Computing Intelligence Technique and Multiresolution Data Processing for Condition Monitoring

Khalid Rabeyee

School of Computing and Engineering

This thesis is submitted to the School of Computing and Engineering, University of Huddersfield, in partial fulfilment of the requirements for the degree of Doctor of Philosophy

March 2019

COPYRIGHT STATEMENT

- ✓ The author of this thesis owns any copyright in it (the "Copyright") and s/he has given The University of Huddersfield the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.

- ✓ Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the University Library. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

- ✓ The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

ABSTRACT

Condition monitoring (CM) of rotary machines has gained increasing importance and extensive research in recent years. Due to the rapid growth of data volume, automated data processing is necessary in order to deal with massive data efficiently to produce timely and accurate diagnostic results. Artificial intelligence (AI) and adaptive data processing approaches can be promising solutions to the challenge of large data volume. Unfortunately, the majority of AI-based techniques in CM have been developed for only the post-processing (classification) stage, whereas the critical tasks including feature extraction and selection are still manually processed, which often require considerable time and efforts but also yield a performance depending on prior knowledge and diagnostic expertise.

To achieve an automatic data processing, the research of this PhD project provides an integrated framework with two main approaches. Firstly, it focuses on extending AI techniques in all phases, including feature extraction by applying Componential Coding Neural Network (CCNN) which has been found to have unique properties of being trained through unsupervised learning, capable of dealing with raw datasets, translation invariance and high computational efficiency. These advantages of CCNN make it particularly suitable for automated analyzing of the vibration data arisen from typical machine components such as the rolling element bearings which exhibit periodic phenomena with high non-stationary and strong noise contamination. Then, once an anomaly is detected, a further analysis technique to identify the fault is proposed using a multiresolution data analysis approach based on Double-Density Discrete Wavelet Transform (DD-DWT) which was grounded on over-sampled filter banks with smooth tight frames. This makes it nearly shift-invariant which is important for extracting non-stationary periodical peaks. Also, in order to denoise and enhance the diagnostic features, a novel level-dependant adaptive thresholding method based on harmonic to signal ratio (HSR) is developed and implemented on the selected wavelet coefficients. This method has been developed to be a semi-automated (adaptive) approach to facilitate the process of fault diagnosis. The developed framework has been evaluated using both simulated and measured datasets from typical healthy and defective tapered roller bearings which are critical parts of all rotating machines. The results have demonstrated that the CCNN is a robust technique for early fault detection, and also showed that adaptive DD-DWT is a robust technique for diagnosing the faults induced to test bearings. The developed framework has achieved multi-objectives of high detection sensitivity, reliable diagnosis and minimized computing complexity.

DECLARATION

I am the author of this thesis and to the best of my knowledge it contains no materials previously published or written by another person, I also declare that no portion of the presented work in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

DEDICATION

I would like to dedicate my work to those who have supported and encouraged me during my studies- my mother, my wife, my children and other family.

ACKNOWLEDGEMENTS

I am deeply indebted to Professor Fengshou Gu and Professor Andrew Ball for their valuable help, unlimited support and encouragement to make this dream true.

LIST OF CONTENTS

LIST OF TABLES

TABEL OF FIGURES

LIST OF ABBREVIATIONS

ANC	Adaptive Noise Cancellation
ADC	Analog-To-Digital Converter
AI	Artificial Intelligence
ANN	Artificial Neural Network
ADI	Average Discrimination Index
CA	Charge Amplifier
CCNN	Componential Coding Neural Network
CM	Condition Monitoring
CWT	Continues Wavelet Transform
DAQ	Data Acquisition System
DSP	Digital Signal Processing
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
DDFT	Double DFT
DDD-DWT	Double-Density Dual-Tree DWE
DD-DWT	Double-Density DWT
DT-DWT	Dual-Tree DWT
EMD	Empirical Mode Decomposition
FFT	Fast Furrier Transform
FFNT	Feed Forward Neural Network
HSR	Harmonic To Signal Ratio
HMM	Hidden Markov Model
HOS	Higher Order Statistics
HFRT	High-Frequency Resonance Technique
I/O	Input / Output
IFFT	Invers Fast Furrier Transform
KNN	K Nearest Neighbours
KFDA	Kernel Fisher Discriminant Analysis
LDA	Linear Discrimination Index
LP	Linear Prediction
MED	Minimum Entropy Deconvolution
MSB	Modulation Signal Bispectrum
PS	Power Spectrum
PCA	Principle Component Analysis
REM	Reconstruction Error Based Detection Model
RUL	Remaining Useful Life
REBs	Rolling Element Bearings
RMS	Root Mean Square
SGWT	Second Generation Wavelet Transform
SANC	Self-Adaptive Noise Cancellation
SOM	Self-Organizing Feature Maps
STFT	Short-Time Fourier Transform
SNR	Signal-To-Noise Ratio

SK	Spectral Kurtosis
SVM	Support Vector Machine
TRB	Tapered Roller Bearings
TSA	Time Synchronous Averaging
UDWT	Undecimated DWT
WT	Wavelet Transform
WVD	Wigner-Ville Distribution

LIST OF NOTATIONS

$x(t)$	Raw vibration signal
$X(f)$	Fourier transform
$X_{an}(f)$	Fourier transform of analytic signal
$p(f)$	Power spectrum
$\psi(t)$	Wavelet function
$W(x)$	Wavelet transform
σ	Standard deviation
a	scaling parameters
τ	translation parameters
$\psi^*(t)$	Complex conjugate of ψ(t)
$h(n)$	Wavelet Low pass filter
$g(n)$	Wavelet High pass filter
$\phi(t)$	Scaling function
$\psi_h(t)$	Real part of DT-DWT
$i\psi_g(t)$	Imaginary part of DT-DWT
$\psi_2(t)$	Second wavelet of DD-DWT
$h_0(n)$	Low pass filter of DD-DWT
$h_1(n)$	High pass filter of DD-DWT
$d_j^R(k)$	Real wavelet coefficients of DT-DWT
$s_j^R(k)$	Real Scaling coefficients of DT-DWT
$d_j^{Im}(k)$	Imaginary wavelet coefficients of DT-DWT
$s_j^{Im}(k)$	Imaginary Scaling coefficients of DT-DWT
$H[\cdot]$	Hilbert transform
α	Contact angle
d	Rolling element diameter
D	Pitch diameter
z	Rolling elements number
f_s	Shaft frequency
Er	Reconstruction error in CCNN
ξ	Output of neurones in neural networks
$\hat{x}$	Reconstructed signal
w	Weight vector
$cr(\cdot)$	Correlation function
$cv(\cdot)$	Convolution function
$F^{-1}(\cdot)$	Inverse of Fourier Transform
$r(\beta)$	Activation function
δ	Threshold parameter of CCNN neuron
s	Softness parameter of CCNN neuron
$y_j(x)$	The code of neuron function
a_j	Weight scales
E^T	Mean squared error MSE averaged over training datasets

$\langle\cdot\rangle_{\{x\}}$	Averaging over training data operator
λ	Neural network learning rate
R_u	MSE of unseen dataset
R_m	MSE of monitored dataset
ADI	Average Discrimination Index
$\widehat{w}_{li}$	Denoised wavelet coefficients
T	Nonlinear threshold function (soft, hard)
μ	Estimated threshold value
$\hat{s}$	Reconstructed signal
b_k^l	Disjoint blocks of Neighblock threshold
μ_n^*	The minimax risk bound
$W^{-1}(\cdot)$	Reconstruction function of wavelet coefficients
r	Harmonic ratio
a	Ratio of un-thresholded signal
b	Harmonics of thresholded signal
HSR	Harmonic ratio to signal ratio
f_r	Fault frequency
v	Shaft peaks
$\hat{S}$	Envelope spectrum of the reconstructed signal

LIST OF PUBLICATIONS

1) Khalid Rabeyee, Xiaoli Tang, Fengshou Gu, Andrew D. Ball. (2018) '*The Effect of the Wear Evolution on Vibration based Fault Detection in Tapered Roller Bearings*'. In: Fifteenth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, 10th - 12th September 2018, Nottingham, UK.

2) Khalid Rabeyee, Xiaoli Tang, Yuandong Xu, Dong Zhen, Fengshou Gu, Andrew D. Ball. (2018) '*diagnosing the Change in the Internal Clearances of Rolling Element Bearings based on Vibration Signatures'*. In: Proceedings of the 24th International Conference on Automation & Computing, Newcastle University, Newcastle upon Tyne, UK, 6-7 September 2018.

3) Khalid Rabeyee, Yuandong Xu, Aisha Alashter, Fengshou Gu, Andrew D. Ball. (2019) '*A Componential Coding Neural Network based Signal Modelling for Machinery Condition Monitoring*'. In Proceedings of the 32nd International Congress and Exhibition on Condition Monitoring and Diagnostic Engineering Management. 3-5 September 2019, University of Huddersfield, West Yorkshire, UK.

4) Khalid Rabeyee, Yuandong Xu, Samir Alabied, Fengshou Gu, Andrew D. Ball. (2019), *Extraction of Information from Vibration Data using Double Density Discrete Wavelet Analysis for Condition Monitoring'.* In Proceedings of Sixteenth International Conference on Condition Monitoring and Asset Management, 25th - 27th June 2019, Glasgow, UK

5) Khalid Rabeyee, Yuandong Xu, Fengshou Gu, Andrew D. Ball, *A Novel Wavelet Thresholding Method for Vibration Data Denoising and Diagnostic Feature Enhancement in Condition Monitoring'.* In: Proceedings of the 25th International Conference on Automation & Computing, Lancaster University, Lancaster UK, 5-7 September 2019.

6) Yuandong Xu, Dong Zhen, James Xi Gu, Khalid Rabeyee, Fulei Chu, Fengshou Gu, Andrew D. Ball, *Early Fault Detection in Rolling Bearings based on Autocorrelated Envelope Signals*, Mechanical Systems and Signal Processing, (under review).

7) Aisha Alashter, Yunpeng Cao, Khalid Rabeyee, Fengshou Gu and Andrew D. Ball' *Bond Graph Modelling for Condition Monitoring of Electromechanical Systems Based on Motor Current Analysis*'. In Proceedings of the 32nd International Congress and Exhibition on Condition Monitoring and Diagnostic Engineering Management. 3-5 September 2019, University of Huddersfield, West Yorkshire, UK.

8) Samir Alabied1, Alsadak Daraz, Khalid Rabeyee, Ibrahim Alqatawneh, Fengshou Gu and Andrew D. Ball, *Motor Current Signal Analysis Based on Machine Learning for Centrifugal Pump Fault Diagnosis'*. In: Proceedings of the 25th International Conference on Automation & Computing, Lancaster University, Lancaster UK, 5-7 September 2019.

CHAPTER ONE

INTRODUCTION

In this chapter, the importance of condition monitoring and fault diagnosis are outlined in association with various monitoring techniques. Vibration based monitoring is paid more attention in terms of data processing techniques. Moreover, this chapter presents the motivation of the research, on which the aim and objectives of this research are put forward. Finally, it presents the structure of the thesis.

1.1 Background

Data processing has become a key factor in almost every industrial process due to the massive quantity of data produced by modern machines and instruments. With the development of communication network and information technology, machinery systems are becoming more systematic, automated, complicated and expensive with a lower tolerance for performance degradation, safety hazards and productivity decline. Condition Monitoring (CM) is defined as the procedure of monitoring and analysing some parameters of a system condition[1]. CM has been the subject of interest in various industries to ensure the reliability and state of machine health. In fact, condition monitoring can be considered as a data processing system as it can be carried out through modelling, signal processing and artificial intelligence. However, finding useful information in a given dataset with the increase of quantity and complexity is becoming more critical and challenging task[2]. The so-called feature extraction and selection has gained more importance and attention in recent years. Feature extraction and selection can be seen as a general method and is applied for several reasons depending on the goal of the application, it can be used to reduce the dimensionality of the data or as pattern recognition, also for classification or predictions tasks, removing irrelevant and redundant data etc. I.e. extracted features can be used as a representation of the data. Conventional manual approaches are thought to rely entirely on the prior knowledge of the expertise to carry out the task of extracting useful and informative information. With the development of AI approaches and adaptive data processing algorithms, the process of extracting and selecting relevant and informative information can be facilitated and also automated[3, 4]. In the field of vibration-based CM, feature extraction and selection are critical and have a real impact on the diagnosis process results with respect to effectiveness and efficiency. Therefore in this research, automated and semi-automated approaches are investigated as an integrated framework for vibration data analysis. Based on AI technique, the features are

automatically extracted and used to build a data model for anomaly detection, whilst for fault diagnosis, features are adaptively extracted using multiresolution data analysis based adaptive thresholding method.

1.2 Condition Monitoring Steps

Generally, as illustrated in Figure 1-1, CM consists of three main key steps namely; data acquisition, data processing and decision making[5]. Over recent years, an extensive research effort has been done in each of the CM steps. This has led to the emergence of a verity of methods, techniques and algorithms. This thesis is considering data processing step and particularly, to automate the task of diagnostic feature extraction for anomalies detection and facilitate the task of fault diagnosis.

1.2.1 Data Acquisition

In the acquisition step, the data is collected to obtain data relevant to the system condition. Several CM techniques are available for the data acquisition process such as vibration, oil monitoring, thermography monitoring, ultrasonic monitoring, and radiographic monitoring[1]. Mainly CM data can be collected in the data acquisition process within two main categories as following:

Value type: this type of data can be recorded at a particular time and analysed for CM such as oil analysis, pressure, and temperature and humidity data.

Waveform type: this type of data is collected in a timely manner for CM and normally is a time series called time signal. Such as acoustic and vibration data. The most widely used waveform in CM is vibration signal, acoustic emissions signal. Other types of waveform data are; motor current signature, ultrasonic signals, etc. [5].

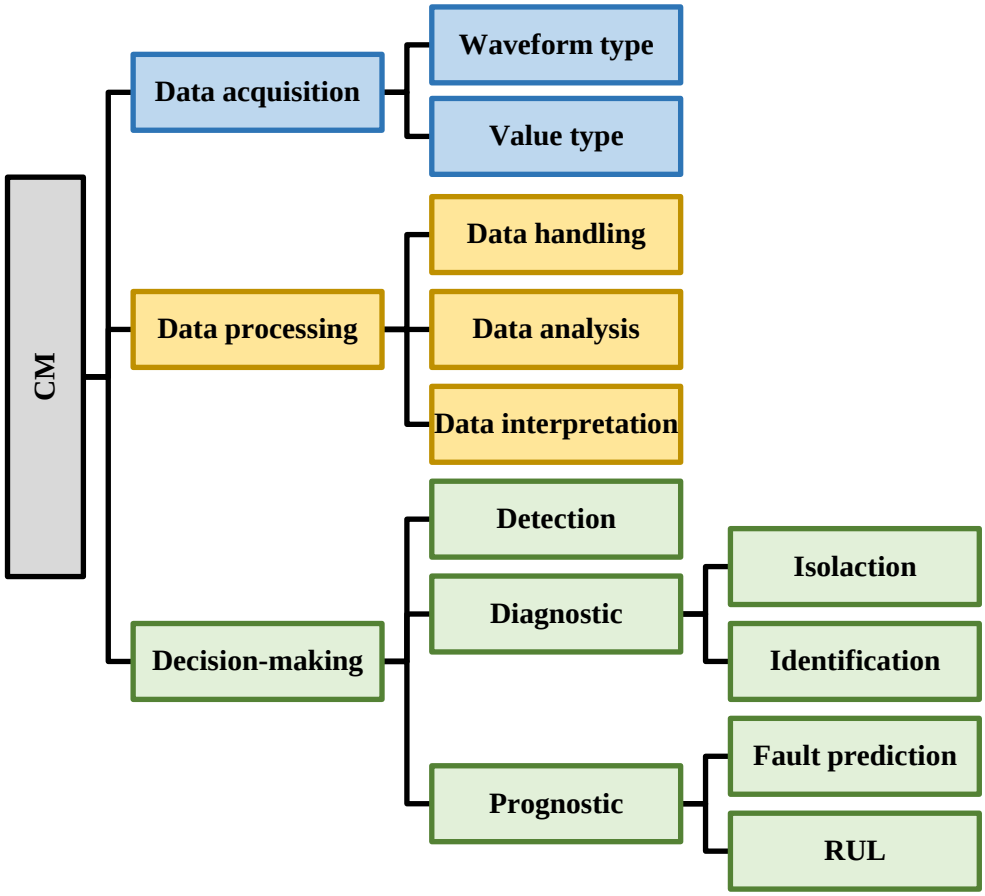

Figure 1-1 Condition monitoring steps

1.2.2 Data Processing

In this step, the collected data in the previous process is to be handled, analysed and interpreted. There are various DSP techniques, modelling methods and knowledge-based algorithms available, and the selection of the analysis technique depends on the type of data [6]. This process is to detect any anomalies occur in the analysed data that can affect the efficiency of the system or lead to a potential breakdown.

1.2.3 Decision-Making

This step is to recommend an efficient maintenance policy depending on the defect prognosis or severity of faults. A number of techniques have been developed over the past decades for decision making in CM strategy. They can be generalised into the following main categories: detection, diagnostics and prognostics. Fault detection focuses on detecting the abnormality in the monitored data as early as possible. Fault diagnostics focus on isolation and identification of faults when they occur. Fault prognostics attempts to predict the remaining useful life of the monitored machine or

predict faults or failures before they occur. However, it was claimed that prognostics cannot entirely replace diagnostics as in reality not all faults and failures are predictable[7].

1.3 CM Approaches

Generally, studying CM is carried out through three main approaches as seen in Figure 1-2, namely as data analysis focus, data gathering technology focus and application focus. In this study, the data analysis focus is considered.

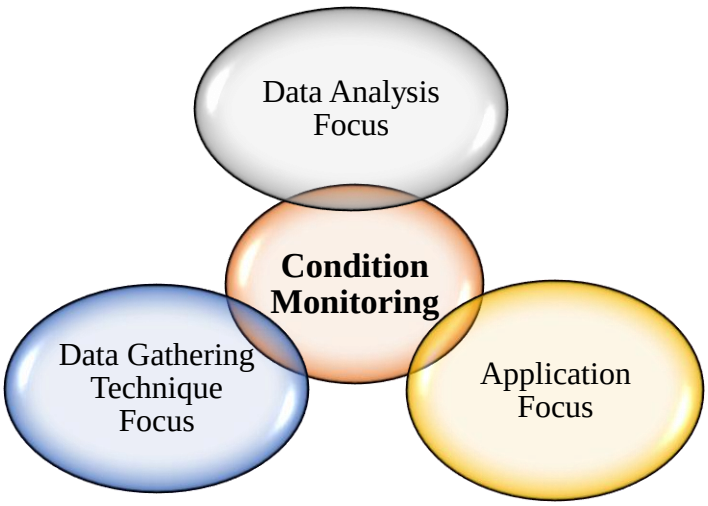

Figure 1-2 CM approaches

1.3.1 Application Focus

In CM, application focus discusses diverse fields of machinery such as induction motors, manufacturing tools, gearboxes, engines, bearings, centrifugal pumps, electro-hydraulics, etc.[8].

1.3.2 Data Gathering Technology Focus

This prospect of CM is focusing on developing instruments, methods and techniques for data sensing and data gathering such as Acoustics, Vibration, Motor Current Signature, Wireless Communications, Impact Analysis, Controller Behaviour, Emissions, etc.

1.3.3 Data Analysis Focus

Data processing is a primary key factor in the CM process and a reliable and effective data processing technique has been for a long time of interest for both academia and industrial word. An ideal data processing technique for CM should result in a low-dimensional, noise-free representation of the processed data with the aim of discovering useful information. Also, it should be robust to time-varying and nonlinearity nature of the monitored data, which can be used to

accurately interpret the condition of the monitored system. In general data processing in CM considering model-based methods, signal processing, high order statistics, demodulation methods, adaptive algorithms, feature extraction, feature selection and AI-based classifiers.

This thesis focuses on data analysis for CM and as shown in Figure 1-3, it can be said that the main approaches for CM data processing are: firstly, conventional signal processing, secondly, adaptive methods and thirdly, automated (AI) methods. Conventional signal processing methods are; time domain analysis, statistical parameters, and frequency domain analysis, also, demodulation analysis technique.

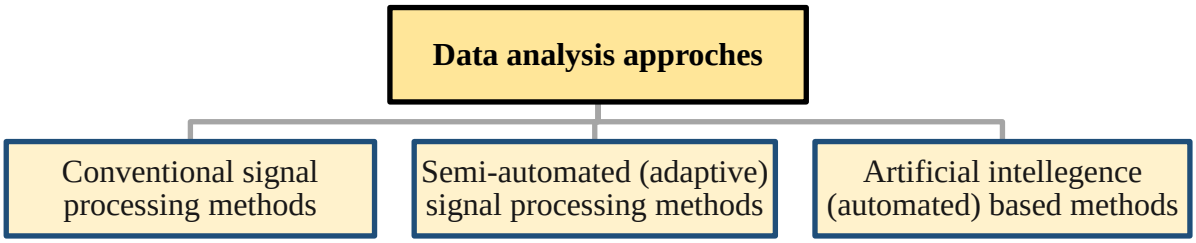

Figure 1-3 Data processing approaches

The semi-automated approach covers all the techniques that adopt adaptive algorithms to process the data. It can be said that in semi-automated techniques only a small number of parameters may need to be adjusted for the studied case[9]. Automated data process approach considers the implementation of artificial intelligence algorithms to minimize human intervention in the process of condition monitoring. Several AI techniques in computer science were implemented for machinery CM such as neural networks and support vector machine, etc.[10]. In this study **automated** and **semi-automated** data analysis approaches are considered for condition monitoring. The aim of developing an automated approach based AI is to make the task of anomaly detection independent of well-trained technical staff. Whilst the aim of developing a semi-automated approach is to facilitate the task of data analysis for diagnostic and make the task less-dependent on well-trained technicians.

The main aim of CM is to avoid systems breakdown and the consequence of catastrophic failures by detecting and diagnosing initial faults as early as possible. For a successful CM and from data analysis standpoint, as depicted in Figure 1-4, data analysis methods can be classified according to the task of processing into two main methods ; **Feature Extraction** by extracting the relevant

features from a dataset or **Data filtering** by removing the irrelevant components from the signal and use the residual signal.

Feature extraction task can be conducted through decomposing the sensors data into several components and then, to identify the condition of the machine. The sensor data components are explored based on prior knowledge or by investigating the change in the behaviour of a dataset compared to the reference data taken from an ideal condition. Several methods have been developed to decompose the sensor data, they can be classified based on the domain of the analysis into the frequency domain and time-frequency domain.

Data filtering is the second method of analysis, it is performed by removing the unwanted component and consider the residual sensor data as an output of interest in the process. This task is also called denoising and data enhancement. Several methods have been developed to carry out sensor data filtering and enhancement. Filtering methods fall into two main categories, namely, non-adaptive methods and adaptive methods. Moreover, data filtering can be performed in the time domain or in the transformation domain. Time domain methods are used for strictly periodic signals, a typical time domain filtering method is signal averaging.

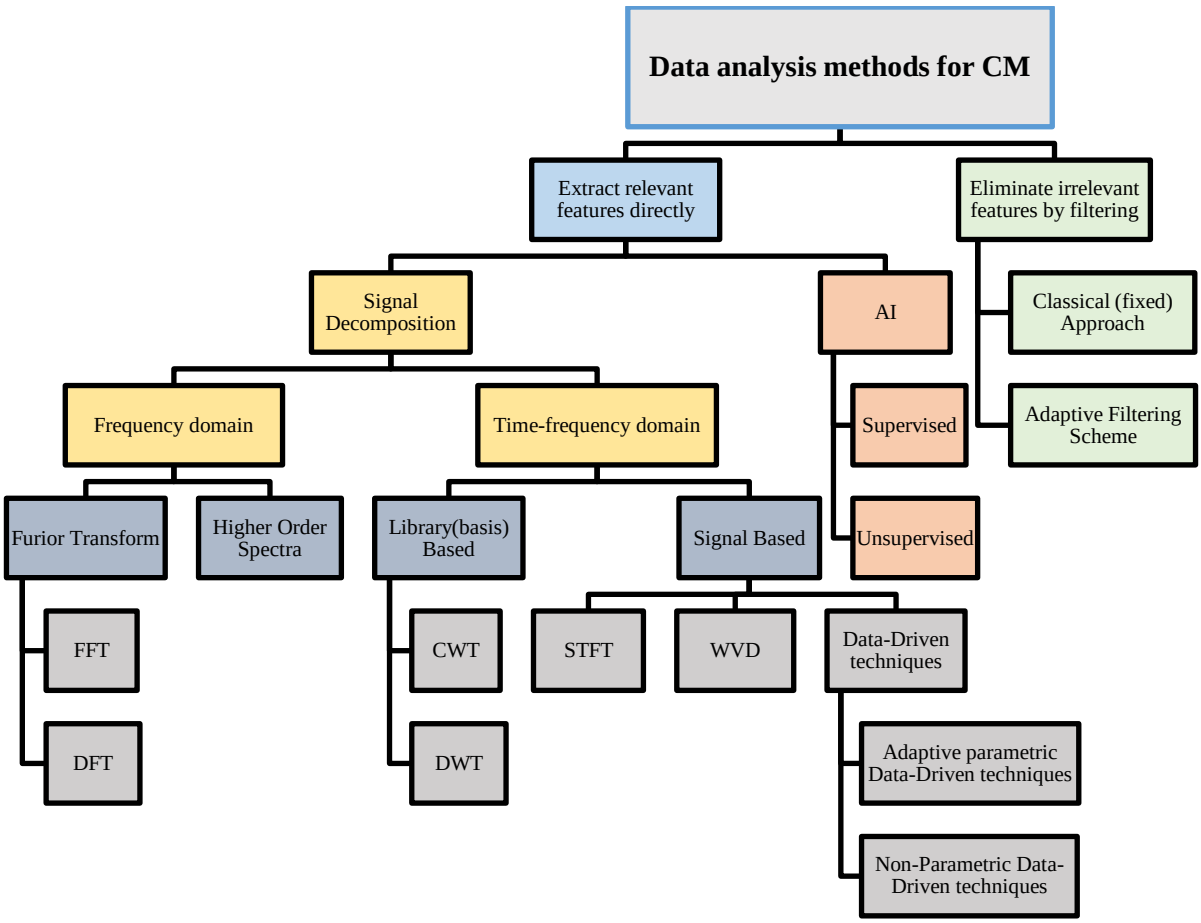

Figure 1-4 Data analysis methods for CM

1.4 Condition Monitoring of REBs

Condition Monitoring Applications to REBs have been extensively studied due to several reasons e.g., REBs play an important role in almost all types of rotary machines and REBs failures are the most cause of machines breakdown [11]. A survey was carried out by the Electric Power Research Institute concluded that about 40% of most common faults in an induction motor are related to bearing [12, 13]. Moreover, it has been stated statistically in [14], that the most common faults occur due to rolling contact fatigue after a certain running time. This issue starts with the presence of tiny cracks underneath the surface of the bearing components. It was also reported that inner or outer race flaw are amongst about 90% of the various REBs defects, whilst cage and rolling element flaw is the cause of the remaining malfunction[15].

In the REBs, the presence of faults such as cracks or pits located at bearing raceway surfaces or fatigue may lead eventually to machine breakdown. Also, failure of the bearings caused by, misalignment, etc., may cause catastrophic failure of the machinery system. Moreover, bearing failure was found to be one of the most common reasons for a breakdown in rotary machines[16].

Such failures can lead to catastrophic and usually result in long-lasting industrial downtime that has usually economic consequence [17]. Also, the bearing fault detection in the early stages will decrease the cost of unwanted shut down [18]. In order to prevent such unexpected bearing failures, several techniques were developed to monitor REBs. Among them, vibration analysis has been for a long time one of the most widely used as an effective method and popular strategy[13, 19]. Also, the vast majority of the advanced signal processing techniques is related to vibration measurements [11, 20-23].

Therefore, in this research based on vibration, an integrated framework is proposed to implement an AI approach for early fault detection, also adaptive time-frequency data analysis for fault diagnosis. Thus, the decision could be taken as early as possible to increase the life of the bearings and reduce the maintenance cost.

Several methods based on vibration have been applied to bearing fault detection as shown in Figure 1-5. These methods can be categorised into three main groups namely: signal-based, model-based and knowledge-based. [10, 18].

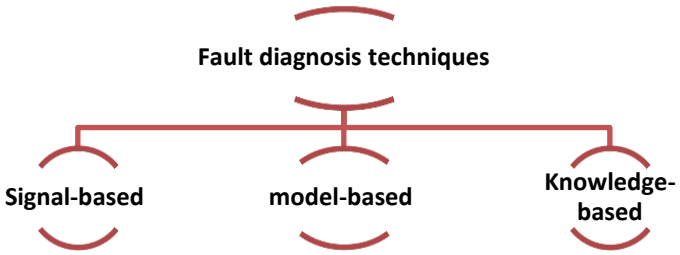

Figure 1-5 Fault detection techniques

1.4.1 Signal-based methods

Signal-based techniques for fault detection are based on analysing the time domain and spectral components of measured data. The signal-based techniques applied to process vibration data for CM of REBs can be classified into the following main categories as depicted in Figure 1-6: time-domain analysis, frequency domain analysis, time-frequency analysis and demodulation analysis methods [24]. Vibration signal contains enormous information, therefore, a number of vibration signal analysis techniques have been developed to highlight some components of interest in a signal. These techniques analysis signals in the time domain or in the transformation domain (frequency or time-frequency domain)[25].

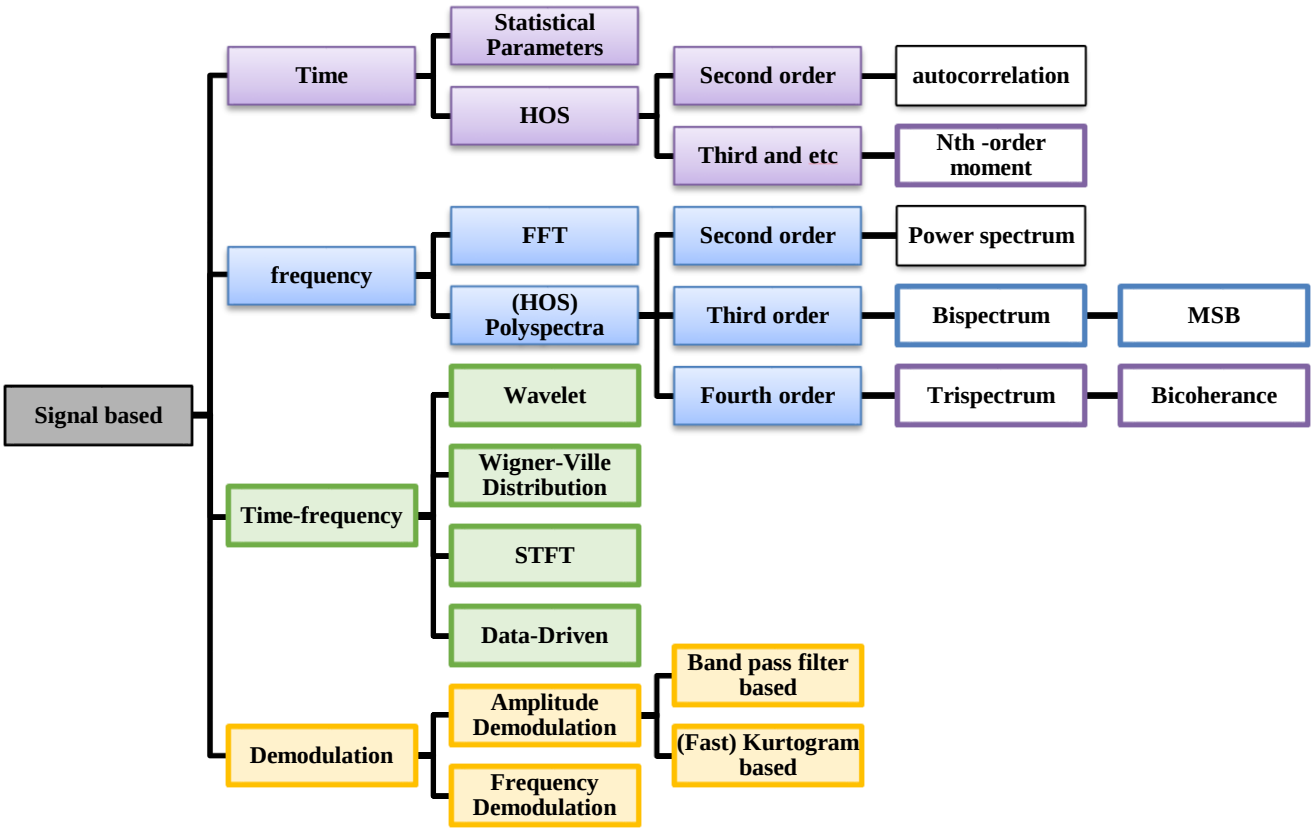

Figure 1-6 Signal based Fault diagnosis using vibration

1.4.1.1 Time-domain

Time domain analysis is the process of analysing and displaying the vibration data as a function of time. In the time domain, methods that sensitive to oscillations involve statistical parameters are used to highlight some trends in the vibration data. The analysis in the time-domain is usually carried out to explore the statistical characteristics of the monitored data[26]. Many statistical parameters used for vibration analysis to describe the monitored data include root mean square (RMS), kurtosis, crest factor, peak value, Kurtosis, peak-to-peak interval, high-order statistics, skewness, etc.[21]. The derived statistical parameters are usually known as time-domain features. Kurtosis and crest factors are said to be more sensitive to the shape of the signal because they increase with the increase of the spikiness of the vibration signal [27]. On the other hand, Skewness measurement is claimed to be an effective measure only for asymmetrical signals [28]. However, limited success has been reported when using time domain analysis methods for diagnosis of REBs and they were found to be a poor measurement of fault features in REBs [29]

RMS is a simple statistical method to detect the gradual or sudden changes in the energy of a vibration signal in the time domain. RMS for sinusoidal signal cab derived [30] as:

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i)^2} \tag{1.1}$$

Where x_i is a time series signal n is the total number of data points and i is time index. RMS can be used as a measure of the signal energy, however, it does not identify in which component the defect occurs [31].

Kurtosis is a measure of the impulsiveness of a signal, it calculates the normalized fourth centred moment. It can be derived [32] as:

$$Kur = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^4 \tag{1.2}$$

Where μ is the mean of the signal and σ is the standard deviation of the signal. The kurtosis value is about 3 for a healthy bearing in a good condition, when an initial fault occurs the kurtosis value will increase, however, the value declines as the severity increase [33].

Peak value can be calculated as the sample index for the maximum amplitude of a sinusoidal signal as it can be calculated [34] as:

$$peak\ value = \sqrt{2} \times RMS \tag{1.3}$$

The Crest factor computes the ratio of peak value to the root mean square, it allows to estimate the shape of the vibration signal waveform. Crest factor value of a signal less than 3.0 will indicate to a sinusoidal signal, however, a higher value will indicate to the impulsive signal so can be used to assess the bearing condition. Crest factor is computed [32] as:

$$crest\ factor = \frac{peak\ value}{RMS} \tag{1.4}$$

It was reported that the peak value will highly depend on the section of the signal being analysed, thus, crest factor may not be stable. Analysing the distribution of the acceleration can give more details. Healthy bearings are expected to have a Gaussian distribution, on the other hand, a relative increase in the tail levels can be expected from a damaged bearing. High order statistics with a variety of moments is another measure and fourth order has been found in somewhat useful. Bandpass filtering also developed to examine only the band of interest in a signal.

1.4.1.2 Frequency-domain

Frequency analysis is a widely used technique to analyse vibration signal, Fast Fourier Transform (FFT) has facilitated the spectrum analysis and made it more efficient. It gives the ability to investigate the changes in sub frequency bands of a signal. Both the amplitude and frequency of the spectral lines can be investigated. Knowing the speed of rotation and the calculated frequencies for the machine components, some of the peaks can be identified. The frequency-domain analysis is a well-known and broadly used data analysis technique based on the idea of transforming time series data into the frequency domain. FFT was developed to efficiently perform the Fourier transformation process with remarkably reduced complexity. FFT has been for a long time the most commonly used method for transforming raw vibration signals from the time domain into the frequency domain. Frequencies of interest in the transformation domain can be easily identified and isolated, this property is one of the main advantages of FT analysis over time-domain analysis. Moreover, features can be extracted from the whole spectrum or from just certain frequencies of interest [5]. Power spectrum is amongst the most commonly used tool in spectrum analysis. However, REBs usually produce complex vibration signals due to the effects of the background noise. Therefore, identifying bearing defects only by observing the spectrum signals is difficult, besides the effects of the sidebands and the harmonics of the fault frequencies. Moreover, the non-stationary nature of signals makes applying the FFT method, which is in fact based on the assumption of the periodic signal, not suitable[35]. Therefore, it is very unlikely to identify the faulty peaks in the spectrum as the energy of impact vibrations usually will be distributed over a wide range of frequencies. In addition, the frequency of the defect has low energy and hence get easily masked by other low frequencies or noise, also, interpreting and quantifying some other peaks of the spectrum may not be an easy task in some cases.

1.4.1.3 Time-frequency Domain

The nonlinearity and nonstationarity characteristics of some vibration data make extracting useful and sensitive features from the data, not an easy task. To address this issue, several Time-frequency analysis methods have been developed, they have been popular methods to process non-stationary signals. One of the well-known methods is Short-Time Fourier transform (STFT), the STFT technique divides the entire waveform signal into with short-time segments using a sliding window and then apply FFT to each segment. However, STFT suffers from low-resolution problems. Another time-frequency analysis method is Wigner-Ville Distribution (WVD) which used to overcome both the low-resolution problems and time information. Moreover, Wavelet Transform

(WT) is another time-frequency analysis based on the idea of multiresolution analysis. There are a number of wavelet transform methods available such as continues wavelet transform, discrete wavelet transform. Wavelet packet transform, etc. Wavelet transforms have been extensively used for REBs fault diagnosis in the past two decades[36]. For example, Rubini et al.[14], used wavelet transform to diagnose bearings affected by an incipient surface fault.

1.4.1.4 Higher Order Statistics (HOS)

For several years the first and second-order statistics, such as variance, mean, power spectrum and autocorrelation have been extensively applied for vibration analysis. However, they just used to characterise Gaussian and linear signals. Whilst in practice, non-Gaussian and nonlinear signals can be studied using higher order statistics such as kurtosis and skewness [37].

1.4.1.4.1 HOS Time Domain

In the time domain, the autocorrelation function is the second order measure, and it is known that the third-order moment depends on two independent lags m1 and m2, so in a similar way by adding lag terms to the third-order, higher order moments can be formed [38]. It should be known that the term of moments is used to denote first/second order whilst, cumulants are used for HOS.

1.4.1.4.2 HOS Frequency-domain

Polyspectra is used to refer to HOS in the frequency domain, including the 2nd order and so on.

Power Spectral (PS) is a second-order measure and it can be simply computed by multiplying the signal Fourier Transform together with its complex conjugate as:

$$p(f) = E\langle X(f)X^*(f)\rangle \tag{1.5}$$

where $X(f)$ Fourier Transform and $X^*(f)$ it's conjugate of the $x(t)$, and $E\langle\cdot\rangle$ the expectation operation or by computing a Discrete Fourier Transform (DFT) of the autocorrelation function [38].

However, Thomas [39] reported that in the presence of high background noise from rotating machinery, the use of power spectrum become ineffective and will affect the fault diagnosis accuracy[20].

Bispectrum, most implementation of HOS in the frequency domain, focusses on the bispectrum and the trispectrum known as third-order and fourth-order measures respectively [38, 40]. At the third-order, the bispectrum $b(f_1, f_2)$ can be also computed by taking a **D**ouble **DFT** (DDFT) of the third-order cumulants or as a product of the FT at different frequencies as:

$$b(f_1, f_2) = E\langle X(f_1)X(f_2)X^*(f_1 + f_2)\rangle \quad (1.6)$$

Where $X^*(f)$ is the complex conjugate and $E\langle\,.\,\rangle$ is the statistical expectation operator. To reduce Gaussian noise and to preserve some of the non-Gaussian information of the measured vibration, HOS, has been implemented to solve detection and classification problems [41]. it was claimed that using non-linear features motivated by the higher order spectra is a promising solution to analyse the non-Gaussian and non-linear vibration signals, thus, it can extract more diagnostic features than power spectrum does[42]. Collis et al.[43], claimed that the trispectrum is a more powerful tool because it represents a decomposition of kurtosis over frequency. Unlike the PS, it may be considered that the bispectrum and trispectrum are functions of multiple frequencies. They contain phase information as well as magnitude information about the original vibration signal. It has been reported that the bispectrum analysis provides a much better feature for the diagnosis of different faults simulated on an experimental rig when compared to the spectrum analysis alone, without the phase and orbit analysis [44].

Bicoherence is realised as a normalised bispectrum. Bicoherence takes bounded values between 0 and 1, this makes it a suitable measure to quantify the extent of phase coupling in the vibration signal. The normalisation is arisen due to variance issues of the bispectral estimators **[40].**

Several studies used bispectrum in the fault diagnosis have been reported in the literature. For instance, Saidi et al. [37] Stated that higher-order spectra are a promising approach to extract non-linear features used to analyse the non-linear and non-Gaussian characteristics of the vibration signals. Liang et al. [45], applied power spectrum, bispectrum and neural network to extract fault pattern from a vibration signal of induction motors, the study showed that bispectrum suppressed the noise and showed some useful information in signals. Pineyro et al. [46] examined the implementations of power spectral and bispectral on the bearing fault signals. However, the study reported that when implementing second order power spectral, the resonances cannot be distinguished from periodic signals, while Bispectrum was found highly sensitive in detecting the phase coupling peaks in the spectrum, however, the main disadvantage is the high memory consumption needed for data processing[20].

