(\overline{B}^l\right) \right)} = \frac{\sum_{l=1}^{M_c} \overline{B}^l \left(\prod_{i=1}^{n} \mu_{A_i^l}(x_i) \right)}{\sum_{l=1}^{M_c} \left(\prod_{i=1}^{n} \mu_{A_i^l}(x_i) \right)} = \frac{u^N(x)}{u^D(x)} \tag{A.13}$$

where $u^N(x)$ and $u^D(x)$, respectively, denote the numerator and denominator of the FLS output and will be employed in the following section for theoretical verification.

A.3 Theoretical Verification

Based on the technique discussed in Section A.2, an FLS can be constructed from the given nonlinear system in (A.1). The objective of this section is to prove that the proposed methodology can guarantee the resulting FLS performs identically to the original nonlinear system.

A.3.1 Demonstrative Example

Consider a 3-input 1-output system described as

$$u(x) = F\big(f_1(x_1),\ f_2(x_2),\ f_3(x_3)\big), \qquad x = [x_1,\ x_2,\ x_3]^T \tag{A.14}$$

where $f_i(x_i)$ is a sub-function of the state variable x_i, i = 1, 2, 3. Assume that two fuzzy sets cover the possible input space $[\overline{A}_i^0\ \ \overline{A}_i^1]$ for each input x_i. Based on the proposed technique, the input MFs are constructed as per (A.3) and (A.4),

$$\begin{aligned} \mu_{A_i^0}(x_i) &= \frac{f_i\left(\overline{A}_i^1\right) - f_i(x_i)}{f_i\left(\overline{A}_i^1\right) - f_i\left(\overline{A}_i^0\right)} \\ \mu_{A_i^1}(x_i) &= \frac{f_i(x_i) - f_i\left(\overline{A}_i^0\right)}{f_i\left(\overline{A}_i^1\right) - f_i\left(\overline{A}_i^0\right)} \end{aligned}. \tag{A.15}$$

The centroids of output fuzzy sets are computed as per (A.11),

$$\begin{aligned} &\overline{B}^0 (or\ \overline{B}^{000}) = F\left(f_1\left(\overline{A}_1^0\right), f_2\left(\overline{A}_2^0\right), f_3\left(\overline{A}_3^0\right)\right) \\ &\quad\cdots \qquad\qquad\qquad \cdots \\ &\overline{B}^7 (or\ \overline{B}^{111}) = F\left(f_1\left(\overline{A}_1^1\right), f_2\left(\overline{A}_2^1\right), f_3\left(\overline{A}_3^1\right)\right) \end{aligned}. \tag{A.16}$$

The FLS output $u(x)$ is calculated as per (A.13) where $u^D(x) = 1$ for any input x (the proof can be found in [139]). Thus, $u(x) = u^N(x)$ and $u(x)$ can be subsequently derived as

$$
\begin{aligned}
u(x) &= \mu_{A_1^0}\mu_{A_2^0}\mu_{A_3^0}\overline{B}^0 + \mu_{A_1^0}\mu_{A_2^0}\mu_{A_3^1}\overline{B}^1 + \cdots\cdots + \mu_{A_1^1}\mu_{A_2^1}\mu_{A_3^0}\overline{B}^6 + \mu_{A_1^1}\mu_{A_2^1}\mu_{A_3^1}\overline{B}^7 \\
&= \mu_{A_1^0}\mu_{A_2^0}\mu_{A_3^0}\overline{B}^0 + \mu_{A_1^0}\mu_{A_2^0}(1-\mu_{A_3^0})\overline{B}^1 + \cdots\cdots \\
&+ (1-\mu_{A_1^0})(1-\mu_{A_2^0})\mu_{A_3^0}\overline{B}^6 + (1-\mu_{A_1^0})(1-\mu_{A_2^0})(1-\mu_{A_3^0})\overline{B}^7 \\
&= \mu_{A_1^0}\mu_{A_2^0}\mu_{A_3^0}(\overline{B}^{000} - \overline{B}^{001} - \overline{B}^{010} + \overline{B}^{011} - \overline{B}^{100} + \overline{B}^{101} + \overline{B}^{110} - \overline{B}^{111}) \\
&+ \mu_{A_1^0}\mu_{A_2^0}(\overline{B}^{001} - \overline{B}^{011} - \overline{B}^{101} + \overline{B}^{111}) + \mu_{A_1^0}\mu_{A_3^0}(\overline{B}^{010} - \overline{B}^{011} - \overline{B}^{110} + \overline{B}^{111}) \\
&+ \mu_{A_2^0}\mu_{A_3^0}(\overline{B}^{100} - \overline{B}^{101} - \overline{B}^{110} + \overline{B}^{111}) + \mu_{A_1^0}(\overline{B}^{011} - \overline{B}^{111}) \\
&+ \mu_{A_2^0}(\overline{B}^{101} - \overline{B}^{111}) + \mu_{A_3^0}(\overline{B}^{110} - \overline{B}^{111}) + \overline{B}^{111}
\end{aligned}
\quad . \qquad \text{(A.17)}
$$

Then, the following two cases will be considered:

1) Sub-functions $f_i(x_i)$ in (A.14) are manipulated by the operator "addition": In this case, the consequent of each rule in (A.16) is represented by

$$
\begin{aligned}
&\overline{B}^0 (or\ \overline{B}^{000}) = f_1\left(\overline{A}_1^0\right) + f_2\left(\overline{A}_2^0\right) + f_3\left(\overline{A}_3^0\right) \\
&\quad \cdots \qquad\qquad \cdots \\
&\overline{B}^7 (or\ \overline{B}^{111}) = f_1\left(\overline{A}_1^1\right) + f_2\left(\overline{A}_2^1\right) + f_3\left(\overline{A}_3^1\right)
\end{aligned}
\quad . \qquad \text{(A.18)}
$$

The output of the designed FLS in (A.17) becomes

$$
\begin{aligned}
u(x) = \mu_{A_1^0}\mu_{A_2^0}\mu_{A_3^0} & \left\{ \begin{array}{l} [f_3(\overline{A}_3^0) - f_3(\overline{A}_3^1)] + [f_3(\overline{A}_3^1) - f_3(\overline{A}_3^0)] \\ + [f_3(\overline{A}_3^1) - f_3(\overline{A}_3^0)] + [f_3(\overline{A}_3^0) - f_3(\overline{A}_3^1)] \end{array} \right\} \\
& + \mu_{A_1^0}\mu_{A_2^0}\left\{[f_2(\overline{A}_2^0) - f_2(\overline{A}_2^1)] + [f_2(\overline{A}_2^1) - f_2(\overline{A}_2^0)]\right\} \\
& + \mu_{A_1^0}\mu_{A_3^0}\left\{[f_3(\overline{A}_3^0) - f_3(\overline{A}_3^1)] + [f_3(\overline{A}_3^1) - f_3(\overline{A}_3^0)]\right\} \\
& + \mu_{A_2^0}\mu_{A_3^0}\left\{[f_3(\overline{A}_3^0) - f_3(\overline{A}_3^1)] + [f_3(\overline{A}_3^1) - f_3(\overline{A}_3^0)]\right\} \\
& + \mu_{A_1^0}[f_1(\overline{A}_1^0) - f_1(\overline{A}_1^1)] + \mu_{A_2^0}[f_2(\overline{A}_2^0) - f_2(\overline{A}_2^1)] \\
& + \mu_{A_3^0}[f_3(\overline{A}_3^0) - f_3(\overline{A}_3^1)] + [f_1(\overline{A}_1^1) + f_2(\overline{A}_2^1) + f_3(\overline{A}_3^1)] \\
= \mu_{A_1^0}\mu_{A_2^0}\mu_{A_3^0} & \{0\} + \mu_{A_1^0}\mu_{A_2^0}\{0\} + \mu_{A_1^0}\mu_{A_3^0}\{0\} + \mu_{A_2^0}\mu_{A_3^0}\{0\} \\
& + [f_1(x_1) - f_1(\overline{A}_1^1)] + [f_2(x_2) - f_2(\overline{A}_2^1)] + [f_3(x_3) - f_3(\overline{A}_3^1)] \\
& + [f_1(\overline{A}_1^1) + f_2(\overline{A}_2^1) + f_3(\overline{A}_3^1)] \\
= & \sum_{i=1}^{3} f_i(x_i) \qquad .
\end{aligned} \tag{A.19}
$$

It is seen from (A.19) that the constructed FLS is functionally identical to the given nonlinear system, i.e., $u(x) = f_1(x_1) + f_2(x_2) + f_3(x_3)$.

2) Sub-functions $f_i(x_i)$ in (A.14) are manipulated by the operator "multiplication": In this case, the consequent of each rule in (A.16) is represented by

$$
\begin{array}{l}
\overline{B}^0 (or\ \overline{B}^{000}) = f_1\left(\overline{A}_1^0\right) \times f_2\left(\overline{A}_2^0\right) \times f_3\left(\overline{A}_3^0\right) \\
\quad \cdots \qquad\qquad\qquad \cdots \qquad . \\
\overline{B}^7 (or\ \overline{B}^{111}) = f_1\left(\overline{A}_1^1\right) \times f_2\left(\overline{A}_2^1\right) \times f_3\left(\overline{A}_3^1\right)
\end{array} \tag{A.20}
$$

Consequently, the output of the designed FLS in (A.17) becomes

$$
\begin{aligned}
u(x) &= -\prod_{i=1,2,3}[f_i(\overline{A}_i^1) - f_i(x_i)] + \begin{cases} f_3(\overline{A}_3^1)\prod_{i=1,2}[f_i(\overline{A}_i^1) - f_i(x_i)] \\ f_2(\overline{A}_2^1)\prod_{i=1,3}[f_i(\overline{A}_i^1) - f_i(x_i)] \\ f_1(\overline{A}_1^1)\prod_{i=2,3}[f_i(\overline{A}_i^1) - f_i(x_i)] \end{cases} \\
&\quad - \begin{cases} \prod_{i=2,3} f_i(\overline{A}_i^1)[f_1(\overline{A}_1^1) - f_1(x_1)] \\ \prod_{i=1,3} f_i(\overline{A}_i^1)[f_2(\overline{A}_2^1) - f_2(x_2)] + \prod_{i=1,2,3} f_i(\overline{A}_i^1) \\ \prod_{i=1,2} f_i(\overline{A}_i^1)[f_3(\overline{A}_3^1) - f_3(x_3)] \end{cases} \\
&= f_3(x_3)\prod_{i=1,2}[f_i(\overline{A}_i^1) - f_i(x_i)] - \begin{cases} f_3(x_3) f_2(\overline{A}_2^1)[f_1(\overline{A}_1^1) - f_1(x_1)] \\ f_3(x_3) f_1(\overline{A}_1^1)[f_2(\overline{A}_2^1) - f_2(x_2)] \end{cases} \\
&\quad + f_3(x_3)\prod_{i=1,2} f_i(\overline{A}_i^1) \\
&= f_3(x_3)\{-[f_1(\overline{A}_1^1) - f_1(x_1)] f_2(x_2)\} + f_3(x_3)[f_1(\overline{A}_1^1) f_2(x_2)] \\
&= \prod_{i=1}^{3} f_i(x_i)
\end{aligned}
\qquad \text{(A.21)}
$$

Again, it is seen from (A.21) that the constructed FLS is functionally identical to the given nonlinear system, i.e., $u(x) = f_1(x_1) \times f_2(x_2) \times f_3(x_3)$.