Modulation Signal Bispectrum (MSB)

Due to the observed efficiency of bispectrum and to compensate for the deficiencies of Polyspectra techniques, moreover, to enhance the conventional bispectrum in characterizing the vibration signals, Gu et al. [47], has examined a new modified form of the conventional bispectrum, called a Modulation Signal Bispectrum (MSB) as:

$$B_{ms}(f_1, f_2) = E\langle X(f_2 + f_1)X(f_2 - f_1)X^*(f_2)X^*(f_2)\rangle \tag{1.7}$$

Where, $X^*(f)$ is the complex conjugate of $X(f)$ and E⟨.⟩ is the statistical expectation operation. f_1, f_2 and $f_1 + f_2$ represent three individual frequency components derived from Fourier series integral. Since then it has been as a promising technique for detecting nonlinear components by detecting phase coupling in modulation signal. It was also found to be an effective tool to suppress random noise.

1.4.2 Model-based methods

Model-based methods have been widely implemented for CM of machinery systems. An accurate model of the system is needed to imitate the real process behaviour [18]. Implementing model-based methods for CM of mechanical systems will require mechanistic knowledge and relevant theories of the monitored systems[5]. Over the years several model-based methods have been devolved and they can be categorized into two main groups namely: System dynamic modelling, Fault dynamic modelling[48].

1.4.2.1 System Dynamic Models

System dynamic models are developed to simulate the dynamic of REBs in order to investigate their behaviour and features. Also to understand the transmission of vibration through bearings structure and the influence of load distribution on the dynamic of the bearings[11], etc.

1.4.2.2 Fault Dynamic Modelling

Analytical and numerical fault models are being used for CM of REBs in several ways. A fault dynamic model can be used to simulate the faulty conditions of REBs, also, it is used to evaluate the capability of data analysis methods in extracting the diagnostic features. In addition, it can be used to determine the severity of the faults [11]. In literature, a number of model-based methods developed for fault diagnosis such as Pennacchi et al. [49] presented a model-based transverse crack identification method where Vania et al. [50] proposed model-based fault diagnosis method based on the frequency domain.

Most of the model based have studied localized defects using a variety of modelling techniques. For more details, El-Thalji et al [11] provided a comprehensive comparative review of system and fault modelling for rolling bearings. However, it was claimed that model-based techniques have several limitations such as how to experimentally verifying the expected results [51]. Moreover, the model-based techniques fail to deliver satisfactory performance due to the presence of disturbances, noise, modelling uncertainties and/or parameter variations [52]. Developing an accurately a mathematical model which describes a physical system in real-world applications is usually not an easy task. Therefore, the implementation of model-based methodologies is still limited [53].

1.4.3 Knowledge-based methods

The industrial world becoming more automated and the amount of data that systems can produce has increased massively. Hence, analytical modelling and the traditional digital signal processing cannot handle such huge, diverse and rapid data and they may not be able to perform to a sufficient diagnosis, In addition, the fault diagnosis of a machine normally requires technical skills, experience and knowledge of the machine's structure, where in practice, the experts are either too busy or costly. Thus, in order to automate the diagnostic procedures and provide the engineer with aid to make a decision about the REB's health state, An expert system or artificial intelligent system (AI) that support parallel processing can be utilised[10]. There are several knowledge-based methods for automatic fault diagnosis have been used such as Artificial Neural Network (ANN), Expert Systems, Fuzzy Logic, Support Vector Machine (SVM) [10].

This research focuses on knowledge-based techniques which will be discussed in the upcoming sections in more details. Recently, the use of methodologies such as diagnostics and prognostics aided by AI tools such as ANN, KNN, SVM etc. have witnessed an increase for assessing the health of the REBs [54]. Generally, intelligent Condition monitoring approach involves three main steps, firstly, signal acquisition step secondly features extraction step and finally, faults classification [55]. Feature extraction is the task of extracting the most informative features which represent the data from the gathered signals using signal processing techniques. Currently, there are several manual ways to carry out the task of feature extraction such as statistical parameters, spectrum analysis and time-frequency analysis. However, obtained features may contain useless or redundant information and consequently, affect the diagnosis outcomes as well as increase the computation cost. Thus, feature selection becomes an important task to reduce data dimensionality by selecting only informative and sensitive features. In the final step, the selected features are used and fed to

train AI algorithm. As a result, by using these techniques the conditions of the monitored system can be determined [56]. The quality of extracted features from the monitored REBs signals mostly play an important role in the effectiveness of these approaches.

It seems that AI plays a vital role in CM as in the last decades, the application of AI to REBs has been gaining more attention in the industrial world, such as ANN, SVM, discriminant analysis etc. for instance, Tyagi et al. [28] has presented a comparative study of SVM and ANN in the application of REB fault detection. The reliability of REBs diagnosis can be improved by utilising the automated approaches, which also can be cost effective and save time. Furthermore, automated REBs CM does not depend on expertise judgment [57]. Several attempts adopting AI Approaches have been carried out, most of the studies used AI approaches for post-processing (classification) while some studies used AI approaches for data processing process such as anomaly detection task.

1.4.3.1 Anomaly detection

Anomaly detection can be defined as the process of identifying that the test data differ in some ways from the data used in the training the model. Anomaly detection has attracted many researchers and received lots of attention in many applications such as processing massive datasets arises form critical systems. The abnormal modes of a system are not always known priori, this may make the use of conventional multi-class methods invisible. The anomaly detection approach carries the potential solution in which the prior knowledge of abnormal modes are not required and normal mode can be learnt by building a data-model with the available data of the normal conditions. The unseen monitored data then is compared with the model of the normal condition and the resultant novelty score can be used to measure the difference between the monitored data and the normal condition based on a predefined threshold. If the score reached above the threshold, the condition is deemed to be a deviation from normality which indicating to a physical change in the system. Anomaly detection is classified into five general categories as seen in [58].

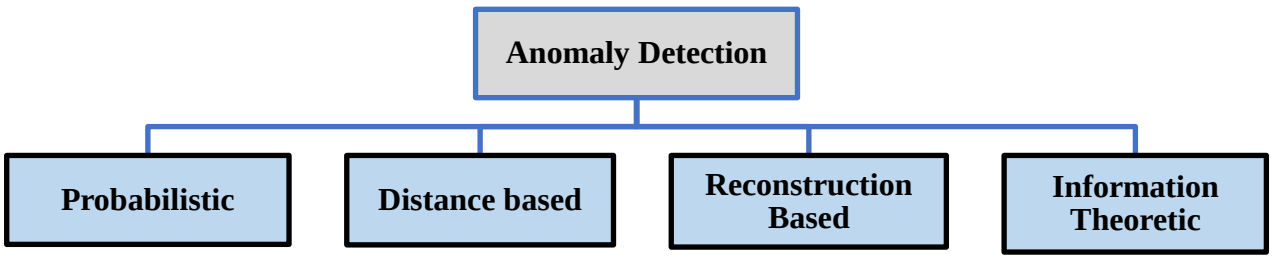

Figure 1-7 Main methods of anomaly detection

In probabilistic methods usually, density estimation of the normal class is involved based on the idea that having low density areas in the data used for training means the probability of containing normal components in these areas is low. Distance-based methods are based on the concept that normal condition is tightly grouped while anomaly data is located far from its nearest neighbour these methods involve clustering and nearest neighbour techniques. The reconstruction based method is realised by training a model using training data, and then use the trained model to map the unseen data. If the unseen data is abnormal, the reconstruction error between the unseen data and the training data will raise the value of the novelty score and the anomaly then is detected. Information theoretic based methods based on the idea that the information content will be altered significantly if the anomalies exist in the datasets. This approach use theoretic measures like entropy to compute the information contents in the training data set[59].

1.4.3.2 Shallow Features learning methods

Learning intrinsic structure of data has attracted researcher's attention for many years. In order to extract a representation which can precisely describe a set of data, valuable features are to be extracted and selected effectively. Feature learning is an important task in AI-based approaches for both novelty detection and classification and has delivered good results in many fields and applications. According to the literature, many AI-based techniques have been applied in CM for the classification task whilst the features are still empirically extracted such as calculating statistical parameters and then to be used as an input features in a classifier. However, with the increase in the complexity and high-dimensionality of nowadays data, manually crafted parameters cannot effectively describe and represent the data. Ideally, extracting an informative representation from high-dimensional and complex data or discovering valuable information becomes an important and more challenging task. In AI-based approaches, features extraction is typically implemented based on two main paradigms known as supervised learning and unsupervised learning as depicted in Figure 1-8. Both supervised and unsupervised learning can be implemented based on shallow learning methodology and deep learning methodology. The selection of the paradigm is based on the aim of the application and the availability of labelled data for the task at hand. Supervised learning is used to build a classifier that learns the means to predict the output y from given input x by discovering the intrinsic structure in the given dataset. I.e. identify and then assign an unknown hidden pattern to a previously defined class. This method is visible when the labelled data is available, the earliest supervised technique for feature learning is Principle component analysis (PCA) and it was developed in 1901 by Pearson[60]. On the other hand, unsupervised feature learning is used to discover hidden patterns in a given set of unlabelled data to be used for

clustering by grouping similar items together or as a data representation for classification task, the earliest technique implemented the unsupervised learning paradigm is linear discrimination index (LDA) which was developed by Fisher in 1936. [60]. PCA and LDA are well-known as the earliest features learning methods.

1.4.3.2.1 Deep Features Learning

Besides shallow learning, deep learning was proposed in 2006 by Hinton by [61] and since then, it has been investigated in many domains such as computer vision, image processing, voice recognition and natural language processing etc. Deep learning is achieved by building a deep architecture network with extended hidden layers to learn multiple levels of data representations. A number of deep learning models are available such as Deep Belief Network, Stacked Auto-Encoder, Convolutional Neural Network, and Recurrent Neural Network [62].

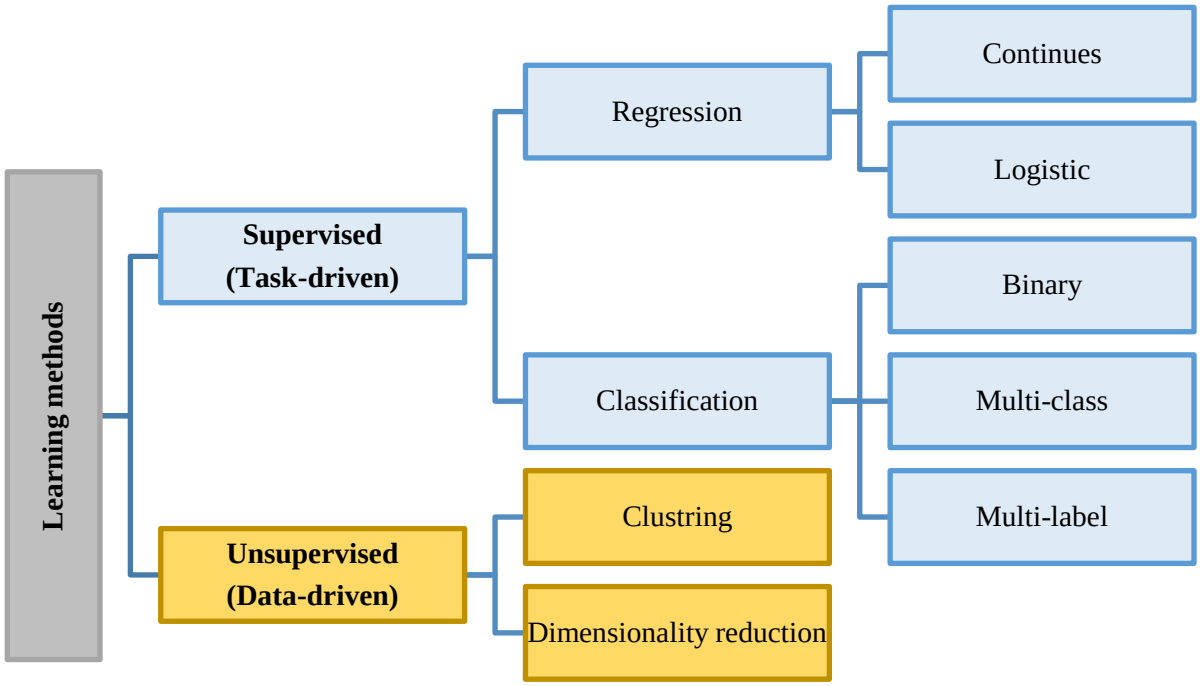

Figure 1-8 Feature learning methods in AI-based approach

1.5 Aim and Objectives

The aim and the objectives of this research are listed below:

1.5.1 Aim

The main aim of this study is to develop a framework for machinery CM. This aim will be carried out by automating the task of anomaly detection adopting unsupervised machine learning algorithm. Moreover, to make the task of diagnostic procedure less dependent on well-trained

labours by developing an adaptive algorithm with wavelet transform for feature extraction and enhancement by denoising the decomposed coefficients.

Intelligent data analysis approach based on unsupervised ANN will be investigated for efficiently characterising large raw datasets arisen from condition monitoring systems and thereby for the automated fault detection at the very early stages.

The second part of the developed framework is to enhancing the features of interest using an adaptive algorithm with an expensive wavelet transform (DD-DWT) and hence carry out the diagnosis of critical machine components such as rolling element bearings.

1.5.2 Objectives

To achieve the aim, the research sets up the following key objectives:

Objective one: To explore and gain insight into the methods of current CM and their applications.

Objective two: To review the current analysis techniques of the experimental vibration signals, also, Artificial intelligence based techniques for anomaly detection in vibration data. Furthermore, to carry out a critical review of the wavelet-based data analysis techniques and their existing benchmark thresholding methods used for experimental data denoising and feature enhancement.

Objective three: To automate the task of early fault detection and severity estimation by implementing an automated technique based AI approach.

Objective four: To implement DD-DWT for vibration data analysis in field of CM and develop a thresholding algorithm for vibration data denoising and feature enhancing.

Objective five: To investigate the impact of changes in internal clearance, due to inevitable wear, on the richness of diagnostic signal information and fault detection and diagnosis. Moreover, to design a test rig and develop an adjustable clearance mechanism in which the radial clearances can be controlled and the defects can be seeded into tapered bearings.

Objective six: To evaluate the performance and capability of the CCNN using simulated data in order to explore its reliability and effectiveness on bearing fault detection in comparison with conventional techniques.

Objective seven: To evaluate the performance and capability of the CCNN using the vibration data gathered from baseline and defective bearings.

Objective eight: To evaluate the performance DD-DWT with comparison to both DT-CWT and DDD-DWT in features extraction with the developed thresholding technique.

Objective nine: To evaluate the performance of the developed thresholding technique (HSR) in features enhancement and data denoising against the benchmark thresholding techniques.

1.6 Flowchart of Data Analytics

Figure 1-9 shows the integrated framework with two main implementation approaches adopted in this PhD project. Firstly, it focuses on extending AI techniques as an automatic data processing for early fault detection by applying Componential Coding Neural Network (CCNN). Then, once an anomaly is detected, a further analysis technique to identify the fault is proposed as a semi-automatic data processing procedure using a multiresolution data analysis approach based on Double-Density Discrete Wavelet Transform (DD-DWT) with a novel level-dependant adaptive thresholding method.

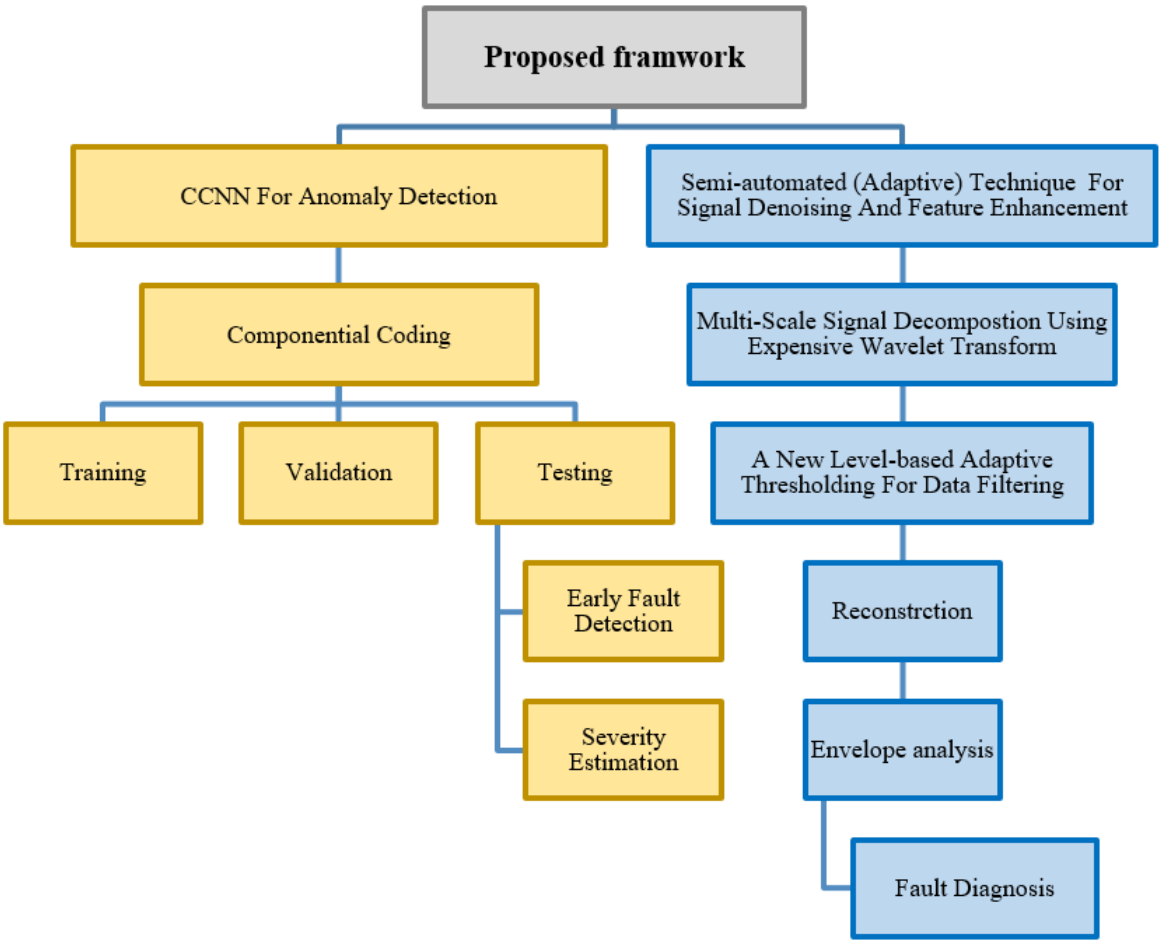

Figure 1-9 Implementation framework

1.7 Summary

In this Chapter, a brief background is given in (section 1.1), whilst, the condition monitoring steps explored in (section 1.2). Moreover, the condition monitoring focuses were listed in (section 1.3) and Condition monitoring of REBs were discussed in (section 1.4) with focus on signal based, model based and knowledge based. Finally, the aim and the objectives alongside with flowchart of this research were presented in (section 1.5).

CHAPTER TWO

FEATURE EXTRACTION AND ENHANCEMENT TECHNIQUES FOR CONDITION MONITORING - THE LITERATURE REVIEW

Adaptive and iterative methods are reviewed with critical comments in this chapter. Signal processing based feature extraction and enhancement techniques including AI approaches are explored in line with their applications to condition monitoring. With emphasises on vibration-based monitoring, both critically sampled and oversampled or expensive wavelet are discussed in details.

2.1 Introduction

A variety of signal denoising and feature enhancement techniques have been developed and applied in CM. These techniques are categorised and discussed in this research according to the methodology of processing the acquired data, hence, grouped as conventional, adaptive and AI-based (automated). This categorization is adopted because the aim of this study is to automate the detection task and to facilitate the diagnosis task, in addition, in [63] Cerrada et al. reviewed the recent work 2010-2016 in vibration signal analysis and fault diagnosis and stated that the techniques can be categorised into two main categories, signal-based and AI-based techniques.

2.2 Demodulation Signal Based Approach

Amplitude modulation is identified as a multiplication of a low frequency modulating signal by a high-frequency carrier signal. This phenomenon will produce frequency components (peaks) in the spectra of the modulated signal, these peaks appear with sidebands located at the carrier (high) frequency band, and spaced with the modulation (low) frequency. Therefore, demodulating the signal to extract the modulation signal from the carrier frequency will be useful [64]. For bearing fault detection, modulating frequency is the frequency of interest as the resonance frequency is high and has few spectral lines. However, by using the conventional spectrum, it is not possible to recover characteristic defect frequencies from the resonance frequency. Fortunately, a technique called high-frequency resonance technique (HFRT), also called envelope analysis, allow extracting the modulating frequency from the resonance frequency band. HFRT can extract periodic pulses, also it can extract the amplitude-modulated signals from vibration signals with less sensitivity to the influences of slippage. Envelope analysis has been proposed and comprehensively used,

particularly for bearing fault diagnosis [65-67]. Envelope analysis can extract amplitude-modulated signals from vibration signals.

2.2.1 Envelope analysis

For a defective bearing, the pulse generated by the contact between the defective area and the bearing rollers will excite resonance frequency. The structural resonant frequency caused by pulse excitations is considered as an amplitude modulated signal. The amplitude modulation of the exited bursts can occur due to two reasons, the first reason is when the rolling elements passing through the load zone with so the modulation will be at the same rate, the second scenario, it can happen when the defect is rotating, the transmission path will vary with respect to the fixed sensor. The envelope analysis has the ability to provide more diagnostic information than raw data or spectrum analysis. It has been widely applied to detect and diagnose defects in bearing and gears. As seen in Figure 2-1, the application of envelope analysis is applied through three main steps, bandpass filtering, rectification, and spectrum analysis.

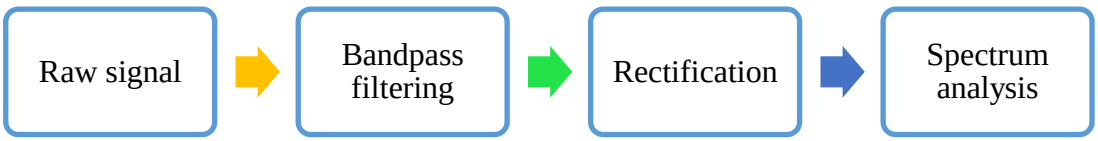

Figure 2-1 Envelope analysis procedure

In more details as depicted in Figure 2-2, envelope analysis is applied by firstly to ensuring maximum signal-to-noise ratio, a bandpass filtering step around the excited frequency band to exclude the undesired low frequencies. These frequencies may be associated with imbalance and the misalignment. The next step is rectification by the demodulation process that extracts the signal envelope using Hilbert transform and smooth the signal by low pass filter to remove the resonance (carrier) frequency. Finally, the spectrum of the demodulated signal is calculated [68]. The diagnostic information, including the repetition rate of the fault peaks, can be found in the derived spectrum of the envelope. Envelope analysis of a signal is applied through several steps [69] as is depicted below:

$$X(f) = FFT(x(t)) \tag{2.1}$$

$$X_{ant}(f) = \begin{cases} X(f), & f = 0, N/2 \\ 2X(f), & 1 < f < \frac{N}{2} - 1 \\ 0, & \frac{N}{2} + 1 < f < N - 1 \end{cases} \tag{2.2}$$

$$x_{ant}(t) = FFT^{-1}(X_{ant}) \tag{2.3}$$

$$x_{en}(t) = \sqrt{x_{ant}(t)^{*}x_{ant}(t)} \tag{2.4}$$

Where $x(t)$ is the raw signal, $X(f)$is the Fourier transform of the raw signal, $X_{ant}(f)$ is the Fourier transform of the analytic signal derived from raw signal $x(t)$, and $x_{ant}(t)$ is the analytic signal, $x_{en}(t)$ is the calculated envelope. The spectrum analysis of the envelope function $x_{en}(t)$ can be expressed as

$$X_{en}(f) = \left| FFT\left(x_{en}(t)\right)\right| \tag{2.5}$$

$X_{en}(f)$ is Fourier transform of $x_{en}(t)$.

The process of amplitude demodulation can be carried out digitally either using full base-band rectification or by using Hilbert transform techniques. A considerable improvement in performing the envelope analysis has been achieved using the Hilbert transformer because it reduces the whole number of data samples to be processed [70].

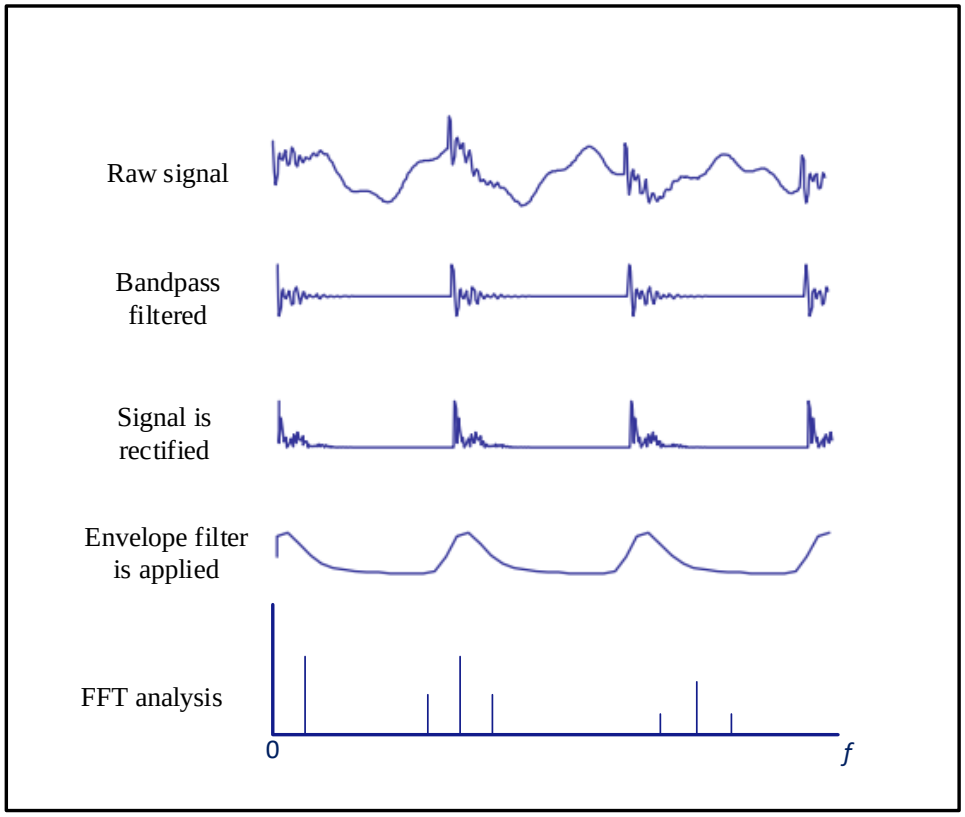

Figure 2-2 Envelope analysis process [71]

However, it was claimed that as bearing damage develop and increases to a sever level, the vibration signal may exhibit more random features and become irregular. This can make envelope technique inaccurate for identifying diagnostic features from the vibration signal [35]. Moreover, it is not an easy task to specify the frequency band which contains the highest signal-to-noise ratio (SNR). This has led to the emergence of other techniques such as Spectral Kurtoses and Kurtogram and Fast Kurogram etc. to specify the best frequency band and select the central frequency where the highest SNR in allocated in the spectrum.

2.2.2 Hilbert transform

It was first introduced by David Hilbert to solve some integral equations, Gabor in [72] applied the Hilbert transform to associate the real signal with the complex signal. Given time series $x(t) = \alpha \cos \omega t + b \sin \omega t$, the complex signal can be derived as (2.6).

$$\hat{x}(t) = x(t) + j\sigma(t) \tag{2.6}$$

Where $j\sigma(t)$ is an imaginary signal added to the real signal $x(t)$, and $\sigma(t)$ is a function yielded from $x(t)$ by converting $(sin\, \omega t)$ into $(-cos\, \omega t)$ and converting $(cos\, \omega t)$into$(sin\, \omega t)$. In this case, $\sigma(t)$ is a quadrature to $x(t)$thus the oscillating is transformed into a rotating vector. The $\sigma(t)$ associated with $s(t)$ can be yielded as:

$$\sigma(t) \frac{1}{\pi} \int_{-\infty}^{+\infty} x(\tau) \frac{d\tau}{t - \tau} \tag{2.7}$$

Where τ is the transformer parameter and t represents time. To verify the sufficiency of (2.6) in converting $\sin \omega t$ into $-\cos \omega t$ and $\cos \omega t$ into $\sin \omega t$

$$x(t) = -\frac{1}{\pi} \int_{-\infty}^{+\infty} \sigma(\tau) \frac{d\tau}{\tau - t} \tag{2.8}$$

According to Gabor [73] $x(t)$ and $\sigma(t)$ with satisfying reciprocal relations are recognized as a pair of the Hilbert transform. Analytic signal is a complex signal whose imaginary part is a Hilbert transform of the real part.

Randall and Bond [9] stated that Hilbert transform of a time function $x(t)$ can be obtained as

$$H[x(t)] = \hat{x}(t) \frac{1}{\pi} \int_{-\infty}^{+\infty} x(\tau) \frac{1}{t - \tau} dt \tag{2.9}$$

It can be interpreted as a convolution of the function $x(t)$ with $1/\pi$ of signal as in and will result in another time domain signal.

$$\hat{x}(t) = \frac{1}{\pi t} * x(t) \tag{2.10}$$

It shifts the input signal phase by 90º. Hilbert transform can be realized in the frequency domain by taking the Fourier transform of $\hat{x}(t)$ as

$$\hat{X}(f) = X(f)(-jsgn(f)) \tag{2.11}$$

Thus, it can be achieved in the frequency domain by shifting the phase $-\pi/2$ for the positive frequency and $+\pi/2$ for the negative frequency elements. And then taking the inverse of the Fourier Transform to get the signal back to time domain[9].

Hilbert transform can be applied to either nonstationary or nonlinear signals, it can produce sharper output more than other conventional methods. Hilbert transform is used to compute the envelope of the signal when it applied to the modulated signal, the output will be the modulating frequency. In the case of using Hilbert transform for a defective bearing, it will yield the impulses response matches the roller passing frequency. However, to perform envelope analysis to a signal, selecting the optimum frequency band is usually set manually, and it has been for a long time a tedious task. Therefore, several attempts have been carried out to facilitate the selection of the optimum frequency band, among them spectral kurtosis, kurtogram and the later developed version Fast Kurtogram.

2.2.3 Spectral kurtosis

Selecting the best suitable band for demodulation has been a real challenge when using envelope analysis, with many claiming that it is difficult. To overcome this issue, Spectral Kurtosis (SK) has been proposed to find the most impulsive band and has been used as a tool for extracting transients buried in noise [74-76].

Spectral kurtosis is developed to solve the problem of selecting the suitable frequency band for bandpass filtering. It has been proposed to find the most impassive frequency band, the first reported use of SK was in [77] to detect impulsive parts in a sonar signal. The frequency-domain kurtosis was derived as a function of frequency based on STFT. The first reported use of SK for bearing faults was in [78], it was applied to a vibration data collected from a defective bearing and SK was used, based on STFT, to find the highest SNR region. SK is an indicator which indicates

to the distribution of the impulsiveness of a signal in the frequency transformation using fourth-order statistics. This gives SK the ability to identify the transients and in which frequency band these transients occurred even in the presence of high additive noise. SK can be obtained from STFT by sliding the window along the signal and calculate the local Fourier transform for each window at the time t. The magnitude $\langle|X(t,f)|^2\rangle$ is squared so can represent the local power spectrum at the time $_t$ as a function of frequency and averaged over time. However, $\langle|X(t,f)|^2\rangle$ can be seen as a function of time and interpreted as a complex envelope of $x(t,f)$, if the frequency bandpass filtered contains pulses, it can be detected by taking the kurtosis of the complex envelope $X(t,f)$ as:

$$K(f) = \frac{\langle|X(t,f)|^4\rangle}{\langle|X(t,f)|^2\rangle^2} - 2 \tag{2.12}$$

Where (-2) is used to normalize the result to be zero when $X(t,f)$ is a complex Gaussian, and the operator $\langle . \rangle$ is time averaging[74]. For maximum SK, the window of STFT has to be smaller than the spacing between pulses and larger than the pulse itself.

2.2.4 Kurtogram and Fast Kurtogram

As stated above, the STFT length, as well as the bandwidth of the bandpass filter, have an important impact on the SK value obtained. For this, Antoni et al.[79] proposed Kurtogram to show a two-dimensional map contains SK obtained from different STFT window lengths. The results will be calculated for all potential combinations of bandwidths, used for bandpass filtering, and centre frequencies.

In practice, it will be a high computation cost to calculate the Kurtogram for all potential combinations of bandwidths, used for bandpass filtering, and centre frequencies. Thus, Fast Kurtogram was developed, an as alternative to Kurtogram, by Antoni in [79] to reduce the computational cost of calculating the kurtogram by adopting the fast multi-rate filter-bank procedure. The bandwidths will be iteratively halved and the process will begin with the entire signal spectrum as a one-sample window.

2.3 Adaptive and Iterative Signal Based Approaches

Signal-based methods are widely implemented to the task of feature extraction in CM and the extracted features are used for process monitoring. However, due to the presence of strong

background noise in the acquired data, identifying the faults from raw data has not always been possible, hence, denoising and enhancing the desired features are fundamental steps in CM for accurate and effective detection and diagnosis procedure[63].

2.3.1 Signal Averaging

Signal averaging has been for a long time and widely used to reduce the noise and enhance experimental signals[80, 81]. Signal averaging is achieved by using the available time for the measurement in taking many identical successive measurements, instead of using all the available time for a single measurement. And then with a shorter time constant, the optimum filter applied. By adding successive signals together, because if its incoherence, the noise will tend to average, whilst, signals will tend to add coherently[82]. Averaging can be done in both time domain and in the frequency domain[83], also in the time-frequency domain [84]. McFadden and Toozhy [85] suggested the averaging technique of the envelope signal for rolling element bearing diagnosis.

2.3.2 Linear Prediction

Linear prediction is a well know technique uses Autoregressive (AR) model and has been implemented as a way of extracting a signal of interest from a contaminated signal. Linear prediction can be derived [9]as in (2.13)

$$\hat{x}(t) = -\sum_{k=1}^{p} a(k)x(t-k) \tag{2.13}$$

Where $\hat{x}(t)$ is the predicted value and is derived as weighted sum of p value. Sawalhi in [76], Claimed that it is possible to use autoregressive methods in linear prediction to remove deterministic components, by using an AR-based linear prediction filter to separate the impulses originates from a defective bearing from the measured signal.

2.3.3 Adaptive Noise Cancellation (ANC)

ANC is used to separate two uncorrelated components from a primary signal. The procedure uses a reference signal contains only one of them. The ANC will try adaptively to find a transfer function, and then the modified reference signal will be subtracted from the primary signal, while the other components are left unchanged[86]. The output of the ANC is given as in

$$C = x + n_0 - y \tag{2.14}$$

Where x is the signal path, n_o is the primary input noise path, y is the output of the filtered reference input and will be subtracted from $x + n_0$ to give the canceller output C [87]. In [88, 89], ANC was used to extract a faulty bearing signal corrupted by severe gear meshing noise. The results showed considerable success in detecting a seeded defect-bearing signal corrupted by gear background noise. In [88], ANC was implemented in bearing fault detection and was used to denoise the signal and improve the SNR. The study showed that the spectral and statistical analysis techniques have become more effective in the diagnostic roles after the application of ANC.

2.3.4 Self-Adaptive Noise Cancellation (SANC)

SANC is another signal enhancement technique used to separate a deterministic frequency from other random frequencies. In this method a delayed version of the primary signal is made as a reference signal, then if the correlation length of the random signal is shorter than the delay Δ, the SANC will not be able to identify the relationship, and the transfer function between the delayed version of itself and the deterministic part of the signal and will be found [74]. Many adaptation rules are used to minimize the total output noise power, however, the most widely used algorithm is the least mean square (LMS). The output of the filter is derived [90] as below

$$y(k) = W^T(k).X(k - \Delta) \tag{2.15}$$

Where W is the a vector represents the H weighting coefficients w_i, $X(k - \Delta)$ a vector with the delayed version of the signal and the output can be then derived from the below :

$$s(k) = x(k) - y(k)) \tag{2.16}$$

In [90], the study investigated the implementation of SANC in order to denoise the bearing faulty signal, the study claimed that the results obtained showed that SANC is capable to eliminate the unwanted noise and facilitate the recognition of the different components in the spectrum of the signal. For more details, [91] has reviewed and evaluated the adaptive algorithms for noise cancellation. However, when using SANC, the convergence stage may last for a long time period, especially for filters of high orders[74].

2.3.5 Time Synchronous Averaging (TSA)

TSA based on the idea of exploiting the natural periodicity of vibration signals, this means averaging a signal over many rotations can remove almost all the components of a signal which are not at a frequency related to the rotation[92]. TSA cab be derived [8] as:

$$y_a(t) = \frac{1}{N}\sum_{n=0}^{N-1} y(t + nT) \tag{2.17}$$

Where $y(t)$ is the averaged signal, T is the averaging period, N stands for the average segments number. It has been widely applied to denoise and enhance vibration signal for two desirable reasons: firstly, it can be used to reduce the components of a vibration signal that are asynchronous with rotating shaft frequency, secondly, as result, this will amplify the amplitude of the important component in the vibration signal relative to the noise. McFadden and Toozhy applied TSA to the envelope signal collected from a bearing [85], vibration signal was synchronized relative to the cage with the rotation speed of the shaft. The study claimed that it revealed the spalls which were induced already on the inner race. TSA was extensively studied and applied as a filtering process to vibration signal of gears and bearing and in some studies was combined with other techniques, for instance, [93] has applied TSA to envelope signal combined with Support Vector Machines for bearing diagnosis and it was claimed that lead to efficient bearing fault diagnosis. Another application of TSA was reported in [94], the proposed technique is based on analysis of the jerk energy gradient of the synchronously averaged vibration signal collected from a faulty bearing with inner race fault and outer race fault.

2.3.6 Empirical Mode Decomposition (EMD)

EMD is a self-adaptive signal processing technique and one of the time-frequency analysis techniques for nonlinear and non-stationary data processing., EMD can decompose any input data set, based on its local characteristic time scale, into a set of a finite and small number of components called intrinsic mode function (IMF) [95]. EMD decomposes data $x(t)$ into Intrinsic Mode Function (IMFs) k_i [96] as in

$$x(t) = \sum_{i=1}^{n} k_i + r_n, \tag{2.18}$$

In which, after extracting n of IMFs, r_n is the residue of $x(t)$. EMD has been used for enhancing bearing signals, [97] has applied EMD to extract the fault features and remove the noise from the data.

2.3.7 Minimum Entropy Deconvolution (MED)

MED was developed by Wiggins [98], and since then it has been widely applied in signal processing. It more likely, especially in the case with high-speed bearings, individual fault pulses will be modified by the transmission path, thus, sharp impacts, which are travelling through the transmission path between the bearing and the sensor, may be extremely misshapen [99]. The MED aims to find the optimal filter $g[\,]$ to invert impulse response function of the system $h[n]$ as$(h * g)[n] = \gamma[n - l_m]$, where $\gamma[n - l_m]$ represents Kronecker delta function. The filter $g[n]$ is derived with k coefficients [100]as in (2.19)

$$y(n) = \sum_{k=1}^{K} g[k]x(n-k) \tag{2.19}$$

The objective function is sought by maximizing the kurtosis value. It is used to enhance the fault pulses by removing the effect of the transmission path. MED technique has been used in [76] by Sawalhi et al. for maximizing the capability of Spectral Kurtosis. The study claimed that MED effectively de-convolved the effect of the transmission path and clarified the impulses also greatly enhanced the results of envelope analysis in diagnosing the bearing fault. Jiang et al. [101] implemented MED technique to seek an optimal set of filter coefficients, to improve the fault impulses, in order to make the filtered signal containing clearer fault information, and envelop spectrum analysis was used to demodulate the fault frequencies. Barszcz et al. [99] presented the usage of the MED technique to enhancement the fault features for both fault detection and diagnosis. Despite the successful implementation of MED in some cases, [102] it has been reported that MED is unable to handle band-limited data properly. Thus, when analyzing noisy data, it is difficult to overcome this limitation.

2.3.8 Wavelet Transform

It has to be noted that due to time-varying environments, the CM data is usually complex in reality[26]. Time-Frequency domain analysis, it has been an effective technique to analysis vibration data collected from rotatory machines. Both stationary and non-stationary signal can be effectively analysed using time-frequency domain techniques. This can be considered as the main advantage compared to frequency domain techniques. The well-known time-frequency techniques are Wavelets, Short Time Fourier Transform and Wigner-Ville Distribution [103]. This research focuses on Wavelet Transform and its application to vibration signal analysis.

2.3.8.1 Wavelet Transform for Feature Extraction

In the conventional filter-based signal denoising methods, the frequency components outside a predefined range are normally set to zeros, which may lead to losing some useful required information from the signal. The fault impulses appear in signals normally cover a wide frequency band, thus, the filtering methods may smooth some of the fault impulses. Features extracted from time or frequency domains cannot include all useful information. Thus, time and frequency domain are combined both in Time-Frequency domain method such as Short-time Fourier Transform (STFT), where a signal is decomposed into frequencies and corresponding time resolution and then applying FFT to each window to monitor frequencies over time, however, STFT suffers from the fixed window whose time-frequency resolution is constant. For more details, the performances of the different time-frequency domain methods are compared and can be found in[104].

Besides the ability of wavelet in analysing non-stationary data which is considered as the original intention of developing wavelets, another successful implementation of wavelet transform if feature extraction. The compact support property of wavelet gives it the feature of energy concentration, this results in yielding many coefficients with small energy which can be excluded without losing the important and informative components in the analysed signal. Hence, few coefficients can be used to represent the diagnostic features. The key issue is to identify the best coefficients that represent the diagnostic features. Thresholding has been known as a promising solution and widely accepted to shrink the uninformative components from the analysed signal[104].

2.3.8.2 The adaptive property of Wavelet transform

Wavelet transform was developed, to overcome STFT problems, as an advanced technique of signal and image processing [105]. Compared to the constant resolution property of the time-frequency in STFT, the time-frequency resolution of the wavelet transform is adaptive and depends on the frequency of the signal, hence, it's considered as a multiresolution analysis. The adaptive property gives wavelet the ability to obtain high time but low-frequency resolution at high frequencies, whilst, at low frequencies, it can obtain low time but high-frequency resolution[106]. The wavelet transform is well known and widely used in data processing for several applications such as image processing and signal processing.

In condition monitoring, there are different applications of wavelet such as the analysis of time-frequency domain, feature extraction, signal enhancement and denoising, signal compression etc.

[107]. In terms of vibration signal, wavelet gives an excellent representation for several types of signals that containing jumps and spikes (singularities), it provides optimal sparse representation for such signals, the sparsity comes from the fact that since wavelets oscillate locally, only wavelets overlapping a singularity will have large wavelet coefficients whilst the rest of other coefficients will have smaller coefficients [108].

Wavelet is obtained by scaling factor a and translation factor τ from mother wavelet $\psi(t)$ as:

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}}\left(\frac{t-\tau}{a}\right) \ (a > 0) \tag{2.20}$$

Where a, τ scaling and translation parameters respectively[109]. Several kinds of wavelets are available for a different of types the applications can be found in the literature. Comparison study based on their properties can be found in [110]. Kunpeng et al., in [111] reviewed the state-of-the-art of wavelet transform methods with some results.

In this research we consider wavelets that are analogous to Daubechies' construction for orthogonal and compact support, DWT, DT-DWT[108], DD-DWT [112], and double density dual tree DDD-DWT [113], as they suit the application of analysing vibration bearing signals for several reasons, Daubechies wavelets provide the best match to vibration signal produced from a defected bearing. Also, it has a given number of vanishing moments and it supports FIR filters and allows the use of the fast algorithm [110].

Unlike filtering-based methods, the wavelet de-noising method does not corrupt the important components of the signal, because the wavelet shrinks the noise using simultaneous re-scaling in both domains frequency and time [114]. Wavelet transforms has been widely studied and proposed for experimental vibration signal processing.

2.3.8.3 Continues Wavelet Transform (CWT)

The continues wavelet transform has been investigated to map the data into two-dimensional coefficients to identify the presence of impacts in a signal due defects[115]. CWT is represented as:

$$cwt(a,\tau) = \frac{1}{\sqrt{a}}\int x(t)\psi^*\left(\frac{t-\tau}{a}\right)dt \tag{2.21}$$

Where $\psi^*(t)$ denotes the complex conjugate of $\psi(t)$ and $\psi^*\left(\frac{t-\tau}{a}\right)$ is the basis wavelet function[109]. However, in the application of CWT, a very redundant transform will be produced

which leads to the increase of computational time[116]. This led to the emergence of widely implemented dyadic wavelet called Discrete Wavelet Transform (DWT).

2.3.8.4 Discrete Wavelet Transform (DWT)

In order to overcome the redundant transform produced form the CWT, a discretization method is applied to dilation and translation parameters. This can be done by changing the dilation parameter $a\ by\ 2^a$ and the parameter $b\ by\ 2^a b$ [109]. When the choice of scales and translations based on powers of two, the analysis will be more efficient and with the same accuracy of CWT. This type of analysis is called DWT [117]. The DWT was developed by Mallat [118]and it can be expressed as:

$$dwt(a,b) = \frac{1}{\sqrt{2^a}} \int x(t)\psi^* \left(\frac{t - b2^a}{2^a}\right) dt \tag{2.22}$$

This transform is orthogonal and non-redundant wavelet[119]. DWT is critically sampled wavelet using FIR perfect reconstruction filter banks [120]. Wavelet and dilation functions at multi-scales are generated as shown in (2.23) and (2.24).

$$\emptyset(t) = \sqrt{2} \sum_n h(n)\emptyset(2t - n) \tag{2.23}$$

$$\psi(t) = \sqrt{2} \sum_n g(n)\emptyset(2t - n) \tag{2.24}$$

Where $h(n)$ represents low pass filter and $g(n)$ high pass filter, $\phi(t)$ scaling function and $\psi(t)$wavelet function [111]. The interest in using DWTs method comes from the fact that signal impulses can be identified from the high frequencies of the wavelet with a good resolution [121]. Low time resolution and high-frequency resolution can be obtained at low frequencies, whereas, a high time resolution but a low-frequency resolution can be obtained at high frequencies [104].

Several DWT denoising approaches have been developed amongst them coefficient modelling and wavelet shrinkage methods. Wavelet shrinkage method is the most widely adopted method for signals denoising, as it requires low computational complexity. This method is based on thresholding technique to shrink the noise and preserve the important components in the residual signal. Therefore, wavelet-based de-noising techniques have been extensively implemented in recent years.

The sequence of applying wavelet-based de-noising is to:

Firstly, decompose noisy data into *L* levels, where *L* is the number on decomposition levels.

Secondly, perform a nonlinear thresholding after selecting the optimal thresholding value, to remove noise from the data up to *L* levels.

Finally, a reconstruction of the de-noised signal is carried out through the inverse wavelet transform of the shrunk detail coefficients. Using the nonlinear shrinkage in the transformation domain makes this method distinctive from other linear denoising methods [122].

It is possible to denoise contaminated vibration signal with sharp transients through the thresholding function in wavelet transformation domains [123, 124]. Several wavelet-based methods for the denoising have been available, for instance, [125] implemented the wavelet transform scale space filtering techniques and Bayes shrinkage for noise estimation and denoising. Altman [126] used wavelet packet analysis based multiple band-pass filtering to denoise bearing vibration signals and good results were obtained. In [127] the signal is denoised first to eliminate unwanted noise and spikes using wavelet de-noising, and the obtained wavelet coefficients were fed to non-linear PCA algorithm as an input vector. Although wavelet has been widely applied for signal processing, the denoising process affected by the base function where it cannot change adaptively according to the signal characteristics.

Second generation wavelet (SGWT) emerged in recent years and has been widely implemented in signal processing. It is based on a lifting scheme to construct biorthogonal wavelets introduced by Daubechies and Sweldens [128], it has some advantages and faster than the conventional wavelet DWT. For instance, Li et al. [129], proposed a technique named adaptive morphological gradient lifting wavelet based on SGW for bearing vibration signal denoising and feature extraction. However, some shortcoming has been reported when using second generation wavelet, it suffers from frequency aliasing problem due to the splitting and merging operations process[130]. To overcome frequency aliasing, Bao et al. [131], proposed a redundant second-generation wavelet (RSGW). In this method, split and merge operations are avoided in the transformation process, thus, it does not suffer from the frequency aliasing problem. Lu et al. [130] proposed an adaptive redundant SGWT denoising method for vibration signal. The extracted features from the de-noised signal used as inputs into the SVM for fault detection. Feng et al. [132] proposed differential evolution (DE) optimization and antisymmetric real Laplace wavelet as a filter to eliminate the interferential vibrations and remove stochastic noise from the original vibration signal. Then used envelop analysis to detect the bearing defects. In [133] Su et al. applied optimal Morlet wavelet

filter to eliminate the frequency associated with interferential vibrations in rolling element bearing and an autocorrelation enhancement algorithm is applied to the denoised signal to shrink the residual in-band noise and highlight the periodic faulty features. Despite the wide implementation of DWT and SGWT, when it comes to the non-stationary and non-linear vibration signal, conventional wavelet suffers from two main disadvantages as following:

- High shift sensitivity, a small shift in the signal may result in a major variation in the energy distribution of DWT coefficients at different levels. The DWT is shifting-sensitive because the coefficients have unpredictable behaviour when the input signal shifts in time [134].
- Frequency aliasing which may lead to loss of important components of a signal.

For the mentioned reasons above, it turns out the idea of using an expensive wavelet instead of a critically-sampled one. An expansive wavelet transforms (N)-point signal into (M) coefficients with (M > N). A number of expansive DWTs types have been developed such as Undecimated DWT, Dual-Tree Complex DWT and Double-Density DWT.

2.3.8.5 Expansive DWTs

The expansive or redundant wavelet transform results in a redundant representation of a signal in the wavelet domain. The wavelet coefficients are longer in length than the original signal, by increasing the sampling plane of the time-frequency using oversampled filter banks. As a result, this will give more flexibility in the wavelet design and this type of wavelet has more advantages compared to critically sampled DWT.

2.3.8.5.1 Undecimated DWT

The undecimated discrete wavelet transforms (UDWT) also known as redundant wavelet transform is an improved version of DWT. UDWT have the advantages that it is a shift-invariant transform and there is no down-sampling operation involved Unlike the DWT, which downsamples the detail and approximation coefficients at each level [135]. Therefore, the approximation and detail coefficients of UDWT equal in length to the original signal at each level. UDWT up-samples the coefficients at each level of both low and high pass filters. The up-sampling operation is equivalent to dilating wavelets and then down-samples in the reconstruction process. The coefficients resolution declines with growing levels of decomposition. However, in the case of j scales are implemented, UDWT is redundant (expensive) by $j + 1$ and therefore considered as a highly redundant transform [136].

UDWT has been studied for vibration signal analysis, for instance, Hao et al., [137] have implemented undecimated wavelet (MUDW) for noise smoothing and feature extraction of the bearings signal in rotating machines. [138] used undecimated wavelet transformation based on lifting scheme for denoising vibration signal gathered from a gearbox, the results showed the capability of the method in enhancing the vibration signal. [139] implemented UDWT to denoise gear vibration signal as a pre-processing technique, the approximation coefficients were used and have shown to be most suitable to denoise the signal. Zhang et al.[140] investigated the application of UDWT to feature extraction of impulsive vibration signal and claimed that the UDWT is effective in fault diagnosis. In [141], UDWT has been applied to bearing vibration signal and, the results showed that it can diagnose bearing failure quickly and effectively. Qin et al.[142] proposed a higher-density dyadic WT and several wavelet transforms were investigated for vibration signals collected from faulty roller bearings. The study claimed that the presented technique outperforms the other typical wavelet transforms, however, this method produces high redundant wavelet, which as a result, increases the complexity of computation.

Despite the reported successful application of UDWT compared to the critically sampled DWT, especially for non-stationary signals, it has an expansion-factor of log N, therefore it is considered as a highly redundant expensive wavelet and therefore computationally high expensive [131, 143, 144].

2.3.8.5.2 Double Density DWT (DD-DWT)

Double Density DWT was introduced by [112] and is grounded on over-sampled filter banks, to reduce the translation sensitivity, instead of critically sampled conventional DWT. It is analogous to well- known Daubechies orthonormal wavelets and now in the oversampled case with tight frames. It is called oversampled as the overall rate of the subband signals is larger than the input by 3/2. the DD-DWT uses scaling function $\phi(t)$ and two distinct wavelets ψ_1 and ψ_2, where one wavelet is set to be offset by half from the other wavelet [113] as shown in (2.25) .

$$\psi_2(t) \approx \psi_1(t - 0.5) \tag{2.25}$$

DD-DWT has one scaling function (t) , two wavelet functions $\psi_1(t), \psi_2$ and they should satisfy scaling function (2.26) and wavelet function (2.27):

$$\emptyset(t) = \sqrt{2} \sum_n h_0(n) \emptyset(2t - n) \tag{2.26}$$

$$\psi_i(t) = \sqrt{2}\sum_n h_i(n)\emptyset(2t-n)\ , i = 1{,}2 \tag{2.27}$$

In the equations, $h_0(n)$ represents the low pass filter while $h_i(n),\ i = 1{,}2$ are the high pass filters [112].

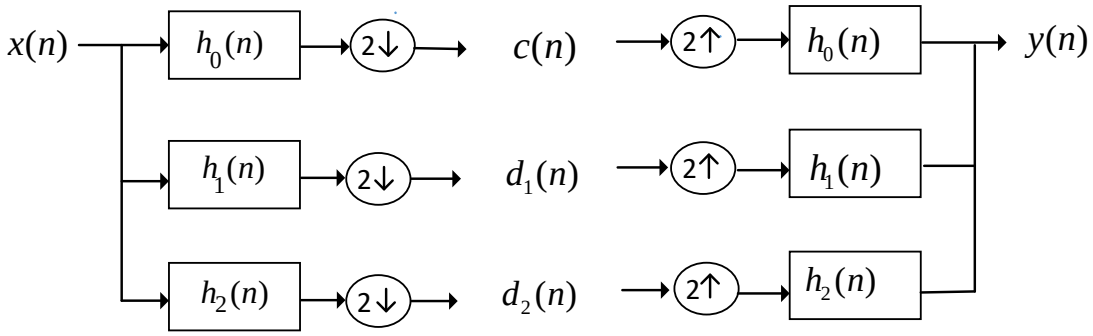

Figure 2-3 DD-DWT oversampled analysis and synthesis filter banks

For satisfying perfect reconstruction condition where $Y(z) = X(z)$, filters should satisfy the conditions (2.28) and (2.29) when $H_i(z)$ is the Z transform of $h_i(n)$ as:

$$H_0(z)H(1/z) + H_1(z)H_1(1/z) + H_2(z)H_2(1/z) = 2 \tag{2.28}$$

$$H_0(z)H_0(-1/z) + H_1(z)H_1(-1/z) + H_2(z)H_2(-1/z) = 0 \tag{2.29}$$

With the design of having more wavelets, a narrower spacing between adjacent wavelets within each scale will be obtained [145]. DD-DWT is constructed with decomposing and reconstruction three filter banks oversampled by 3/2 as shown in Figure 2-3. Motivated by the success of adopting an overcomplete expansion by a factor of 2 redundancy in dual tree discrete wavelet transform DT-ℂWT[144], which improves the shift-sensitivity of the DWT, Double density DWT was proposed. DD-DWT has several advantages that make it outperform critically sampled DWT and undecimated DWT. The double density DWT is a less expansive version of the undecimated DWT. Also, DD-DWT has very smooth wavelets and it is nearly shift-invariant. This property is important for extracting periodical peaks. Another property is the reduced frequency aliasing effects which claimed to be effective for detecting harmonic features and makes the DD-DWT well suited for applications such as non-stationary signal processing that rolling bearing produces. DD_DWT has improved time-frequency bandwidth product. It has more wavelets than necessary which give a narrower spacing between adjacent wavelets within the same scale and is less redundant than

undecimated wavelet [113]. In UDWT, the redundancy grows with the number of levels, however, DD-DWT is just two times expensive [120].

DD-DWT has been implemented in image processing and denoising, for instance, Sveinsson. et al, [146] applied DD-DWT to denoise Synthetic Aperture Radar (SAR) images by reducing the speckle of SAR images and claimed that the method was able to remove the speckles and enhanced the performance of detection for SAR based recognition. In [147], DD_DWT applied for image denoising in order to derive texture feature of the images, the results showed the potential capacity of DD-DWT in performing the task. Arfia et al [148] used DD-DWT to filter image components and experimental results showed the effectiveness of this image denoising method. In [149], a comparative study has been carried out using different wavelets for Ground Penetrating Radar (GPR) signals and it was found that the same level of processing, the DD-DWT outperforms the Haar mother wavelets or Daubechies order 6 when using soft thresholding. Another comparison study carried out [150], between DWT and DD-DWT in image denoising, It was found that with the same level of decomposing, the DD-DWT outperforms the DWT.

DD-DWT has several advantages that make it outperform critically sampled DWT and undecimated DWT. The double density DWT is a less expansive version of the undecimated DWT. Also, DD-DWT has very smooth wavelets and it is nearly shift-invariant. This property is important for extracting periodical peaks. Another property is the reduced frequency aliasing effects which claimed to be effective for detecting harmonic features and makes the DD-DWT well suited for applications such as non-stationary signal processing that REBs produce. DD_DWT has improved time-frequency bandwidth product. It has more wavelets than necessary which give a narrower spacing between adjacent wavelets within the same scale [108]. However, according to my knowledge, DD-DWT has never been explored to the scenarios of detecting and diagnosing faults from machine components such as bearings.

2.3.8.5.3 Dual-Tree Complex DWT

The Dual-Tree Complex DWT (DT-ℂWT) is an enhancement to the DWT to overcome the limitation in DWT and to mitigate the expensive cost of the undecimated DWT.

DT-ℂWT was initially proposed by Kingsbury [144] as an expensive wavelet and investigated later by Selesnick [108]. DT-ℂWT can be expressed as:

$$\psi(t) = \psi_h(t) + i\psi_g(t) \tag{2.30}$$

Where $\psi_h(t)$ is the real and, $i\psi_g(t)$ is the imaginary part [151]. It uses two critically sampled DWT trees in parallel, one tree to generate the real part and second tree to generate the imaginary part of the wavelet coefficients separately. The wavelet coefficients of the real part can be expressed as:

$$d_j^R(k) = 2^{\frac{j}{2}} \int_{-\infty}^{+\infty} x(t)\psi_h\left(2^j t - k\right)dt, j = 1,2,\dots J \tag{2.31}$$

Where j is the level, while the scaling coefficients can be expressed as:

$$s_J^R(k) = 2^{\frac{J}{2}} \int_{-\infty}^{+\infty} x(t)\varphi_h(2^J t - k)dt \tag{2.32}$$

Similarly, wavelet $d_j^{Im}(k)$ and scaling $c_j^{Im}(k)$ coefficients for the imaginary part can be derived from $\psi_g(t)$ and $\varphi_g(t)$. And then the wavelet coefficients and scaling coefficients can be derived by combining the dual-tree [152] as in (2.33) and (2.34):

$$d_j^c(k) = d_j^R(k) + id_j^{Im}(k), \qquad j = 1,2,\dots\dots J \tag{2.33}$$

$$s_J^c(k) = s_j^R(k) + is_J^{Im}(k) \tag{2.34}$$

When setting other coefficients to zero, the decomposition coefficients are individually reconstructed [153] as shown in (2.35) and (2.36).

$$d_j(t) = 2^{(j-2)/2}\left[\sum\nolimits_m d_j^R(k)\psi_h(2^j t - m) + \sum\nolimits_n d_j^{Im}(k)\psi_{\mathrm{g}}(2^j t - n)\right], j = 1,2,\dots,J \tag{2.35}$$

$$s_J(t) = 2^{(J-1)/2}\left[\sum\nolimits_m s_J^R(k)\varphi_h(2^J t - m) + \sum\nolimits_n s_J^{Im}(k)\varphi_{\mathrm{g}}(2^J t - n)\right] \tag{2.36}$$

m and n represents the filter lengths, $d_j(t)$ and $s_j(t)$ are real and equal in length with $x(t)$. The analytic of $\psi(t)$ gives DT-$\mathbb{C}$WT excellent properties, to achieve this, $\psi_h(t)$ and $\psi_{\mathrm{g}}(t)$ should form a perform Hilbert transform as:

$$\psi_g(t) = H\{\psi h(t)\} \tag{2.37}$$

where $H[\cdot]$ represents the Hilbert transform and the two DWTs form an approximate Hilbert transform pair, so the second pair is viewed as a Hilbert transform of the first pair, and as demonstrated in [153, 154] the associated scaling filter of one pair should be approximately a half-sample delayed version of the scaling filter of the other pair as:

$$g_0(t) \approx h_0(t - 0.5) \tag{2.38}$$

When designed in this way, this wavelet is called Dual-Tree Complex DWT [155]. It was demonstrated that, by using a pair of wavelet transforms, significant improvements can be achieved in signal processing.

DT-ℂWT comes with some additional important properties, it is approximately shift-invariant, perfect reconstruction using short linear-phase filters and limited redundancy, 2:1 for 1-D with independent of the number of scales [113, 143]. This doubled redundancy offers more information about the data for analysis with limited extra computation cost. The implementation of DT-ℂWT, using separable filter banks, significantly improved denoising capability[156]. In the DT-ℂWT, implementation of decomposition and reconstruction is carried out using two parallel DWTs with different low-pass and high-pass filters in each scale as shown in[153].

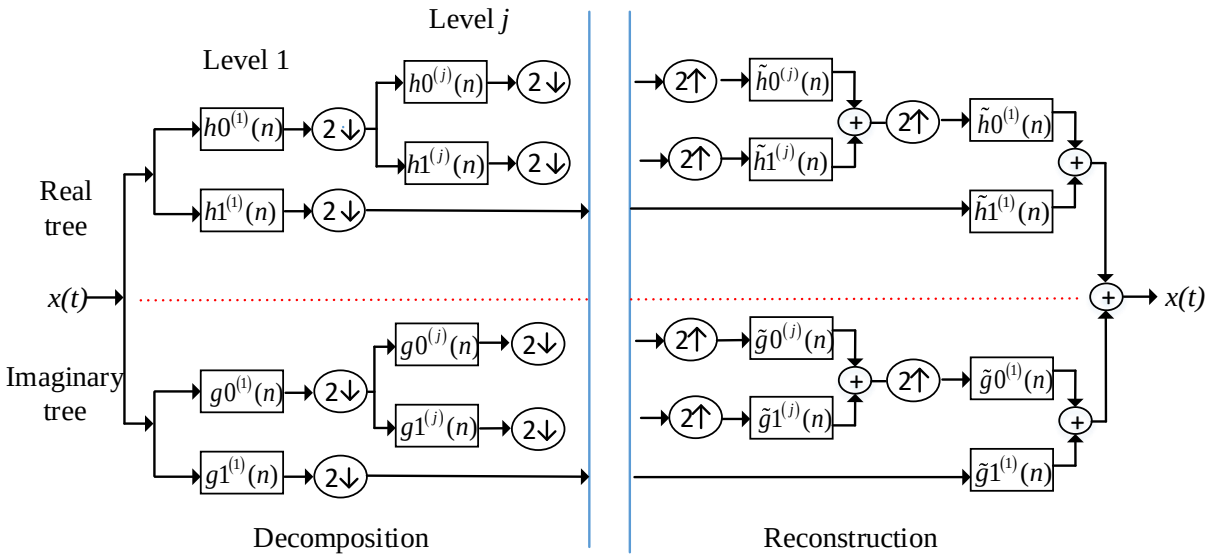

Figure 2-4 dual-tree complex wavelet transform [153]

The tree of DT-ℂWTs uses a different set of filters, and each set satisfies the perfect reconstruction condition. The perfect reconstruction property of the DT-ℂWT, high shift sensitivity, low computational and perfect reconstruction make it a good candidate for denoising signals [157]. The design of the filters of DT-ℂWT has some important characteristics such as the low-pass filters in the two trees differ by 0.5 a sample period. All filters are from the same orthonormal set. Reconstruction filters are the reverse of analysis. The filters of Upper-tree are the reverse of the second-tree filters. The two trees have the same frequency response. All of this comes with a limited

redundancy cost of 2N in 1-D, which is much lower than the redundancy of a perfectly shift-invariant DWT. DT-ℂWT has been extensively investigated in image and signal processing, for instance, Wang et al [153] used the DT-ℂWT with the NeighCoeff thresholding for denoising gearbox vibration signal. The study claimed that the developed technique outperformed DWT and fast kurtogram. [158] DT-ℂWT was applied to denoise audio signals contaminated with additive white noise of different intensity. The study claimed that DT-ℂWT outperforms conventional DWT in the case of optimal selection of threshold level.

2.3.8.5.4 The Double-Density Dual-Tree DWT (DDD-DWT)

Double Density Dual Tree DWT was developed by Selesnick in [155]. This wavelet is an overcomplete wavelet and intended to have the properties of both DT-ℂWT and DD-DWT into one transform. DD-DWT and DT-ℂWT share some properties such as both are overcomplete by a factor of 2 in 1D, approximately shift invariant, both adopt FIR filter banks with perfect reconstruction. However, they are different in some aspects such as two wavelets form a Hilbert transform in Dual Tree whilst in Double density they set to be offset by one half from each other, they use a different structure of filter bank. DT-ℂWT is considered as a complex while DD-DWT not viewed as such. These differences and similarities motivated the idea of combining the two wavelets in a single transform which have the advantages of both, dual tree and double density wavelets. The new function is a dyadic wavelet with tight frames based on two scaling functions $\phi_u(t), \phi_l(t)$ and four distinct wavelets designed in a specific way as in [159] :

$$\psi_{h,j}(t), \psi_{g,j}(t), \qquad j = 1,2 \tag{2.39}$$

From the four wavelets, one pair is set to be offset from the other pair by one half as:

$$\psi_{h2}(t) \approx \psi_{h1}(t - 0.5), \qquad \psi_{g2}(t) \approx \psi_{g1}(t - 0.5) \tag{2.40}$$

While $\psi_{g1}(t)$ will form approximate Hilbert transform of $\psi_{h1}(t)$ and $\psi_{g2}(t)$ will form an approximate Hilbert transform $\psi_{h2}(t)$ as:

$$\psi_{g1}(t) = H[\psi_{h1}(t)], \qquad \psi_{g2}(t) = H[\psi_{h2}(t)] \tag{2.41}$$

As seen in Figure 2-5, two separate filter banks $h_1(n)$ and $g_1(n)$ are used with $i = 1,2$. The upper tree represents the real part whilst the lower tree represents the imaginary part of the complex transform [159, 160]. The DDD-DWT uses two oversampled iterated filter banks working in a parallel manner. The time-revers of the analysis filters are used as synthesis filters. The filter banks

satisfy the properties: perfect reconstruction, Hilbert transform pair property, specified vanishing moments, short support.

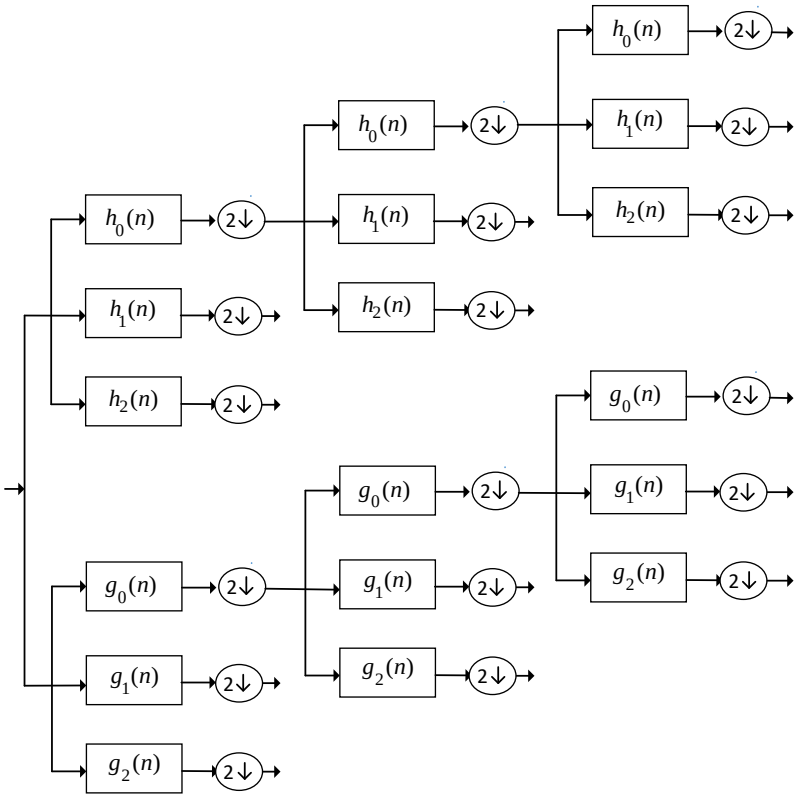

Figure 2-5 Double Density Dual Tree DWT [155]

When H_i(z) is the Z transform of $h_i(n)$ and g_i(z) is the Z transform of $g_i(n)$. The perfect reconstruction condition as the following:

$$\sum_{i=0}^{2} H_i\,(z)H_i(1/z) = 2, \qquad \sum_{i=0}^{2} H_i\,(z)H_i(-1/z) = 0 \tag{2.42}$$

And the same for $g_i(z)$. The wavelet and dilation functions through (2.43) and (2.44) and (2.45).

$$\phi(t) = \sqrt{2} \sum_{n} h_0(n)\phi h\,(2t-n) \tag{2.43}$$

$$\psi_{h1}(t) = \sqrt{2} \sum_{n} h_1(n)\phi h\,(2t-n) \tag{2.44}$$

$$\psi_{h2}(t) = \sqrt{2} \sum_{n} h_2(n)\phi h\,(2t-n) \tag{2.45}$$

And the same for $\phi_g(t)$, $\psi_{g_i}(t)$, $i = 1,2$. The new function is compact support with vanishing moments. The application of DDD-DWT includes image processing, data denoising and enhancing etc. Data Analysis Based feature enhancement and denoising Methods

Studying experimental signals normally is not an easy task, because some weak signals are non-stationary and suffer from having a low signal to noise ratio (SNR). In practice, signals do not exist without noise, however, when the noise level corrupts a signal, the noise removing become a necessary process. The denoising process can take place in the original signal (time domain) or in the transformation domain (Fourier or Wavelet)[122, 161]. for experimental signals, conventional signal processing methods cannot be used effectively to extract the true signal [162]. As a result, enhancing and denoising those signals become a real challenge [80]. Enhancement of vibration signal measured from machines has been carried out by suppressing the background noise so as to increase the sharpness of the fault impulses. Noise is normally random signals with broadband frequency and this band will overlap with the signals of interest. Thus, with general filtering methods, it is challenging to effectively remove the unwanted noise from the signals [104]. An optimal denoising method should preserve the required signal features which are fundamental to the application while eliminating unwanted noise as much as possible. Many techniques have been proposed in signal enhancement for fault diagnostics and to denoise experimental vibration signal, here some of the well-known and widely used techniques.

2.4 Automated Data Analysis based on Artificial Intelligence

The main aim of AI-based approaches is to learn the pattern from the treated signal in order to associate it to a predefined or known condition. The learned pattern is considered as data representation and used for a multi-classification task when a supervised paradigm is adopted. In this case, the entire knowledge about the expected condition must be known. The vast majority of AI-based approached addressed in this manner. On the other hand, unsupervised paradigm aims to learn the pattern of the normal condition and then use the learned model to monitor the process and detect any changes occur in the data due to a physical change. In this paradigm, AI-based technique is used for features extraction, features selection from the signals generated by the monitored system also for dimensionality reduction.

2.4.1 AI for Data Post-Processing (Classification)

The vast majority of AI-based approaches are implemented for post-processing process in CM, the expected conditions have to be pre-defined and the algorithm learns how to classify each input

instance to a predefined condition. Using AI for classification in CM has been of interest to researchers for a long time. The features are extracted based on human experience and are crafted based on the expertise knowledge[2]. For instance, Zhang et al. [163] have implemented SVM to diagnose REBs, features were extracted based on some statistical parameters from the signal in the time and the frequency domain. Saidi et al.[37], has used bi-spectrum as feature extraction and PCA to reduce the dimensionality of the data. The extracted principal components were fed to SVM to detect four types of bearing defects with different severities for each fault type. Yuwono et al. [164] proposed an automatic bearing fault diagnosis method using Hidden Markov Model (HMM) fed with the extracted fault frequency signatures by Wavelet Kurtogram and Cepstral Liftering. In [165] simulation data generated by high resolution simulates is used to train machine learning classifiers instead of using historical data, the generated data were used as inputs to train and investigate SVM and KNN, also CNN was applied to the generated data to diagnose simulated defects. The study claimed that high accuracy was achieved in all cases and applied methods. However, simulation data cannot simulate the reality and all possible operating condition of nowadays complex systems. Although AI approaches have witnessed an increase in the field of CM of REB, however, the accuracy of these methods highly depends on a set of suitably selected feature vectors as input to classifiers in order to accurately detect and identify the bearing faults.

2.4.2 AI for Data Processing (feature extraction)

The quantity of machine-readable data rapidly increases but fortunately, machine learning offers techniques by which the massive data can be automatically processed. As the performance of machine learning techniques is highly depending on the feature extraction and selection step, hence, signal processing techniques widely implemented to extract and select the input features of AI-based methods. However, they have to be very efficient in the isolation of fault characteristics from the raw signals. In the literature, several techniques based on time domain [24, 55, 166, 167], frequency domain [168]and time-frequency domain[164, 169-171] have been implemented to extract features for REBs fault diagnosis using neural networks. However, in most cases, the feature extracted from these domains were high-dimensional and redundant and as a result in gaining poor diagnostic information [172]. The task of dimensionality reduction is deeply connected to the feature extraction and selection, so the aim is to capture the significant components of a dataset [173] Thus, robust diagnostic feature extraction and analysis techniques are needed to extract the discriminative and informative fault features in a given feature vector. This is can be considered as means of either dimensionality reduction of the feature space or feature selection of the feature vector. To obtain the representative features from the data, a number of the AI algorithms in CM

were developed to carry out the task of feature extraction. Thus, many methods have been developed for diagnostic feature extraction and analysis. Among them, component analyses such as principal component analysis (PCA) [37, 174] and linear discriminant analysis (LDA)[175], Wavelet Kurtogram [164], kernel Fisher discriminant analysis (KFDA)[58], these techniques have been widely utilized in extracting features for fault diagnosis.

PCA is unsupervised analysis method and was reported to be effective for diagnostic feature analysis and the produced principal components can provide discriminative faulty features for diagnosis. However, it was claimed that PCA is limited to preserve the discriminative properties of the analyzed data as it lacks intercategory separability estimation process [176]. Feature learning methods may hold the potential solution to overcome the aforementioned limitation in conventional AI diagnosis methods. To overcome shortcoming mentioned above, adaptively learn the feature from raw data can be highly desirable which can accurately represent the data by using advanced AI techniques. In this research, a neural network based adaptive autoencoder is used for feature extraction and selection.

2.4.3 Artificial Neural Network (ANN)

Several attempts have been carried out in order to automate the fault diagnosis of REBs, ANN amongst the most AI methods applied for fault diagnosis. Also, some variants of ANN also investigated in CM such as polynomial neural networks, dynamic wavelet neural networks, self-organizing feature maps (SOM) [177], multilayer perceptron neural network[53]. ANN is a parallel data processing unit consists of an assembly of grouped connected neurons, ANN implements a training algorithm either feedforward or back propagation to perform specified functions to adjust the interconnection weights and biases value until the error between the predefined predicted output and the actual network output reaches the possible minimum value[165].

ANN has the advantages that it can readily process nonlinear, high-order, and non-stationary dynamics. The structure of ANN consists of three layers: the input layer, the hidden layer, and the output layer. Each layer consists of a number of neurons act as processing elements linked with each other in a way that they interact by using numerically weighted connections[53]. Mainly, the implementation of ANN consists of three main steps namely, training, testing, and implementation. In the training stage of a model, feature extraction and selection are the most important and critical stages.

2.4.3.1 Fundamentals of ANN

Artificial neural network (ANN) is an information processing paradigm based on biological nervous systems of the human brain. ANN consists of a predefined number of interconnected processing neurons that work together in order to solve a specific issue[28]. ANN learn by example, it simulates the learning process in human biological systems includes the adjustments of the synaptic connections that exist between the neurons. The earliest artificial neuron was developed by the neurophysiologist Warren McCulloch and the logician Walter Pits in 1943. But due to the limitation of the available technology at that time, they did not do too much[178]. To understand neural networks, as an example, we will describe the simplest possible neural network, one which comprises a single "neuron." As shown in Figure 2-6.

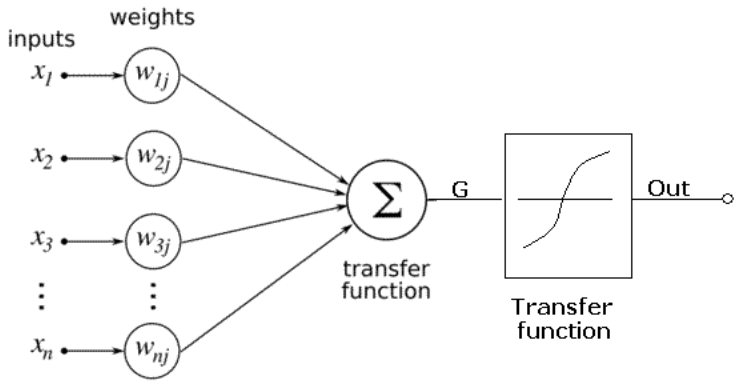

Figure 2-6 Neuron Structure

The input signals are multiplied with individual weights, and the multiplication results are summed up together into the value g as illustrated below:

$$g = \sum_{i=0}^{n} (x_i \,.\, w_{ij}) \tag{2.46}$$

The sum is fed into an activation function; one example is the sigmoid function $f(x)$, as illustrated below:

$$f(x) = \frac{1}{1 + exp^{-x}} \tag{2.47}$$

The sigmoid function output results $f(x)$ lie between 1 and 0 as shown in function, however, there are several versions of activation functions such as tang activation function where the results lie between -1 and 1.

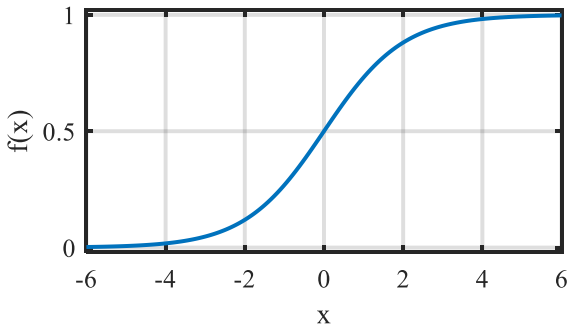

Figure 2-7 Sigmoid function

ANN can be configured for pattern recognition or data classification or any other specific application. Through a learning process, ANN has been extensively used to solve several problems such as in prediction, classification, data dimensionality reduction, self-control, function approximation, pattern recognition etc. [179]. Data processing using ANN has first to go through a training stage which can be carried out by one of several three main methods:

- **Supervised**
- **Semi-supervised**
- **Unsupervised**

In supervised training, the network is trained with input values against known output values, i.e there is a target output for each input pattern, and the network tries to learn to create the desired outputs.

In semi-supervised training is the case where a few labelled data or unlabelled data are available and one is extended in order to increase the accuracy of the network.

Unsupervised training algorithm, on the other hand, is the case where there is only training data without any corresponding target values. However, by looking for the input patterns sharing common features, the network can allow us to discover hidden patterns in the data. This is achieved by approaching problems with little or no idea what our results should look like, and the network will be able to recognize the features across the range of input patterns.

As ANN has the ability to extract meaningful information from complicated or imprecise data, thus it can be used to perform patterns recognition and detect trends that are too awkward to be noticed by other computer techniques[178]. When a neural network is well trained, it can be considered as

an "expert" in the task of analyzing the data which has been given to the network. Some other advantages can be achieved when using ANN like:

- Adaptive learning: the algorithm can adapt itself based on the input data during training, it has the ability to learn how to perform some tasks.
- Self-Organization: based on the data an ANN receives while training, it can create its own organization or representation.
- Parallel Processing: ANN has the advantages of parallel processing as the network consists of a predefined number of highly interconnected neurons working as processing elements in a parallel way to solve a specific issue.

2.4.3.2 ANN Training Methods

Generally, training methods are classified into three categories supervised, semi-supervised and unsupervised. Each of which has its advantages and drawbacks[172]. Recently, deep learning methodology has been of interest in CM and witnessed increased implementation in feature learning. Several studies applied a deep learning approach to avoid the need for using feature extractors[180]. Among them, Convolutional Neural Network is a well-known and widely implemented deep learning methodology and many studies attempted to apply CNN to machinery CM.

Supervised Training, in supervised training methods, for each input pattern there is a target output, and the network learns during the training stage how to produce the required outputs[181].

Semi-Supervised Training, semi-supervised learning method is a hybrid approach in which extending either unsupervised or supervised learning using labelled and unlabeled data.

Unsupervised Training, in unsupervised training methods, the network tries to find a hidden pattern in the input data to learn the most important features and then to be used to classify network inputs. Common unsupervised learning tasks include[179]:

- Data Clustering or grouping similar items together in each cluster, so the main task is to separate the input instances into groups.
- Anomaly detection, which identifies the abnormality in the data with reference to training data and how much the new data is different from the majority.
- Dimensionality reduction, in many cases the data needs to be processed is too large or it has a high dimensional space, unsupervised algorithms are widely utilized in the task of

dimensionality reduction which aims to represent efficiently the input data with a lower dimensional feature space.

2.4.3.3 Application of ANN to REB Condition Monitoring

The application of artificial neural networks (ANN's) to the CM of REBs (CM) offers substantial financial rewards and has consequently been the subject of considerable research recently. ANN has been extensively studied as a fault detection and diagnosis technique. ANN has been applied in several manners for REBs fault detections and diagnosis. It's used as data processing (features extraction and selection) and as data post-processing (classification task).

2.4.3.4 ANN for Post-processing

ANN has been applied as a post-processing tool so far, while the inputs were derived by conventional feature extraction and selection techniques. For example, in [182], for bearing outer race and inner race defect detection, supervised pattern recognition technique based on neural network is proposed and used both time domain and frequency domain features as the inputs of the neural network. The study reported that time domain features more accurate after adopting statistical parameters to extract the features include RMS, Variance, Skewness, and Kurtosis and used as inputs to the neural network. Unal et al. [183] proposed bearing fault diagnosis system using envelope accompanied with FFT and Hilbert Transform methods for extracting the diagnostic features, and ANN network optimized by GA was implemented to classify the features extracted from of REBs vibration data.

In order to improve the precision of fault description, Chen et al. [184] used dependent feature vector (DVA) for extracting and selecting features, and probability neural network (PNN) is proposed to denote the fault symptom attributes. In [185], a Feed Forward Neural Network (FFNN) with Levenberg Marquardt training algorithm has been proposed to diagnose the defect of bearings. A single dataset collected from one bearing was used for both the network training stage and the validation stage. However, the method cannot be generalized as they used just the same bearing for training and testing stages. Another attempt has been proposed by [186] by using Weibull Failure Rate Function in order to reduce the effect of noise factors and used Artificial Neural Network (ANN) for bearings defect diagnosis. However, the error between estimated and actual damage severity was high at the beginning of the experiment and it declined only at the end of the experiment where the REB damage becomes very large. In [187], authors have presented a rolling bearing fault diagnosis using supervised neural networks and the features extracted using

time/frequency-domain. Diagnosis results around 70% for each for three different faults, inner race, ball, and looseness, by using Computer-simulated and real data. Sreejith et al. [24], developed automated diagnosis using a feed-forward neural network for REBs. The FFNN fed with features extracted from time-domain vibration signal by using both Normal negative log-likelihood value and kurtosis value. Levenberg-Marquardt (LM) algorithm was used for supervised training. Single point faults on inner race, outer race and ball are introduced into the test bearings using electron-discharge machining. However, although the result claimed to be good, vibration signals measured from only a single location.