A.3.2 Generalized Nonlinear Systems Described in (A.1)

For the general case, $u^D(x) = 1$ still holds and thus $u(x) = u^N(x)$. The output of the constructed FLS with *n* inputs is derived as

$$
\begin{aligned}
u(x) &= \mu_{A_1^0}\mu_{A_2^0}\mu_{A_3^0}\cdots\mu_{A_{n-2}^0}\mu_{A_{n-1}^0}\mu_{A_n^0}\overline{B}^0 + \mu_{A_1^0}\mu_{A_2^0}\mu_{A_3^1}\cdots\mu_{A_{n-2}^0}\mu_{A_{n-1}^0}\mu_{A_n^1}\overline{B}^1 \\
&\quad + \cdots\cdots \\
&\quad + \mu_{A_1^1}\mu_{A_2^1}\mu_{A_3^1}\cdots\mu_{A_{n-2}^1}\mu_{A_{n-1}^1}\mu_{A_n^0}\overline{B}^{2^n-2} + \mu_{A_1^1}\mu_{A_2^1}\mu_{A_3^1}\cdots\mu_{A_{n-2}^1}\mu_{A_{n-1}^1}\mu_{A_n^1}\overline{B}^{2^n-1} \\
&= \mu_{A_1^0}\mu_{A_2^0}\mu_{A_3^0}\cdots\mu_{A_{n-2}^0}\mu_{A_{n-1}^0}\mu_{A_n^0}\overline{B}^0 + \mu_{A_1^0}\mu_{A_2^0}\mu_{A_3^1}\cdots\mu_{A_{n-2}^0}\mu_{A_{n-1}^0}(1-\mu_{A_n^0})\overline{B}^1 \cdot \\
&\quad + \cdots\cdots \\
&\quad + (1-\mu_{A_1^0})(1-\mu_{A_2^0})(1-\mu_{A_3^0})\cdots(1-\mu_{A_{n-2}^0})(1-\mu_{A_{n-1}^0})\mu_{A_n^0}\overline{B}^{2^n-2} \\
&\quad + (1-\mu_{A_1^0})(1-\mu_{A_2^0})(1-\mu_{A_3^0})\cdots(1-\mu_{A_{n-2}^0})(1-\mu_{A_{n-1}^0})(1-\mu_{A_n^0})\overline{B}^{2^n-1}
\end{aligned}
\tag{A.22}
$$

Equation (A.22) can be represented in a concise form as:

$$
u(x) = \sum_{\lambda=0}^{n}\left\{\sum_{m=1}^{\binom{n}{\lambda}}\left[\left(\prod_{\left\{i\in[1,n]\,|\,i\notin\binom{n}{\lambda,m}\right\}}\mu_{A_i^0}\right)\left(\sum_{\left\{b_1\cdots b_n\in[0\cdots0,1\cdots1]\,|\,b_i=1\,\forall i\in\binom{n}{\lambda,m}\right\}}(-1)^{-\lambda+\sum_{i=1}^{n}b_i}\overline{B}^{b_1\cdots b_n}\right)\right]\right\}, \tag{A.23}
$$

where λ is the number of the input indices that do not appear in the term $\prod\mu_{A_i^0}$, i = 1, 2, …, n. The combination without repetition $\binom{n}{\lambda}$ in (A.23) represents the number of un-ordered selections of λ bits from a group of n bits, where $\binom{n}{0}$ is set to be 1; m denotes the mth combination, and $\binom{n}{\lambda,m}$ is a set of indices whose bits are chosen in the mth combination of $\binom{n}{\lambda}$. Consequently, $i\in[1,n]\,|\,i\notin\binom{n}{\lambda,m}$ represents only those indices irrelevant to mth combination. $b_1\cdots b_n\in[0\cdots0,1\cdots1]\,|\,b_i=1\forall i\in\binom{n}{\lambda,m}$ means that $b_1\cdots b_n$ cover all the possible n-bit binary

assembly whose $b_i\left(\forall i \in \binom{n}{\lambda, m}\right)$ bit is 1. In addition, the sign of $\overline{B}^{b_1 \cdots b_n}$ is determined by the number of 1's in the assembly $\{b_1 \cdots b_n\}$ except those in $b_i\left(\forall i \in \binom{n}{\lambda, m}\right)$.

The validation for the generalized nonlinear systems described in (A.1) is addressed via the following two steps:

1) Sub-functions $f_i(x_i)$ in (A.1) are manipulated by the operator "addition": In this case, the consequent of each rule in (A.11) can be represented by

$$\begin{aligned} &\overline{B}^0 (or\ \overline{B}^{0\cdots 0}) = f_1\left(\overline{A}_1^0\right) + \cdots + f_n\left(\overline{A}_n^0\right) \\ &\quad \cdots \qquad\qquad\qquad \cdots \\ &\overline{B}^{2^n-1} (or\ \overline{B}^{1\cdots 1}) = f_1\left(\overline{A}_1^1\right) + \cdots + f_n\left(\overline{A}_n^1\right) \end{aligned} \tag{A.24}$$

The output of the constructed FLS in (A.23) becomes

$$u(x) = \sum_{\lambda=0}^{n} \left\{ \sum_{m=1}^{\binom{n}{\lambda}} \left[\left(\prod_{\left\{ i \in [1,n] \| i \notin \binom{n}{\lambda,m} \right\}} \mu_{A_i^0} \right) \left(\sum_{\left\{ b_1 \cdots b_n \in [0\cdots 0, 1\cdots 1] \| b_i = 1 \forall i \in \binom{n}{\lambda,m} \right\}} (-1)^{-\lambda + \sum_{i=1}^{n} b_i} \left(\sum_{i=1}^{n} f_i\left(\overline{A}_i^{b_i}\right) \right) \right) \right] \right\}. \tag{A.25}$$

When the mth combination of $\binom{n}{\lambda}$ is chosen, the left $(n-\lambda)$ bits can be combined in terms of any possible $(n-\lambda)$-th binary assembly $c_1 \cdots c_{n-\lambda} \in [0\cdots 0, 1\cdots 1]$, where c_j corresponds to a bit in $b_1 \cdots b_n \in [0\cdots 0, 1\cdots 1]$ and points to a particular $f_i\left(\overline{A}_i^{b_i}\right)$ (as per (A.24), now denoted as $f_j\left(\overline{A}_j^{c_j}\right)$ in each designated rule consequent. It can be seen that when $0 \le \lambda \le n-2$, the summation

$(-1)^{\sum_{j=1}^{n-\lambda} c_j}\left(\sum_{j=1}^{n-\lambda} f_j\left(\overline{A}_j^{c_j}\right)\right)$ is zero because all the associated terms $f_j\left(\overline{A}_j^{c_j}\right)$ have been cancelled out. This leads the FLS output $u(x)$ in (A.25) to be

$$
\begin{aligned}
u(x) = & \sum_{\lambda=0}^{n-2}\left\{\sum_{m=1}^{\binom{n}{\lambda}}\left[\left(\prod_{\left\{i\in[1,n]\,|\,i\notin\binom{n}{\lambda,m}\right\}}\mu_{A_i^0}\right)0\right]\right\} \\
& + \sum_{\lambda=n-1}^{n}\left\{\sum_{m=1}^{\binom{n}{\lambda}}\left[\left(\prod_{\left\{i\in[1,n]\,|\,i\notin\binom{n}{\lambda,m}\right\}}\mu_{A_i^0}\right)\left(\sum_{\left\{b_1\cdots b_n\in[0\cdots0,1\cdots1]\,|\,b_i=1\,\forall i\in\binom{n}{\lambda,m}\right\}}(-1)^{-\lambda+\sum_{i=1}^{n} b_i}\left(\sum_{i=1}^{n} f_i\left(\overline{A}_i^{b_i}\right)\right)\right)\right]\right\} \\
= & \sum_{i=1}^{n}\mu_{A_i^0}\left(f_i\left(\overline{A}_i^0\right)-f_i\left(\overline{A}_i^1\right)\right)+\sum_{i=1}^{n} f_i\left(\overline{A}_i^1\right) \qquad \text{(A.26)} \\
= & \sum_{i=1}^{n}\left(f_i(x_i)-f_i\left(\overline{A}_i^1\right)\right)+\sum_{i=1}^{n} f_i\left(\overline{A}_i^1\right) \\
= & \sum_{i=1}^{n} f_i(x_i)
\end{aligned}
$$