Moura et al.[188], applied hybrid methods based on signal processing and pattern recognition techniques to diagnose the severity of bearing defects. The features were extracted by using both detrended-fluctuation analysis and rescaled-range analysis techniques. The extracted features were fed into ANN. Supervised learning algorithm was used to train the classifier. Three different levels of bearing fault severities were introduced at the outer race and have been classified yielded reasonably good results.

Although the application of AI has been gaining importance in the area of automated REBs fault detection and diagnosis, there are still a lot of work has to be done in order to maximize the benefits of it. In the literature, most of the AI approaches have been applied using supervised learning techniques to train the network. However, supervised learning requires labelled data of each of the known faults and this needs a lot of human efforts to prepare this sort of data. Furthermore, most of the AI approaches have been applied for post-processing process (classification task), while a number of conventional techniques have been implemented for processing stage (features extraction and selection) to extract discriminative features such as time domain techniques, frequency domain, and time-frequency domain techniques. However, when adopting conventional techniques for data processing stage, extracting the features task, need to be carried out manually, and that relies on prior knowledge of expertise, moreover, this highly dependent on the advantage of human ingenuity, thus, time-consuming, costly and labour intensive[56]. Moreover, using conventional techniques in the processing stage when applying ANN, the post-process stage will suffer from being affected by background noise which degrades the measurement quality and led to the high-dimensionality of features vectors. Traditional AI techniques are unable to extract diagnostic and discriminative features from raw bearings data directly. Thus, It is can be useful to remove such random noise before proceeding with bearing diagnostic analysis[74] In unsupervised feature learning, network learning is considered as a nonlinear function, and by this function, the

raw data will be transformed from the original space into a feature space. So no prior knowledge or labelled data are required to train the network.

2.4.3.5 Recent Work Applied ANN for Feature Learning

Several studies reported in the literature that applied ANN to feature learning process, some of the recent work applied ANN to feature learning such as following:

- Gan et al.[189] in 2016 proposed a hierarchical diagnosis network constructed with two layers deep belief network combined with wavelet packet used to extract the representative features. The extracted features were fed to train the network using supervised paradigm. The health conditions were classified using the proposed network.
- Li et al. in 2016 developed a deep statistical feature learning method using stacked Gaussian-Bernoulli deep Boltzmann machine for deep feature learning. Vibration measurements of rotating machinery were used as inputs to the developed method.
- Chen and Li [190] in 2017 proposed feature learning method based on multi-sensor data fusion technique. The feature vectors were constructed by extracting statistical features from the vibration signals of different sensors. The feature vectors used as inputs to multiple two-layer sparse autoencoder networks for feature fusion. Finally, the fused feature vectors for each health condition were used to train Deep Belief Network for bearing fault classification task using supervised paradigm.
- Zhang et al. [180] in 2017 presents a new method for features learning called a transfer learning approach for fault diagnosis with neural networks. In this method, the network was fed with massive data to learn the features and the network parameters adjusted accordingly, also the structure of the network responds to the distribution changes in the training data. The data used in the training are different from the target data to improve the performance of the network.
- Shao et al.[189] in 2018 proposed an improved convolutional deep belief neural network with compressed sensing technology. Features were learned from vibration data collected from bearings based on supervised deep learning paradigm and exponential moving average technique was implemented to enhance and generalising the performance of the constructed model. In this study, all studied cases (expected defects) were made artificially to the test bearings and this can limit the methodology applied to learn patterns from the known existed cases.

- Feng et al.[190] in 2018 developed a local connection network based sparse autoencoder neural network for intelligent fault diagnosis. The developed methods were applied to vibration data collected from bearings with the different healthy conditions. Supervised paradigm is implemented using 10 class classification problem. Although the features were learned from the raw data, the accuracy of the classification achieved is limited to the artificially made defects in the studied conditions and it may not be valid to be implemented to a different unseen machinery system.
- Liu et al. [191] in 2018 presented recurrent neural network-based autoencoders for fault diagnosis, it was applied to vibration data collected from bearings with different health conditions. The autoencoder was used to denoise the vibration data and a supervised training paradigm is used to lean the features and classify the inputs into predefined class conditions. The classification accuracy achieved was compared to the results obtained from SVM and it was claimed that the proposed method outperformed SVM with good results.
- Jiang et al. [62] in 2018 developed an intelligent fault diagnosis based on an improved deep recurrent neural network. Supervised deep learning paradigm is used to train the proposed method, spectrum data used as inputs and was fed into the developed network. Stacked hidden layers were constructed for deep learning the features. Experimental vibration data was used to evaluate the effectiveness of the proposed methods and it was reported to be more effective than traditional intelligent diagnostic methods.
- Pan et el. [192] in 2018 developed a novel deep learning neural network named liftingNet to adaptively learn features directly from vibration data. It was constructed with several layers, split layer, prediction layer, updating layer, pooling layer and finely connection layer. A supervised paradigm is used to train the network using two datasets of motor bearings. The study claimed that the features were layer-wise learned and good classification results were achieved.
- Hoang et al.[63] in 2019 applied deep structure of convolutional neural network for bearing fault diagnosis, each vibration signal was transformed from 1-D into one corresponding 2-D vibration image and used as input to the network, the supervised paradigm was used to train the network. The study claimed that a very high classification accuracy was achieved using the proposed method.
- Waziralilah et el.[2] in 2019 reviewed the application of deep learning using Convolutional Neural Network in bearing fault diagnosis, the study concluded that in the literature only a

few works were reported that has proposed the architecture of CNN to learn features for bearing fault diagnosis with its severity.

2.4.3.6 Drawbacks of existing methods

- It can be seen that most of the works above applied to manual simulated health conditions which were made according to a specific diagnosis issue and probably unsuitable for other issues.
- The above methods based on the assumption that data will be collected in the same operating conditions and exhibit similar distribution and feature space. Whilst, in reality, could not be applicable for real-world working conditions.

2.5 Research motivation

Studying experimental signals normally is not an easy task, because some weak signals are non-stationary and suffer from having a low signal to noise ratio (SNR). Data collected from vibration sensors mounted on a defective bearing usually have the nature of non-stationarities and they are instead considered as cyclo-stationary signals [74]. Furthermore, bearing fault signals are always relatively contaminated by background noise and often is higher than the amplitude of the incipient anomalies. Developing a reliable algorithms to effectively detect anomalies and diagnose the health condition of such a complex systems is the main motivation of this PhD project.

To address the aforementioned weaknesses, an integrated framework with two main approaches is presented in this research. Firstly, an unsupervised feature learning CCNN is concentrated on in this study for machinery condition monitoring to carry out the task of early fault detection and severity estimated. As the CCNN has the advantage of translation invariance to tackle the cyclo-stationary nature of bearing signals, this study applies it to Tapered roller bearing fault detection and this might hold the potential solution to overcome the mentioned obstacles mentioned above in the early fault detection task.

For the diagnosis to localise the defects occur in REBs, Multiresolution data analysis approaches applied using DD-DWT to analyse the vibration data alongside with a novel thresholding technique to denoise and enhance the extracted features from the collected data. The double density DWT is used in this research as it is a less expansive version of the undecimated DWT. Also, DD-DWT has very smooth wavelets and it is nearly shift-invariant. This property is important for extracting periodical peaks. Another property is the reduced frequency aliasing effects which claimed to be effective for detecting harmonic features and makes the DD-DWT well suited for applications such as non-stationary signal processing that REBs produce. DD_DWT has improved time-frequency

bandwidth product. It has more wavelets than necessary which give a narrower spacing between adjacent wavelets within the same scale [108]. However, according to my knowledge, DD-DWT has never been explored to the scenarios of detecting and diagnosing faults from machine components such as bearings.

2.6 Summary

In this Chapter, feature extraction and enhancement based adaptive and iterative approach reviewed in (section 2.2). In in (section 2.3) automated methods based AI approach were discussed. The drawbacks of the most widely applied recent methods were listed. The approaches are explored in line with their applications to condition monitoring.

CHAPTER THREE

ROLLING ELEMENT BEARINGS AND FAILURE MODES

This chapter presents the fundamentals of REBs including REBs types, REBs components and its applications. Moreover, bearing failure modes and their potential root causes are discussed with particular interests of various slow effects such as wear and erosions. Fault frequencies appear in vibration signal signature of the bearing due to the presence of local defectives are also illustrated in this chapter.

3.1 Introduction

REBs have been widely utilized in the vast majority of rotating machines to reduce the friction between rotating adjacent parts. Bearings are one of the most important elements in these machines due to their relatively lower price and operational ease. The reliability of REBs depends on the smooth and quiet running within the machines. However, REBs are recognized as a common reason for failures in rotatory machines. The literature shows that approximately half of failures in induction machines are due to bearing faults [193, 194]. Different kinds of REBs have come into use in industrial applications, each type is designed to support specific task. Before discussing the CM of bearing, this chapter will present the basic types of REBs and their components.

3.2 REB Types

REBs come in different sizes and shapes and they can be classified into different categories according to the application such as the load they support or according to the shape of the roller elements. However, in Table 2, REBs are classified according to their standard geometric shape [177]. The main two categories are roller shape and ball shape, each of which includes a verity of types.

Generally, ball bearings can sustain lower load capacity but they are used in high-speed applications as the ball-race contact area is very small (point contact) and hence, results in smaller friction force [195]. Whilst, roller bearing can support higher load capacity because they have a larger contact area (line contact), hence, applied load will produce smaller contact stresses. On the other hand, the angular contact roller bearings can support simultaneously high radial and axial loads [196]. According to Table 2, it’s clear that bearings are classified depending on the physical shape and the kind of load direction they support, combined load, radial load or axial load. Another

classification can be rotation speed, temperature resistance and so on. For instance, ball bearings are found in precision applications and good at the application of high speed under moderate axial and radial loads. Thus, they are very popular and widely used in the industrial world. Whilst, roller bearings are utilized in diverse types of machinery such as high load or temperature, also, the application requiring simultaneously support of axial and radial load and so on [197]. Therefore, this research will focus on tapered roller bearing and they will be examined in the real experiment. Moreover, studies based on tapered bearings can be extended to other types of Roller bearings.

Table 3-1 Types of rolling element bearings[177]

REBs	**Ball**	Radial Ball	Single-row • deep-grove • angular-contact Double row angular-contact Duplex angular-contact Four-point-contact ball Self-aligning
		Trust Ball	Single-direction trust • with flat back face • with sealing ring Double-direction trust • with flat back face • with sealing rings • contact axial
	Roller	Radial Roller	Cylindrical • Single-row • Double-row Needle RBs Tapered RBS • Single-row • Double-row Spherical
		Trust Roller	Cylindrical Needle Tapered Spherical

There are several kinds of bearings available and they have their individual advantages and applications. The commonly used REBs according to their shape are illustrated in the following, ball bearings are widely used and they are available in designs to sustain radial and axial loads independently or simultaneously.

Cylindrical bearing uses rollers as rolling elements. The contact area is line-contact between the roller and the race, the load is distributed over the larger area, this gives them the ability to carry a high capacity of radial loads, as well aas high-speed capability. Moreover, rollers are designed in

such a way that their length and diameter are not much different to avoid their tendency to skew. They are available with or without ribs, they can carry the light axial load when designed with rips, while bearings without ribs can facilitate the assembly and gives them the ability to absorb the shaft expansion.

Tapered bearings are designed to carry simultaneously radial and thrust loads. For this purpose, the rollers and the races are made in a tapered shape. Their axial load capacity ratio has a linear relation with the contact angle between rollers and cones. Tapered bearings usually come into components apart, so the outer race is separate from the inner which comes with rollers and cage. The components are assembled when mounted preloaded or with the amount of clearance depends on the application.

Spherical roller bearing normally has two barrel-shaped rows of rollers in separate raceways rolling around two raceways with a spherical outer race. They are capable to support large radial and thrust load capacity and can be used for heavy industrial equipment.

Needle bearing uses cylindrical needle rollers with a small diameter. They can be used in limited radial space conditions. The large ratio between the rollers length and diameter makes them able to carry the radial load for their size. There are different designs of needle bearings depends on the application.

3.3 REB Components

Rolling bearing consists of four basic parts: inner race, outer race, rolling elements, and cage. Other special bearings have added parts such as seals and guide race. For supporting the bearing load, inner race, outer race, and rolling elements are used, while the cage is used to separate adjacent rolling elements from each other to avoid the friction. Bearing components are described in more details below.

3.3.1 Inner race

The inner race is mounted on the rotating shaft. As a result, it will be rotating with the shaft. The raceways shape will depend on the type of bearing.

3.3.2 Outer race

The outer race is usually the fixed part and it is mounted in the housing of the machine. Depending on the type of bearing, the shape of the raceways will have different forms such as cylindrical, spherical, or tapered.

3.3.3 Rolling Elements

There are two major types of rolling elements and can be classified to ball bearings and roller bearings. In balls bearings the load transfers by point contact with the raceways. Where in the roller bearing, the load transfers via line contact with the raceways. The rolling elements are often made of carbon chromium steel.

3.3.4 Cage

The cage also called separator is made from cold rolled steel strip and it is used to separate adjacent rolling elements from each other and keeps them evenly spaced to avoid the friction while the machine is operating.

3.4 Bearing Failure Modes and Their Causes

REBs failure is one of the foremost factors in rotary machinery failure, such bearing failures may occur in a very rapid time due to several reasons. The prediction of rating fatigue life of a bearing is stated as L_{10} life based on LP which theory was developed by Lundberg and Palmgren [198] as in (3.1).

$$L_{10} = \left(\frac{C}{P}\right)^{l} \tag{3.1}$$

With L_{10} means 90% reliability bearing fatigue life in 10^6 revolutions, where C represents the basic dynamic capacity, P is the equivalent load and l is the load life exponent. The LP theory is based on the hypothesis that the primary cause of bearing failure will be normal fatigue. This means about 10% of the identical bearing population running in normal operating conditions will fail within about one year, whilst, 90% of the same population will equal or exceed the one year time before showing any sign of fatigue damage. However, half of the survived bearings may fail within five years of use and referred to as L_{50} [199, 200].

Common premature failure of the bearing can be caused by several factors such as fatigue, a lack of lubrication or contaminated lubricant, excessive loading, plastic deformation, corrosion, improper installation and manufactures error [51]. Therefore, the investigation of these factors and vibration they produce is significant for the condition monitoring of bearings.

The bearing defects generally can be classified into two main categories, distributed defects also called distributed wear and localized defects also called localized wear [201]. Distributed defects is an important category of bearing faults and can affect the entire structure of the bearing. To some

extent, a deviation from the ideal design of a bearing in operation cannot be avoided and the shape of the bearing components as a result will change. The severity of change will depend on the tolerance with which the bearing was manufactured, also on the wear which will occur during operation. Distributed defects include raceway roughness, off-size rollers, waviness, and misaligned races, can be caused by several reasons like manufacture errors, improper handling or installation and abrasive wear [202]. These defects are likely will increase the contact forces and which eventually lead to premature bearing fatigue and consequently, machine failure [203].

Localized defects can be caused by fatigue damage on the rolling surfaces and often will lead to bearing failure. The failure caused by rolling contact fatigue is the most common cause of failure in bearings. Contact fatigue has many characteristics such as pits, spalls, cracks and scoring etc. [204, 205].

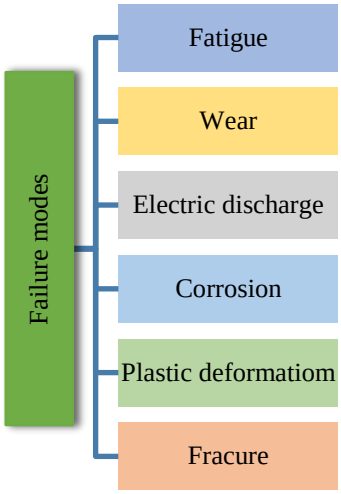

Figure 3-1 Failure Modes

Despite the recent advance made in understanding and analyzing the failure mechanisms of rolling bearings, it is still not reasonably possible to outline a coherent view of the failure modes. Failure modes are in somehow interrelated and in some cases, a developed failure mode can form and considered as another failure mode. Failure modes classification adopted in this research is found consisted with ISO 15243:2017 [206]. The ISO standard states six failure modes with their various sub-modes, these failure modes are Fatigue, Corrosion, Wear, Electrical erosion, Plastic deformation and finally Fracture as it can be seen in Figure 3-1. However, these failure modes do not highlight the causes behind the failure of bearings. According to the SKF study, failure modes can be caused by several reasons as seen in Figure 3-2. The Figure shows almost half of the common causes is lubricant-related. About one-third of causes is fatigue, whilst 1/6 belongs to

other categories which include overloading, improper handling, false design, and fitting or dismounting related issues.

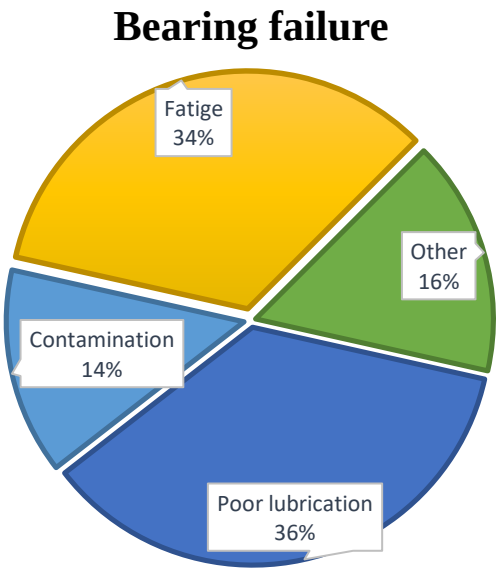

Figure 3-2 Bearing failure reasons [207]

Figure 3-2 may vary according to the application of the bearing as it may encounter several damages due to any of these reasons. The characteristics of each failure mode will be discussed in more details below:

3.4.1 Rolling contact Fatigue

Fatigue is one of the most common factors of the failure in bearings. It can be classified as sub-surfaces fatigue and surface fatigue. Subsurface fatigue results from the heavy load applied in the Hertzian contact zone. This repeated stress can make structural changes to the underneath surface of the raceways. These changes may propagate into surface fatigue after sufficient time and initiating localized defects in several forms, especially in the presence of other factors such as lubricant-related factor. Eventually, these defects may lead to severe contact fatigues. Such fatigue is very likely to increase friction and may have a serious impact on the bearings of the running machinery [208]. Fatigue can affect a bearing in a form of their sub-failure modes as following:

3.4.1.1 Cracks

In running loaded bearing, the sub-surfaces (below the surface) will experience the maximum shear stress. At some point, cracks initially occur at the sliding contact points under maximum stress and eventually propagate to reach the surfaces to form small pits. Several reasons can create cracks on

bearings raceway or roller surfaces, the most common causes are overloading or the excessive force used in mounting or dismantling the bearing,

3.4.1.2 Spalling

Spalling, also called flaking, is well-known damage in bearings and it can be produced as a result of classical contact fatigue, in other words, when the bearing reaches the end of expected operating life. Also, it can occur due to other factors such as excessive load, tight preloaded fitting, expansion due to excessive heat as well as housing or shaft related issues. Spalling can impose damage to rolling bearing before reaching their calculated rating life. Usually, spalling initiate below the surfaces in the area of maximum shear stress, then propagate to reach the component surfaces. When spalling damage reaches a certain level, it can be identified from the produced vibration and noise. This can be an indicator to replace the bearing to prevent the system of breakdown. Spalling can as a result of developed damages such as electric erosion smearing or pitting[209].

3.4.1.3 Smearing

Smearing can occur when the loaded bearing is running under inadequate or unsuitable lubrication, so removed material from one side will be deposited onto the other surface. The transferred metal will get hardened by the high operating temperature [210]. Smearing can also occur as a result of sliding to a radially loaded bearing when the rollers are accelerated due to entering into the load zone.

3.4.2 Corrosion (rust)

Bearing are subjected to get rust during storage, operating if water or corrosive agents enter into bearing assembly, especially in poor lubrication condition where the lubricant fails to protect the bearing component surfaces. Rust can form on raceway, cage or roller surfaces. The risk of corrosion is much higher if the bearing in standstill or poor storage condition

3.4.3 Wear

The wear phenomenon was defined as the removal of operating material surfaces due to the repetitive motion on its surfaces. When two solid surfaces are in repeated contact an interaction either physical or chemical will occurs and will lead to material removal [211]. The typical causes of bearing failure have been investigated for a number of 1400 rolling element bearings. The results showed that almost 30% of rolling bearing failures are caused by wear or lubricant-related matters [212]. In the literature, there are several mechanisms of wear may occur during bearing life. The types of wear are adhesive wear, abrasive wear, corrosive wear and erosive wear [11, 213]. These

mechanisms produce a variety of wear evolution progress with respect to materials and surfaces changes [214].

3.4.3.1 Abrasive wear

Abrasive wear may occur after running-in for a sufficient length of time with respect to the operating conditions, the surfaces of rolling bearing are often subjected to a mild two-body abrasive wear. It is called tow-body abrasive wear when a material is removed from a softer surface by the asperity on a harder surface in the contact area. Another type of abrasive wear is called three-body wear. This type of wear can occur when a foreign particle enters into a bearing and rolling between the opposing contact surfaces. Foreign particles can penetrate into the bearing from worn seals or fine or grinding dust, also during installation or maintenance procedures and roughen the contact surfaces. Similarly, fine particles in lubricants can be released during operation when the metal components contact each other, especially in poor lubrication condition. However, some factors can accelerate the propagation of the wear and such as poor lubrication and leads to premature failure as a result (Olofsson, Andersson, and Björklund 2000). In the case of mild wear, an action can be taken, so the bearing and the housing can be cleaned and readjusted after being lubricated before putting it back to services. Otherwise, bearing subjected to excessive wear should be replaced to avoid breakdown of the system.

3.4.3.2 Adhesive wear

Generally, adhesive wear can occur when two moving surfaces have a molecular attraction between the contacting asperities, and if the internal cohesive strength is less than the adhesive strength of materials. After bearings run for a sufficient time, adhesive wear may take a place. Adhesive wear can also occur due to the breakdown of the protective lubrication film caused by over speed or excessive load, as well as improper lubrication [215]. In these cases, as protective lubricant film will not provide adequate protection, the material will transfer from one side to another side in contact with the friction heat. This type of wear has a linear relationship with pressure applied, so it will increase from mild to severe as the pressure increases. Furthermore, it will be larger if the used materials are identical as the adhesion is more likely to occur between similar atoms and welding them together [213]. In radially loaded bearings, the most critical zone in which slips may occur is that of the rollers entering the loading zone. When the bearing rotates, the rollers run slower in the unloaded area due to friction and then suddenly will be accelerated when they enter to the loading zone. Therefore, the phenomenon of adhesive or smearing wear can be encountered [216].

3.4.3.3 Corrosive Wear

This type of wear is another cause of serious wear defects in a rolling bearing, acid, moisture, and water or other contaminants may penetrate into bearings assembly through worn or damaged seals and cause rust on running surfaces. As the rust particles will interfere with the lubricant, wear will be encountered and the rust pits may develop to flaking and spalling which makes the bearing no longer can be used [51].

3.4.4 Plastic Deformation

If a metal-rolling Hertz contact is overloaded, plastic deformation will take palace, several reasons can cause plastic deformation such as static or shock overload, false handling and mounting, also indentation from foreign particles. Another reason is the unstable thermal balance especially when the high temperature is reached, which will destruct the lubricant, consequently, the friction will increase in the contact points and leads to permanent deformation. It can occur when the applied load is larger than the elastic of the bearing material, this can make a permanent change to the geometry of the bearing.

It can increase the looseness or impose other undesirable consequences in a bearing [217]. Indentation is called *true* brinelling if it caused by false mounting or excessive load. Furthermore, an indentation may also be caused by foreign particles or other contaminants when entering into the bearing assembly, this type is called false brinelling. These foreign particles will roll into the raceways and will cause indentations and harm the bearing. Severe indentations will lead to premature spalling.

3.4.5 Electric Erosion

This type of failure occurs from the passage of current through a bearing used in electrical machinery. The electric erosion is a damage that can cause premature failure to the bearing surfaces and rollers [218]. This potential damage can be caused by either DC current as a result of the electrostatic charge or AC current from inverter-fed drives which may pass through the bearing components [219]. When current proceeds from one surface to another via rollers, it produces very high temperature on the contact points which will result in a series of localized electrical pits. If these electrical pits are severely damaged the surfaces, the bearing should be replaced, however, moderate pits do not usually result in failure [212].

3.4.6 Fracture

The fracture can be a result of excessive force applied to mount or dismount the bearing. When using hummer blows to mount or dismount a bearing, cracks can be initiated and may grow to cause large cracks when the bearing is running. Continue running loaded bearing under this condition, cracks can develop into complete fracture [220]. Moreover, Factors like over preload or excessive shock loads, extreme heat and improper mounting and handling can cause a forced fracture to the bearing components.

The aforementioned failure modes alongside with their potential causes are summarized in Table 3-2. The listed failure modes are matched with their potential causes, some of the failures can be caused by several factors such as fatigue, whilst, a factor can cause several failure modes such as lubrication- related factor.

Table 3-2 Failure modes and their possible causes [221]

Cause / Failure mode symptoms		Reasons												
		Operating condition				**Environmental factor**					**Lubrication**	**Human Error**		
		Overload	**Vibration**	**Excessive speed**	**Shaft or housing issues**	**Inappropriate temperature**	**Dust**	**Water or moisture**	**Electrical leakage**	**Corrosive agents**	**Inadequate /old / contaminated**	**Mounting/fitting**	**Handling /Storage**	**Manufactures error**
Contact fatigue	spalling	*			*	*					*	*	*	*
	Cracks	*	*	*		*	*	*	*		*	*	*	
	Smearing	*		*	*						*	*		*
Corrosion (rust)						*		*		*	*		*	*
Electric erosion									*					*
fracture		*	*	*	*							*	*	*
wear	Abrasive	*	*	*			*				*	*	*	
	Adhesive	*		*		*					*	*		*
	corrosive						*	*		*				
Plastic deformation		*				*						*	*	*

3.5 Summary

In this Chapter, bearing types (section 3.2), bearing components (section 3.3) and bearing failure modes (section 3.4) were discussed. Also, the causes of the failure modes were discovered.

CHAPTER FOUR

VIBRATION ANALYSIS TECHNIQUES FOR CONDITION MONITORING

In this chapter, vibration as the mainstream technique of condition monitoring is explored, the importance of vibration to condition monitoring and fault diagnosis are discussed. Vibration transducer types and placements are overviewed in the course of data acquisition. Vibration responses to bearing defects and their characteristics are exemplified. In addition, vibration data analysis is surveyed. Finally, single row tapered bearing is detailed as the focused bearing of this research.

4.1 Introduction

An object is said to vibrate if it is oscillating with respect to a reference point, while the frequency of the object is the number of motion during a period of one second and its represented in Hertz (Hz) [222]. Several detection techniques are available to monitor machines in the industry; among them vibration technique. Vibration has been for a long time one of the most commonly used techniques for condition monitoring of machinery[48]. Several advantages make vibration-based more likely to be preferable than other monitoring techniques. For instance, it has an immediate reaction to changes in the monitored system, in addition, vibration more likely will highlight directly the defective component. Furthermore, many advanced techniques have been developed and are available to analysis vibration signal [9]. Bearings produce vibration while in operation even if they are geometrically perfect and fault-free due to varying compliance. However, a defective bearing with a localized or distributed fault could generate high vibration levels which include several peaks. These peaks are repeated as a function of rotational speed and the geometrical information of the bearing. Thus, vibration measurement is adopted widely in bearing monitoring. Successful vibration measurement can be carried out by first calibrating and properly mounting the sensors on the monitored machine [223]. Transducers measure the mechanical energy in the form of electric energy by applying a conversion process.

4.2 Vibration-based Condition Monitoring

The vibration-based technology has been successfully applied for fault diagnosis in the monitoring of rotating machinery [224] An effective and reliable CM tools are required to monitor the healthy condition of REBs during their operating life. The vibration signals produced due to defects in them

have been widely studied. Researchers, particularly in the offshore oil and aerospace industries, started during the late 1970s and early 1980s, to employ vibration based fault detection [225]. It has been stated that in industry, vibration is the most commonly applied technique used in CM, and 90% of all machinery failures can precisely be identified by monitoring the vibration signals which can help to make an accurate prediction of future failure [9, 177].

However, it is well known that the CM techniques such as vibration, supply the fault data for bearings in the form of raw signals, which are often contaminated by other system parts noises such as shaft rotational speed, background noise, etc. thus it has to be treated accurately to perform appropriate diagnosis and prognosis of the REBs [20]. Consequently, more robust and reliable signal processing techniques are still needed to accurately extract and select the fault features from the raw signals.

4.3 Vibration Measurements

Measuring Vibration of the rotary machine requires several technical pieces of equipment. In mechanics, the movement can be described by, velocity, displacement or acceleration. Various tools are used to measure vibration, transducers are widely adopted for vibration measurement and they are of three main types; displacement, velocity, and acceleration, and there is a mathematical relationship links them[226].

Displacement refers to the distance that a vibrating object has moved from a reference point, it represented as a sine wave with amplitude and can be quantified by calculating the amplitude of object displacement in a meter $m, mm, \mu m$. Displacement is a preferred measure for low frequency.

Velocity measures the highest speed that a vibrating object reached while oscillating in the direction of motion. It can be quantified by the rate of displacement change per time unit i.e. $m/s, cm/s, mm/s$. Velocity is a preferred measure for the frequency range of 10 Hz-1 KHz.

Finally, acceleration represents the rate of change in object velocity with respect to time. Acceleration is usually quantified by a meter per second squared m/s^2. In this research, accelerometers are considered.

Accelerometers are cost-effective, light in weight with small sizes, also they have high sensitivity with decent wide frequency range, thus, it can be used to measure high frequency[227, 228]. The output signal produced by accelerometers is proportional to the acceleration of the mechanical vibration [71]. The measurement value type can be determined depending on the frequency range

of interest. For bearing monitoring, acceleration is chosen because of the modulation of defects (low) frequencies by the resonance (high) frequency.

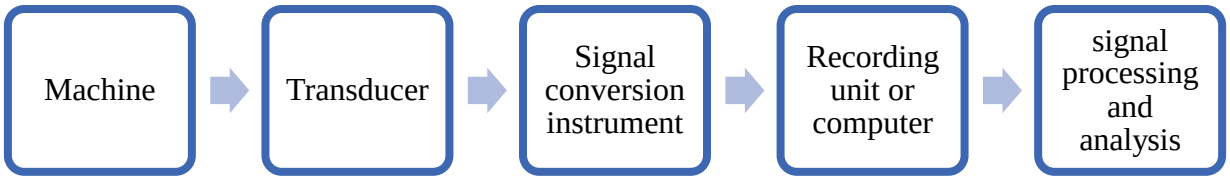

Figure 4-1 Vibration measurement steps

The vibration measurement for condition monitoring consists of several steps, as illustrated in Figure 4-1, it starts with mounting the transducer on the vibrating machine, and signal is amplified and converted into electric current or voltage before it goes to the linked computer or recoding unit, and then the signal can be analyzed by vibration analysis techniques.

In machinery condition monitoring, the widely adopted sensors are piezoelectric accelerometers. These accelerometers use the piezoelectric properties of ceramics and crystals to convert vibration into electric quantity [9].

The accelerometer is used as a basic sensor to measure the vibration on stationary parts of machines, the accelerometer measures absolute vibration by converting the deformation of piezoelectric ceramic plats, which built in the sensor, into electric charge. Generally, larger or greater weight accelerometer (or accelerometer's mass) has better sensitivity and lower frequency range at which it can measure. The generated electric charge cannot travel a long distance, therefore, accelerometers are fitted with pre-amplifiers to transform the electric charge into a voltage and their sensitivity is stated as mV/g. Figure 4-2 shows a typical design of the so-called pressure type accelerometer. The piezoelectric is placed between the upper mass (inertia) and the lower base mass, a spring is used to clamp the components of the sensor. When the monitored machine vibrates the lower mass also will vibrate while the upper mass stays at rest. This will result in a force produced and applied to the piezoelectric plate and electric charge will be created with a quantity proportional to the acceleration of the machine surface.

Several advantages make accelerometers more preferable for machinery condition monitoring, accelerometers are designed in such a way that mounting orientation on machines does not affect their performance, unlike velocity sensors. Another advantage, accelerometers do not contain any

moving components, therefore, no frequent calibration process is needed. This makes them reliable and durable, also mounting and dismounting them on the machine structure is very easy. They can measure a decent wide frequency range from 0.1Hz-30KHz, in addition, they provide a large dynamic frequency range [71]. Therefore, the accelerometer is used for data collection in this research.

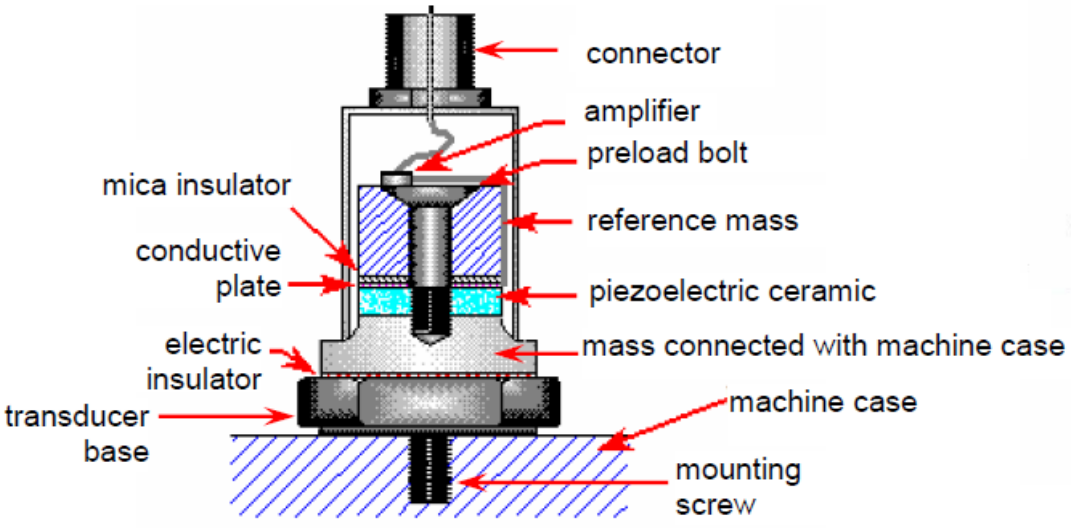

Figure 4-2 schematic diagram for accelerometer [71]

For accurate vibration measurement, an accelerometer attachment into machine structure has to be done in a proper manner. As the accelerometer measures what is vibrates itself, thus, it must be attached directly onto the monitored component of the rotary machine. Incorrect installation of the sensor can ruin the measured signal or it may not fully use the available frequency range of the used sensor. The accelerometer can be mounted in several ways, it can be mounted using a screw, glue (adhesive), magnet, beeswax, and touch needle. The way of mounting depends on the machine application or frequency range of interest. Each way of mounting has its cons and pros. However, screw attachment is considered one of the most reliable methods. The surface area underneath the sensor has to be flat and cleaned before attachment. In addition, there should be a full contact between the sensor base and the machine surface. Adhesive or glue method is another reliable method when the suitable adhesive is applied, also the surface area underneath the sensor is cleaned and roughened. In practice, a base with a screw is attached rigidly with glue onto the machine surface rather than glueing the sensor itself and then the sensor is fitted using a screw. The mounting task must ensure that the surface and the measured vibration are identical [229].

The vibration signal is transmitted, from the sensor to data analyzer through cables for wired sensors. In the industrial environment, factors such as electromagnetic field and static charge are

expected to be present. These factors can strongly affect the measured data if the cables used to link the sensors are not protected and shielded to a high standard quality.

4.4 Vibration Response due to Bearing Defects

For a bearing with well-known localized faults, when the rolling element passes over the defect area, a series of repetitive short transient pulses will be produced that can excite structural resonances. Other vibration sources may include manufacturing errors and background noise. The vibrations induced by either or manufacturing errors or localized defects may contain, in a rather wide band, high energies around the structural resonant frequencies. However, as a result of slippage and sliding caused by the variation of the load angle of the rolling element, the spacing between these impulses is not constant and the impulses are usually described as pseudo cyclostationary [99]. Thus, it has to be treated accurately to perform an appropriate diagnosis and prognosis of the REBs [20].

When a fault occurs in a bearing such as peeling, spalling, galling, subcase fatigue or failure of the bearings as a result of misalignment, surface roughness, shaft slope, or due to high extent of waviness and inclusions, etc., the overall vibration level is affected. The faults should be detected and a decision has to be made on time to avoid, the consequences of fault development which can be a catastrophic collapse of the system, resulting in a financial loss and maybe put the safety of the workers at risk [22]. It is known theoretically that a robust early fault detection method has to be sensitive to impulsive signals and the changes occur in signature [177, 230].

The most popular approach for detecting the defects is vibration monitoring of REBs. Single point defects like cracks, spits, located at the inner race or outer race and at the balls. When one of these defects occurs, the roller element pass over the defect area results in a repeated series of impacts because of the collisions of the metal to metal contact. As these impacts have wideband energy, they will excite the natural frequency resonance of the machine. The repetition rate of these impacts is equal to one of the characteristic fault frequencies previously defined[193].

The vibration signature is characterized by these sharp peaks. The characteristic fault frequencies are normally low frequencies and masked by background noise, therefore, it is very unlikely to diagnose the insipient defect frequency by using conventional spectrum analysis as these impact vibrations distribute their energy over a wide range of frequencies. However, as the impacts have wideband energy, consequently, they will excite some higher frequencies resonance of the system (carriers), the amplitudes of the excited higher resonance frequencies will be modulated by

characteristic fault frequencies. Around those resonance frequencies, the main energy of the fault signal of the REBs can be found [53].

4.5 Characteristic Frequencies of Bearing Faults

Defects in REBs can be mainly categorized into four types: outer race fault, inner race fault, rolling element fault and carriage defects, each of which is characterized by periodic impacts with a characteristic rate. For ideal operations, i.e. without slippage and surface wear, the defect or characteristic frequency can be calculated based on bearing geometric parameters and shaft speed [74] according to Equations (4.1) to (4.4).

Ball Pass Frequency of Outer Race (*BPFO*) —produced from the rolling of all rolling elements across a defect in the outer race:

$$f_{BPO} = \frac{1}{2} z \left(1 - \frac{d}{D} \cos a\right) f_s \tag{4.1}$$

where z stands for the number of rolling elements, d is the diameter of the rolling element, D is the pitch diameter, α is the contact angle, and f_s is the shaft frequency.

Ball Pass Frequency of Inner Race (*BPFI)* —produced from the rolling of all rolling elements across a defect in the inner race:

$$f_{BPI} = \frac{1}{2} z \left(1 + \frac{d}{D} \cos a\right) f_s \tag{4.2}$$

Ball Spin Frequency (*BSF*) — the spin frequency generated from the spins of each rolling element, the Ball Fault Frequency (BFF) is calculated as 2×BSF:

$$f_{BS} = \frac{D}{2d} = \left[1 - \left(\frac{d}{D} \cos a\right)^2\right] f_s \tag{4.3}$$

Fundamental Train Frequency (*FTF*) —is generated due to a defect in the cage:

$$f_c = \frac{1}{2} \left(1 - \frac{d}{D} \cos a\right) f_s \tag{4.4}$$

A localised damage of a rolling element will impact the inner race and the both outer race once a spin and two transients are generated. Thus, $2f_{BS}$ is used as the fault frequency of rolling element. The theoretical fault frequencies of bearings used in this study are calculated for the outer race, rolling elements, carriage and inner race.

4.6 Summary

In this Chapter, vibration technique as the mainstream technique of condition monitoring was discussed. Vibration-based Condition Monitoring and vibration measurement were explored in (section 4.2 and 4.3), whilst, vibration response due to defective bearing was discussed in (section 4.4). In addition, characteristic frequencies of bearing faults were presented in details in (section 4.5).

CHAPTER FIVE

COMPONENTIAL CODING NEURAL NETWORK

This chapter presents the fundamentals of ANN. Both supervised and unsupervised schemes are understood based on a specified neural network: auto-encoder. Finally, as a promising algorithm for effectively processing vibro-acoustic signals, the componential coding neural network (CCNN) is elaborated with its theoretical basis and the implementation procedure.

5.1 Introduction

The application of artitficial intellegance techniques to machine fault detection has been gaining a remarkable success, especially when applied to REBs condition monitoring. One of the successful implemented Artificial intelligence (AI) approaches is the Artificial Neural Network (ANN)[182]. This chapter presents the use of a componential coding neural network for machinery fault detection and potential diagnosis.

5.2 Unsupervised Features Learning and Auto Encoder

Supervised learning has been a successful tool of AI, and has gained a successful implementation to several domains and application such as automatic zip code recognition, speech recognition, driverless cars, and so on. However, in some domains or applications, supervised learning still considered as severely limited as the input features need to be specified manually before it can be fed into the network and force the network to learn a model to match the output. In Such domains for example like computer vision, signal, and audio processing, their features-engineering work is very clever. Also, the design of feature extractors is extremely expensive and the generalization of these extractors can be really poor [231]. Therefore, in such cases, learning better feature representations than the hand-engineered ones can be optimized by adopting unsupervised algorithms[232]. One of the neural network unsupervised learning algorithms is autoencoder that

adopts backpropagation. Autoencoder was introduced in the 1980s by[233]. Autoencoder consists of an input layer, hidden layer, and output layer as shown in Figure 5-1.

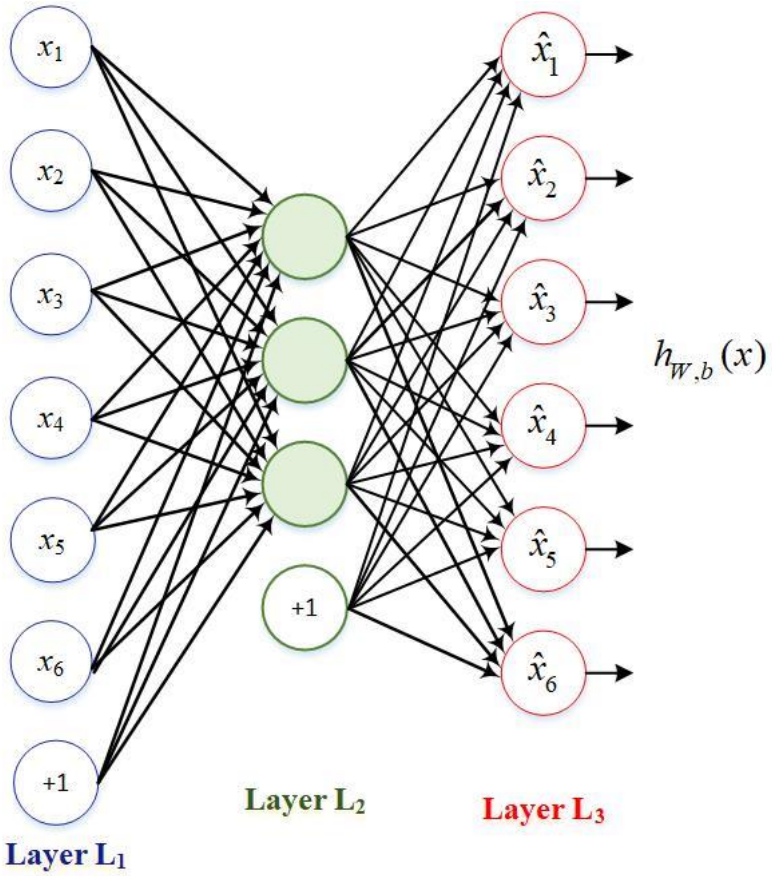

Figure 5-1 Auto encoder

Autoencoder tries to make the target values equal to the input by learning a function $h_{W.b}(x) \approx x$ In other words, it is trying to get an approximation to the identity function, in order to get the output $\hat{x}$ that is very similar to input x by minimizing the reconstruction error as in shown equation (5.1).

$$E_r = \langle |x - \hat{x}|^2 \rangle \tag{5.1}$$

Autoencoder is based on the principle of encoding a module followed by decoding the module, by reducing the number of units in the hidden layer. Thus, If there are correlated features in the inputs, the algorithm can learn a compressed representation of the data and extract structures (patterns) from the data[231]. This method is highly recommended where a low dimensional representation of the data needs to be modelled.

However, several studies advocated that the use of sparse representations for data, in which the dimension of the hidden layer is equal or bigger than the dimension of the input layer [234-236].

They stated that sparse- representations can provide many advantages. For example, adopting high-dimensional representations in image processing raises the possibility that image categories will be easily separable, and it can present a simple analysis of the input data by extracting the hidden structure in the data. Moreover, in a human brain, there are some pieces of evidence that the biological vision uses sparse representations in early visual areas[237, 238]. Therefore, so as to preserve the useful information in the input, considering a complete representation to reconstruct the input data to it can be really reasonable. Based on the principles of ANN and unsupervised learning adopting autoencoder, this research presents the use of optimized autoencoder approach called Componential coding neural network to bearings fault detection and severity measurement.

5.3 Componential coding neural network (CCNN)

Componential Coding Neural Network was originally introduced and investigated for high-dimensional image processing in [57, 187, 239]. CCNN is an unsupervised neural network and based on the idea that the unsupervised network is able to discover related repetitive component substructures, which appears frequently in the different input data. CCNN architecture has been developed based on the idea of autoencoder architecture of one-dimensional network, a basic data pattern is specified as a one-dimensional column vector of length N windowed from a long time sequence. The input layer is equal to the data pattern length. The training algorithm used in CCNN was developed to optimize the accuracy of the reconstruction process when the model during training tries to reconstruct the input signal on average, mean square error (MSE) is minimized based on how much differ model-based reconstructed signal from the input signal[57].

CCNN was developed for image data processing. Limited work has been reported of apply this network to the machinery condition monitoring, Webber et al. [240], have developed a one-dimensional network architecture, for novelty detection and has been applied to both induction motor three-phase and to a transverse flux motor current data, for the induction motor, Voltage imbalance was studied by seeding of a 20 V imbalance and a 40 V imbalance to the power supply system. The study claimed that the network has a good capability in novelty detection and discrimination to detect seeded voltage imbalances, also performed well when compared to conventional techniques [241]. For the transverse flux motor, flux data from the rotor were collected and CCNN has been applied to detect five different fault cases, CCNN performed well and has given good results [241]. However, CCNN has never been explored to the scenarios of detecting and diagnosing faults from other machine components such as bearings. The vibration

signals from faulty bearings are representatives for many different components but with significant differences from that studied in [241].

5.3.1 Novelty Detection and Diagnosis Models

Novelty detection and diagnosis are essential tasks in CM and from the network model and training fundamentals, data reconstruction is an approach that can be followed to generate data models in CM tasks.

5.3.2 Error Based Detection

If the CCNN has been trained with a baseline dataset, the novelty or the changes of a new dataset from the baseline training data set can be found through the reconstruction error of the new data set using the trained neural network model. The averaged error E_r, as it shown in (5.1), represents the difference between the current input data and the data model obtained during training stage. This averaged error then gives a general measurement of the novelty. This method is called the reconstruction error based detection model (REM) [240].

CCNN tries to capture the characteristics of the training signal and then subsequently measure how much is the difference between the characteristics of the unseen signal and the original signal used fir training. In neural network terminology, this can be defined as an anomaly detection task. In [58], the anomaly detection defined as the task of recognizing that the testing data is different in some respect from that data used during the training process. The capability of CCNN in detecting novelty characteristics is because CCNN was developed on the basis that it will not reconstruct all signals with the same accuracy, however, it will only reconstruct accurately signals that share very similar characteristics with signals used for training the network. So the optimal accuracy of the reconstruction will be only for the data that share similar characteristics with the training signals, otherwise, there will be a certain difference in the reconstruction error and it can be determined by measuring how imprecisely each data-set reconstructed [240].

CCNN algorithm will be training using bearing signals recorded under healthy condition. The anomaly will be detected when monitoring data differ from the data that was recorded under healthy operating condition during the training process. Thus, fault detection can be realized as anomaly detection, where anomaly arises as the result of a fault in the monitored system. The severity of the change is inferred from the reconstruction error measured value between the signal and its reconstruction. The first stage of using CCNN is to train CCNN by benchmark a date set under a certain operating condition as a baseline i.e. free of faults, the training is carried out in order to

model a training signal by minimizing the MSE between the data set and its reconstruction, as less as possible. The second stage is the validation stage, in this stage; in order to test how the data model matches the unseen baseline data, the model will be tested using unseen baseline data. This is conducted by acquiring a new dataset, known to be healthy, under the same condition used in recording the training data. Using a new data set, the MSE will be measured to determine how new data set differ from the model. No training or error minimization is involved in this stage. The previous sages are considered as a calibration process to calibrate the CCNN. The final stage is the monitoring stage, in this stage a new data set is recorded and measure how MSE differs from the model obtained in the first stage and then from the difference between the MSE of the validation stage and MSE of monitoring stage, it possible to know how does the monitoring data differ from the model than the validation data does. The anomaly can be detected using the Average Discrimination Index (ADI), which represents the ratio of MSE for the monitoring stage to the MSE for the validation stage minus 1. If the ADI value greater than zero by a fraction of 1, this clearly means that the monitoring data much differ from the model than the validation data set does and a clear indication of a physical change in the system, i.e., a fault is detected. Obviously, a higher amplitude of the error would indicate to a greater physical change in the monitored system from its healthy condition and the amplitude of the ADI is used as an indication of the fault severity. Moreover, the idea behind considering the amplitude of the ADI is that a high change in the amplitude would indicate a larger change in the monitored system from its healthy condition.

A conventional autoencoder is designed to transform input data into a coded form while applying some constraints such as minimizes the dimensionality of the data to a certain predefined level. However, CCNN algorithm was designed to have the sparseness property when coding the inputs to keep the most important components of the dataset, then retains the maximum information about that input dataset when decoding it back.in the reconstruction process. Thus, the reconstructed data will be similar to the original data by minimizing the reconstruction error to a minimum possible value

Compared to similar algorithms such as PCA, Componential coding neural network has some advantages such as. Firstly, when using CCNN, there is no need to pre-process the data, also there is no need to prior-knowledge about the data, as it extracts its own features automatically from dataset directly through an adaptive training algorithm. Secondly, CCNN offers the advantage of time translation invariance, for instance, discovering time-localized features in a signal that have time-invariance (non-stationary) is one of the profits of CCNN, while linear PCA can find only

periodic eigenvectors. The sensitivity of CCNN to time translation invariance arises from the non-linearity neural output function of CCNN. Thus, CCNN is capable to detect faults with greater sensitivity.

Essentially, CM of the bearings using CCNN may be implemented up to different stages:

- Novelty detection can be accomplished through the reconstruction errors made by comparing the original one with model prediction.
- Fault severity can be detected by the characteristics of the amplitudes of the ADI.
- Fault identification can possibly be realised by characterisation of the model parameters such as weight vectors and scale parameters.

5.3.3 Componential Coding neural network Architecture and Theory

CCNN has as a single layer of neurons, every neuron will receive the same input vector x and computes a different output y. The output is derived as a function of the input x and the network parameters. Most of the parameters values are obtained adaptively in the training stage. The neurons encode the input data as a non-linear function of the input data into encoding coordinates $y = (y_1, y_2, \dots y_n)$. The reconstruction process $y \rightarrow \hat{x}$, however, is a linear inverse transformation reproduces a model-based $\hat{x}$ of the input vector x based on the output . The transformation process $y \rightarrow \hat{x}$ is loos because of the constraints imposed on the process, In other words, the output y will not contain all the information available in the input data . Thus, $\hat{x}$ will not be exactly equal to x, however, instead, they will be different by the value of the error $(|x - \hat{x}|^2)$.

The data model is derived from all the neural values the output encode $y(x)$, so the reconstruction error or the inaccuracy of the reconstruction of the input data is a function of the data model as deriving $(\hat{x})$ relies on (y). The weight vectors are the neural parameters which define the data features that the model extracts to describes the data. The difference between any new data and the data model is computed by MSE as $\langle|x - \hat{x}|^2\rangle$ averaged over all that data set [239, 240].

The training algorithm used in CCNN is called gradient descent algorithm. This algorithm trains the model in order to match the training data set as close as possible. It works by searching around the combination of all parameters space until it reaches the lowest level in the error surface and where MSE cannot be minimized anymore. At this level, it can be said that training is reached it's optimum and the data model is said to be best matches the training data set. The parameters present in the function $y(\mathrm{x})$ derived during the training process are named adaptive parameters, as they

have been derived in an adaptive process. These parameters are weight vectors and scale parameters. Other parameters also present in $y(\mathrm{x})$ are non-adaptive as they are not generated by the gradient descent training algorithm. Those parameters include the threshold value, the softness value, the number and the size of weight vectors. Those parameters are manually chosen and adjusted by trial and error to make ADI value larger as possible for a known anomaly data.

The derived features that form the data-model in the training phase are encoded in the adaptive process and will be called weight vector $W = [w_1, w_2, \dots w_j]$, where the index j runs from 1 to J where J represents the total number of components in the weight set.

5.3.3.1 The Components Coding Stage

The output of neurones in any neural network is usually computed as a scalar product between the input data (x) and weight vectors (w) as $\xi = x \times w_j$. However, instead of one scalar product, the output of the neurones in CCNN is being computed N products for each w_j so as to form a periodic correlation function $cr = (x, w_j)$

The periodic correlation function can be computed as a circular convolution operator very efficiently by using FFT, as an order of $N\log(N)$ operations, alternative of the N^2 product operations which is required in calculating the N offset scalar products. The correlation function for data samples originates from a sensor is computed as shown in

$$cr(x, w_j) = F^{-1}\big(F(x) \times F(w_j)^*\big) \tag{5.2}$$

Where x is the input data, the function $F()$ represents one dimensional FFT transformation, $F^{-1}()$ is the inverse of Fourier Transform, $(\,)^*$ stands for complex conjugate and $\times$ represents the point-wise multiplication to get a vector of coordinate products.

5.3.3.2 Threshold Activation Function

In this step, a non-linear threshold activation function $r(\beta)$ is applied to the output of the circular correlation between the input pattern x and the weight vector w_j so $\beta = cr(x, w_j)$. This will generate J output code vectors $y_j(x)$. The non-linearity is adopted to add the sparseness property to the output code, so the output of a few neurons should contribute more than the rest. If the value of the threshold is set properly, only a few neurons will produce large output $x, w > \delta$ as all neurons output will be thresholded. Hence, the neurons will fire only if their weight vectors match well the input data pattern.

Neuron function used as depicted in (5.3) where the symbol r denotes the network neuron function is a non-linear function [239]:

$$r(\beta) = \frac{1}{2}\left\{s\ log\left[1 + exp\left(\frac{\beta - \delta}{s}\right)\right]\right\}^2 \tag{5.3}$$

Where β is the correlation between the inputs and the weight vector, which is derived by .

$$\beta = cr(x, w_j) \tag{5.4}$$

The threshold δ and softness s are the parameters of the network neuron and hence, the initial values of these parameters, as well as the number of weight vectors, need to be determined at the early stage of network design.

The code $y_j(x)$ is yielded from the neuron function applied to cr correlation function as in

$$y_j(x) = r(cr(x, w_j)) \tag{5.5}$$

The sets of the output $y_j(x)$ constitute the output code of the CCNN for the input x [239].

5.3.3.3 Model-Based Data Reconstruction

The output code derived using Equation (5.5) is used to perform the reconstruction process ($\hat{x}$) of the input (x), by firstly convolve each of $y_j(x)$ with the corresponding w_j and then the resulting convolution J is combined by the weighted summation [57, 240].

$$\hat{x} = \sum_{j=1}^{J} a_j\ cv\left(w_j, y_j(x)\right) \tag{5.6}$$

Where a_j represents the adaptive parameters called weight scales, the values of weight scales are determined by optimization procedure and will be illustrated later.

The convolution function can be efficiently computed by using FFT as the w_j is only one-dimensional time signals belongs to the input. This convolution function is computed as

$$cv\left(w_j, y_j(x)\right) = F^{-1}\left(F(w_j) \times F\left(y_j(x)\right)\right) \tag{5.7}$$

Where cv is a circular convolution operator.

The network scale parameter a_j in Eq (5.6) is a one-dimensional array$a = \{a_1, \dots, a_J\}$. Its elements a_j are scalars, each corresponding to a weight vector w_j. Where J denotes the number of weight vectors. The scale parameter is found using equation.

$$a_j \equiv \sum_{j'}^{J} (M^{-1})_{jj'} \langle x \cdot cv(w_{j'}, y_{j'}(x)) \rangle_{\{x \in w\}} \tag{5.8}$$

Where the inverse M^{-1} is a square matrix of size $J \times J$ and its elements are calculated by:

$$m_{jj'} \equiv \langle cv\ (w_j, y_j(x)) \cdot cv(w_{j'}, y_{j'}(x)) \rangle_{\{x \in W\}} \tag{5.9}$$

In both of the above equations, the scalar product involves summation over all the time samples and $cv\ (w_j, y_j(x))$ is N dimensional vectors.

The random formation of training data patterns is one of the important steps for the realisation of the translation invariant mechanism. Moreover, with the definitions of correlation and convolution functions, the reconstruction process is invariant with respect to translation of any weight vector by any arbitrary time displacement. On the other hand, with respect to constant weight vectors, any translation (wrap-around) of input *x* will result only in the translation of the reconstruction $\hat{x}$ accordingly but its shape will not be altered, thus, the reconstruction accuracy will not be affected by the absolute angular position of the revolving component[240].

5.3.3.4 Obtaining matched weight vectors by minimizing the MSE

The adaptive training process involves matching the weight vectors to the training signals in order to minimize the reconstruction errors on averages over the training dataset T [239]. The MSE of the reconstruction is computed as in.

$$E^T = \langle |x - \hat{x}\ |^2 \rangle_{\{x \in T\}} \tag{5.10}$$

The operator $\langle \cdot\cdot \rangle_{\{x \in T\}}$ denotes averaging process overall x in training dataset T.

The optimization of the reconstruction is realised by minimizing E^T with respect to $\hat{x}$ implicit adaptive parameters w_j of , by the iterative gradient descent algorithm on E^T in the vector space of the w_j. The gradient descent algorithm updates the network weights by implementing (5.11) [240]:

$$\frac{dw_j}{dt} = \lambda_j(t)\,\langle a_j\, cr\left(x - \hat{x}, y_j(x)\right)\rangle_{\{X\}} + \lambda_j\,\langle\, cr\left(x, y'_j(x) \times g_j(x)\right)\rangle_{\{X\}} \quad (5.11)$$

Where $g_j(x)$ is obtained [240] as:

$$g_j(x) = cr\left((x - \hat{x}), w_j\right) \quad (5.12)$$

In the training process, the best match of weight vectors w_j to training data features will be achieved when the scale parameters a_j reach the values that reduce the reconstruction error E^T. Therefore, in order to accomplish the best model of the data, weight vectors and scale parameters will be optimized jointly together. The other notations are defined below.

The network size and parameters are as following:

- The network weights are a two-dimensional array as $W = [w_1, w_1, .. w_j$. The size of each w_j is equal to the size of the data pattern x.
- The scalar parameter is a one-dimensional array as $a = \{a_1, .., a_j\}$ and for each weight vector w_j there is a corresponding scalar parameter a_j.

The purpose of training the CCNN is to optimise the weights and the scale parameter on a statistical basis. These optimal values are identified through the adjustment of the network and network training parameters given below.

5.3.3.5 Network Parameters:

1) The length of each batch of the input x or the length of the weight vector w_j, as the input layer equal to the hidden layer and equal to the output layer.
2) The number of weight vectors J.
3) The non-linear function parameters values: theta ϑ and sigma s.
4) To define the value of softness more accurately, [239] proposed to use the following rule: $s = 0.05 * \vartheta$ this rule is more practical since it provides a higher potential for the neuron to move away from the initial configuration at both the lower and the higher threshold value.
5) The network learning rate λ is selected to be as small as possible in order to obtain a smooth minimisation process.

5.3.4 Componential Coding neural network Implementation Stages

The application of CCNN in the context of condition monitoring involves, capturing the features of the training dataset and then measure how much does differ any new dataset from the original training dataset. In this study, MATLAB is used to implement CCNN. The application of componential coding neural network is consisting of three stages namely training, validation and monitoring as discussed below:

5.3.4.1 Training Stage

The first stage of applying CCNN is to first benchmark a baseline data-set (free of faults) under a known operating condition. Then train the CCNN algorithm to model the baseline dataset as healthy operating condition, by minimizing the MSE averaged over the baseline training dataset. A small MSE value means that the data model is precisely matched to the training dataset. The training procedure by minimizing the error should result in a data model which is matched to the dataset used in training as accurate as possible.

Training Algorithm

The process of training CCNN involves the following basic steps:

- In this step, generate initial weight vectors with random values and scale into unit length.
- The optimal scale parameters are calculated for the weight vectors, a weight-vector gradient for each weight is obtained, and these gradient vectors are scaled to a constant value of learning rate.
- The weight vectors are updated by the scaled gradient vectors and scale the new weight vectors to unit length.

The training stage starts by:

- Segment the input data samples into a sub-pattern array (segments) each segment length L as the same length of the weight vector to form x of size $N \times L$ from the data set, which is a long time sequence. The patterns number L is larger than weight vectors number C so duplication of the weight vector is carried out to be $C \times L$, thus, the size of the arrays is identical. In addition, each of the patterns is truncated from the long sequence with random entries. These entries also vary from iteration to iteration. This random formation of training data patterns is one of the important steps for the realisation of the translation invariant mechanism of the CCNN.

- Calculate the correlation of the L training patterns one by one to each of the weight vectors in W.
- Implement the non-linear neuron function element by element to the correlation results. These produce C numbers of array R of size $N \times L$. Each R belongs to a weight vector.
- Run the convolution of the C weight vectors over the corresponding matrix R to produce the corresponding convolution array C_n.
- Calculate the scale factors averaging over all L values as in (5.8) and (5.9).
- Calculate the raw gradient vectors, normalise it to unit length and multiply the unit-length vector by the specified learning rate so that the length of the vector is re-scaled into the value of the learning rate. Upgrade the old weight vector by adding the re-scaled gradient vector and then normalise the new weight vector so that its norm equals 1.

5.3.4.2 Validation Stage

Validation is conducted to measure how accurately does the trained model matches another new healthy dataset. This is carried out by feeding the CCNN with a new (unseen) healthy dataset, from the same operating condition as training data set operating condition, and then measuring MSE to determine how close is the new healthy dataset to the data model derived during the training stage. This validation will result to obtain the MSE for the healthy dataset without performing training or minimizing. Validation dataset must be unseen dataset and not the same data set used during the training process. Training and validation process constitute a calibration process and has to be done with health (free of faults) condition.

5.3.4.3 Monitoring Stage

This stage is the final process of applying CCNN and is conducted by feeding the CCNN with new datasets from the same configuration under the same conditions during the training, and then measuring, using Mean Squared Error MSE, how the new dataset differs from the data model derived during training. By comparing the MSE for both baseline and new data set, it can be known how much each new monitoring dataset differs from the training model than the difference of the healthy calibration (unseen) dataset. The novelties can be detected using the Average Discrimination Index (ADI) and it is calculated as shown in (5.13).

$$ADI = \frac{R_m - R_u}{R_u} \tag{5.13}$$

where R_m is Mean Squared Error (MSE) of the new monitored dataset. R_u is the MSE of the unseen healthy dataset [240].

If the ADI of R_m is larger than R_u by a fraction of 0.1, this means the model has a poor match to a new dataset than it does to the baseline dataset used in the validation. This indicates that a physical change (fault) has occurred in the monitored bearing.

5.4 Summary

In this chapter, unsupervised features learning and auto encoder was explored in (section 5.2), componential coding neural network was discussed in (section 5.3) with the theoretical background, architecture and the implementation stages.

CHAPTER SIX

WAVELET THRESHOLD FOR DENOISING, A PROPOSED METHOD

This chapter presents the fundamentals of Wavelet Transform (WT) for enhancing the diagnostic information that usually is contaminated by noise. The mechanisms and corresponding algorithms of WT based de-noising are examined firstly. Then a new method is proposed to achieve effective noise reduction for the signals dealt with in this study.

6.1 Introduction

To enhance signals of bearing features, noise suppression is a critical step. Shrinkage denoising in the transformation domain is the process of removing the noise or unwanted components from a number of wavelet coefficients. The use of the shrinkage method has proven its ability as an effective way to suppress the noise of noisy signals with low computational complexity [242]. The selection of threshold method and estimation of the value are vital steps to the success of the wavelet regression, thus, it has been receiving an intense research effort. Nason [243] reviewed various methods of thresholding methods including their selection and estimation. The thresholding rule decides the components of the coefficient that needs to be retained or eliminated. Thresholding of the detail coefficients can be applied into two categories, as universal for all decomposition levels or level-dependent approach [244]. Wavelet thresholding method introduced firstly by Donoho et al. [123] and two thresholding were presented, namely as **hard** and **soft** threshold. In the case of hard thresholding, as seen in (6.5), all the coefficient with absolute values $\leq \mu$ will be put to zeros, while the soft thresholding rule additionally subtracts the set threshold value from the all the coefficient towards zero, as depicted in (6.6) where $T(\cdot)$ is a thresholding function[245].

6.2 Data Denoising by Thresholding-A Review

More analytically, assume there is a given N samples of a noisy signal $x(t)$, where $s(t)$ is the unknown function of interest and $n(t)$ is the Gaussian white noise with zero mean and noise level $\sigma > 0$ [246]:

$$x(t) = s(t) + \sigma n(t) \tag{6.1}$$

where $t = 1,2, \dots, N$ then the steps of wavelet denoising can be represented as [80]:

- Choose the $W(\cdot)$ type and the decomposition level L with $1 \leq L \leq log_2(N)$.

- Apply the wavelet transform as

$$w_{li} = W(x) \tag{6.2}$$

where $W(\cdot)$ wavelet transform, w_{li} noisy coefficients, lth is a level of $l = (1,..,L)$ and ith is a detail component location of $i = (1,..,2^L)$.

- Estimate the noise level σ based on the thresholding technique.

- Apply the threshold as

$$\widehat{w}_{li} = T(w_{li}, \mu) \tag{6.3}$$

Where $\widehat{w}_{li}$ denote the denoised wavelet coefficients, T denote nonlinear threshold function (soft, hard) and μ stands for estimated threshold value.

- Reconstruct the denoised signal $\hat{s}$ as

$$\hat{s} = W^{-1}(\widehat{w}_{li}) \tag{6.4}$$

Where $W^{-1}(\cdot)$ denote reconstruction process (invers wavelet transform)
The reconstructed signal $\hat{s}$ quality can be influenced by the type of $W(\cdot)$ or by the shrinkage process by either thresholding function $T(\cdot)$ or estimated threshold value μ. Several thresholding methods have been proposed to shrink the noise of experimental signals. As shown in (6.5) and (6.6) respectively, hard and soft thresholding among others is the most widely used in the wavelet transformation domain for removing the noise from noisy data, [247].

$$T_{hard}{}^{(w_{li})} = \begin{cases} w_{li}, & |w_{li}| \geq \mu \\ 0, & |w_{li}| < \mu \end{cases} \tag{6.5}$$

$$T_{soft}{}^{(w_{li})} = \begin{cases} sign(w_{li})\,(|w_{li}| - \mu), & |w_{li}| \geq \mu \\ 0, & |w_{li}| < \mu \end{cases} \tag{6.6}$$

For visual clarity, Figure 6-1 shows both functions which scaled to the interval [−1, 1]. The dotted lines indicate the original single. X-axis represented detail wavelet coefficients w_{li} obtained in (6.2), and Y-axis for the corresponding thresholding function as it is obtained in (6.3). Figure 6-1 (a) shows $T_{hard}{}^{(w_{li})}$ function and it clarifies how all observations $\geq \mu$ are shrunk to zeros, whilst,

only large observations have remained unchanged. Figure 6-1 (b) shows how $T_{soft}{}^{(w_{ii})}$ can shrink all the observations by the estimated value of μ.

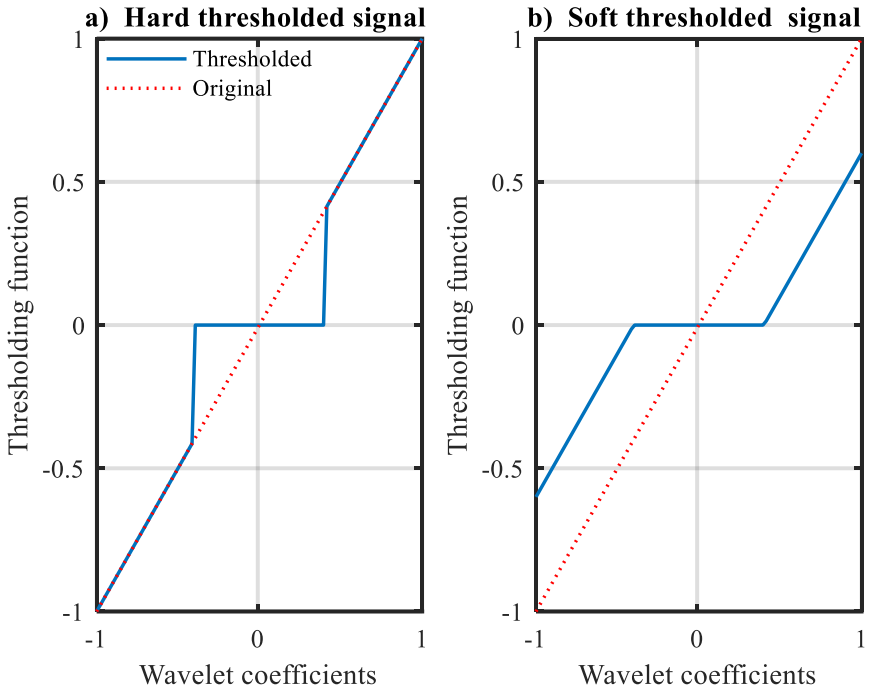

Figure 6-1 Hard and Soft Thresholding Function

It has been reported that, in wavelet transformation domain, the energy of signal tends to be concentrated into a relatively few numbers of large coefficients whilst, the noise will be spread at a large number of the coefficients with relatively low energy [80, 116, 248, 249]. This increases the options to illuminate the noise while retaining the important information of the signal as much as possible. As a result, a signal can be enhanced by removing components smaller than an estimated threshold [250]. Based on this principle, wavelet coefficients thresholding for data denoising has been an extensive research domain since the first pioneering work by Donoho and Johnstone [251]. Since then, several thresholding techniques and estimators have been developed, this research focuses on some of the most widely adopted techniques, such as VisuShrink, SureShrink, HeurSure, NeighBlock, BayesShrink and Minimaxi thresholding technique.

6.2.1 VisuShrink

Also considered as **universal** threshold or **sqtwolog** threshold, it was proposed by Donoho and Johnstone [124] with regardless of the thresholding function type and represented by

$$\mu = \sigma\sqrt{2\log(N)} \tag{6.7}$$

Where σ is the noise variance obtained by median absolute deviation (MAD) of the coefficients as:

$$\sigma_l = \left[\frac{MAD(w_l)}{0.06745}\right] = \left[\frac{median(|w_l|)}{0.06745}\right] \quad (6.8)$$

and (N) is the length of the observed signal while (w_l) is the empirical wavelet coefficient to the level (l) [252]. The value 0.6745 was selected for calibration with the Gaussian distribution[253]. Due to the fact that STD of the signal may not be a reliable estimator, later on, Donoho and Johnstone suggested estimating the σ in the wavelet domain, instead of original signal, at the finest levels as they hold most of the noise. However, using VisuShrink with soft thresholding, it was reported that VisuShrink is more likely to over smooth the signal and result in missing some genuine features of a signal [254]. Whilst, genuine features (peaks) more likely to be preserved when it is applied in conjunction with hard thresholding, sometimes at the cost of less smooth fits [255]. Another concern was reported because of its dependence on data size, it can be unwarrantedly large and lead to unknown signal distortion [256].

6.2.2 SureShrink

Also called **Rigrsure**, this threshold is generated under a risk rule by minimizing Stein's Unbiased Risk Estimate (SURE). For each detail level, a sub-band threshold is calculated based on SURE rule [257]. This technique is a subband adaptive, level dependant. It is represented by

$$\mu_i = \sigma_i\sqrt{w_m} \quad (6.9)$$

With (w_m) is the mth coefficient wavelet square at the lowest risk which selected from wavelet coefficient squares vector, sorted in ascending way [m1,m2,m3,…mn], and (σ_i) is the level-dependant standard deviation of a noisy signal [258]. Although it was claimed that SureShrink results in relatively large noise when the SNR is too small, it is still one of the most widely adopted thresholds [259].

6.2.3 HeurSure

HeurSure technique was developed by Donoho, this method is an automatic procedure and hybrid approach combines the VisuShrink with SureShrink. It can be applied using one of two scenarios automatically, first, it uses the SureShrink technique shown in (6.9), however, if a test of signal

coefficients at a level l proved that the signal is deemed too small, as a second scenario, a fixed threshold based on $\mu_2 = \sigma\sqrt{2\log(N)}$ will be applied instead [255, 260].

Both VisuShrink and SureShrink can be considered as functions of the data length and the variance of the noise [261]. However, as SureShrink is a level-dependent technique, the local neighbourhood of each coefficient is not taken into consideration. Thus, lots of the components from the wavelet coefficients could be removed [161].

6.2.4 NeighBlock

As noted above, in SureShrink, the local neighbourhood of each coefficient is not taken into account, which may lead to losing many important components from the coefficients. On the contrary, NeighBlock threshold, proposed by Cai et al. [262], as a block thresholding technique, increases the accuracy of estimation by including the influence of neighbouring coefficients. This is performed by grouping the coefficients with similar properties, assuming they contain important parts of the signal, and then calculate the threshold based on the grouped block instead of level-dependant technique. it was claimed that when applying NeighBlock technique, the estimation accuracy is enhanced compared to SureShrink [263]. More analytically:

After transforming the signal into the wavelet domain, wavelet coefficients are grouped at each level l into disjoint blocks b_k^l of length $g_1 = [(logN)/2]$, and then extending each block b_k^l by points of $g_2 = max[(1, g_1/2)]$ in both directions, in this way, overlapping bigger blocks B_k^l of length $g = g_1 + 2g_2$ will be formed [264]. For each level, each block b_k^l will be indexed k, the coefficients will be estimated by a shrinkage rule as $\widehat{w}_{li} = T\left(w_{li}, \mu_k^l\right)$ for all $(l, k) \in (b_k^l)$ where the μ_k^l is obtained from the larger block b_k^l as

$$b_k^l = \max(0, (\frac{1 - \lambda g \sigma^2}{{\delta^2}_{l,k}}) \tag{6.10}$$

Where λ is a constant of 4.50524, and the variance of the extended block is represented by σ^2 and ${\delta^2}_{l,k}$ is the sum of the squared coefficients in each sliding window denoted as

$${\delta^2}_{l,k} = \sum_{(l,i) \in B_k^l} {w^2}_{l,k} \tag{6.11}$$

In this way, B_k^l can be seen as a sliding window translates each time by the amount of $(1.5 * g_1)$ of the coefficients in the centre of the window for each window will be estimated. Finally, reconstruct the thresholded coefficients to restore the signal [262].

6.2.5 BayesShrink

The BayesShrink technique implements Bayesian mathematical framework to derive the optimal level-dependant threshold for soft threshold function [256] and it can be calculated for a subband B as:

$$T_B = \frac{\sigma^2}{\sigma x} \tag{6.12}$$

With σ^2 represents noise variance estimation and σ_x is the signal variance estimation for the considered subband and can be calculated as:

$$\sigma x = [\max(\sigma_w^2 - \sigma^2, 0)]^{1/2} \tag{6.13}$$

Where σ_w^2 can be derived as:

$$\sigma_w^2 = \frac{1}{N^2} \sum_{l,i=1}^{N} {w^2}_{li} \tag{6.14}$$

Where N is the wavelet coefficients number w_{li} on the considered subband whilst σ can be calculated as:

$$\sigma = \frac{MAD(w_{li})}{0.06745} \tag{6.15}$$

6.2.6 Minimaxi

In this technique, based on the statistical minimax principle for estimator designing, a constant threshold value is chosen to produce minimax performance for MSE against an ideal procedure. The minimax estimator is the procedure of realising the minimum of the maximum MSE over specific function classes, see Donoho and Johnstone [265].

More analytically, To recover the unknown function $s(\cdot)$ from $x(t) = s(t) + \sigma n(t)$, Minimaxi technique measures the performance of the estimation of $\hat{s} = (\hat{s}(t))$ from $s = (s(t))$, with regard to a quadratic loss at the sample points by minimizing the risk as small as possible as:

$$r(\hat{s}, s) = n^{-1}\, E\|\hat{s} - s\|_{2,n,}^2 \tag{6.16}$$

$$\mu_n^* = \underset{\mu}{inf}\ \underset{\theta}{sub}\ \left\{\frac{r_\mu(\theta)}{n^{-1} + (\theta^2\ \mu 1)}\right\} \tag{6.17}$$

Where μ_n^* is the minimax risk bound and will be the value of μ.

6.3 Limitation of the Current Thresholding Methods

Thresholding methods suffer from some limitations, the choice of threshold type usually based on an ad-hoc manner. Another challenge is that the signal and noise distribution may not be well matched at all levels [261]. Moreover, the conventional hard thresholding sometimes becomes highly sensitive to little changes in the experimental signals and exhibits some discontinuities [266]. In the case of soft thresholding, all the coefficients are equally shrunk through a threshold value, it may lead to the deviation in the reconstruction process[162, 267]. Furthermore, although thresholding is considered as an effective and simple task, estimating its effective value is not an easy task. A trial and error approach to find the optimal value (μ) can be a tedious and challenging task. Hence, noise level (σ) estimation has received intensive research to simplify the selection of the optimal threshold value of (μ) and effectively eliminate noise from signal coefficients. Despite the much recent effort that has sought to effectively estimate and shrink the noise in the wavelet transformation domain, however, limited work was reported that specifically targeted experimental signals. Generally, successful thresholding method in signal denoising highly relying on the accurate estimation of the noise level [268], thus noise level must be estimated correctly to obtain good performance denoising.

6.4 Proposed Method

Based on these considerations and challenges, in this research, an adaptive data-driven based on hard thresholding compound with envelope analysis is proposed and tested against the well-known thresholding techniques Sqtwolog, Rigrsure, HeurSure and Minimaxi. Hard thresholding is adopted as it was claimed that, it can lead to a better reproduction of peak heights and discontinuities [94]. Moreover, the developed technique is level-dependant and takes into account the coefficients with high energy in the process of estimating the noise value. The developed technique can adaptively find the optimal threshold value effectively by obtaining the optimal ratio of the first three harmonics of the demodulated signal. Also, the developed method outperforms the conventional wavelet thresholding methods. The proposed technique estimates the noise for each level in the wavelet domain and iteratively find the optimum threshold as described in the following algorithm:

6.4.1 Denoising Algorithm

The denoising algorithm is used to denoise the signal by estimating the thresholding value adaptively and level-dependant, the algorithm starts with decomposing the signal into L levels and then shrinks the noise based on the estimated threshold value. The iterative process continue until the highest signal to noise ratio is reached using HSR as a measure of signal improvement. The steps of the denoised algorithm are as below:

i. Select the input signal
ii. Decompose the signal into L levels w_{li} =$W(x)$, where $L < log2(N)$
iii. Calculate the coefficients energy using STD
iv. Coefficients with similar high energy are considered to be selected and put the rest of the coefficients to zeros
v. For each level l sort in descending order the data points within the range of (1:2000), this range covers one rotation period and it can be calculated as F_s/f_s where F_s sampling rate, f_s shaft frequency (50000/25=2000).
vi. Estimate the initial threshold value by calculating the mean value for the selected data range for each lth level as

$$m = (\sum_{j=1}^{n} w_l(j))/n \tag{6.18}$$

vii. Apply the hard threshold level-dependant using the estimated initial value $\hat{w}_{li} = T(w_{li}, \mu)$ with μ=zero for the first step as the output will be compared with the output of the next steps.
viii. Reconstruct the decomposed signal $\hat{s} = W^{-1}(\hat{w}_{li})$.
ix. Calculate the envelope spectrum of the reconstructed signal $\hat{S} = FFT(env(\hat{s}))$
x. calculate b as the sum of the first three harmonics for the thresholded signal and obtained as:

$$b = \sum_{j=1}^{3} \hat{S}(j \cdot f_r) \tag{6.19}$$

xi. where f_r is the fault frequency j is the harmonic number
xii. calculate the sum of data points within the range of band B (from 100Hz to 600Hz) as

$$a = \sum_{j=\min(B)}^{\max(B)} \hat{S}(j) \tag{6.20}$$

xiii. Find the shaft harmonics v within the selected band range B of the demodulated signal as

$$v = \sum_{j=1}^{\max(B)/f_s} \hat{S}(j \cdot f_s) \tag{6.21}$$

xiv. calculate the ratio r of the first three harmonics to the rest of the demodulated signal as

$$r = \frac{b}{(a - b - v)} \tag{6.22}$$

xv. To reach an optimal threshold using automatic adjustment , iteratively increase the threshold μ value in each step by a fixed small value and repeat all the steps until the reconstructed signal becomes too spars i.e. when the length of the data points=zero in reconstructed signal reaches 95% compared to the original signal then stop and select the best threshold value with the best SNR by selecting the highest harmonics ratio reached from all the steps as:

$$hsr = \left(\left(\frac{\max(r)}{a_1}\right) - 1\right) * 100 \tag{6.23}$$

Where a_1 is ratio of the un-thresholded reconstructed signal within the band B,

xvi. end

Using the proposed threshold, several types of WT will be investigated to enhance the experimental vibration signals. Moreover, this research shows that considering an accurate statistical description of both the signal and noise components can lead to an effective denoising technique that can reduce the noise level as well as preserve the required features.

The flowchart shown below Figure 6-2 describes the implementation steps of the algorithm starts with decomposing the input signal and ends with calculating the best harmonic to signal ratio achieved.

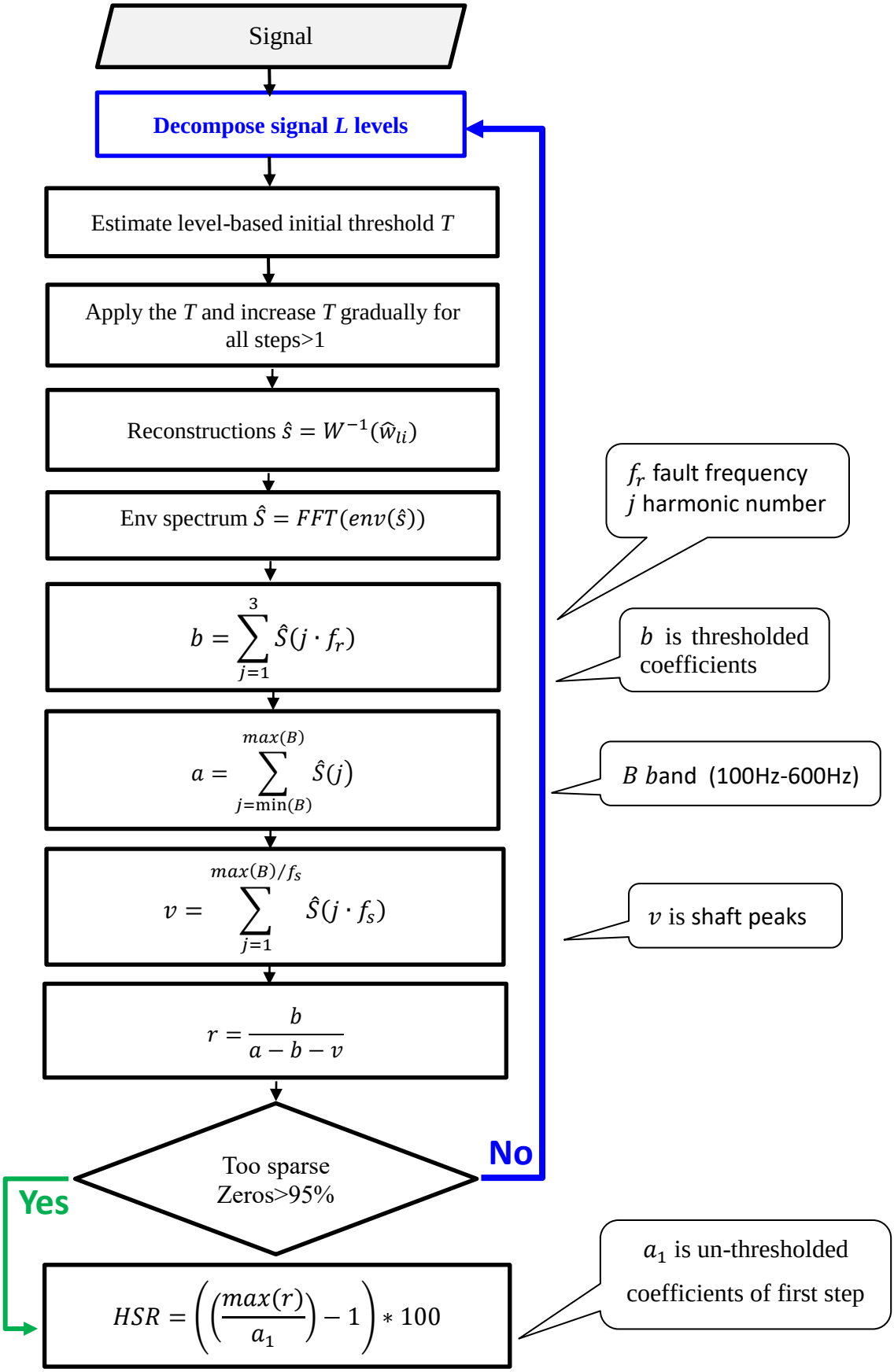

Figure 6-2 Implementation flowchart

6.5 Summary

In this Chapter, the fundamentals of Wavelet Transform for enhancing the diagnostic information were discussed. The existing thresholding methods with their limitations were reviewed in (section 6.2 and section 6.3) respectively. Also, the proposed thresholding algorithm was presented in (section 6.4).