2) Sub-functions $f_i(x_i)$ in (A.1) are manipulated by the operator "multiplication" In this case, the consequent of each rule in (A.11) can be represented by

$$
\begin{aligned}
&\overline{B}^0 (or\ \overline{B}^{0\cdots0}) = f_1\left(\overline{A}_1^0\right)\times\cdots\times f_n\left(\overline{A}_n^0\right) \\
&\quad\cdots \qquad\qquad\qquad \cdots \quad . \qquad \text{(A.27)} \\
&\overline{B}^{2^n-1} (or\ \overline{B}^{1\cdots1}) = f_1\left(\overline{A}_1^1\right)\times\cdots\times f_n\left(\overline{A}_n^1\right)
\end{aligned}
$$

The output of the constructed FLS in (A.23) becomes

$$u(x)=\sum_{\lambda=0}^{n}\left\{\sum_{m=1}^{\binom{n}{\lambda}}\left[\left(\prod_{\left\{i\in[1,n]\,|\,i\notin\binom{n}{\lambda,m}\right\}}\mu_{A_i^0}\right)\left(\sum_{\left\{b_1\cdots b_n\in[0\cdots0,1\cdots1]\,|\,b_i=1\,\forall\, i\in\binom{n}{\lambda,m}\right\}}(-1)^{-\lambda+\sum_{i=1}^{n}b_i}\left(\prod_{i=1}^{n}f_i\left(\overline{A}_i^{\,b_i}\right)\right)\right)\right]\right\}. \tag{A.28}$$

By substituting $\mu_{A_i^0}$ with $\dfrac{f_i(\overline{A}_i^1)-f_i(x_i)}{f_i(\overline{A}_i^1)-f_i(\overline{A}_i^0)}$ as per (A.3), and by some mathematical manipulations, (A.28) becomes

$$\begin{aligned}
u(x)&=\sum_{\lambda=0}^{n}\left\{\sum_{m=1}^{\binom{n}{\lambda}}\left[(-1)^{n-\lambda}\left(\prod_{\left\{i\in[1,n]\,|\,i\in\binom{n}{\lambda,m}\right\}}f_i\left(\overline{A}_i^1\right)\right)\left(\prod_{\left\{i\in[1,n]\,|\,i\notin\binom{n}{\lambda,m}\right\}}\left(f_i\left(\overline{A}_i^1\right)-f_i(x_i)\right)\right)\right]\right\}\\
&=f_n(x_n)\sum_{\lambda=0}^{n-1}\left\{\sum_{m=1}^{\binom{n-1}{\lambda}}\left[(-1)^{n-1-\lambda}\left(\prod_{\left\{i\in[1,n-1]\,|\,i\in\binom{n-1}{\lambda,m}\right\}}f_i\left(\overline{A}_i^1\right)\right)\left(\prod_{\left\{i\in[1,n-1]\,|\,i\notin\binom{n-1}{\lambda,m}\right\}}\left(f_i\left(\overline{A}_i^1\right)-f_i(x_i)\right)\right)\right]\right\}\\
&=f_n(x_n)f_{n-1}(x_{n-1})\sum_{\lambda=0}^{n-2}\left\{\sum_{m=1}^{\binom{n-2}{\lambda}}\left[\left((-1)^{n-2-\lambda}\left(\prod_{\left\{i\in[1,n-2]\,|\,i\in\binom{n-2}{\lambda,m}\right\}}f_i\left(\overline{A}_i^1\right)\right)\right)\left(\prod_{\left\{i\in[1,n-2]\,|\,i\notin\binom{n-2}{\lambda,m}\right\}}\left(f_i\left(\overline{A}_i^1\right)-f_i(x_i)\right)\right)\right]\right\}\\
&=\cdots\cdots\\
&=\prod_{i=1}^{n}f_i(x_i)
\end{aligned} \tag{A.29}$$

From the results of (A.26) and (A.29), it is seen that the constructed FLS is functionally equivalent to the given nonlinear system as described in (A.1).

Appendix B
A GP-Based Feature Reconstructed Approach

A genetic programming based technique has been developed in this work to further enhance the feature's characteristics. This technique would be useful to specific fault diagnostic applications. However, more advanced research needs to be conducted to make these reconstructed features physically interpretable.

B.1 Overview

In recent years, some research efforts have been devoted to the feature reconstruction by using the evolutionary computing tools, such as genetic algorithms [99], particle swarm optimization [100], and genetic programming (GP) [101]. Although the formulated features are constructed by some data-driven "blackbox" procedures, the investigation results show that they can provide more discrimination information than the classical features [102] to the problems of fault detection. In this work, a new GP-based feature reconstruction technique is proposed for bearing fault detection. To facilitate demonstration, the feature reconstruction is based on two classical monitoring indexes: the kurtosis and the RMS [13, 37]. The investigation results show that although these two monitoring indexes have been used for some special bearing applications [15, 18, 103-106], they are not robust yet and cannot provide consistent bearing fault detection information, particularly when the machine operates in noisy environments. By the proposed GP-based technique, the formulated features can be robust in certain applications, that is, they are more sensitive to faulty signatures but less sensitive to variations in the speed, load, and the geometry of bearings.

B.2 Tested Bearing Conditions

The experimental setup used in this work is shown in Figure 1.4. Both the ball bearings (MB ER-10K) and the cylindrical roller bearings (NJ 303 ECP) are tested. To simplify analysis, only the cases as summarized in Table B.1 are considered in this work: healthy (HY), inner-race defect (ID), and outer-race defect (OD) for the ball bearings; HY, ID, and rolling element damage (RD) for the roller bearings.

Table B.1. Fault type and size

	Fault Type	Fault Size (mm)	No.		Fault Type	Fault Size (mm)	No.
ball bearing	HY	no fault	1	roller bearing	RD	0.89	7
	ID	0.56	2		ID	0.43	8
	OD	0.44	3			0.56	9
		0.57	4			0.84	10
roller bearing	HY	no fault	5			1.12	11

B.3 Performance Evaluation of Classical Features

Bearings will generate vibrations due to the interactions between bearing components and possible defects [13, 37]. If the bearing is damaged, whenever a local defect on a bearing element strikes or is struck by the corresponding mating element, an impulse is generated. This impulse in turn excites structural resonances. Indicators such as the kurtosis and RMS have been applied for bearing fault detection in the literature [15, 18, 103-106]. The kurtosis essentially describes the shape of the probability density function of a signal, and can be estimated using

$$kurtosis = \frac{\sum_{i=1}^{N}(x_i - \mu)^4}{\sigma^4}, \tag{B.1}$$

where N is the total number of the samples, μ is the sample mean, and σ is the sample standard deviation. The RMS is the square root of the sample variance representing the energy level of the signal. RMS can be calculated using

$$RMS = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N-1}}. \tag{B.2}$$

Fault related signatures are usually modulated by the signals at resonance frequencies of the bearing and its surrounding structures, and these resonance signals may cover a wide bandwidth [26]. To investigate the effects of signals of different frequency band to the performance of these two

classical indicators, in this work, the signals will be processed by using the low-pass, band-pass, and the high-pass filters, respectively. These filters are in eighth order Butterworth IIR type, and their bandwidths are determined by examining the full spectra of the vibration signals and the characteristic defect frequencies of the bearings [27]. The filter characteristics are summarized in Table B.2.

Table B.2. Filters used for signal preprocessing

Filters	Raw Signal	LP	HP	LP	HP
Cutoff Frequency (Hz)	[0 8000]	300	300	1000	1000
Filters	LP	HP	BP		BP
Cutoff Frequency (Hz)	2500	2500	[1000 2500]		[300 3000]

LP: low-pass; HP: high-pass; BP: band-pass.

Two accelerometers are employed to measure vibration along the horizontal direction (S_1) and the vertical direction (S_2), around the tested bearing. A new signature is constructed by $\sqrt{S_1^2 + S_2^2}$. Furthermore, because the signals are usually amplitude modulated by defects, the signal envelope is extracted by using the Hilbert transform [107],

$$\bar{S}(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{S(\tau)}{t-\tau} \, d\tau \ , \tag{B.3}$$

where $S(t)$ is a real-valued signal. The analytic signal of $S(t)$ is determined by $S(t) + j\bar{S}(t)$. The envelope of the signal $S(t)$ is the magnitude of this analytic signal,

$$\left| H(S(t)) \right| = \left(S^2(t) + \bar{S}^2(t) \right)^{1/2} . \tag{B.4}$$

Table B.3 summarizes the signatures to be investigated in this work, where $|S_1|$ and $|S_2|$ are the rectified vibration signals.

Table B.3. Signatures used to calculate the classical features

S_1	S_2	$\sqrt{S_1^2 + S_2^2}$	$\lvert S_1 \rvert$	$\lvert S_2 \rvert$	$\lvert H(S_1) \rvert$	$\lvert H(S_2) \rvert$

A comprehensive investigation has been conducted in this work corresponding to various bearing health and operation conditions. The analysis results show that although the kurtosis and RMS can provide some useful diagnostic information under certain situations, these classical indexes are not robust and reliable for bearing fault detection and fault type classification, no matter which signal-filtering combination is applied. As an example, the best classification result from the kurtosis is illustrated in Figure B.1, and the following observations can be obtained:

- It is possible to distinguish a damaged bearing from the healthy one for ball bearings. There is a clear separation between curve 1 and the curves 2, 3 and 4. This distinction is insensitive to load and speed.
- For a roller bearing, however, it is possible to detect faults in the bearing only if the fault dimension is sufficiently large, such as the curves 10 and 11.
- It is impossible to diagnose the fault type for both the ball bearings and roller bearings.

Correspondingly, the best classification result based on RMS is demonstrated in Figure B.2, and it can be observed that:

- The RMS is highly dependent on the shaft speed.
- The RMS for the healthy ball bearings is quite different from that of the healthy roller bearings.
- For all the tested bearings, if the defect size is small, the RMS is close to that of the healthy bearing.