CHAPTER SEVEN

EXPERIMENTAL FACILITIES AND PROCEDURES

In this chapter, the test system is described in details. It starts with the design of the test rig. In addition, the bearings used to carry out the experiment as well as their geometrical information is described. Furthermore, this chapter illustrates the mounting settings of bearing and the clearance adjustment mechanism, also describes all the instruments required to carry out the vibration measurement such as, the data acquisition system, transducers used to collect the vibration data, temperature coupling and the encoder to monitor the motor speed.

7.1 Introduction

One of the important elements of CM studies is the introducing of the faults into a practical system that simulates the real industrial world in a controlled manner which will provide real data. Several reasons behind of idea of using an existing test rig that was developed and used in CEPE lab at the University of Huddersfield. The rig offers isolation from other sources of vibration and offers also the ease of accessing the instruments and the other experimental facilities due to the optimal construction of the test rig.

In order to investigate the proposed methods for experimental signal analysis, an experiment was carried out using tapered roller bearings. The test rig is carried out in a simple way with relevance to the industrial world in which faults can be introduced in a controlled way and the vibration can be accurately measured. The choice of a TRB is due to the widespread of using this component in industries.

The test rig was developed on the basis of an adjustable clearance mechanism. The test rig consists of a motor, coupling, two bearing housing, shaft, data acquisition system, two piezoelectric accelerometers (CA-YD-104T) which have been mounted on the test bearing housing, one was horizontally and the other was mounted vertically. A clearance adjustable mechanism was developed to adjust the clearance within the typical maximum life range provided by the manufacturer. A slip metric gauge box set, type Matrix Pitter 8075 C, was used to precisely measure the clearance. The tapered bearing is chosen in this study because it is well-known as a noise bearing and they come apart, hence, the clearance is adjustable while mounting. The test rig is designed in such a way to be easy to be assembled and dismounted. Thus, the bearings used with

the different conditions can easily be fitted in the test rig. The experiment facilities and procedure will be explored in details in the following:

7.2 Test Rig Development

The experimental work was conducted using two main parts, the first part is the bearing rig, as shown in Figure 7-1 and Figure 7-2, which was developed to facilitate the running of bearing in a similar way as the real world application. In addition, it was developed to facilitate the adjusting procedure of the internal clearance using the precise tools to measure the internal clearance. The second part is the instruments required to collect the vibration data, the shaft speed and the temperature. This simple structure was adopted to avoid possible noise influences of additional components such as radial load devices.

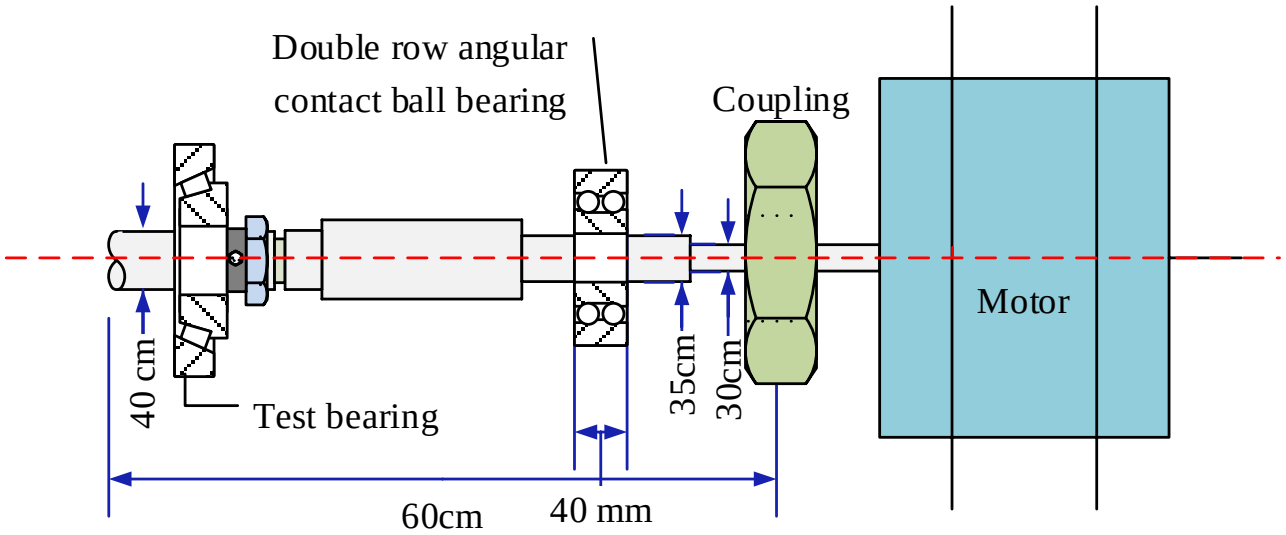

Figure 7-1. Schematic diagram of the test rig

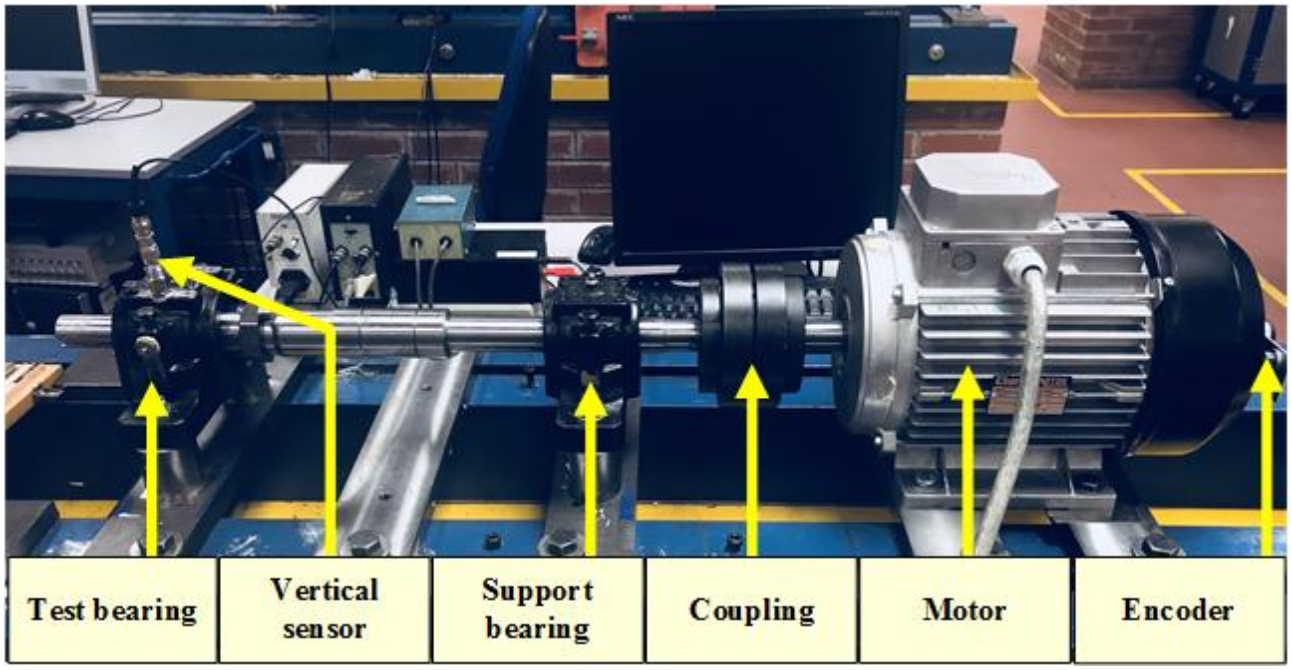

Figure 7-2. Test Rig

7.2.1 Motor

An induction motor is used as a drive power, as seen in Figure 7-2, it is a three-phase and 4kw. A spider flexible coupling is used to couple the motor to the shaft. The motor speed is controlled by a controller type Siemens at 1500 rpm. The specifications of the motor are illustrated in the Table 7-1.

Table 7-1 technical specifications of Clarke induction motor [269]

No	Technical Features	Value
1	Number of phases	3
2	Number of Poles	4
3	Voltage	230/400 V
4	current	5.9-9.2 A
6	Motor type	112M/4
7	Shaft Speed	1500 RPM
8	hp	5.5

7.2.2 Data Acquisition System (DAQ)

Data acquisition is a process, Figure 7-3, in which a physical phenomenon is transformed into a digital data format using DAQ system and its accessories. A DAQ system is used to sample, acquire, convert and store data on a computer. The data acquisition device can collect different types of data such as vibration, current temperature, voltage, pressure, etc. DAQ system converts

the input signal from analog format into digital format using analog-to-digital converter (ADC) and then send the acquired data to a computer for storage and analysis task.

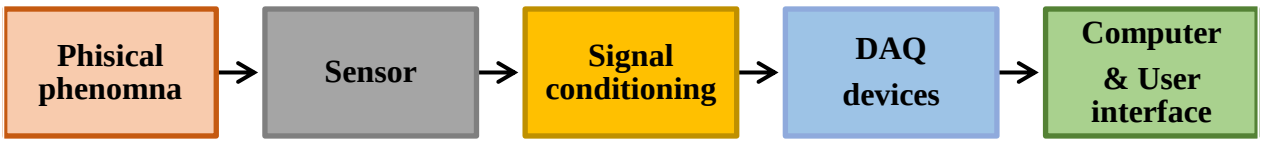

Figure 7-3 DAQ process steps

Several instruments and accessories are used for a complete data acquisition procedure including transducers, data acquisition device, controller card, cables, amplifiers, power supply devices, a computer, data acquisition software for receiving, storing and analysing the acquired data. These basic elements of the DAQ system used in this research will be explored below.

7.2.2.1 DAQ system devices

The data acquisition Accessories used in this research consists of two main devices, SCB-68 Shielded Desktop Connector Block 68-Pin and Multifunction I/O (MIO), both SCB-68 and MIO are connected via 68-pin cable.

The SCB-68 is a Shielded Desktop Connector Block with 68-Pin screw terminal type National Instruments as shown in Figure 7-4, this device has a general breadboard which can be customised to interchanging electrical components. This device can support conditioning to signals up to 16 analog-input channels. It has been equipped with 5 input channels [270]. The technical specifications of the SCB-68 are listed in Table 7-2.

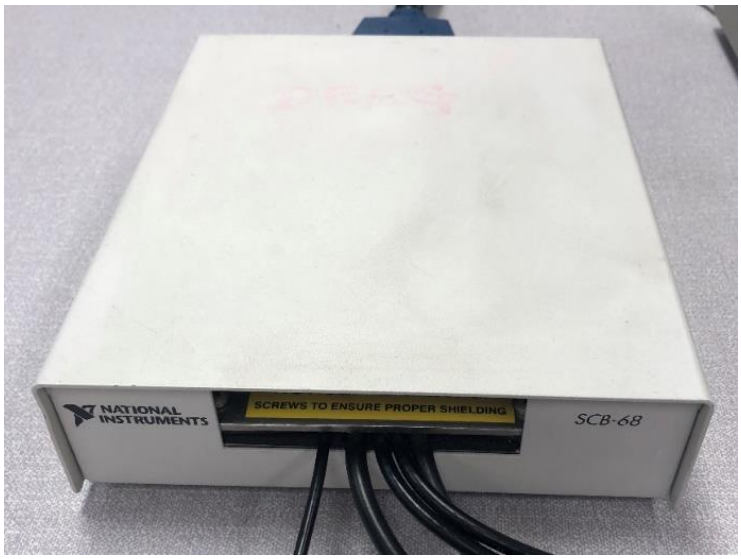

Figure 7-4 SCB-68 connector

Table 7-2 Technical specifications of the SCB-68 device [271]

No	Feature	Value
1	Number of screw terminals	68 all I/O signals are available at screw terminals
2	I/O connector	One 68-pin male SCSI connector
3	Maximum Working Voltage	≥42 Vpk/60 VDC
4	Operating temperature	0 to 70 °C
5	Weight	828 g
6	Wire gauge	14–30 AWG

Multifunction I/O devices offer a mix of I/O with varying channels, sample rates, output rates, and other features to meet many common measurement requirements. These devices are ideal for a wide variety of industrial applications such as laboratory automation, research, and design verification. A multi-function IO type NI 6221, as it can be seen in Figure 7-5, is used in this research. The technical specifications are shown in Table 7-3. It has been fitted in a computer and linked to SCB-68 connection block via SCSI-68 connector [270].

Figure 7-5 Multi-function IO NI 6221

Table 7-3 Technical specifications of Multifunction I/O

No	Feature	Value
1	Number of channels	8 differential or 16 single ended
2	ADC resolution	16 bits
3	Sample rate	250 kS/s
4	Timing accuracy	50 ppm of sample rate
5	Input coupling	DC
6	Input range	±0.2 V, ±1 V, ±5 V, ±10 V
7	Input FIFO size	4,095 samples

7.2.3 Accelerometer

The accelerometer is used as a basic sensor to measure the vibration on stationary parts of machines, the accelerometer measures absolute vibration by converting the deformation of piezoelectric ceramic plats, which built in the sensor, into electric charge. The generated electric charge cannot travel a long distance, therefore, accelerometers are fitted with pre-amplifiers to transform the electric charge into a voltage and their sensitivity is stated as mV/g or pC/unit. The selection of sensors has to be made based on the frequency range of interest and the anticipated level of the signals. Piezoelectric accelerometers are well-known and widely used to measure the vibration in condition monitoring of machinery. They have a wide dynamic frequency and a decent frequency range as well as good sensitivity

Figure 7-6 Piezoelectric accelerometer

In this study, two accelerometers as seen in Figure 7-6, type (CA-YD-104T), are used to collect the vibration data from the test bearing. One sensor was mounted vertically on the top of the hosting housing of the test bearing, whilst, another sensor was mounted horizontally on the side of same the housing.

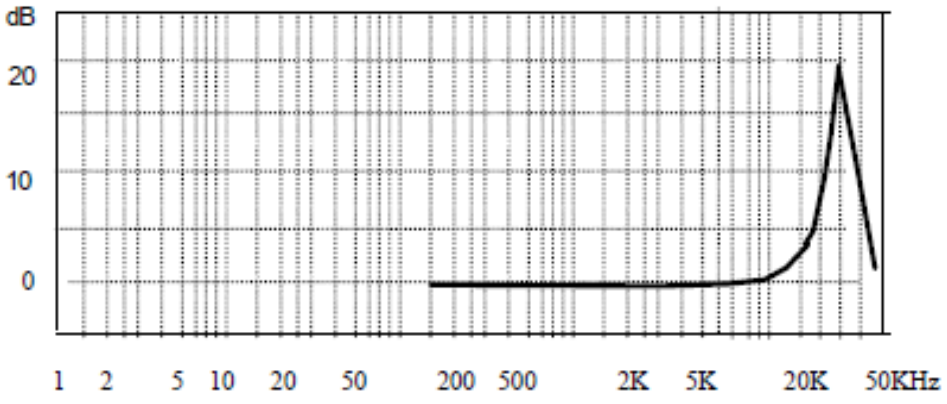

Figure 7-7 Typical frequency response

The frequency response of the accelerometer has a linear trend in the range of 0.1 KHz to 7 KHz as seen in Figure 7-7, in addition, the technical information is listed in Table 7-4.

Table 7-4 Sensors technical specification

No	Feature	Value
1	Model no	CA-YD_104T
2	Charge Sensitivity	3.640 pC/m/s^2
3	Transverse Sensitivity	≤5 %
4	Frequency Range 5%	0.5 - 7,000 Hz
5	Capacitance	1722 pF
7	Polarity	Positive direction
8	Operating Temp Range	-20 to +120 °C
9	Weight	32g
10	Mounting Method	M5
11	Sensing Element	PZT-5
12	Structural Design	Central Compression
13	Output Type	TNC

7.2.4 Charge Amplifier (CA)

The Charge Amplifier is used to condition the output vibration signal from the piezoelectric sensor and then transfer it to the DAQ system. A general purpose charge amplifier type YE5856 is used to link the sensors to the DAQ system. This amplifier has several features, the upper-frequency limit is 500 kHz, small in size, and with high integration. Three decimal-system uniformization output with a Built-in multiple lowpass and high pass filters with low noise and high precision. It can accept charge and voltage inputs with overload Indication. The technical specifications of the CA are listed in Table 7 5. The charge amplifier is used in the experiment to condition the signal. The CA accepts sensor output signals in a form of charge and convert them into a form that the data acquisition system can manipulate which is voltage in this experiment. The CA YE5856 has the ability to adjust the output using the switcher. The CA Switch is Selectable with values of (1, 10, 100 to 1000 mV/unit). The vibration data collected by the accelerometer with a value of (3.640 pC/m/s2) is transferred to the CA to be amplified and converted from charge into a voltage with an adjusted switch value of (10 mV/m/s2) and then transferred into DAQ system.

Table 7-5 Technical specification of Charge Amplifier

No	Feature	Value
1	Input Range	Voltage: ±10VP (Max); Charge: ±106 PC (Max)
2	Output Range	Voltage: ±10VP (Max)
3	Noise	≤ 5 μ V
4	Gain (Selectable)	0.1, 1, 10, 100, 1000 ,10000 mV/Unit
5	Accuracy	±1%
6	Low Pass Filter (selectable)	1k, 3k, 10k, 30k, 100k,500kHz (-3dB±1dB),-12dB/Oct
7	High Pass Filter (selectable)	0.3, 1, 3,10,30,100Hz (-3dB±1dB), -6dB/Oct
8	Temperature Operating	0-40°C
9	Max Humidity	95%R.H
10	Power Supply	DC: ±18 to ±27V; AC: 220V50Hz/110V60Hz
11	Dimensions	70mm(W) - 132.5mm(H) - 200mm(D)
12	Weight	1.5kg
13	Connections	Input: L5; Output: BNC; Power Supply: 3GTJE3(AC)
14	Input Cable	Double-ended L5 STYV-1 low noise cable(2m)
15	Output Cable	Double-ended BNC 50Ω output cable (2m)

7.2.5 Encoder

The instantaneous angular speed of a rotating shaft can be measured using encoders. To monitor the rotating speed of shaft a Hengstler Incremental Encoder (type RI32) is fitted on the fan cowl using a flexible coupling. The maximum speed of this encoder, as seen in

Figure 7-8, is 6000 rpm and can produce a series of pulses with a number of 100 pulses per revolution. The specifications of the Encoder provided by the manufacturer are illustrated in Table 7-6.

Figure 7-8 Hengstler Incremental Encoder [272]

Table 7-6 Technical specifications of the Encoder RI32 [272]

No	Attribute	Value
1	Pulses Per Revolution	100
2	Encoder Technology	Incremental
3	Maximum Speed	6000rpm
4	Output Signal Type	Push Pull
5	Shaft Type	Solid
6	Supply Voltage	10 → 30 V Dc
7	IP Rating	IP40
8	Overall Height	27mm
9	Minimum Operating Temperature	-10°C
10	Maximum Operating Temperature	+60°C
11	Series	RI32
12	Switching Frequency	200 kHz

7.2.6 Thermos Couples

Thermocouple RS PRO Type is used to measure the temperature of the tested bearing. It has been mounted on the tested bearing housing and linked to DAQ system. The technical specifications provided by the manufacturer are depicted in Table 7-7.

Table 7-7 Thermocouple technical specification

No	Feature	Value
1	Type	K
2	Probe Length	2m
3	Diameter	1/0.2mm
4	Temperature range	-60 ºC to +350 ºC
5	Termination Type	Miniature Plug
6	Standards Met	IEC
7	Response Time	Fast

7.2.7 Slip Metric Gauge

A slip metric gauge box set, type Matrix Pitter 8075 C, was used to precisely measure the clearance. This is the workshop grade 2 version, typically used for setting up machine tools, positioning milling cutters and checking mechanical widths. Moreover, it can also be used as a length measurement for the regulation and adjustment of indicating measuring instruments and linear dimensions of industrial components[273].

Figure 7-9 Matrix Pitter 8075 C metric gauge

7.2.8 Dial Indicator

Dial indicators, as seen in Figure 7-10, are designed specifically to measure relative position. It consists of the dial (face), case, and the plunger. The plunger is spring loaded part that can be depressed into the case making the dial needle to move clockwise. A dial indicator with a magnetic base is used to measure the shaft endplay[274].

Figure 7-10 Dial indicator [274]

When there is no pressure is applied to the plunger, it is fully extended out of the case. The total distance that s plunger travels in or out varies depending on the model of the indicator. The face

can be rotated to set the indicator needle to zero. It is flexible and hence can be mounted in too many options.

Table 7-8 Moore & Wright Dial Indicator

No	Feature	Value
1	Brand	Moore & Wright (Metric)
2	Range	10 mm
3	Dial reading	0.01mm
4	Graduation	100
5	Stem diameter	8mm
6	Mounting type	Magnetic base

7.2.9 Support Bearing

A double angular contact bearing type NSK 3307 is used in the experiment as a support bearing. Double row angular contact ball bearing, as seen in Figure 7-11, consists of solid inner race, outer race, balls assemblies with polyamide or sheet steel cages. Due to the higher quality of balls, high quality raceway surfaces and the more precise dimensional and running tolerances, their internal friction is low and the noise level of the application during operation is reduced. They can support axial loads in both directions and high radial loads. They are particularly suitable for applications where rigid axial guidance is required. The axial load carrying capacity they can support depends on the contact angle. Hence, the larger the contact angle, the higher the axial load they can be subjected to. The choice of this bearing to act as a support bearing in the experiment is based on the high axial guidance which it can provide.

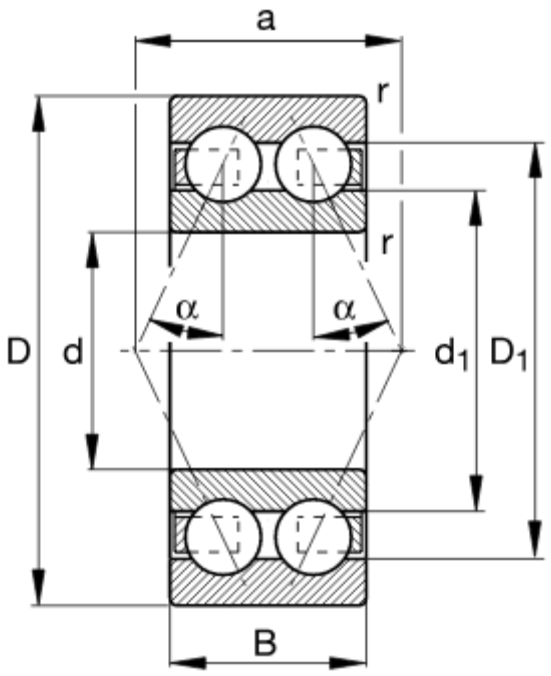

Figure 7-11 Double-row angular contact ball bearing 3307

Table 7-9 Manufacture specifications of double angular contact ball bearing NSK 3307

No	Dimensions	Measurements
1	d	35 mm
2	D	80 mm
3	B	34,9 mm
4	a	47,986 mm
5	D_1	68,9 mm
6	d_1	51,8 mm
7	r	1.5 mm
8	Contact angle **α**	30°
9	Mass	0,73 kg
10	Basic dynamic load rating, radial	55000 N
11	Basic static load rating, radial	36500 N
12	Fatigue limit load, radial	2460 N

7.2.10 Tapered Roller Bearing

TRB have tapered outer race and tapered inner race as well as tapered rollers. They have a high capacity to support combined load simultaneously, radial and thrust load. The capacity of axial load support has a linear relationship with the contact angle α. Thus TRBs are used in the high load applications, such as differential gears and wheels of vehicle, gas turbine engines, helicopter transmissions, milling machine spindle, etc. Long service life is expected when operated under

good condition[275]. A tapered roller bearing consists of outer and inner rings with tapered rollers in a window cage and tapered raceways as seen in Figure 7-12.

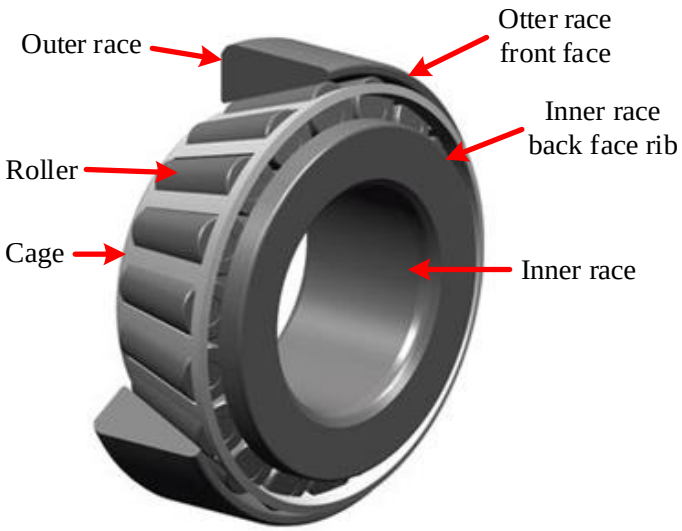

Figure 7-12 TRB 31308

The geometrical dimensions are illustrated in the schematic diagram in Figure 7-13 and the geometrical information is shown in Table 7-10.

When the radial load is applied, it will induce force in the axial direction, thus, it has to be handled by using a pair of bearing. As the cup and the inner race with rollers come apart, mounting the bearing with a preloaded condition or with clearance is possible. TRB can be set at initial machines assembly to any desired axial or radial clearance. This helps to control a bearing to meet expected application operating conditions, and as a result, optimum bearing performance can be achieved. There are several advantages of the flexibility to setting TRB such as longer life can be expected by setting the bearing based on the application performance requirements, easy to be assembled as the inner ring with rollers and the outer ring come separate, moreover, it can be set at the time of machine assembly.

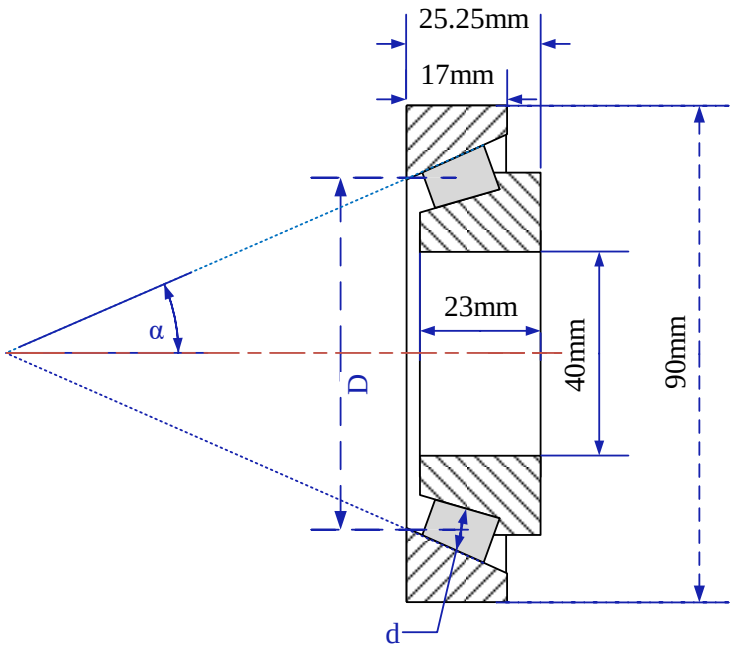

Figure 7-13 TRB schematic diagram

Table 7-10 Tapered bearing 31308 geometrical information

No	Specifications	
1	Series	31308
2	Cone Part Number	X31308
3	Cup Part Number	Y31308
4	Design Units	Metric
5	Bearing Weight	0.700 Kg
6	Cage Type	Stamped Steel
7	**Dimensions**	
8	d - Bore	40.000 mm
9	D - Cup Outer Diameter	90.000 mm
10	B - Cone Width	23.000 mm
11	C - Cup Width	17.000 mm
12	T - Bearing Width	25.250 mm
13	**Basic Load Ratings**	
14	C90 - Dynamic Radial Rating (90 million revolutions)[4]	26200 N
15	C1 - Dynamic Radial Rating (1 million revolutions)[5]	101000 N
16	C0 - Static Radial Rating	88100 N
17	C_{a90} - Dynamic Thrust Rating (90 million revolutions)[6]	37000 N

7.2.11 TRBs Setting Methods

TRBs can be assembled by several methods; supplied as pre-set assemblies, automatic techniques and manual method [276]. Pre-set assemblies are available in a variety of styles, forms and arrangements, this type of assemblies are designed for many applications that require the use of close coupled or double-row bearing assemblies. Pre-set bearings are provided by the manufacturer

and usually supplied with a specified internal clearance for given application requirements. As an alternative to manual methods, automated bearing setting techniques. Many of the automated settings techniques have been developed, among them Set-right, Projecta-Set, Acro-Set, Clamp-set, Torque-Set etc. each of which has its own advantages and limitations. In this research, the manual method is adopted due to the ease of access to the test bearing and the less complexity and tools that are needed to manually set the bearing. Moreover, a settings variation from the optimum settings is expected when using automated methods. The devices which are needed are costly and need a certain level of knowledge to be utilised [277].

Manual bearing setting method is the most widely used to set RRBs on different equipment. Manually method does not require any special tooling or charts or fixtures etc. the successful mounting depends on the assembler's skill and judgment. This method involves tightening the adjustable nut while rotating the shaft until a slight bind in the rotation is felt. Then the desired settings can be achieved by slightly adjusting the nut back or forward turn. A magnetic base indicator can be used with a pry bar to measure the endplay of the bearing [278].

7.2.12 Wear Simulation Using an Adjustable Clearance Mechanism

During the experiment, a mechanism for the preload adjustment was built by controlling the clearance between the outer race, which was fixed in the housing, and rollers assembled with the cage. This was executed by rotating a precision positioning screw nut, to move axially relative to the reference slip, to make the bush, and fit the inner ring into the outer race. The schematic diagram of the clearance measurement system is described in Figure 7-14. The value of the clearance is calculated by measuring the gap between the reference slip and the nut edge based on the thickness of the calibrated slips. The reference slip, as seen in Figure 7-14 and in Figure 7-15, can be firmly fitted to a groove of 5mm width, which was grooved in the shaft, to be used as a reference when measuring the gap.

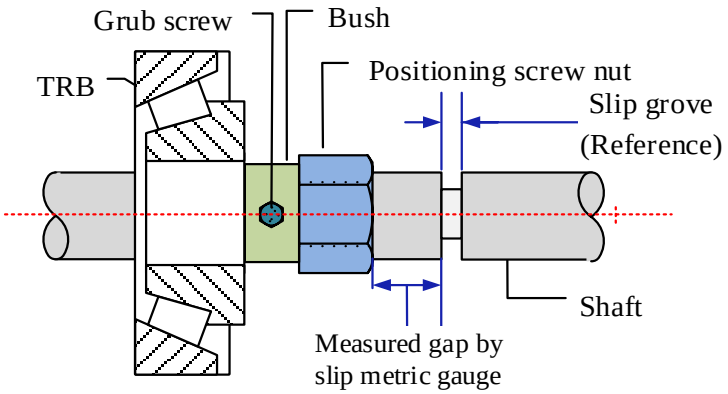

Figure 7-14 Sketch diagram of the clearance adjustment mechanism

The settings of TRBs at assembly are defined mainly as three conditions as following:

- Firstly, the endplay setting is when an axial clearance between rollers and raceways produces a measurable gap.
- Secondly, a line-to-line setting (zero clearance), is the transitional point between endplay and preload settings.
- Thirdly, a preloaded setting is when there is an axial interference between rollers and raceways and there is no measurable gap in-between. This can be carried out by applying pressure to the two surfaces relative to each other in this manner pressing the two surfaces firmly against each other and placing the bearing in a stressed condition[279].

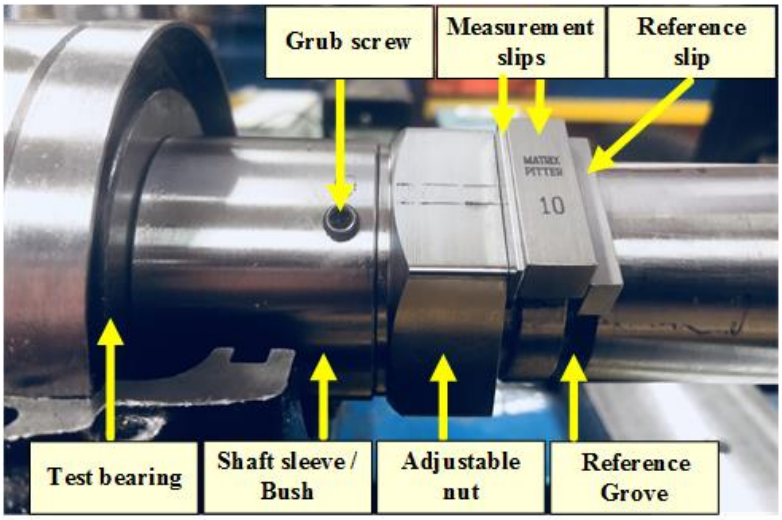

Figure 7-15 Adjustable clearance mechanism

After setting up the bearing to the line-to-line setting which is assumed to have zero clearance, the measurement slips are used to measure the gap between a slip which is fitted in the groove and the positioning nut edge that controls the adjustment of the position of the cone to the outer race.

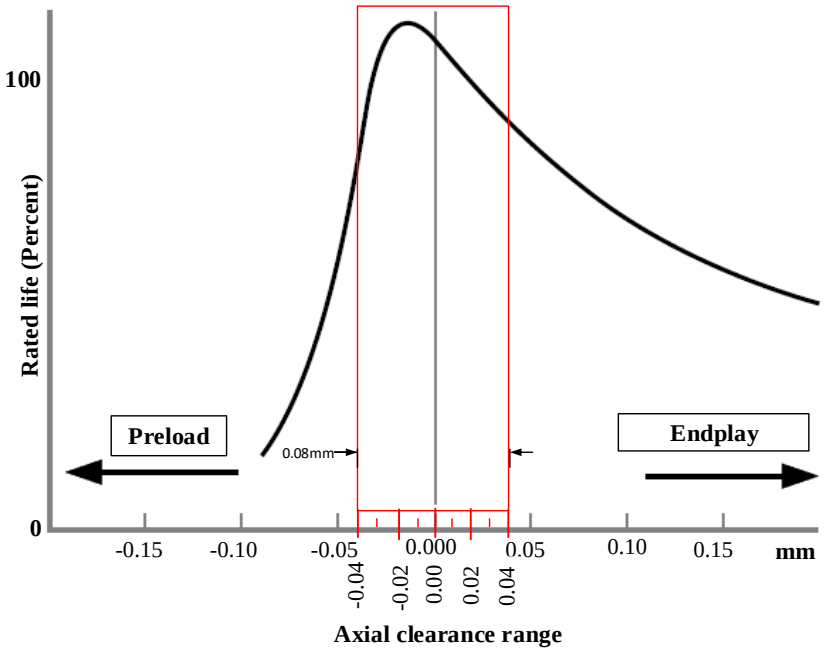

Figure 7-16. Measurement range selection based on typical life[279, 280]

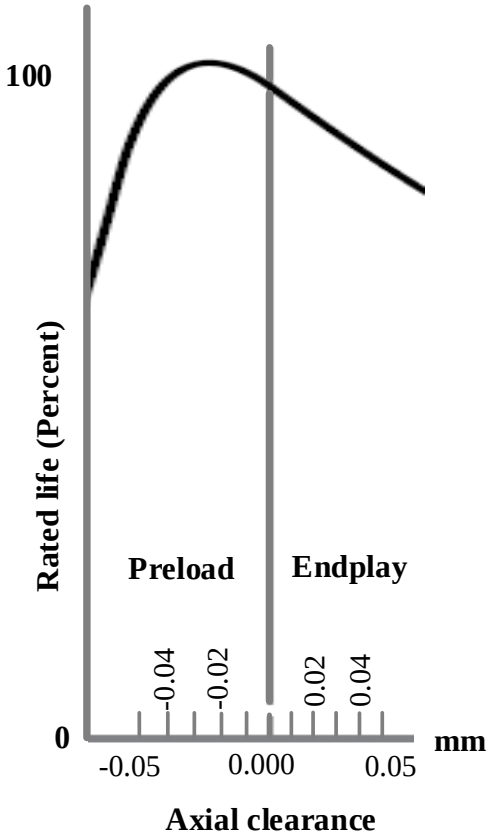

Figure 7-17 Axial clearance range

The total width of the slips fitted in the gap is used as a measure of the internal clearance. So a measure of 11.14 mm is used as a zero-clearance setting, and a smaller gap value than 11.14 mm means an endplay setting, whilst a larger gap value than 11.14 mm indicates a preloaded setting. The clearance range zone was chosen based on the optimal settings range within the maximum typical life, provided by the manufacturer as seen in Figure 7-16. Two endplay cases with measured width of 11.12mm and 11.10mm (with clearance variation of +0.02mm and +0.04mm, respectively) and two preloaded cases with the measured width of 11.16mm and 11.18mm (with clearance variation of –0.02mm and –0.04mm, respectively) were adjusted to study the effect of the clearance.

7.2.13 Fault Seeding

In this experiment, five bearings were used to study the different types and severities of bearing faults. The healthy bearing was used as a baseline or reference. On the other side, rectangular slot defects with two different severities were artificially made using an electro-discharge machine (EDM). Defects were induced into one bearing with small outer race and another bearing with large outer race defect, also one bearing with small fault on the roller and another bearing with large fault on the roller. The fault size and the description can be seen in Table 7-11 and in Figure 7-18, respectively. The defective area was restricted to a small value to simulate initial faults and enlarged in two other cases. In order to exclude the influence of the temperature on the bearing internal clearance, data was recorded after warming up the test rig until the temperature remained stable at around 30ºC. Five recordings were recorded and compared to confirm the stability and accuracy of the experiment.

Table 7-11 Seeded defect size of bearings

Bearing No.	Bearing condition	Defect length (mm)	Defect depth (mm)
1	Healthy	–	–
2	Outer race with small fault	2.0	0.2
3	Outer race with large fault	4.0	0.2
4	Roller fault with small fault	2.0	0.2
5	Roller fault with large fault	4.0	0.2

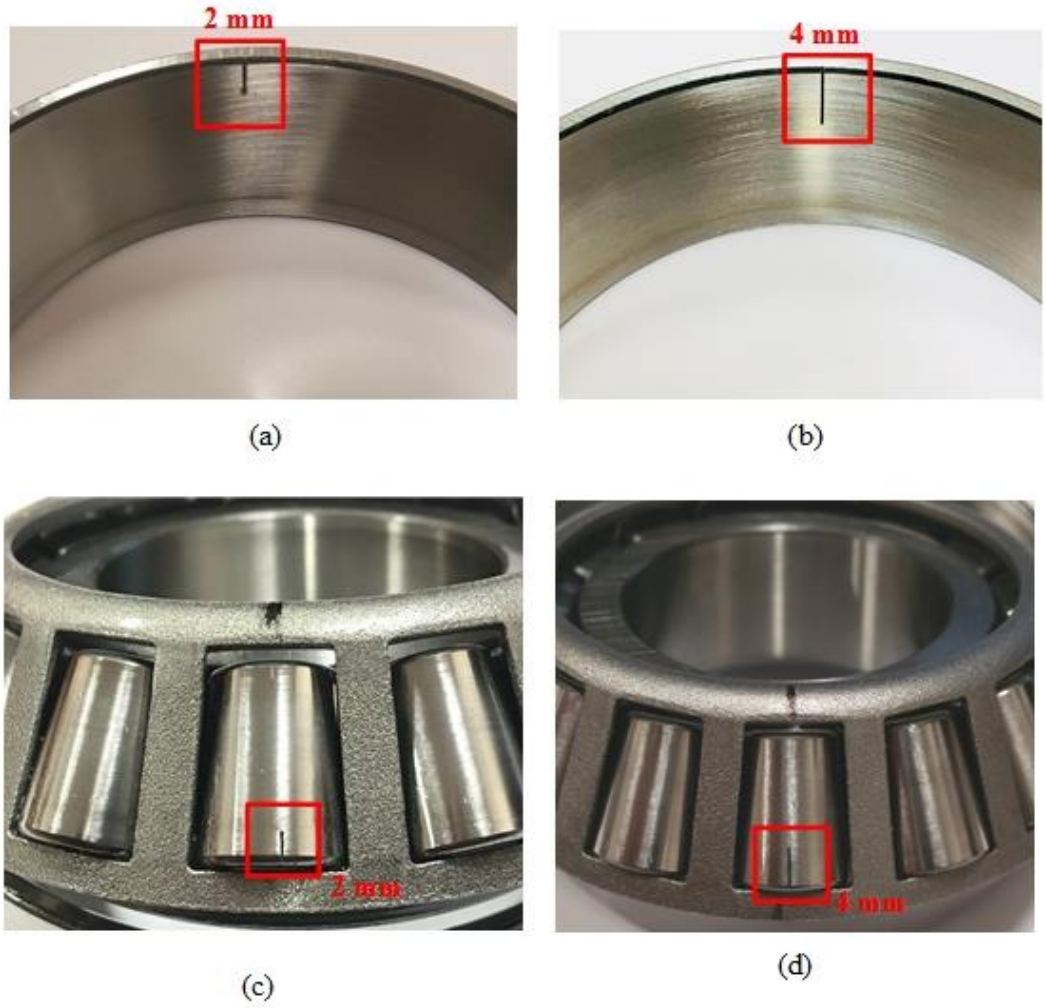

Figure 7-18 seeded defects (a) 2 mm (b) 4mm on outer race, and (c) 2mm (d) 4 mm on roller

7.3 Experimental Procedure

In this study, the vibration data was collected using the following procedure:

1. For warming up the bearing rig, before start recording data, the bearing rig ran 10 minutes.
2. All data were collected under 1500 rpm motor speed.
3. The sampling rate used in the experiment was 50 KHz.
4. The time duration of each data recordings was 30 seconds so this meant each of the recorded data will cover about over 750 shaft rotations (30*25).
5. The length of the recorded data was 50000*30= 1500000 data-points.
6. In order to ensure that the signals obtained are consistent, each case of bearing conditions repeated five times so five data-sets of each condition were recorded at the same day so all experiments were assumingly conducted under the same conditions.
7. The five bearing conditions have been tested with the same speed and operating conditions.

8. The faults-free bearings tested by recording the vibration signals using the accelerometer mounted on the bearing housing in the vertical direction, this data is used as baseline data (reference).
9. The second test for the bearing with outer race fault started with the smaller defect then the larger defect.
10. The next test was for the bearing with roller fault started with the smaller defect then the larger defect.

7.4 Summery

This chapter described the development of the test rig, in (section 7.2) the components with all instruments were explored, also, the adjustable clearance mechanism was described in details. The experimental procedure presented in (section 7.3).

CHAPTER EIGHT

VIBRATION DATA ANALYSIS AND CLEARANCE ESTIMATION BASED ON LOW-FREQUENCY BAND ANALYSIS

In this chapter the collected vibration datasets are analysed and explored in the time domain and frequency domain, also the effect of wear evolution on the condition monitoring of REBs is analysed based on internal clearance changes resulting from the wear effect. Then, an experimental study is ingeniously designed to simulate wear evolution and evaluate its influence on well-known envelope signatures according to measured vibrations from widely used tapered roller bearings. The fault type was diagnosed in two indices: the magnitude variation of characteristic frequencies and the deviation of such frequencies.

8.1 Introduction

In this chapter, data collected from the baseline and defective bearings analysed, the analysis start with identifying the missing parameters that are required to theoretically calculate the fault frequencies of the tested bearing. Moreover, based on the experimentally identified fault frequencies, the deviation of the fault frequencies due to clearance changes is investigated. A clearance adjustment mechanism is developed and used for this purpose. The results are presented and discussed also summarised at the end of this chapter.

Bearing fault frequency can be calculated using the formulas shown in Table 8-1 based on the geometrical information usually provided by the manufacturers[74]. However, TRBs come apart and some of the parameters will be dependent on the bearing settings. Hence, to identify those missing parameters, an experiment is carried out and then the fault frequency can be calculated according to the missing parameters.

Table 8-1 Characteristic Fault Frequencies

Fault type	Calculation formula	
BPFO: Ball Pass Frequency of Outer Race (Hz)	$f_{BPI} = \frac{1}{2} z \left(1 + \frac{d}{D} \cos a\right) f_s$	(8.1)
BPFI: Ball Pass Frequency of Inner Race (Hz)	$f_{BPI} = \frac{1}{2} z \left(1 + \frac{d}{D} \cos a\right) f_s$	(8.2)
BSF: Ball Spin Frequency (Hz)	$f_{BS} = \frac{D}{2d} = \left[1 - \left(\frac{d}{D} \cos a\right)^2\right] f_s$	(8.3)
FTF: Fundamental Train Frequency (Hz)	$f_c = \frac{1}{2}\left(1 - \frac{d}{D} \cos a\right) f_s$	(8.4)

where z stands for the number of rolling elements, d is the diameter of the rolling element, D is the pitch diameter, α is the contact angle, and f_s is the shaft frequency. A localised defect of a rolling element impacts the outer race and the inner race once a spin and two transients are generated. Thus, $2f_{BS}$ is normally used as the fault frequency to indicate rolling element defects. The theoretical fault frequencies of bearings used in this study are calculated for the outer race, rolling elements, carriage and inner race.

8.2 Estimation of Geometric Parameters

Due to the geometry and the component separation characteristics of tapered roller bearings, the parameters, such as rolling element diameter, pitch diameter and contact angle needed for the theoretical fault frequency calculation, are not available from the open data resources. A test approach was adopted to estimate these key parameters. Based on measured data with artificial faults on outer race and a roller, the shaft frequency f_s, f_{BPO}, and $2f_{BS}$ were identified from the experiment results. With these known frequencies, the parameters were estimated, and the parameter β can be estimated according to

$$\beta = \frac{f_s}{2f_{BS}}\left[1 - \left(1 - \frac{2f_{BPO}}{zf_s}\right)^2\right] \tag{8.5}$$

Which is derived by Equations (8.1) and (8.3). Then submitting equation (8.5) to Equation(8.3). Then $\cos a$ can be estimated by

$$\cos a = \left(1 - \frac{2f_{BPO}}{zf_s}\right)/\beta \tag{8.6}$$

Thereafter, at a running speed of 1500 rpm ($f_s = 25Hz$) and clearance of zero, all characteristic frequencies can be obtained as shown in Table 8-2.

Table 8-2. Defect frequencies of bearings tested at 1500 rpm

Fault location	Defect frequency (Hz)
Inner race	217.5595
Outer race	157.1041
Roller	133.5958
Cage	10.4736

8.3 Initial Experimental Results and Discussion

The analysis of the data starts with time domain, frequency domain and envelope spectrum. Also the deviation of the characteristic frequencies is analysed.

8.3.1 Time Domain Analysis

The analysis of the data starts with exploring the time domain analysis of the different cases; baseline, small outer race fault (2mm) and large outer race (4mm). Also, small and large roller fault cases will be discussed. Figure 8-1 and Figure 8-2 explore 0.2 seconds of the vibration raw signals for baseline, small and large outer race faulty signals respectively for all clearances (+0.04 +0.02 0.00 -0.02 -0.04). Three traces are present, the blue trace is from the baseline bearing and the red trace is from the small fault in the outer race while the magenta trace is from the large fault in outer race.

The amplitude of the cases shows a slight increase in the presence of a physical change in the system. The case belongs to outer race fault as seen in Figure 8-2 is larger as it belongs to the large fault with a severity of 4mm on the outer race.

From the time domain analysis of two faulty cases of outer race small and large, it can be noticed that the amplitude of two faulty cases increases slightly with the decrease of clearance with regard to the baseline and it is dominated by the defects signal energy. However, for the cases with large clearance, the amplitude is not clearly shown any changes.

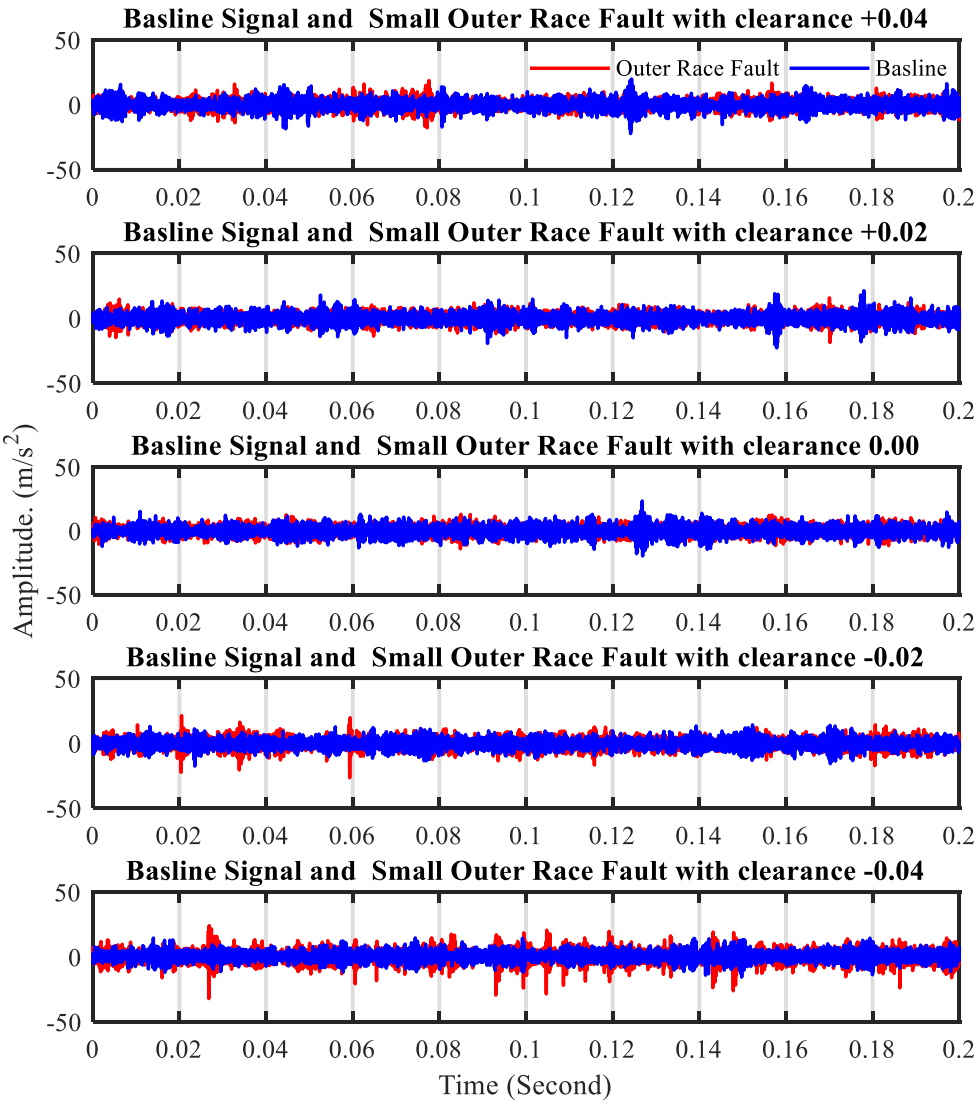

Figure 8-1 Time Analysis of Baseline and Small Outer Race

It can be observed that, in the outer race fault cases, the structures of the data show differences from the baseline data, especially in the smaller clearances. The change occurs in the more detailed parts of the waveforms. The periodic feature, however, remains in the waveforms. Case 2 amplitude clearance -0.04 (magenta trace) is slightly larger than that of case 1 with the same clearance (red trace).

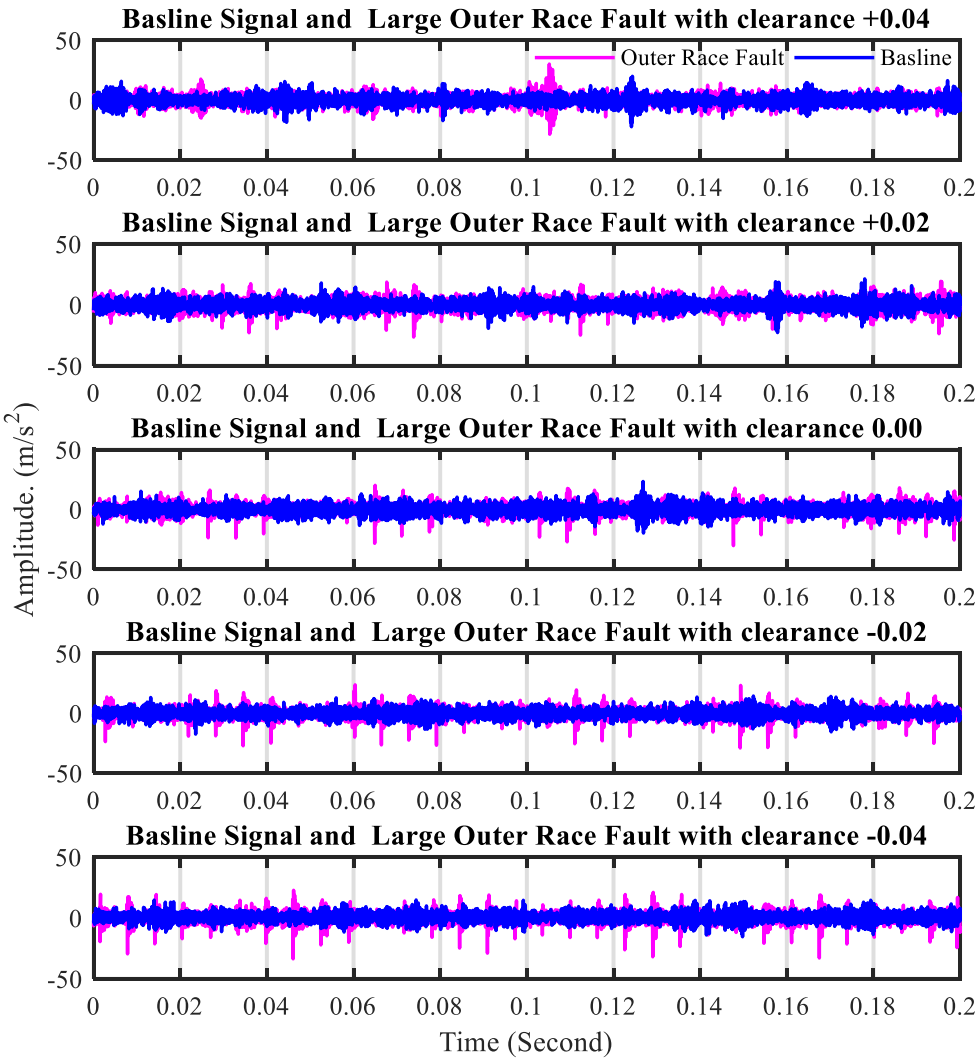

Figure 8-2 Time Analysis of Baseline and Large Outer Race

In Figure 8-3, a healthy bearing signal (blue trace) and a bearing with a small defect on the roller (red trace) are presented. Whilst, Figure 8-4 compares the baseline signal with large roller defect (magenta trace). It can be seen that there is a presence of some peaks which belongs to the defect induced into the roller. The amplitude of the peaks for both small roller and large roller defect increased with the decrease of the clearance. The structures of the data show differences from the baseline data, the change occurs in the more detailed parts of the waveforms.

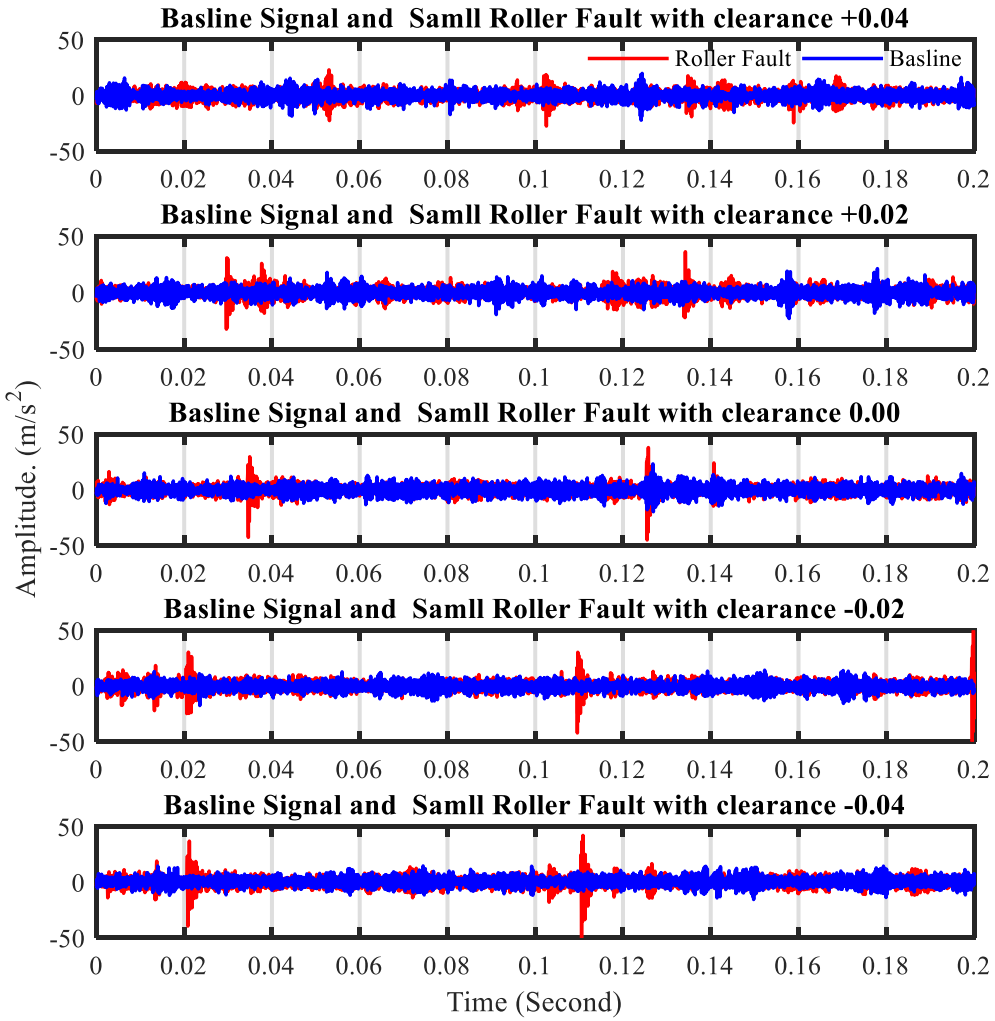

Figure 8-3 Time Analysis of Baseline and Small Roller Defect

The periodic feature, however, remains in the waveforms. Moreover, in the signal signature of the outer race defects, there are repetitive impacts at intervals which corresponding to the time interval between the rolling elements and the defects point when the roller elements pass over the defect area, results in a repeated a series of impacts because of collisions of the metal to metal contact. Also in the case of roller defects as the roller spins, the line contact between the races and defect area on the roller will produce a repeated series of impacts. The magnitude of peaks appeared in the cases of the outer race are smaller than the peaks appeared in the roller cases. This is because of the contact area in the roller cases are larger than the out race cases as the defect induced at the edge of the outer ring, thus, the rollers are not fully in contact with defect area due to the design of the bearing.

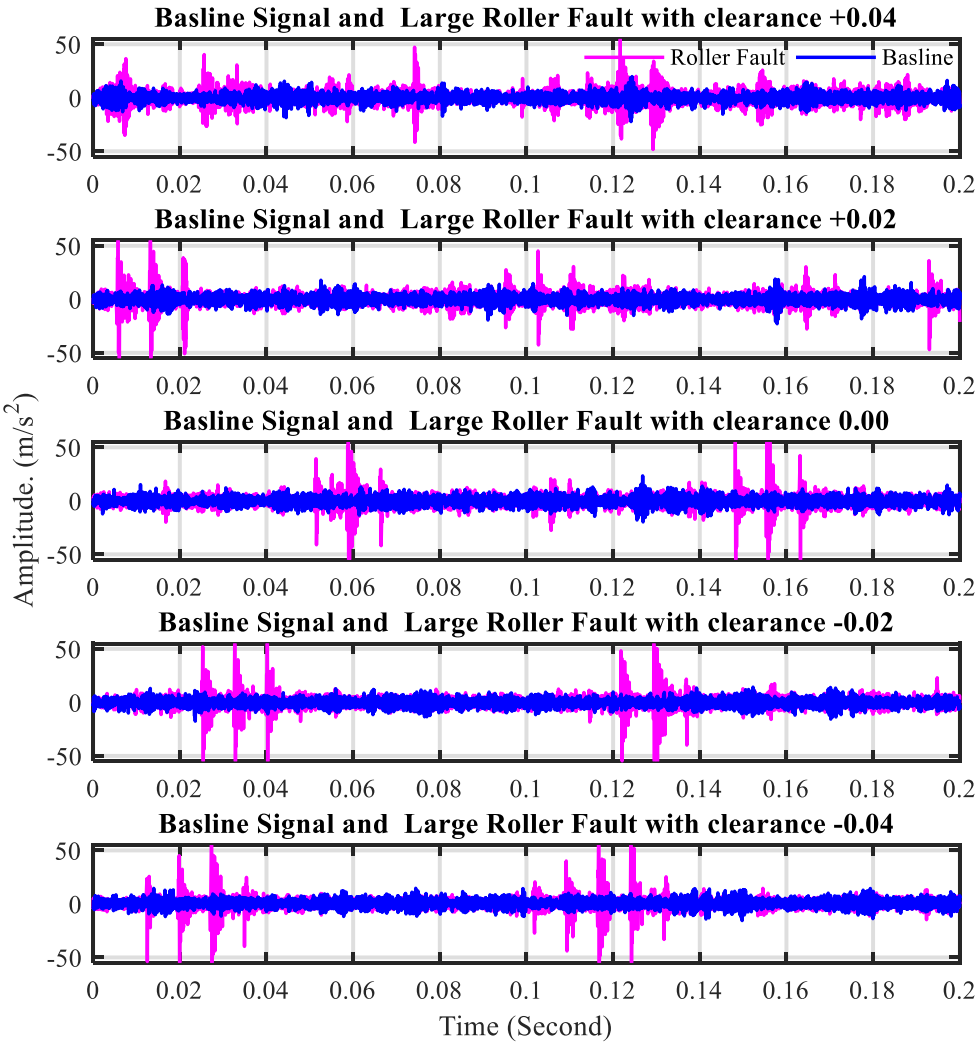

Figure 8-4 Time Analysis of Baseline and Large Roller Defect

Despite the slight change of the signal amplitude in the time domain analysis, it is still not adequate to identify the defects occurred in the two faulty cases. Moreover, raw vibration signals from bearings do not always generate a signal that shows the impacts similar to the clarity as they appear in Figure 8-2 and Figure 8-4, because the raw vibration signal will be the summation of all vibration signals from the entire system, thus, will be containing many components. Moreover, usually, at the early stage of bearing defects like spalling, incipient impulses will be contaminated and masked by the background noise of the entire system components. Therefore, it will be very unlikely to be detected by viewing the time signal. A further investigation will be carried out using statistical analysis in the next section.

For more demonstration, statistical analysis is carried out using RMS and Kurtosis. Figure 8-5 and Figure 8-6 shows the results obtained from the statistical analysis for each bearing. RMS of the baseline and two outer race cases are calculated and shown in Figure 8-5 (a), however, it does not

provide a clear indication of the faults induced into the outer race for the bearing with outer race fault. Figure 8-6 (a) presents the RMS of baseline and the bearings with roller fault, it can be seen that RMS of the large roller fault has a higher amplitude compared to the baseline data while the small fault does not have such clear difference. Kurtosis is another statistical parameter used to investigate the change in the signal structure due to the seeded faults. The results obtained in Figure 8-5 (b) Figure 8-6 (b) from Kurtosis again cannot obviously describe the presence of the faults in both bearings with different severities for all clearance cases.

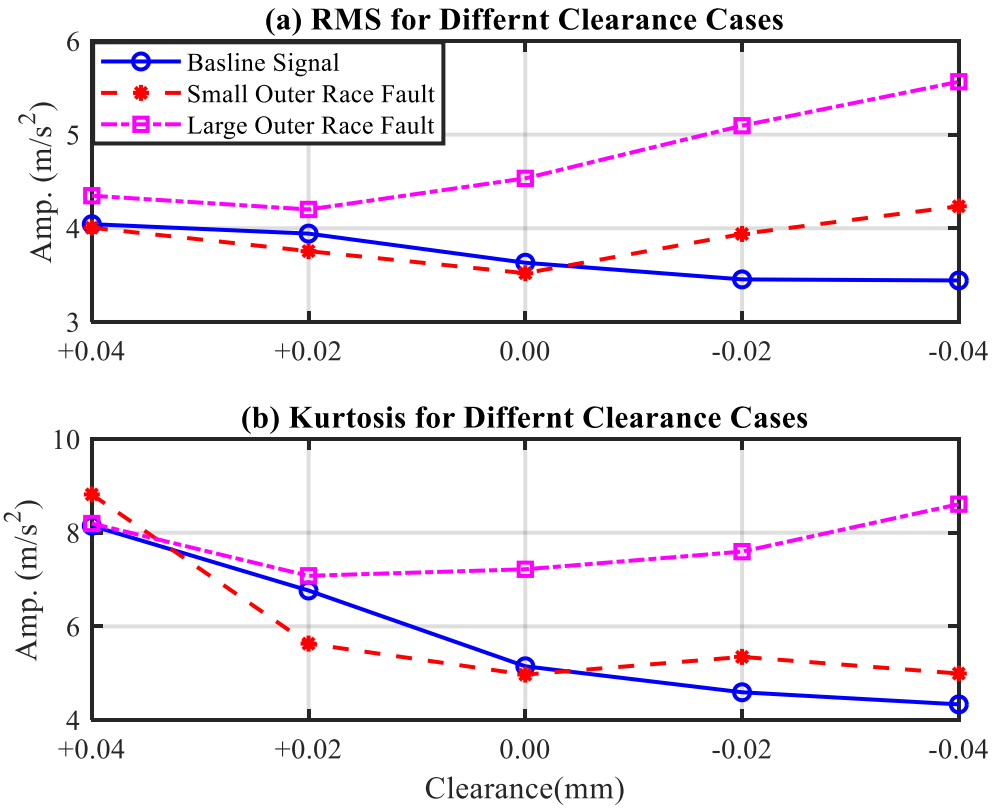

Figure 8-5 RMS and Kurtosis of Baseline and Outer Race Cases

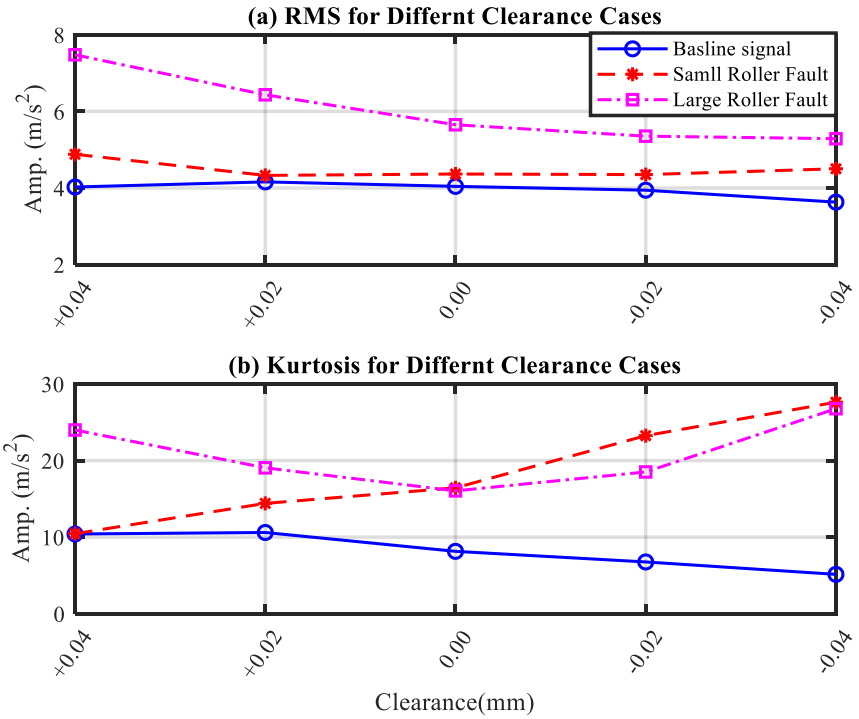

Figure 8-6 RMS and Kurtosis of Baseline and Roller Cases

8.3.2 Spectrum Analysis

The presence of peaks induced by faults in the spectrum analysis can be used to diagnose the faults in rotary machinery. In the case of diagnosing a bearing defect, the characteristic fault frequencies need to be calculated in advance as depicted in Table 8-1 and in Table 8-2. Fault frequencies can be checked by exploring the location of the faults in the vibration spectrum. As seen in Figure 8-7, the spectrum of vibration data is presented for the baseline (a) and two outer race faulty cases, small outer race fault (b) and large outer race fault (c). However, there are no dominant peaks in the spectrum around the outer race fault frequency 156Hz that clearly can be identified by the difference between the baseline and faulty cases.

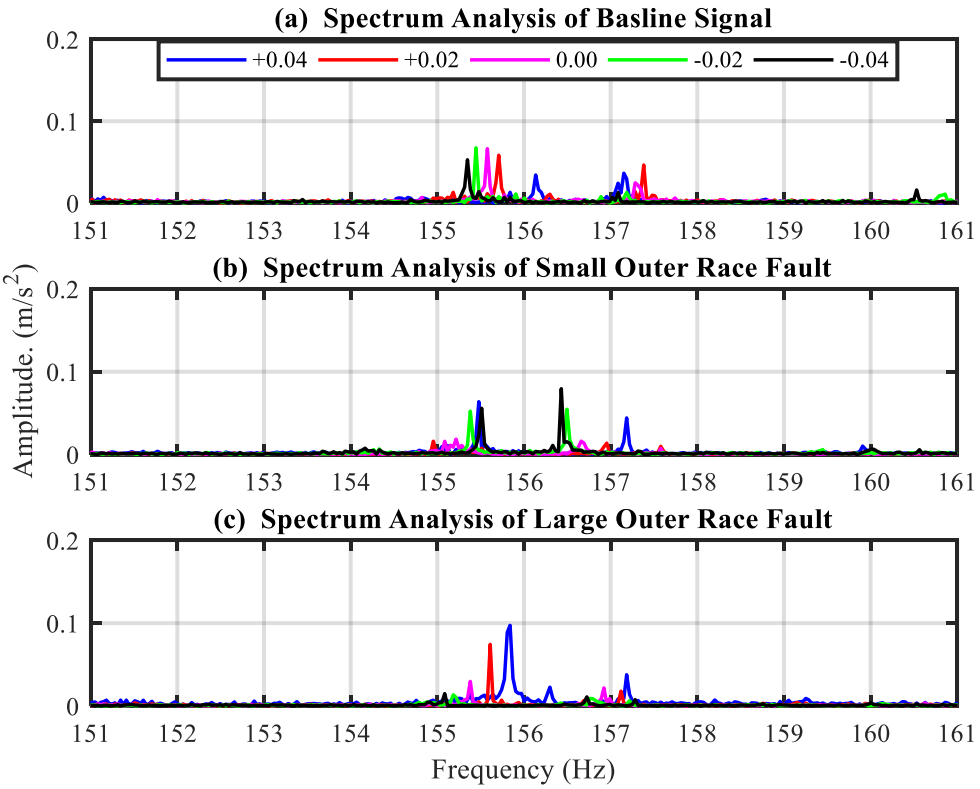

Figure 8-7 Spectrum Analysis of baseline and outer race defects

Figure 8-8 shows the spectrum analysis of the vibration data for baseline (a), small roller fault (b) and large roller fault (c). The calculated characteristic fault frequency for roller as seen in Table 8-2 is around 134 Hz, however, the spectrum analysis does not provide clear peak belongs to the roller defect in the two faulty cases (b) and (c) compared to the spectrum of the baseline signal (a). From the results obtained using Fast Fourier Transform, it can be concluded that it is very unlikely to detect the peaks belong to the calculated faults at 156 Hz and 134 Hz with their harmonics for the outer race defects and the roller defects. It can be seen that there is no significant increase in the amplitudes at the defects frequencies compared with the no-fault case.

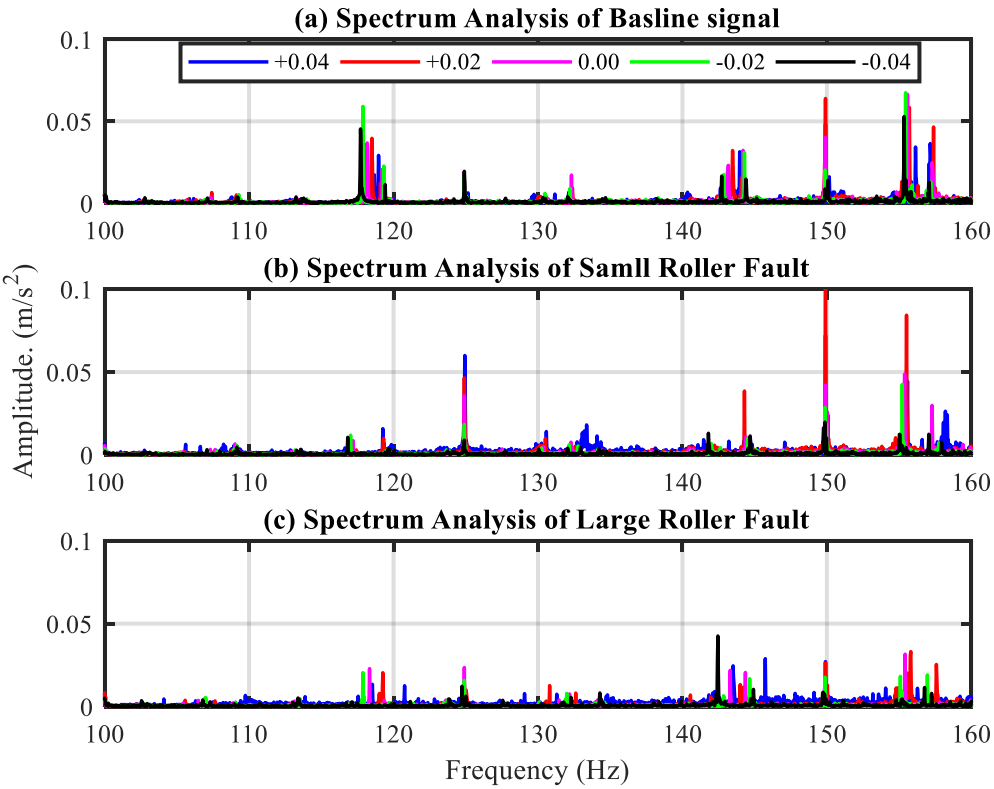

Figure 8-8 Spectrum Analysis of baseline and Roller defects

8.3.3 Envelope Spectrum Analysis

As discussed in Chapter Five, Envelope analysis has been a well-known method for monitoring REBs, to perform envelope analysis, the signal needs to be passed through a band-high-pass filter to eliminate the low-frequency noise, and then rectify the signal, and then in the envelope spectrum, the fault frequency components can be identified. The common envelop analysis has been performed to all bearings data. First, data filtered by passing it through the frequency band 2 KHz to 4 KHz and then envelop analysis performed. Figure 8-9 shows the envelop analysis of baseline signal and the results show that it is very flat and no spectrum signs can be identified for all clearance cases.

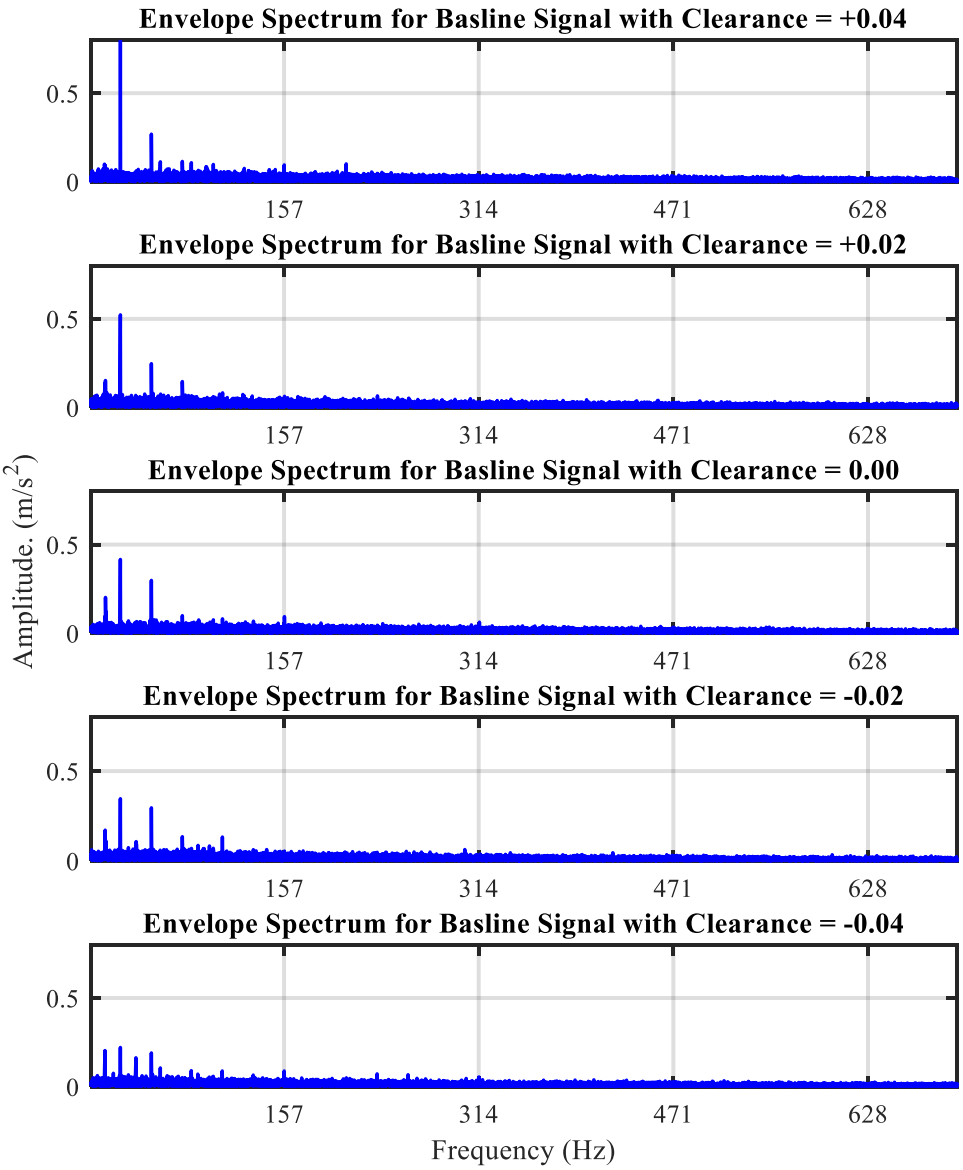

Figure 8-9 Envelope Spectrum Analysis of the Baseline

Figure 8-10 presents the envelope spectrum of vibration data collected from a defective bearing with a small (2mm) outer race fault for all clearance cases. It can be seen that for large clearance cases (+0.04 and +0.02), the fault peaks and their harmonics cannot be easily detected. However, some of the fault peaks associated with preloaded conditions (-0.02mm and -0.04mm) can be detected at 157 Hz with their second harmonics and sidebands spaced with the rotation frequency.

At the same time, peaks at the characteristic defect frequency for large fault severity (4mm) (magenta trace) with some large internal clearance conditions (+0.02mm and +0.04mm) can be seen in Figure 8-11. The faulty peaks for large outer race fault, especially for clearance smaller

than (+0.02), appear clearly at 157 Hz with their second harmonics at 314 Hz and third harmonic at 471 Hz till fifth harmonic spaced with the rotation speed due to the unbalanced shaft.

Figure 8-12 shows the envelope spectrum of vibration data gathered from a defective bearing with a small roller fault (2mm), in the large clearance condition (+0.04), the peaks and their harmonics not clear enough to be used to diagnose the bearing condition. In clearance conditions +0.00 and smaller, the faulty peaks can be detected with their second harmonic spaced with cage frequency (10.5 Hz) as summarised in Table 8-4. The magnitude of the spectral lines increases with the declines of the internal clearance.