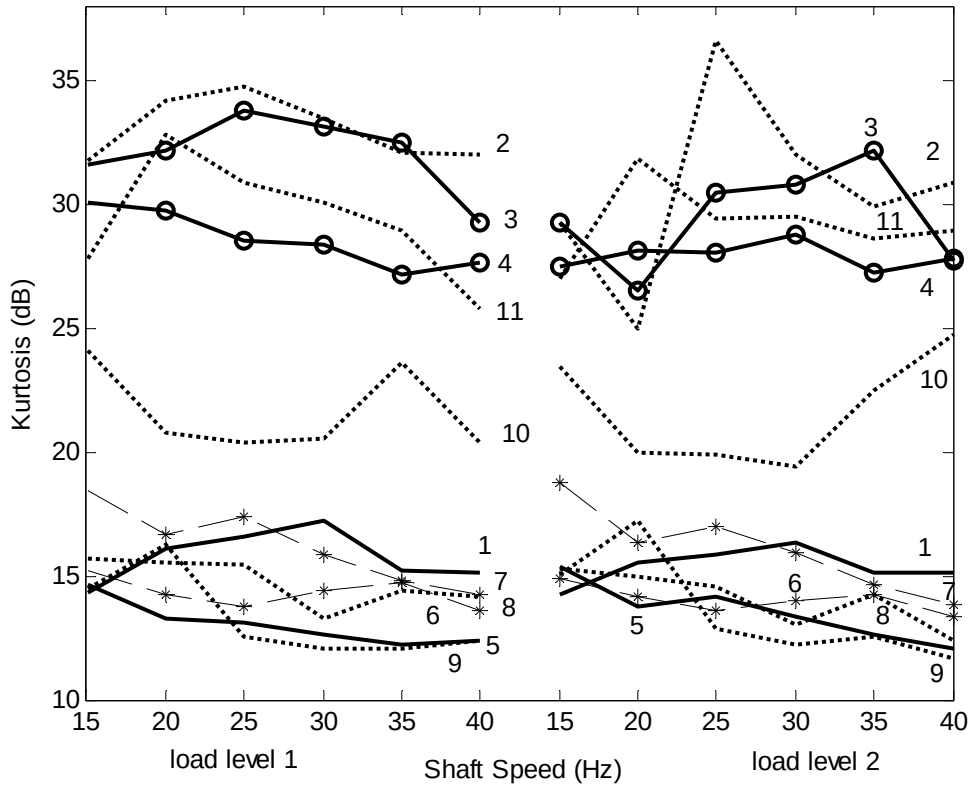

Figure B.1. Kurtosis trends for the ball and roller bearings with high-pass filtered (1000 Hz) signal $|S_1|$: 1-4, ball bearings; 5-11, roller bearings; each number represents one bearing condition as listed in Table B.1.

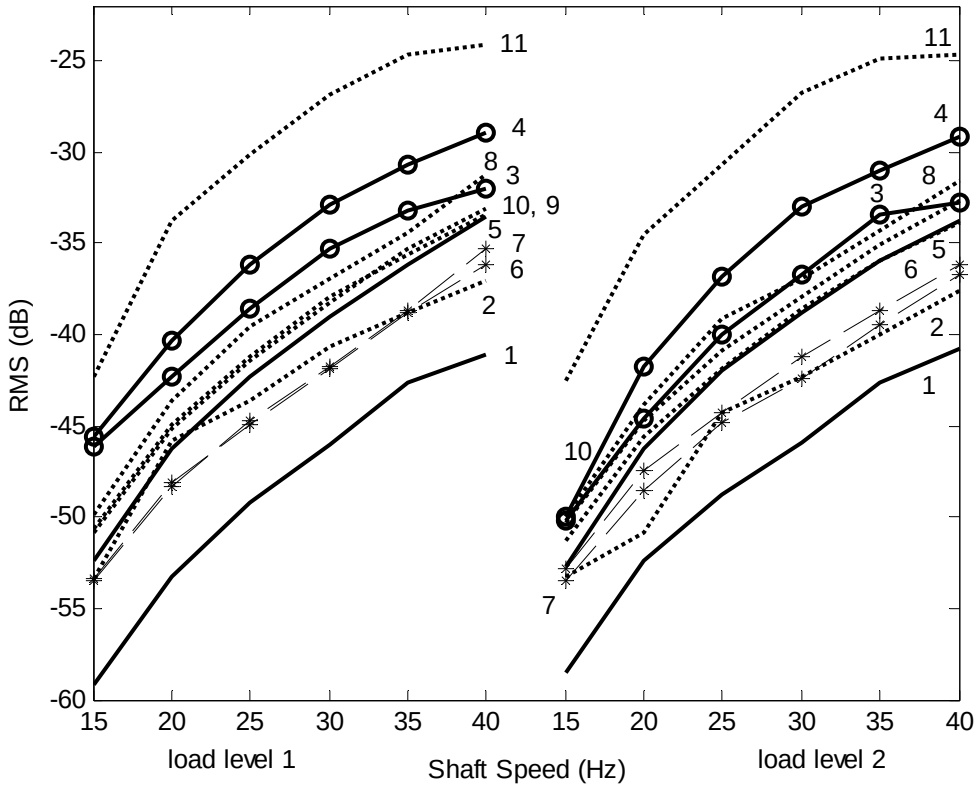

Figure B.2. RMS trends for the ball and roller bearings with high-pass filtered (2500 Hz) signal $|H(S_1)|$: 1-4, ball bearings; 5-11, roller bearings; each number represents one bearing condition as listed in Table B.1.

In the following section, a GP-based feature reconstruction technique is proposed so as to enhance the discrimination properties of these classical indexes.

B.4 A GP-Based Feature Reconstruction Approach

The GP paradigm has been genetically applied to breed computer program populations thanks to its domain-independent characteristics [108]. The GP employs three basic operations to produce a new generation: tree reproduction, crossover, and mutation. In tree reproduction, one individual tree (i.e., a computer program) is entirely selected, with a probability based on a fitness measure for the direct inclusion in the next generation of the population. In tree crossover, two random nodes are chosen from both parent trees, and the respective branches are then swapped to create the offspring; there is no bias towards choosing internal or terminal nodes at the crossing sites. In tree mutation, a

random node is chosen from the parent tree and substituted by a new random tree with the terminals and functions available. The generated random tree also obeys the size/depth restrictions imposed on the trees created for the initial generation. Some detailed description of the GP can be found in [108].

The GP has been applied to fault detection in some simplified mechanical systems [101, 109, 110]. For example, features were reconstructed in [101] by changing the GP parameters, such as generations, population size, and maximum tree depth. This method, however, was subjective; no algorithm was given for the optimization of these parameters. In this work, a novel GP-based approach is suggested to integrate the features from two classical indicators for a more robust feature formulation. An automatic adaptation procedure is applied to compute genetic operation probabilities of occurrence [111]. Each suggested fitness function prompts GP to differentiate one bearing condition from the others, which can lead to a more discriminatory feature. The data obtained from the kurtosis and RMS will be regarded as the GP terminals. The functions to dynamically manipulate the terminals are summarized in Table B.4.

Table B.4. GP operators

Operator	Description	Operator	Description
plus (a,b)	add a and b	log (a)	natural logarithm: 0 if a=0; ln$(\lvert a \rvert)$ (otherwise)
minus (a,b)	subtract a from b	sqrt (a)	square root: 0 (if a<=0); sqrt (a) (otherwise)
times (a,b)	multiply a and b	sin (a)	trigonometric function: sin (a)
abs (a)	absolute value of a	cos (a)	trigonometric function: cos (a)
negator (a)	negation: -a	tan (a)	trigonometric function: tan (a)
square (a)	power 2: a^2	divide (a,b)	division: a (if b=0); a/b (otherwise)

The critical procedure in implementing a GP algorithm is how to appropriately set up the fitness measure, which acts as a primary mechanism to communicate between a high-level statement of the requirements and a problem's solution to the GP paradigm. In this work, the Fisher criterion [112] is adopted to design the fitness functions. The index ρ is suggested as

$$\rho(c_i, c_j) = \frac{\left| m_{c_i} - m_{c_j} \right|}{\sqrt{\sigma_{c_i}^2 + \sigma_{c_j}^2}} \tag{B.5}$$

where m and σ^2 represent the mean and variance of the samples, respectively, and the subscripts denote the classes to be classified. Each class represents one of the four bearing health conditions: HY, RD, ID or OD. Correspondingly, three new features will be reconstructed to separate these four classes. The samples used to calculate the mean m_{c_i} and the variance $\sigma_{c_i}^2$ are the data resulted from the candidate GP program (nonlinear mathematical manipulation of candidate terminals and candidate operators) on the vibration signatures corresponding to the c_ith bearing condition. The index ρ is applied to maximize the distance between the geometric centers of two classes while minimizing the variances within each class. Therefore, ρ can represent the degree of class separability. The fitness measures are proposed as

$$p_1 = Min([\rho(c_1, c_2), \rho(c_1, c_3), \rho(c_1, c_4)]), \tag{B.6}$$

$$p_2 = Min([\rho(c_2, c_3), \rho(c_2, c_4)]), \tag{B.7}$$

$$p_3 = \rho(c_3, c_4), \tag{B.8}$$

where p_i, i = 1, 2, 3, is the fitness measure when the ith new feature is formulated, and $Min(\cdot)$ is the function to find the minimum scalar in a vector. A greater p_i corresponds to a more pronounced class separation. The first class c_1 is chosen as OD, which can be differentiated from the other classes more possibly as discussed in Chapter 2. The classes c_3, c_4 are chosen to represent RD and ID in this case, because the diagnostics of these two types of defects are more challenging due to the non-stationary resonance features generated [26, 98]. The purpose of designing such fitness functions is to facilitate the generation of new features that can make each class of bearing conditions differentiable from the others. This method has a potential to be an effective fault detection technique especially when the bearing conditions cannot be easily classified by using the classical techniques.

Figure B.3 shows that the OD condition can be clearly recognized by using the first reconstructed feature; its robustness against the speed and load variations is also demonstrated.

Figure B.4 shows the performance of the second formulated feature by which the healthy bearing conditions can be separated from the RD and ID bearing conditions.

The most challenging task in this case is how to properly separate the RD bearing condition from the ID condition. This task is eventually accomplished by the third formulated feature, as illustrated in Figure B.5. For the purpose of illustration, the second reconstructed feature is demonstrated in Figure B.6.

To further validate these generated features, a new set of vibration signals are collected at the shaft speed of 32 Hz and the loads of 1.5 and 2.0 N·m, respectively. These signals are subsequently processed using the GP programs to reconstruct new features. The verification results are demonstrated in Figure B.7. It can be seen that the four bearing conditions can be classified clearly.

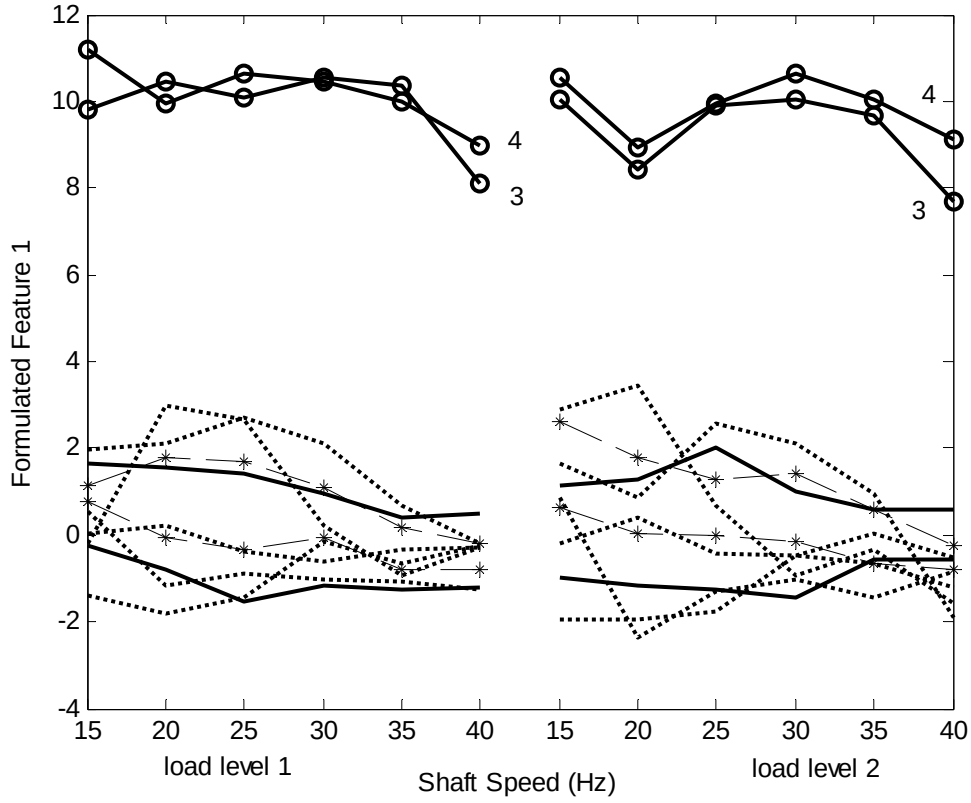

Figure B.3. Classification results using the first formulated feature by (B.6): solid lines, HY bearings; dashed lines with stars, RD bearings; dotted lines, ID bearings; solid lines with circles, OD bearings.

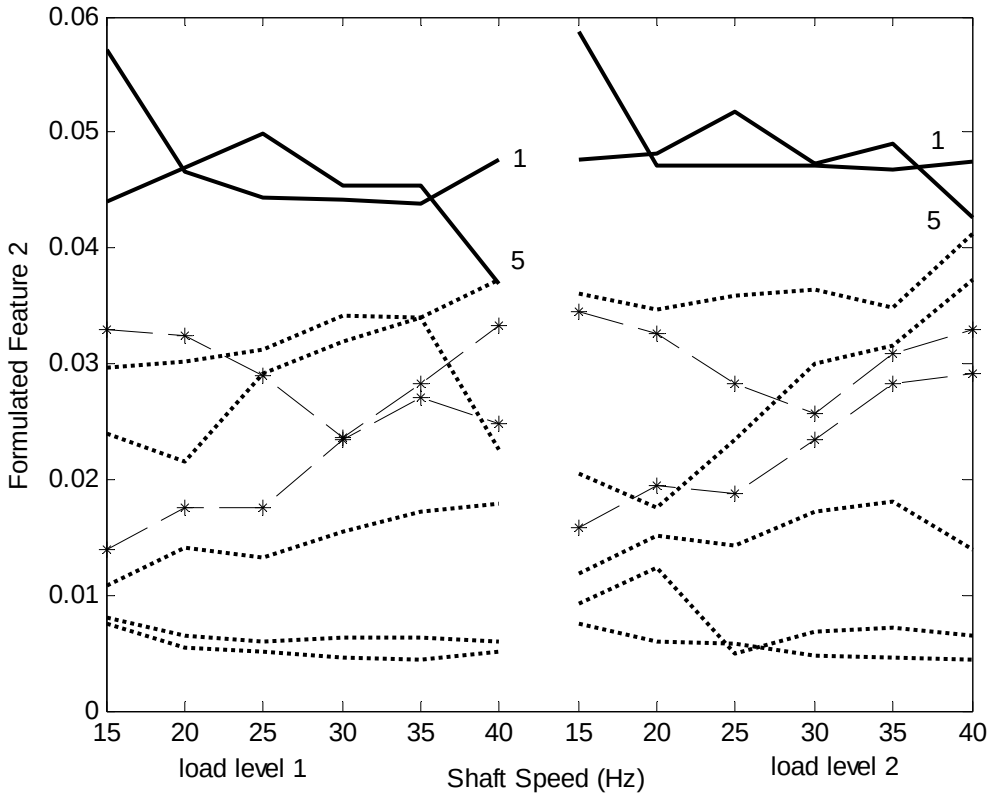

Figure B.4. Classification results using the second formulated feature by (B.7): solid lines, HY bearings; dashed lines with stars, RD bearings; dotted lines, ID bearings.

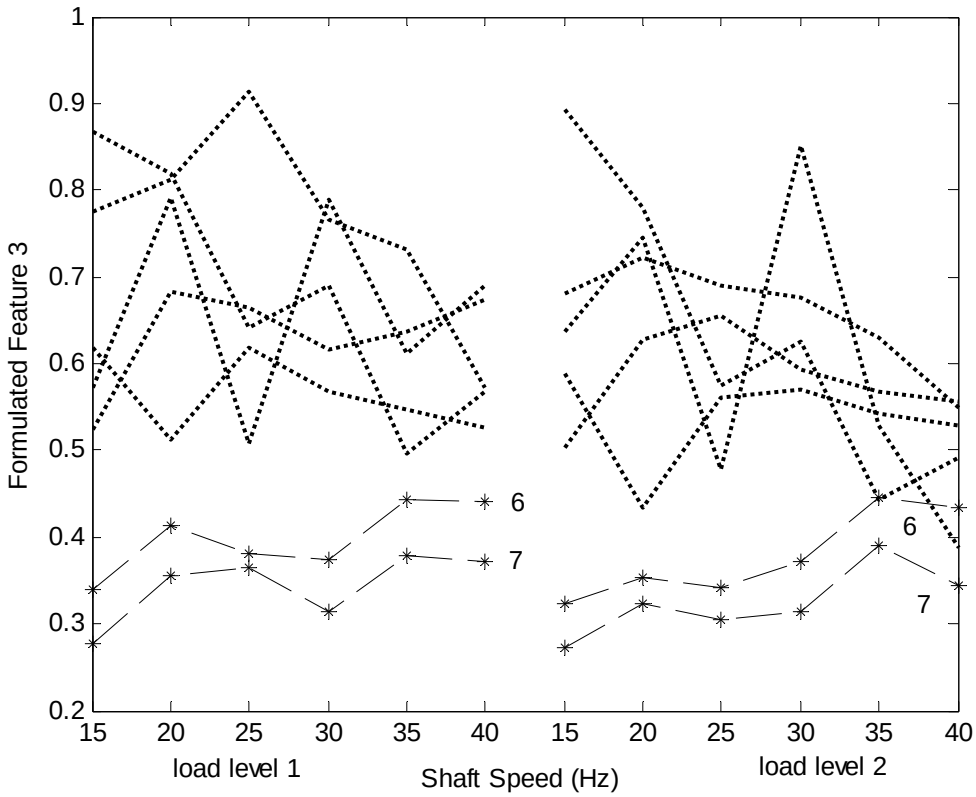

Figure B.5. Classification results using the third formulated feature by (B.8): dashed lines with stars, RD bearings; dotted lines, ID bearings.

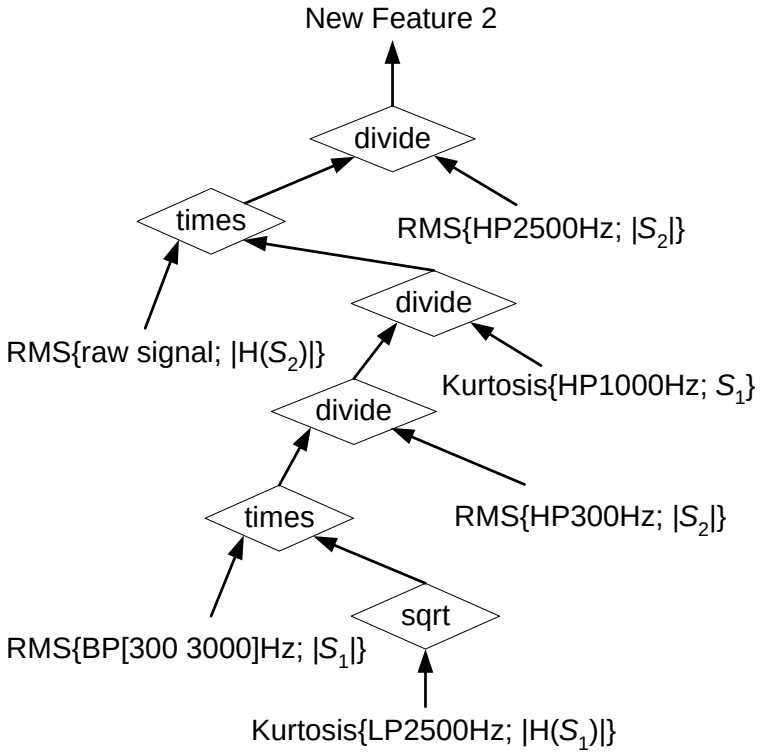

Figure B.6. Program tree that generates the second formulated feature by (B.7).

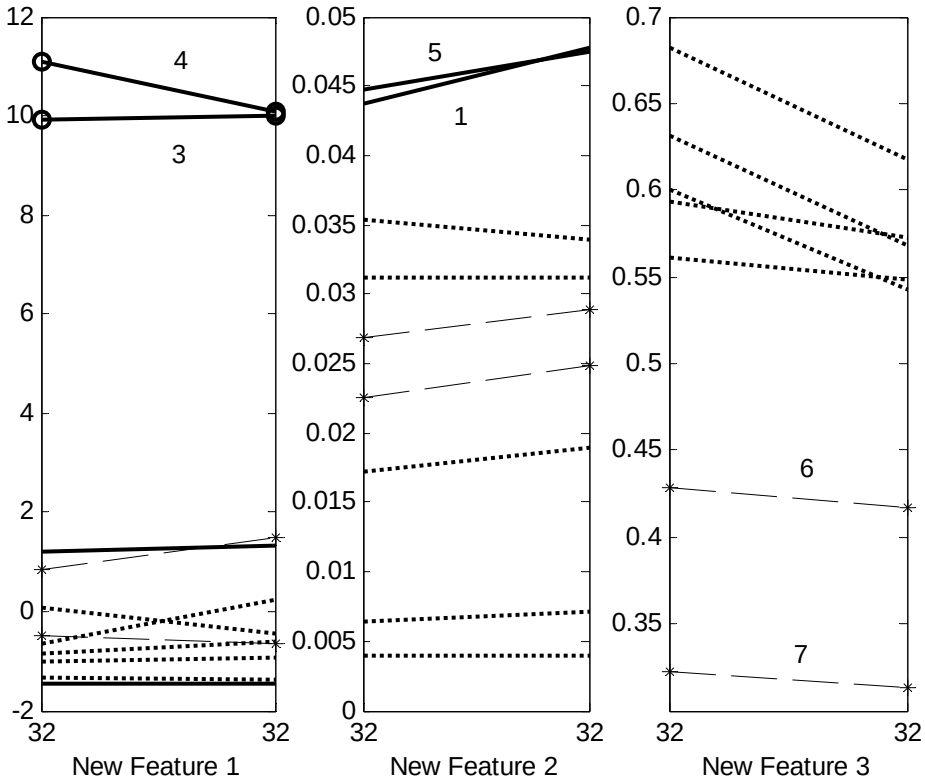

Figure B.7. Classification results using the newly-formulated features by (B.6) - (B.8).

The proposed technique in Section B.4 provides a promising approach to reconstructing features which would be useful to specific fault diagnostic applications. This technique can be used as a post processing technique to further enhance the feature characteristics. However, advanced research needs to be conducted to investigate the analytical relationships between the classical features and the formulated features such that these constructed features can be physically interpretable.

Bibliography

[1] P. Timmins, *Solutions to Equipment Failures*, Materials Park, OH: ASM International, 1999.

[2] A. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, 20, pp. 1483-1510, 2006.

[3] D. Boulahbal, F. Golnaraghi, and F. Ismail, "Amplitude and phase wavelet maps for the detection of cracks in geared systems," *Mechanical Systems and Signal Processing*, 13(3), pp. 423-436, 1999.

[4] W. Wang, F. Ismail, and F. Golnaraghi, "Assessment of gear damage monitoring techniques using vibration measurements," *Mechanical Systems and Signal Processing*, 15(5), pp. 905-922, 2001.

[5] W. Wang, F. Golnaraghi, and F. Ismail, "Condition monitoring of a multistage printing press," *Journal of Sound and Vibration*, 270, pp. 755-766, 2004.

[6] W. Wang, F. Golnaraghi, and F. Ismail, "Prognosis of machine health condition using neuro-fuzzy systems," *Mechanical Systems and Signal Processing*, 18(4), pp. 813-831, 2004.

[7] W. Wang, F. Ismail, and F. Golnaraghi, "A neuro-fuzzy approach for gear system monitoring," *IEEE Transactions on Fuzzy Systems*, 12(5), pp. 710-723, 2004.

[8] W. Wang, "An adaptive predictor for dynamic system forecasting," *Mechanical Systems and Signal Processing*, 21(2), pp. 809-823, 2007.

[9] W. Wang, "An intelligent system for machinery condition monitoring," *IEEE Transactions on Fuzzy Systems*, 16(1), pp. 110-122, 2008.

[10] P. Mcfadden and J. Smith, "Signal processing technique for detecting local defects in a gear from the signal average of the vibration," *Proceedings of the Institution of Mechanical Engineers, Part C: Mechanical Engineering Science*, 199, pp. 287-292, 1985.

[11] A. Choudhury and N. Tandon, "A theoretical model to predict vibration response of rolling bearings to distributed defects under radial load," *ASME, Journal of Vibration and Acoustics*, 120(1), pp. 214-220, 1998.

[12] M. Washo, "A quick method of determining root causes and corrective actions of failed ball bearing," *Lubrication Engineering*, 52(3), pp. 206-213, 1996.

[13] N. Tandon and A. Choudhury, "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," *Tribology International*, 32, pp. 469-480, 1999.

[14] Y. Su and S. Lin, “On initial fault detection of a tapered roller bearing: Frequency domain analysis,” *Journal of Sound and Vibration*, 155, pp. 75-84, 1992.

[15] N. Tandon, “A comparison of some vibration parameters for the condition monitoring of rolling element bearings,” *Measurement*, 12, pp. 285-289, 1994.

[16] B. Samanta and K. Al-Balushi, “Artificial neural network based fault diagnostics of rolling element bearings using time-domain features,” *Mechanical Systems and Signal Processing*, 17, pp. 317-328, 2003.

[17] R. Yan and R. Gao, “An efficient approach to machine health diagnosis based on harmonic wavelet packet transform,” *Robotics and Computer-Integrated Manufacturing*, 21, pp. 291-301, 2005.

[18] R. Heng and M. Nor, “Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition,” *Applied Acoustics*, 53, pp. 211-226, 1998.

[19] L. Rende and T. Dehua, “Using oil analysis to study the wear condition of bearings in trunnion of converter during/after run-in period,” *Proceedings of the 5th International Conference on Quality, Reliability and Maintenance QRM 2004*, pp. 101-104, 2004.

[20] J. Yi, J. Henao-Sepulveda, and M. Toledo-Quinones, “Wireless temperature sensor for bearing health monitoring,” *Proceedings of SPIE 5391*, pp. 368-376, 2004.

[21] R. Schoen, T. Habetler, F. Kamran, and R. Bartheld, “Motor bearing damage detection using stator current monitoring,” *IEEE Transactions on Industry Applications*, 13, pp. 1274-1279, 1995.

[22] R. Randall, J. Antoni, and S. Chobsaard, “The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals,” *Mechanical Systems and Signal Processing*, 15, pp. 945-962, 2001.

[23] J. Stack, T. Habetler, and R. Harley, “Fault-signature modeling and detection of inner-race bearing faults,” *IEEE Transactions on Industrial Application*, 42, pp. 61-68, 2006.

[24] Q. Du and S.Yang, “Improvement of the EMD method and applications in defect diagnosis of ball bearings,” *Measurement Science and Technology*, 17, pp. 2355-2361, 2006.

[25] G. Luo, D. Osypiw, and M. Irle, “On-line vibration analysis with fast continuous wavelet algorithm for condition monitoring of bearing,” *Journal of Vibration and Control*, 9, pp. 931-947, 2003.

[26] P. Mcfadden and J. Smith, “Vibration monitoring of rolling element bearings by the high-frequency resonance technique-a review,” *Tribology International*, 17, pp. 3-10, 1984.

[27] J. Stack, R. Harley, and T. Habetler, "An amplitude modulation detector of fault diagnosis in rolling element bearings," *IEEE Transactions on Industrial Electronics*, 51, pp. 1097-1102, 2004.

[28] Y. Choi and Y. Kim, "Fault detection in a ball bearing system using minimum variance cepstrum," *Measurement Science and Technology*, 18, pp. 1433-1440, 2007.

[29] T. Kaewkongka, Y. Au, R. Rakowski, and B. Jones, "A comparative study of short time Fourier transform and continuous wavelet transform for bearing condition monitoring," *International Journal of COMADEM*, 6, pp. 41-48, 2003.

[30] B. Kim, S. Lee, M. Lee, J. Ni, J. Song, and C. Lee, "A comparative study on damage detection in speed-up and coast-down process of grinding spindle-typed rotor-bearing system," *Journal of Materials Processing Technology*, 187, pp. 30-36, 2007.

[31] X. Li and X. Yao, "Multi-scale statistical process monitoring in machining," *IEEE Transactions on Industrial Electronics*, 52, pp. 924-927, 2005.

[32] X. Fan and M. Zuo, "Gearbox fault detection using Hilbert and wavelet packet transform," *Mechanical Systems and Signal Processing*, 20, pp. 966-982, 2006.

[33] Q. Sun and Y. Tang, "Singularity analysis using continuous wavelet transform for bearing fault diagnosis," *Mechanical Systems and Signal Processing*, 16, pp. 1025-1041, 2002.

[34] C. Wang and R. Gao, "Wavelet transform with spectral post-processing for enhanced feature extraction," *IEEE Transactions on Instrumentation and Measurement*, 52, pp. 1296-1301, 2003.

[35] N. Nikolaou and I. Antoniadis, "Demodulation of vibration signals generated by defects in rolling element bearings using complex shifted morlet wavelets," *Mechanical Systems and Signal Processing*, 16, pp. 677-694, 2002.

[36] J. Cheng, D. Yu, and Y. Yang, "Time-energy density analysis based on wavelet transform," *NDT&E International*, 38, pp. 569-572, 2005.

[37] M. Patil, J. Mathew, and P. Rajendrakumar, "Bearing signature analysis as a medium for fault detection: a review," *ASME, Journal of Tribology*, 130, pp. 1-7, 2008.

[38] J. Gertler, *Fault Detection and Diagnosis in Engineering Systems*, New York: Marcel Dekker, 1998.

[39] B. Paya, I. Esat, and M. Badi, "Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor," *Mechanical Systems and Signal Processing*, 11(5), pp. 751-765, 1997.

[40] R. Isermann, "Supervision, fault-detection and fault-diagnosis methods – An introduction," *Control Engineering Practice*, 5(5), pp. 639-652, 1997.

[41] R. Patton, P. Frank, and R. Clark, *Issues of Fault Diagnosis for Dynamic Systems*, London: Springer, 2000.

[42] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, New York: John Wiley & Sons, 2001.

[43] H. Chin, K. Danai, and D. Lewicki, "Fault detection of helicopter gearboxes using the multi-valued influence matrix method," *ASME, Journal of Mechanical Design*, 117, pp. 248-253, 1995.

[44] W. Wang and L. Song, "A multi-supervision system for the quality control of a CNC system," *Journal of Computer Aided Engineering*, 18(3), pp. 27-35, 1996.

[45] J. Bezdek and L. Kuncheva, "Nearest prototype classifier designs: An experimental study," *International Journal of Intelligent Systems*, 16(12), pp. 1445-1473, 2001.

[46] L. Jack, A. Nandi, and A. McCormick, "Diagnosis of rolling element bearing faults using radial basis function networks," *Applied Signal Processing*, 6, pp. 25-32, 1999.

[47] R. Isermann, "On fuzzy logic applications for automatic control, supervision, and fault diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Human*, 28(2), pp. 221-235, 1998.

[48] T. Liu, J. Singonahalli, and N. Iyer, "Detection of Roller bearing defects using expert system and fuzzy logic," *Mechanical systems and signal processing*, 10(5), pp. 595-614, 1996.

[49] G. Castellano, A. Fanelli, and C. Mencar, "An empirical risk functional to improve learning in a neuro-fuzzy classifier," *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 34(1), pp. 725-731, 2004.

[50] J. Wang and C. Lee, "Self-adaptive neuro-fuzzy inference systems for classification applications," *IEEE Transactions on Fuzzy Systems*, 10(6), pp. 790-802, 2002.

[51] O. Uluyol, K. Kim, and E. Nwadiogbu, "Synergistic use of soft computing technologies for fault detection in gas turbine engines," *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews*, 36(6), pp. 476-484, 2006.

[52] W. Staszewski and K. Worden, "Classification of faults in gearboxes – pre-processing algorithms and neural networks," *Neural Computing and Applications*, 5, pp. 160-183, 1996.

[53] T. Kim and C. Li, "Feedforward neural networks for fault diagnosis and severity assessment of a screw compressor," *Mechanical Systems and Signal Processing*, 9(5), pp. 485-496, 1995.

[54] C. Li and Y. Fan. “Recurrent neural networks for fault diagnosis and severity assessment of a screw compressor,” *Journal of Dynamic Systems, Measurement and Control*, 121(4), pp. 724-729, 1999.

[55] P. Monsen, M. Dwonczyk, and E. Manolakos, “Analog neural networks based helicopter gearbox health monitoring system,” *Journal of the Acoustical Society of America*, 96(6), pp. 3235-3249, 1995.

[56] W. Wang, “An enhanced diagnostic system for gear system monitoring,” *IEEE Trans. Systems, Man and Cybernetics – Part B: Cybernetics*, 38(1), pp.102-112, 2008.

[57] P. Xu and A. Chan, “Fast and robust neural network based wheel bearing fault detection with optimal wavelet features,” *Proceedings of the 2002 International Joint Conference*, 3, pp. 2076-2080, 2002.

[58] D. Ruan and E. Kerre, *Fuzzy If-Then Rules in Computational Intelligence: Theory and Applications*, Boston: Kluwer Academic, 2000.

[59] F. Karray and C. De Silva, *Soft Computing and Intelligent Systems Design: Theory, Tools, and Applications*, New York: Pearson/Addison Wesley, 2004.

[60] L. Kuncheva, “On the equivalence between fuzzy and statistical classifiers,” *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 4(3), pp. 245-253, 1996.

[61] L. Kuncheva, “How good are fuzzy if-then classifiers,” *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 30(3), pp. 501-509, 2000.

[62] Y. Chen, “A fuzzy decision system for fault classification under high levels of uncertainty,” *Transactions of the ASME, Journal of Dynamic Systems, Measurement and Control*, 117(1), pp. 108-115, 1995.

[63] S. Li and M. Elbestawi, “Fuzzy clustering for automated tool condition monitoring in machining,” *Mechanical Systems and Signal Processing*, 10(5), pp. 533-550, 1996.

[64] L. Zeng and H. Wang, “Machine-fault classification: A fuzzy set approach,” *International Journal of Advanced Manufacturing Technology*, 6, pp. 83-94, 1991.

[65] S. Abe, “Techniques in fuzzy rules determination and their applications to pattern classification,” *Fuzzy Theory Systems: Techniques and Applications*, 3, pp. 1051-1079, 1999.

[66] C. Mechefske, “Objective machinery fault diagnosis using fuzzy logic,” *Mechanical Systems and Signal Processing*, 12(6), pp. 855-862, 1998.

[67] X. Lou and K. Loparo, "Bearing fault diagnosis based on wavelet transform and fuzzy inference," *Mechanical Systems and Signal Processing*, 18, pp. 1077-1095, 2004.

[68] J. Altmann and J. Mathew, "Multiple band-pass autoregressive demodulation for rolling-element bearing fault diagnosis," *Mechanical Systems and Signal Processing*, 15(5), pp. 963-977, 2001.

[69] J. Liu, W. Wang, and F. Golnaraghi, "A GA-based feature optimization technique for bearing fault diagnostics," *International Journal of Computer Science and Network Security*, 7, pp. 62-68, 2007.

[70] M. Pourahmadi, *Foundation of Time Series Analysis and Prediction Theory*, New York: Wiley, 2001.

[71] C. Groot and D. Wurts, "Analysis of univariate time series with connectionist nets: A case study of two classical examples," *Neurocomputing*, 3, pp. 177-192, 1991.

[72] H. Tong and S. Lim, "Threshold autoregression, limited cycles and cyclical data," *Journal of the Royal Statistical Society, Series B*, 42(3), pp. 245-292, 1980.

[73] T. Rao, "On the theory of bilinear models," *Journal of the Royal Statistical Society*, *Series B*, 43, pp. 244-255, 1981.

[74] J. Friedman and W. Stuetzle, "Projection pursuit regression," *The Annals of Statistics*, 76, pp. 817-823, 1981.

[75] J. Friedman, "Multivariate adaptive regression splines," *The Annals of Statistics*, 19(1), pp. 1-67, 1991.

[76] D. Brillinger, "The identification of polynomial systems by means of higher order spectra," *Jouranl of Sound and Vibration*, 12, pp. 301-313, 1970.

[77] C. Li and J. Yoo, "Prognosis of gear tooth crack growth," *Proceedings of the 52th Meeting of the Society of Mechanical Failures Prevention Technology*, Virginia Beach, USA, pp. 419-428, 1998.

[78] N. Tandon and A. Choudhury, "An analytical model for the prediction of the vibration response of rolling element bearings due to localized defect," *Mechanical Systems and Signal Processing*, 205(3), pp. 275-292, 1997.

[79] K. Mori, N. Kasashima, T. Yoshioka, and Y. Ueno, "Prediction of spalling on a ball bearing by applying discrete wavelet transform to vibration signal," *Wear*, 195, pp. 612-168, 1996.

[80] R. Yaffee and M. McGee, *Introduction to Time Series Analysis and Forecasting: with applications of SAS and SPSS*, San Diego: Academic Press, 2000.

[81] D. Park, M. El-Sharkawi, R. Marks, L. Atlas, and M. Damborg, "Electric load forecasting using an artificial neural networks," *IEEE Transactions on Power Systems*, 6(2), pp. 442-449, 1991.

[82] A. Refenes and M. Azema-Barac, "Neural network applications in financial asset management," *Neural Computation Applications*, 2, pp. 13-39, 1994.

[83] A. Atiya, S. El-Shoura, S. Shaheen, M. El-Sherif, "A comparison between neural-network forecasting techniques-case study: river flow forecasting," *IEEE Transactions on Neural Networks*, 10(2), pp. 402-409, 1999.

[84] J. Connor, R. Martin, and L. Atlas, "Recurrent neural networks and robust time series prediction," *IEEE Transactions on Neural Networks*, 5(2), pp. 240-254, 1994.

[85] A. Parlos, K. Chong, and A. Atiya, "Application of the recurrent multiplayer perceptron in modeling complex process dynamics," *IEEE Transactions on Neural Networks*, 5(2), pp. 255-266, 1994.

[86] A. Lucifredi, C. Mazzieri, and M. Rossi, "Application of multiregressive linear models, dynamic Kriging models and neural network models to predictive maintenance of hydroelectric power systems," *Mechanical Systems and Signal Processing*, 14(3), pp. 471-494, 2000.

[87] A. Parlos, O. Rais, and A. Atiya, "Multi-step-ahead prediction using dynamic recurrent neural networks," *Neural Networks*, 13, pp. 765-786, 2000.

[88] P. Tse and D. Atherton, "Prediction of machine deterioration using vibration based fault trends and recurrent neural networks," *Journal of Vibration and Acoustics*, 121, pp. 355-362, 1999.

[89] P. Vukovic, "One-step ahead predictive fuzzy controller," *Fuzzy Sets and Systems*, 122, pp. 107-115, 2001.

[90] J. Jang, C. Sun, and E. Mizutani, *Neuro-fuzzy Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Upper Saddle River, NJ: Prentice-Hall, 1997.

[91] E. Walter and L. Pronzato, *Identification of Parametric Models from Experimental Data*, Berlin: Springer, 1997.

[92] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley, MA: Wellesley-Cambridge Press, 1996.

[93] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, 36, pp. 961-1005, 1990.

[94] A. Girgis and B. Guy, "A computer based data acquisition system for teaching transients and switching phenomena and performing research on digital protection," *IEEE Transactions on Power Systems*, 3(3), pp. 1361-1368, 1988.

[95] B. Schneier, *Applied Cryptography: Protocols, Algorithms, and Source Code in C*, New York: Wiley, 1996.

[96] J. Lin and M. Zuo, "Extraction of periodic components for gearbox diagnosis combining wavelet filtering and cyclostationary analysis," *ASME, Journal of Vibration and Acoustics*, 126, pp. 449-451, 2004.

[97] J. Liu, W. Wang, F. Golnaraghi, and K. Liu, "Wavelet spectrum analysis for bearing fault diagnostics," *Measurement Science and Technology*, 19, pp. 1-9, 2008.

[98] J. Liu, W. Wang, and F. Golnaraghi, "An extended wavelet spectrum for bearing fault diagnostics," *IEEE Transactions on Instrumentation and Measurement*, 57(12), (12 pages), 2008.

[99] B. Samanta, "Artificial neural networks and genetic algorithms for gear fault detection," *Mechanical Systems and Signal Processing*, 18(5), pp. 1273-1282, 2004.

[100] H. Firpi and E. Goodman, "Swarmed feature selection," *Proceedings of 33rd Applied Imagery Pattern Recognition Workshop*, pp. 112-118, 2004.

[101] H. Guo, L. Jack, and A. Nandi, "Feature generation using genetic programming with application to fault classification," *IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics*, 35, pp. 89-99, 2005.

[102] H. Firpi and G. Vachtsevanos, "Genetically programmed-based artificial features extraction applied to fault detection," *Engineering Applications of Artificial Intelligence*, 21, pp. 558-568, 2008.

[103] D. Dyer and R. Stewart, "Detection of rolling element bearing damage by statistical vibration analysis," *ASME Journal of Mechanical Design*, 100, pp. 229-235, 1978.

[104] C. Pachaud, R. Salvetat, and C. Fray, "Crest factor and kurtosis contributions to identify defects inducing periodical impulsive forces," *Mechanical Systems and Signal Processing*, 11, pp. 903-916, 1997.

[105] C. Cristalli, N. Paone, and R. Rodriguez, "Mechanical fault detection of electric motors by laser vibrometer and accelerometer measurements," *Mechanical Systems and Signal Processing*, 20, pp. 1350-1361, 2006.

[106] T. Williams, X. Ribadeneira, S. Billington, and T. Kurfess, "Rolling element bearing diagnostics in run-to-failure lifetime testing," *Mechanical Systems and Signal Processing*, 15, pp. 979-993, 2001.

[107] G. Jimenez and A. Munoz, "Fault detection in induction motors using Hilbert and wavelet transforms," *Electrical Engineering*, 89, pp. 205-220, 2007.

[108] J. Koza, *Genetic Programming: On the programming of Computers by Means of Natural Selection*, Cambridge, MA: MIT Press, 1992.

[109] P. Chen, M. Taniguchi, T. Toyota, and Z. He, "Fault diagnosis method for machinery in unsteady operating condition by instantaneous power spectrum and genetic programming," *Mechanical Systems and Signal Processing*, 19, pp. 175-194, 2005.

[110] L. Zhang, L. Jack, and A. Nandi, "Fault detection using genetic programming," *Mechanical Systems and Signal Processing*, 19, pp. 271-289, 2005.

[111] L. Davis, "Adapting operator probabilities in genetic algorithms," *Proceedings of the Third International Conference on Genetic Algorithms*, San Mateo, CA, pp. 61-69, 1989.

[112] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd Edition, New York: Wiley, 2000.

[113] J. Liu, S. Ghafari, W. Wang, F. Golnaraghi, and F. Ismail, "Bearing fault diagnostics based on reconstructed features," *Proceedings of 43rd Annual Meeting of IEEE Industry Applications Society*, Edmonton, Alberta, Canada, Oct. 2008, (to appear).

[114] J. Vass, R. Smid, R. Randall, P. Sovka, C. Cristalli, and B. Torcianti, "Avoidance of speckle noise in laser vibrometry by the use of kurtosis ratio: Application to mechanical fault diagnostics," *Mechanical Systems and Signal Processing*, 22, pp. 647-671, 2008.

[115] P. Estocq, F. Bolaers, J. Dron, and L. Rasolofondraibe, "Method of de-noising by spectral subtraction applied to the detection of rolling bearing defects," *Journal of Vibration and Control*, 12(2), pp. 197-211, 2006.

[116] J. Antoni, "The spectral kurtosis: a useful tool for characterizing non-stationary signals," *Mechanical Systems and Signal Processing*, 20, pp. 282-307, 2006.

[117] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*, Cambridge, MA: MIT Press, 1983.

[118] W. Wang, "An enhanced diagnostic system for gear system monitoring," *IEEE Trans. Systems, Man and Cybernetics – Part B: Cybernetics*, 38(1), pp.102-112, 2008.

[119] C. Lin and C. Lee, *Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems*, Upper Saddle River, NJ: Prentice-Hall, 1996.

[120] J. Liu, W. Wang, and F. Golnaraghi, "An enhanced diagnostic system for bearing fault diagnostics," submitted to *IEEE Transactions on Instrumentation and Measurement*, Ref. No.: IM-08-1177, 2008.

[121] D. Chelidze and J. Cusumano, "A dynamical systems approach to failure prognosis," *Journal of Vibration & Acoustics*, 126, pp. 1-7, 2004.

[122] S. Gupta and A. Ray, "Real-time fatigue life estimation in mechanical structure," *Measurement Science and Technology*, 18, pp. 1947-1957, 2007.

[123] C. Li and H. Lee, "Gear fatigue crack prognosis using embedded model gear dynamic model and fracture mechanics," *Mechanical Systems and Signal Processing*, 9, pp. 836-846, 2005.

[124] A. Atiya, S. El-Shoura, S. Shaheen, and M. El-Sherif, "A comparison between neural-network forecasting techniques-case study: river flow forecasting," *IEEE Transactions on Neural Networks*, 10, pp. 402-409, 1999.

[125] Y. Liang and X. Liang, "Improving signal prediction performance of neural networks through multiresolution learning approach," *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 36, pp. 341-352, 2006.

[126] D. Husmeier, *Neural Networks for Conditional Probability Estimation: Forecasting Beyond Point Prediction*, London: Springer, 1999.

[127] J. Korbicz, *Fault Diagnosis: Models, Artificial Intelligence, and Applications*, Berlin: Springer, 2004.

[128] J. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 23, pp. 665-685, 1993.

[129] M. Figueiredo, R. Ballini, S. Soares, M. Andrade, and F. Gomide, "Learning algorithms for a class of neurofuzzy network and applications," *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Review*, 34, pp. 293-301, 2004.

[130] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," *IEEE Transactions on Fuzzy Systems*, 13, pp. 428-435, 2005.

[131] D. Nauck, "Adaptive rule weights in neuro-fuzzy systems," *Neural Computing and Applications*, 9, pp. 60-70, 2000.

[132] M. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," *Science*, 197, pp. 287-289, 1977.

[133] J. Liu, W. Wang, and F. Golnaraghi, "A multi-step predictor with variable input pattern for dynamic system forecasting," *Mechanical Systems and Signal Processing*, Ref. No.: MSSP-08-128, 2008, (accepted).

[134] F. Bouslama and A. Ichikawa, "Fuzzy control rules and their natural control laws," *Fuzzy Sets and Systems*, 48, pp. 65-86, 1992.

[135] W. Siler and H. Ying, "Fuzzy control theory: the linear case," *Fuzzy Sets and Systems*, 33, pp. 275-290, 1989.

[136] C. Wong, C. Chou, and D. Mon, "Studies on the output of fuzzy controller with multiple inputs," *Fuzzy Sets and Systems*, 57, pp. 149-158, 1993.

[137] L. Wang, "Stable and optimal fuzzy control of linear systems," *IEEE Transactions on Fuzzy Systems*, 6(1), pp. 137-143, 1998.

[138] H. Sun and L. Liu, "A linear output structure for fuzzy logic controllers," *Fuzzy Sets and Systems*, 131, pp. 265-270, 2002.

[139] E. Kubica, D. Madill, and D. Wang, "Designing stable MIMO fuzzy controllers," *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 35(2), pp. 372-380, 2005.

[140] H. Ying, "Constructing nonlinear variable gain controllers via the Takagi-Sugeno fuzzy control," *IEEE Transactions on Fuzzy Systems*, 6(2), pp. 226-234, 1998.

[141] J. Kim and S. Oh, "A fuzzy PID controller for nonlinear and uncertain systems," *Soft Computing*, 4, pp. 123-129, 2000.

[142] A. Haj-Ali and H. Ying, "Structural analysis of fuzzy controllers with nonlinear input fuzzy sets in relation to nonlinear PID control with variable gains," *Automatica*, 40, pp. 1551-1559, 2004.

[143] D. Mba and A.M. AlGhamd, "A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size," *Mechanical Systems and Signal Processing*, 20(7), pp. 1537-1571, 2006.

[144] T.J. Harvey, R.J.K. Wood, and H.E.G. Powrie, "Electrostatic wear monitoring of rolling element bearings," *Wear*, 263, pp. 1492-1501, 2007.

[145] Y. Ohga, K. Moriguchi, S. Honda, and H. Nakagawa, "Fault diagnosis system for hydraulic turbine generator," *Transactions of the Institute of Electrical Engineers of Japan*, Part B, 122(4), pp. 492-497, 2002.

[146] M. Blodt, P. Granjon, B. Raison, and G. Rostaing, "Models for bearing damage detection in induction motors using stator current monitoring," *IEEE Transactions on Industrial Electronics*, 55(4), pp. 1813-1822, 2008.

[147] J.M. Canive, A. Petraqlia, and M.R. Petraqlia, "A 0.8 μm CMOS testable switched-capacitor filter for video frequency applications," *Analog Integrated Circuits and Signal Processing*, 48(2), pp. 133-141, 2006.