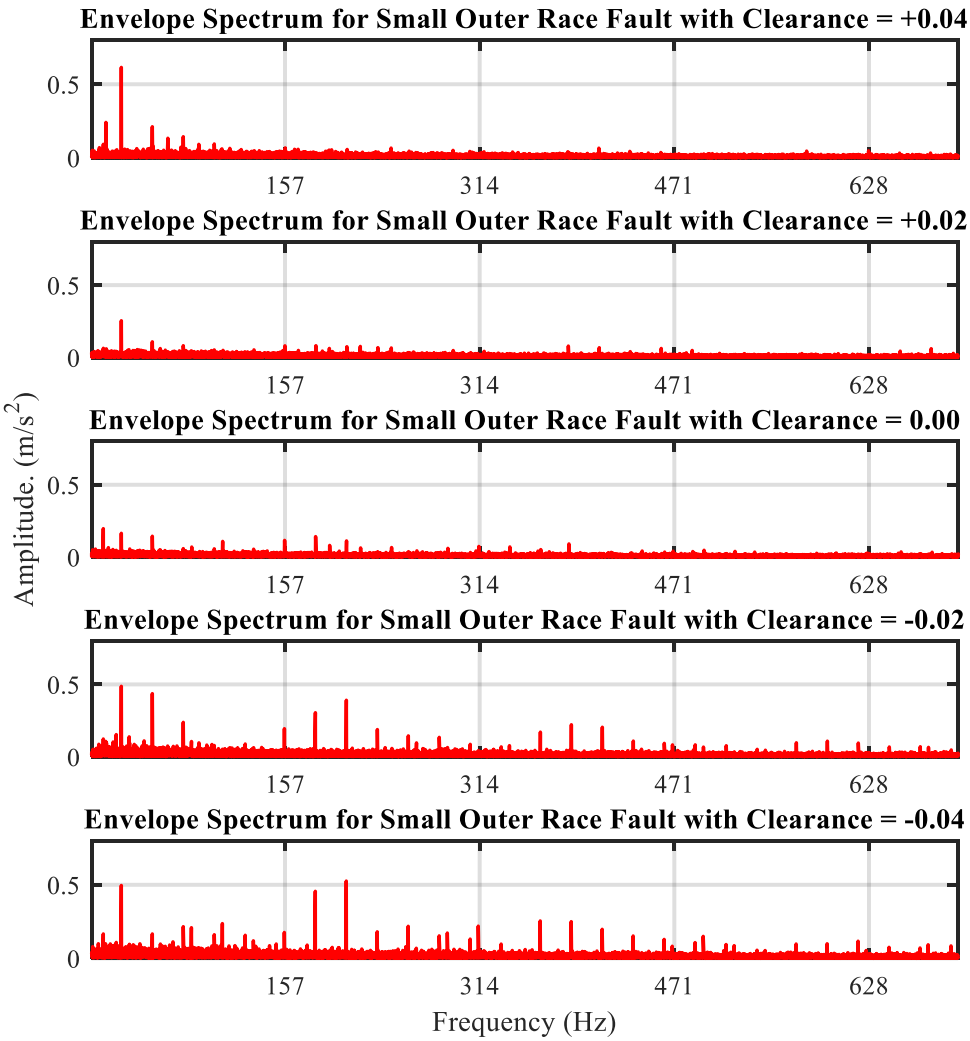

Figure 8-10 Envelope Spectrum Analysis of the Small Outer Race Defect

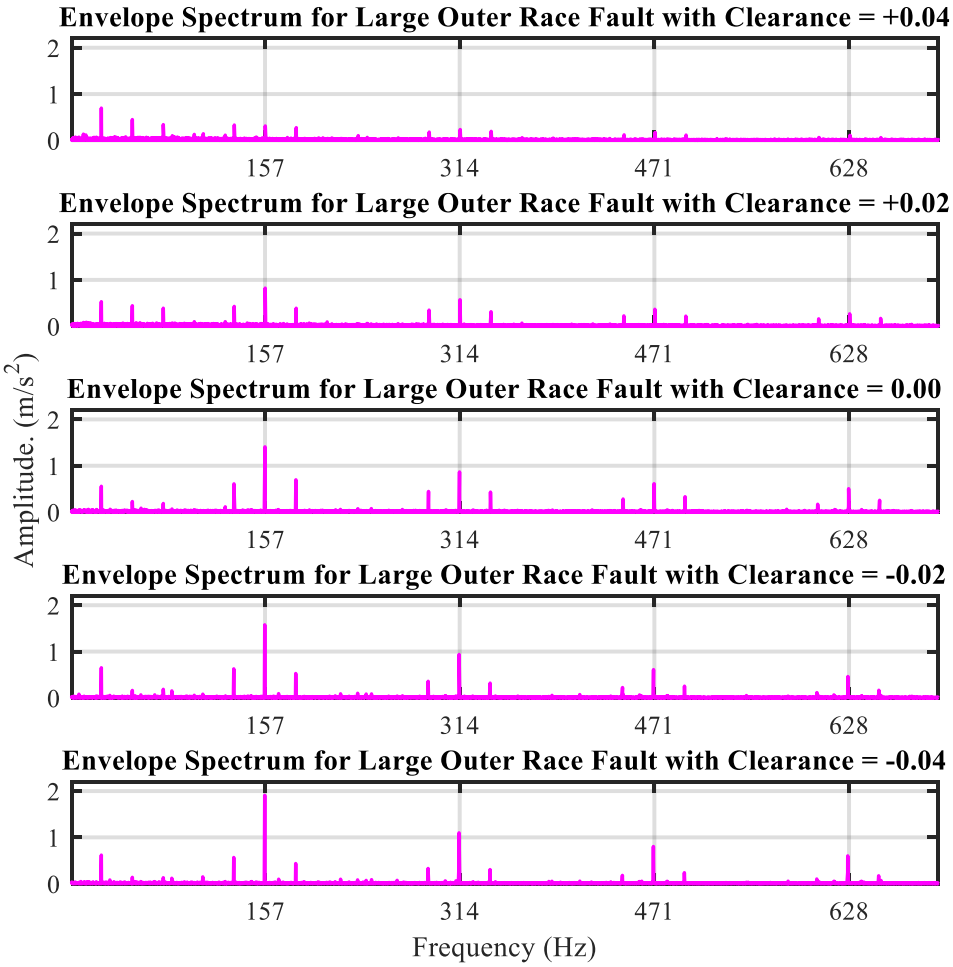

Figure 8-11 Envelope Spectrum Analysis of the Large Outer Race Defect (4mm)

On the other hand, the larger roller fault (magenta traces), shown in Figure 8-13, indicates clearly to the presence of the defects. It can be seen that in the roller case at (134 Hz), the amplitude of the peaks escalated as the severity of the fault increased. Also, the magnitude clearly grows with the decline of the internal clearances. In more details, the frequency of the large roller defect appears with the harmonics at 268 Hz 402 Hz 536 Hz and 670 Hz respectively.

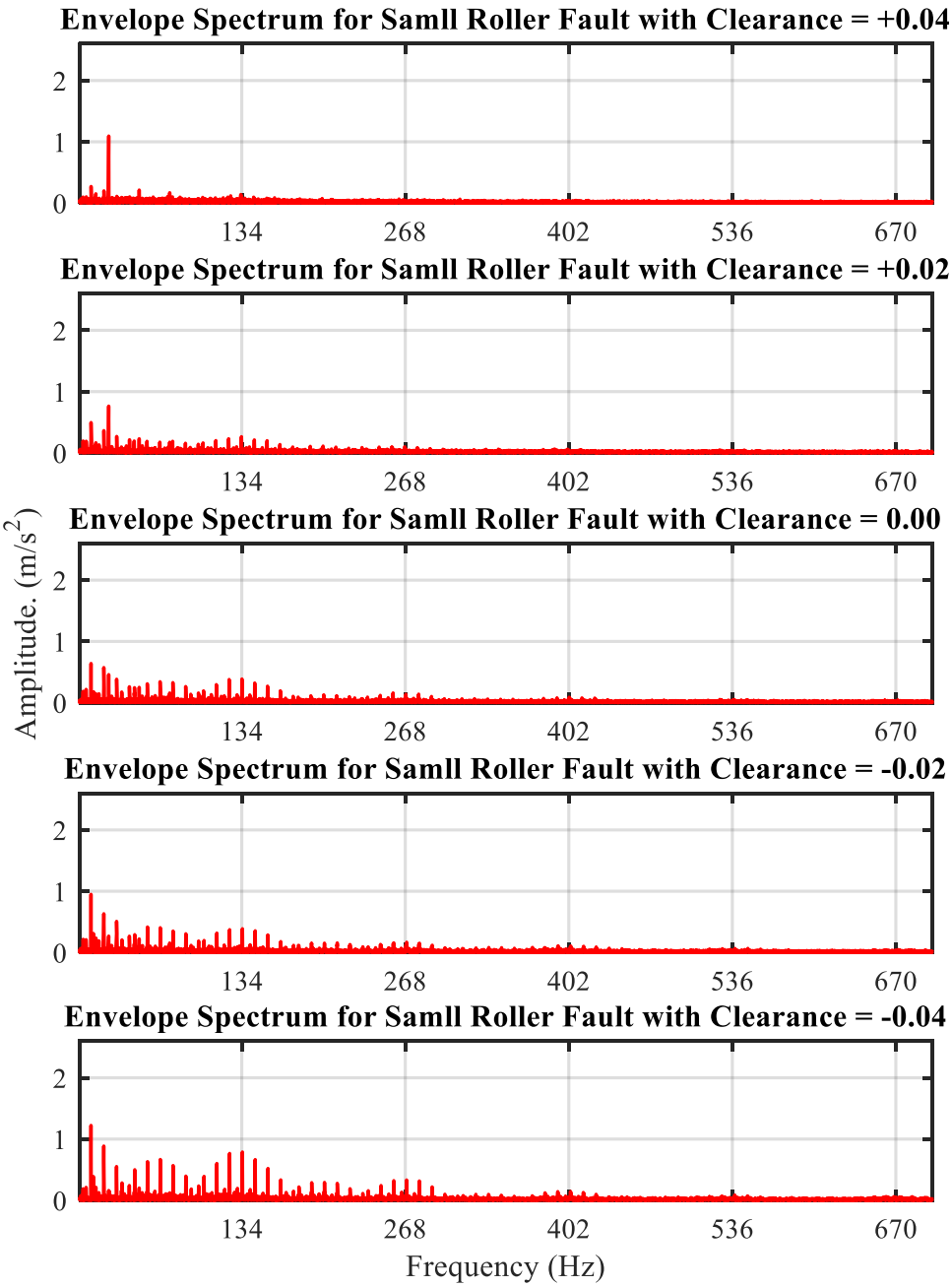

Figure 8-12 Envelope Spectrum Analysis of the Small Roller Defect (2mm)

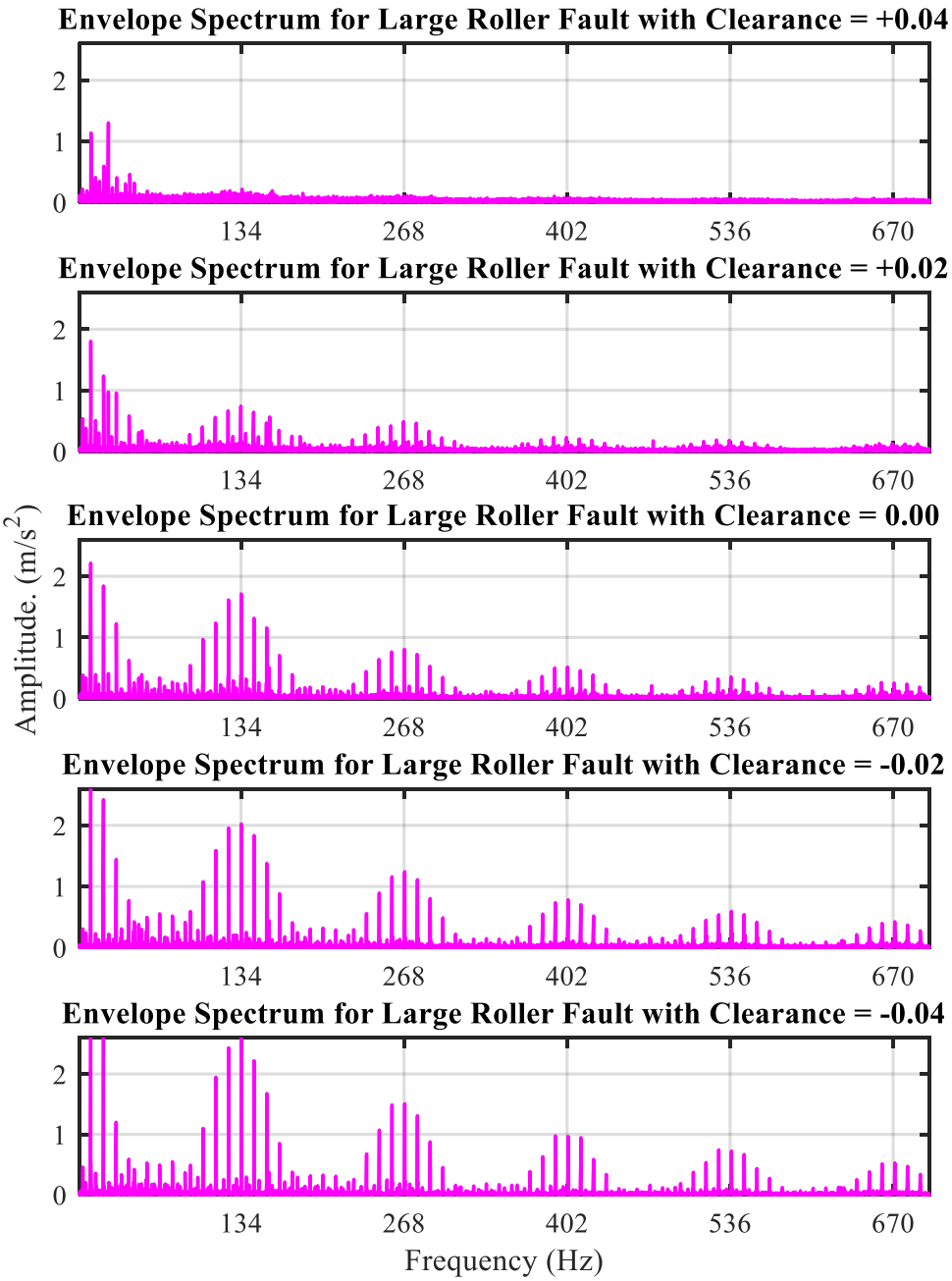

Figure 8-13 Envelope Spectrum Analysis of the Large Roller Defect (4mm)

Although envelope analysis has demonstrated its ability to detect and identify some of the faulty conditions with their harmonics and spaced with the expected sidebands, it cannot detect effectively the faulty conditions in the large clearance cases. Moreover, an important step of applying envelope analysis is to choose the best bandpass filtering, which still not an easy task. Despite the development of advanced techniques to determine the best band such as Fast Kurtogram, these techniques reported to fails in many cases.

8.4 Deviation of characteristic frequencies

Interestingly while exploring the vibration signal collected from defective bearing with an outer race and another bearing with a roller defect, it was found that the fault peaks tend to shift in both cases. This phenomenon motivated us to carry further investigation to identify the reasons behind the deviation of the characteristic frequencies. As it can be seen in Figure 8-14, the outer race fault peaks shifts with the change in the clearance condition. it can be clearly seen that as the clearance declines, the characteristic fault frequency of the outer race declines and the magnitude increase remarkably.

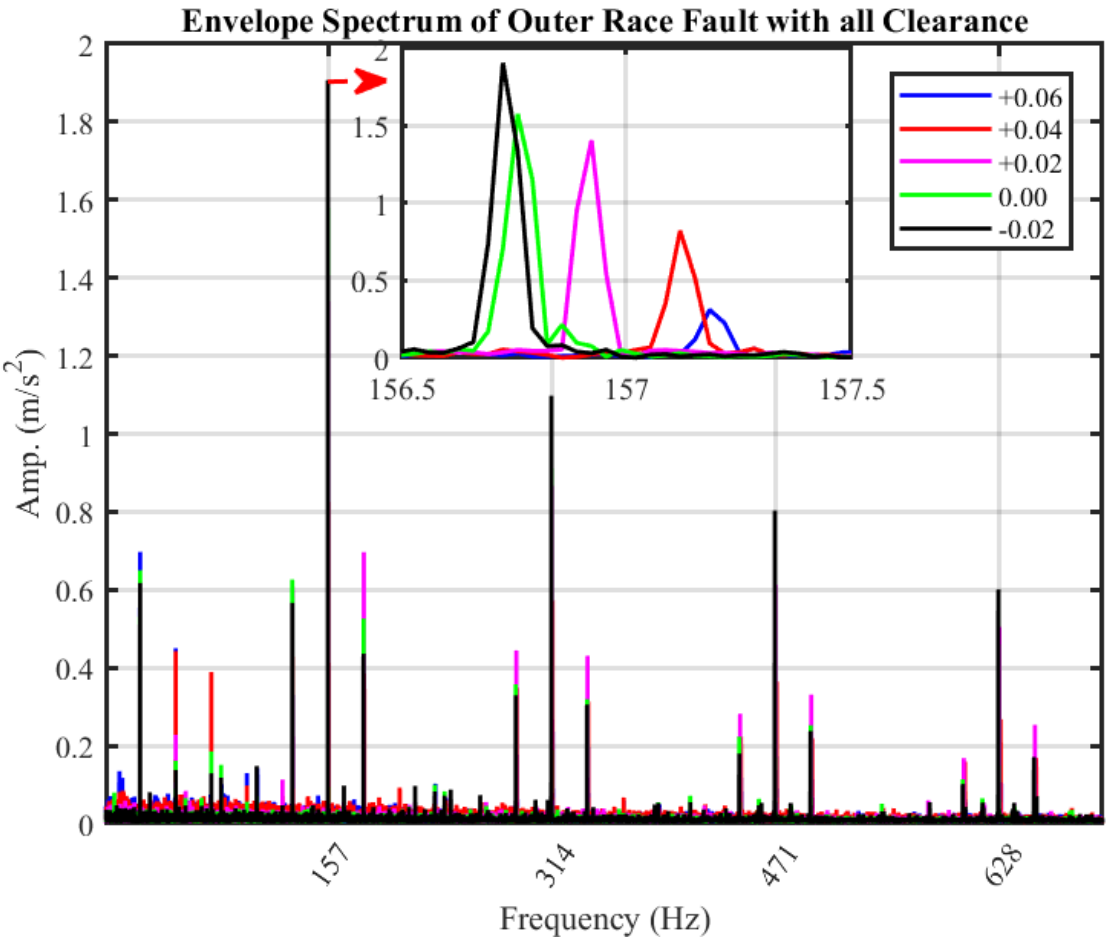

Figure 8-14 Deviation of the Outer Race Characteristic Frequencies

Moreover, in roller fault case, the characteristic frequency tends also to shift, however, in the opposite direction to the outer race fault peak shift. As seen in Figure 8-15, as the clearance decrease the fault peaks frequency of the roller increased and the magnitude also increased.

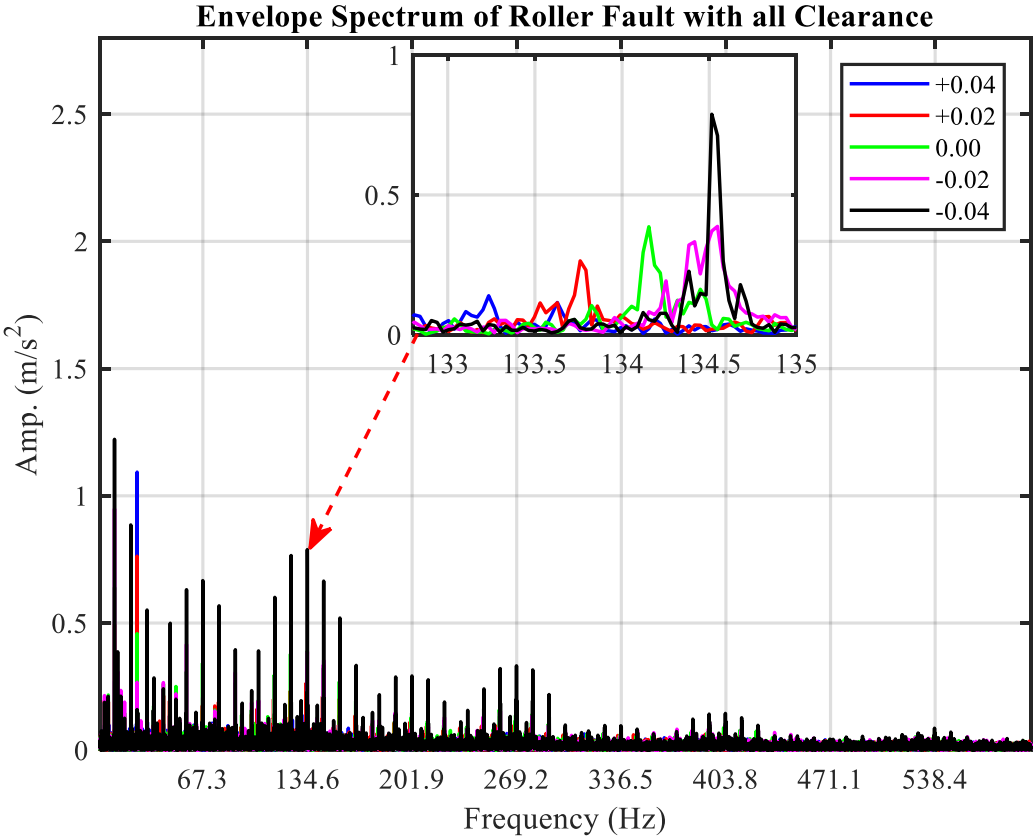

Figure 8-15 Deviation of the Roller Characteristic Frequencies

In reality, tapered bearings, shown in Figure 8-16 (a), are assembled at the location of the application with different clearance setting, the adopted internal clearance will depend on the application they are used for as discussed in chapter three. Moreover, bearing races may undergo inevitable wear in their lifetime [51]. According to Halme and Andersson [281], several mechanisms of wear, such as adhesive, fretting, abrasive, etc., can occur in REBs. Abrasive wear can be classified into two categories, named two-body and three-body abrasive wear. The two-body abrasive wear may be encountered when one of two sliding contact surfaces is harder than the other surfaces causing to remove parts the softer material. Three-body abrasive wear occurs when hard particles penetrate into the sliding surfaces causing material removing from both sliding surfaces. Moreover, the surface texture can be changed due to plastic deformation. The clearance will increase with the occurrence of wear. “Severe wear changes the raceway profile and alters the rolling element profile and diameter, increasing the bearing clearance” [51]. Furthermore, Nguyen-Schäfer [282] pointed out that the internal clearance strongly affects load distribution as well as wear. According to the literature, the effect of internal clearance on REBs’ life has received considerable research interest. However, limited work has been observed that investigated the impact of wear and simulated wear on the condition monitoring of REBs with the assumption of

clearances variance. Rehab et al. [283] investigated vibration amplitude for two ball bearings with two different radial clearances and focused only on the inner race and outer race defects. The study claimed that the vibrations' energy increased with the severity of the outer race defect, while it declined with the severity of the inner race defect. Goerke et al. [284] studied the effects of radial clearance changes on vibration frequencies in double-row self-aligning ball bearings. The experimental study claimed that an increasing clearance leads to a change of the contact angle, and as a result, a linear relationship was found between clearance and vibration frequency. The results showed that, as the clearance decreases, so the contact angle decreases and shrinks the fundamental train frequency. However, this study concluded only the effects of clearance changes on the fundamental train frequency.

Fitzsimmons and Clevenger [285] carried out an experimental study to evaluate the effects of three key parameters of contaminants on the wear of tapered bearing. These parameters are; concentration of contaminants in the lubricant, the size of particles and hardness of the particles. The study stated that tapered bearing not very likely will suffer from the two-body abrasive wear due to the fact that the components of the bearing are made of a similar material with the same hardness level and mechanism. However, they pointed out that tapered bearing can suffer from adhesive wear or three-body abrasive wear. When contaminates penetrate into the bearing, because of the nature of tapered roller, the lubricant is pumped towards the large end of the bearing and the contaminants will be circulated with the lubricant flow, thus, will abrade the surfaces of the bearing. Moreover, the study claimed that wear in the tapered bearing can occur to surfaces that exposed to sliding and rolling contact at the same time. This could be either on the roller ends or the large end rib of the inner race. As excessive wear occurs, it would lead bearing dimensions to change in a form of bearing width reduction, this will change the initial setting of the bearing. Moreover, it can result in an increasing noise levels. The study concluded that excessive wear can occur to tapered bearing if the hardness of the contaminant particles is greater than or equal to the hardness of the bearing material.

In the case of tapered roller bearings (TRBs), however, it has not been reported how an increasing clearance due to mounting setting or unavoidable wear can affect the vibration signature and hence the diagnostic performances.

The analysis shows that the geometric ratio $\beta = d/D$ and contact angle a values will depend on the setting used when mounting the bearing, also will change due to loss of materials during

operation, which will always lead to increased bearing clearances. Eventually, the characteristic frequency will change or vary.

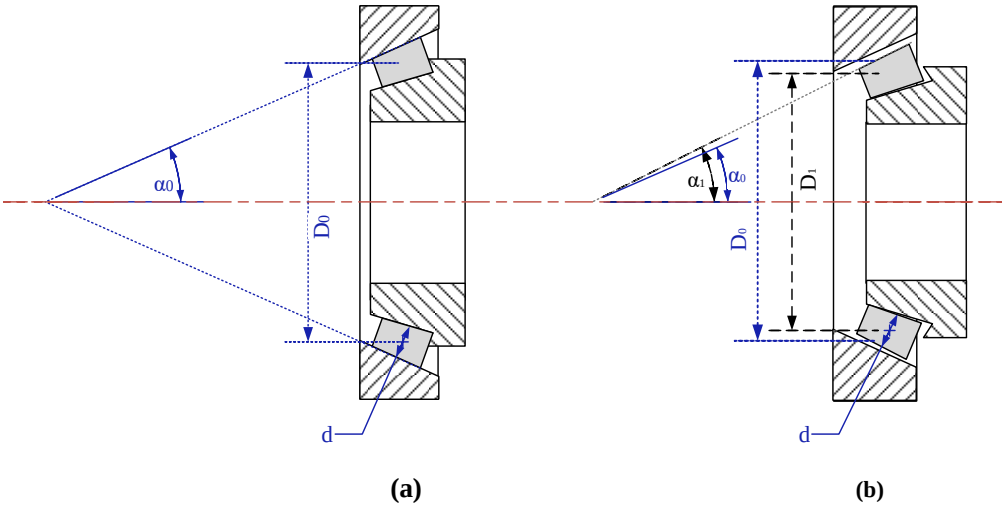

Figure 8-16. (a) Schematic Diagram of a TRB, and (b) a Worn Example

8.4.1 Outer race

In the case of testing a defective bearing with an outer race fault, as depicted in Figure 8-14**,** the fault peaks magnitude increases in each step of minimizing the clearance, however, unlike roller defect frequency behaviour, the outer race fault peaks exhibit a decreasing trend when the internal clearance minimized. The declining trend of the outer race fault frequency f_{BPO} is due to the increase in $\left(\frac{d}{D}\cos a\right)$ value, which according to Equation (8.1), will result in a decrease in the overall value of $\left(1-\left(\frac{d}{D}\cos a\right)\right)$, hence, f_{BPO} will have a declining trend. Interestingly, despite the declining in β value, $(\beta \times \cos a)$ exhibit an increasing trend. As the increase in $\cos a$ (contact angle) more than the declining in β value. It can be concluded that the contact angle has a higher impact on the outer race fault frequency behaviour than the pitch diameter does.

8.4.2 Roller

A defective bearing with roller fault was tested while the internal clearance has been minimized five steps by a value of 0.02 from (+0.04) to (-0.04) in each case. It was found that f_{BS} has increased in each step. Also, $(\beta = d/D)$ value will become smaller and $(\cos a)$ will become larger. As demonstrated in Table 8-3, according to (8.3), although $\left(1-\left(\frac{d}{D}\cos a\right)^2\right)$ will decline, however, the increases of $(D/2d)$ having more influence which makes f_{BS} increases. This is because the roller will slide to the up towards the upper edge of the inner race, the parameter D (pitch diameter) will increase and the contact angle will decrease when clearance decreases, as illustrated in

Figure 8-16 (b). It can be concluded that the change in the pitch diameter value affects the roller characteristic frequency, therefore, the impact of the change in the pitch diameter at f_{BS} is higher than the impact of the change in the contact angle $\cos a$.

There is a clear change in the frequency as the internal clearance changes in the range specified by bearing manufacturers (see Clearance setting range figure).

Table 8-3 Deviation of characteristic frequencies

Clearance (mm)	Outer race fault f_{BPO} (Hz)	Roller fault f_{BS} (Hz)	α	cos(α)	$\beta = d/D$	$\beta\cos(\alpha)$
+0.040	**158.14**	**132.45**	32.05º	0.8475	0.1840	0.1560
+0.020	**157.35**	**133.17**	28.81º	0.8762	0.1828	0.1602
0.0	**157.12**	**133.60**	27.61º	0.8861	0.1821	0.1614
-0.020	**157.05**	**133.89**	27.10º	0.8902	0.1817	0.1617
-0.040	**157.05**	**134.12**	26.91º	0.8916	0.1813	0.1617

8.5 Vibration Response and Spectral Lines Analysis

Calculating the exact dynamic behaviour of a machine is not an easy task especially when the machines are built of many different parts. Thus, a simple model is normally used to represent the machine structure [64]. Intensive theoretical vibration models have been developed to characterise the vibration mechanism and the effects of the parameters' variation such as transmission path and loading, including single and multiple localised defects [51]. Su and Sheen [286] developed a model to characterise the vibrations of a bearing subjected to loading variation and various transmission path conditions with arbitrarily located defects. The developed model specifies the periodic characteristics of loading variations and also the effects of the transmission path on the vibration of the contact energy due to the defects of different bearing components. The magnitude of the contact energy is generally affected by loading related to the unbalanced shaft, misalignment, radial or axial load, manufacturing errors and preload. Table 8-4 illustrates the reasons for periodic characteristics and their effects on each of the bearing defects where f_s is the shaft frequency, f_c is the cage frequency and f_{BS} is the roller spin frequency.

Table 8-4. The periodicity due to various loading and transmission path influences [287]

Cause of periodicities	Defect type		
	Outer race (f_{BPO})	Inner race (f_{BPI})	Rolling element (f_{BS})
Stationary loading	-	f_s	f_c
Loading due to shaft unbalance	f_s	-	$f_s - f_c$
Loading due to roller diameter errors	f_c	$f_s - f_c$	-
Transmission path	-	f_s	f_c and $2f_{BS}$

8.5.1 Outer Race Defect Analysis

Figure 8-17, shows the envelope spectrum analysis of the bearing with outer race defect, the fault peak frequency of outer race (f_{BPO}) in the first harmonic (red dot trace) and the consequent harmonics are spaced with the shaft frequency f_s (25 Hz) as listed in Table 8-4. The sidebands present due to the unbalanced shaft as the shaft used in this test is found to be unbalanced by 0.09 mm when calibrated. The spectral lines with green dot traces represent the shaft frequency f_s and its harmonics. The close spectral line is the sidebands of the harmonics overlapped in some regions.

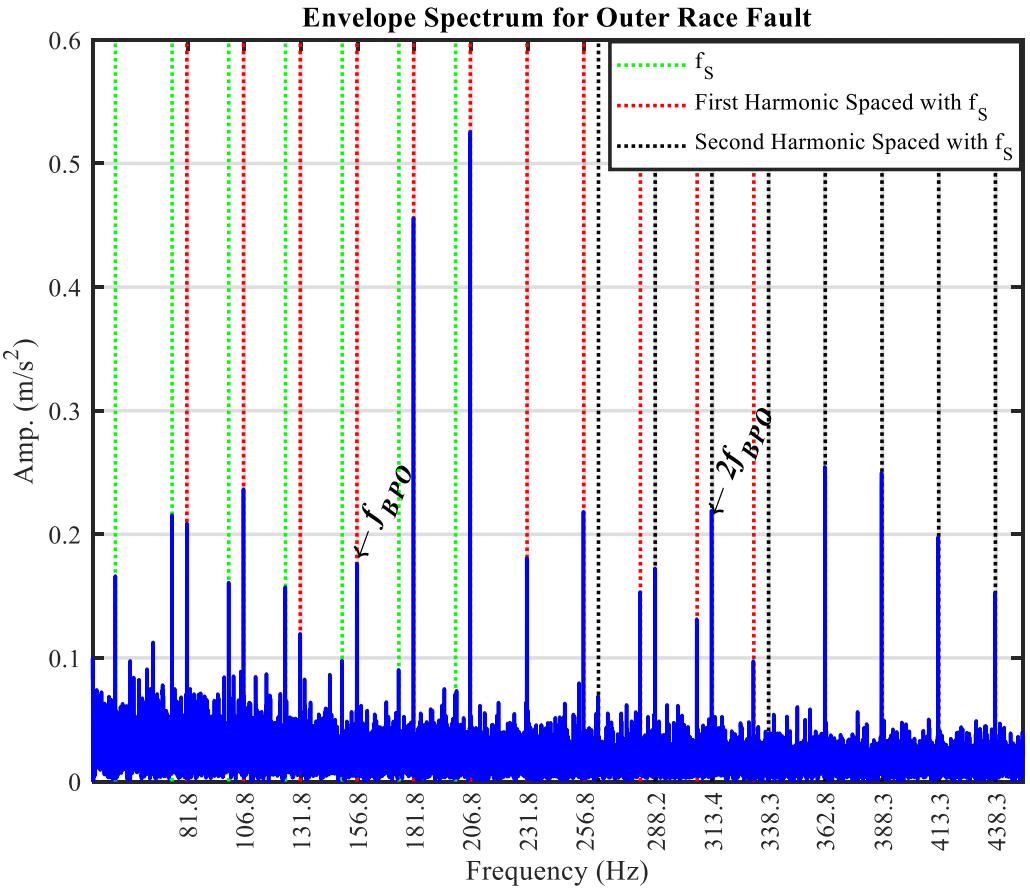

Figure 8-17. Envelope Analysis of Outer Race Fault with Shaft Periodicity Effect

8.5.2 Roller Defect Analysis

The spectra of enveloped vibration signals of roller fault bearings are illustrated in Figure 8-18. The spin frequency f_{BS} (red trace) and roller fault frequency $2f_{BS}$ (black trace) spaced with the cage frequency f_c due to transmission path effect as illustrated in Table 8-4. Interestingly, in this study, it was found that the cage frequency f_c (cyan trace) exists with its harmonics and can be seen

clearly. This because the defective roller when spins it does get a line contact with the cage of bearing and appears with at the cage characteristic frequency.

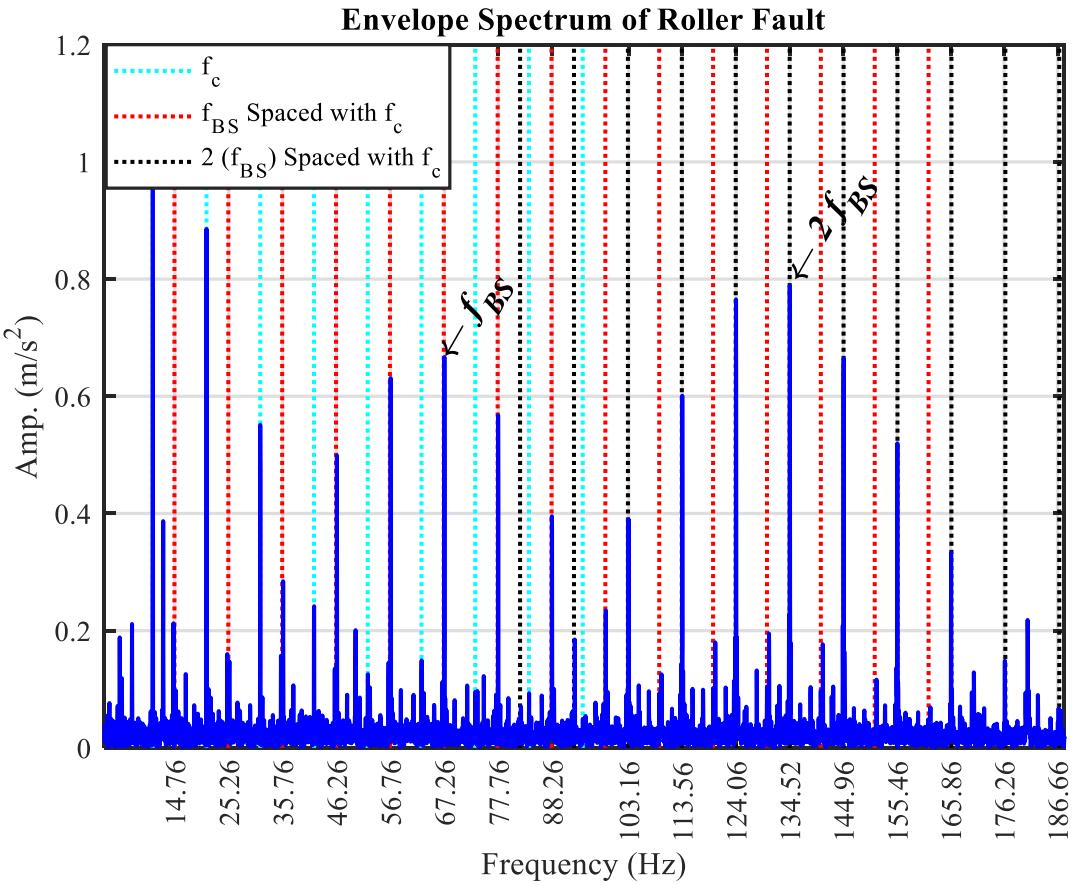

Figure 8-18. Envelope Analysis of Roller Defect with Cage Frequency Effects

The unbalanced shaft effect is illustrated in Figure 8-19, it can be seen that the f_{BS} roller spin frequency is spaced with $f_s - f_c$ (magenta trace) which is 14.5 Hz in this operating condition. This periodicity is consistent with the summary listed in Table 8-4. After the previous analysis and discussion for the three bearings, the results demonstrate that, for accurate severity diagnosis, the effect of the wear evolution for TRBs needs to be taken into account.

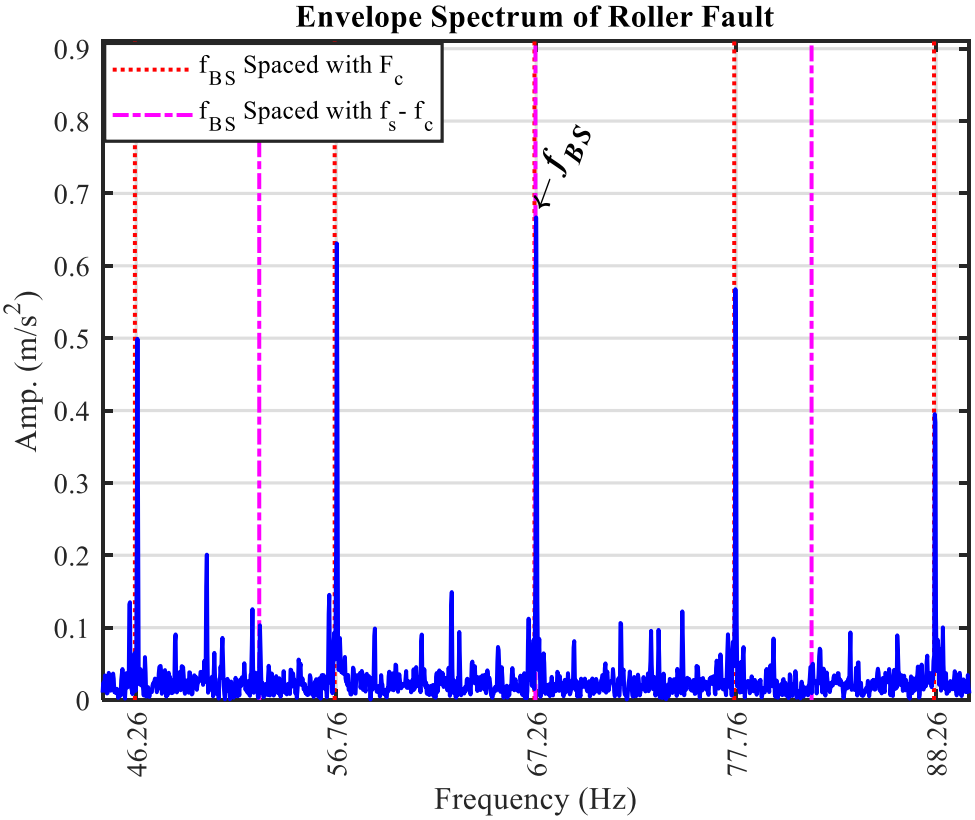

Figure 8-19 Envelope Analysis of Roller Defect with Unbalanced Shaft Effect

8.6 Internal Clearance Estimation Based Low-Frequency Band Analysis

The vibration response to internal clearance changes in the low-frequency band is investigated in this chapter. The faulty peaks were extracted from a defective bearing with outer race and from another defective bearing with roller fault. The results are shown in Figure 8-20 (a) and (b) respectively. The amplitude of the faulty peaks increased with the declining of internal clearances.

Figure 8-21 (a) shows the impact of the change in clearance at the vibrations in a wide band. It can be seen that RMS of whole band vibration for outer race fault with all clearance cases show no clear trend with the declining of the internal clearance of bearings, neither the RMS of data with roller fault does. On the other hand, in a low band frequency of 1:1000 Hz, the RMS values of the vibration signals show a clear declining trend for all cases including baseline data as seen in (b). The amplitude remarkably declined with the decrease of the internal clearances. Whilst interestingly, the fault peaks amplitude of both Outer race and roller increased remarkably with the declining of the internal clearance as depicted in Figure 8-20 (a) and (b).

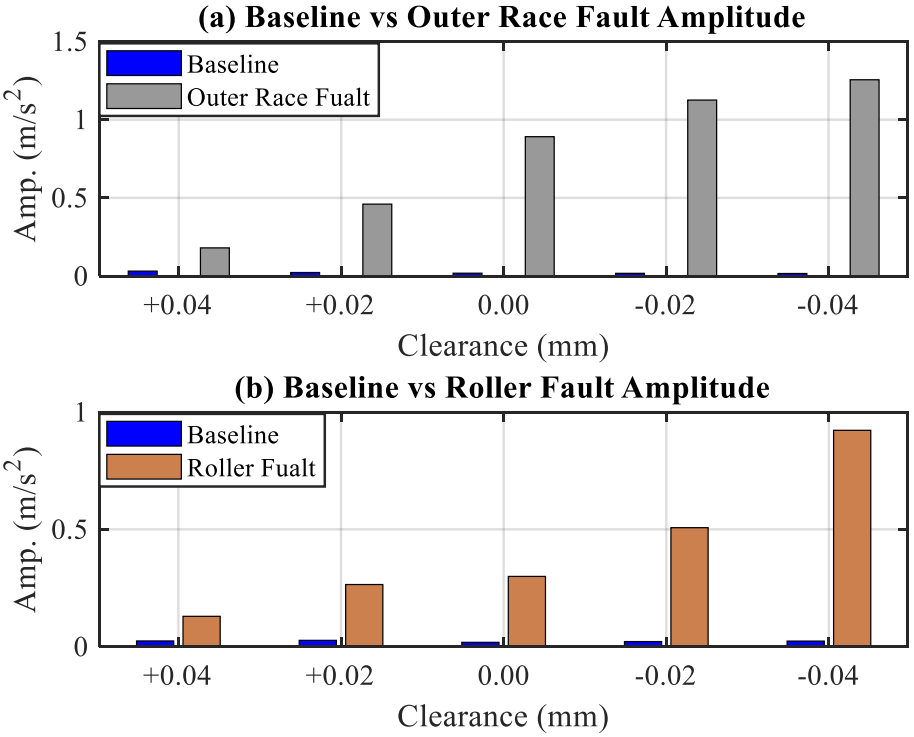

Figure 8-20 Fault Peak Amplitude vs. Clearance

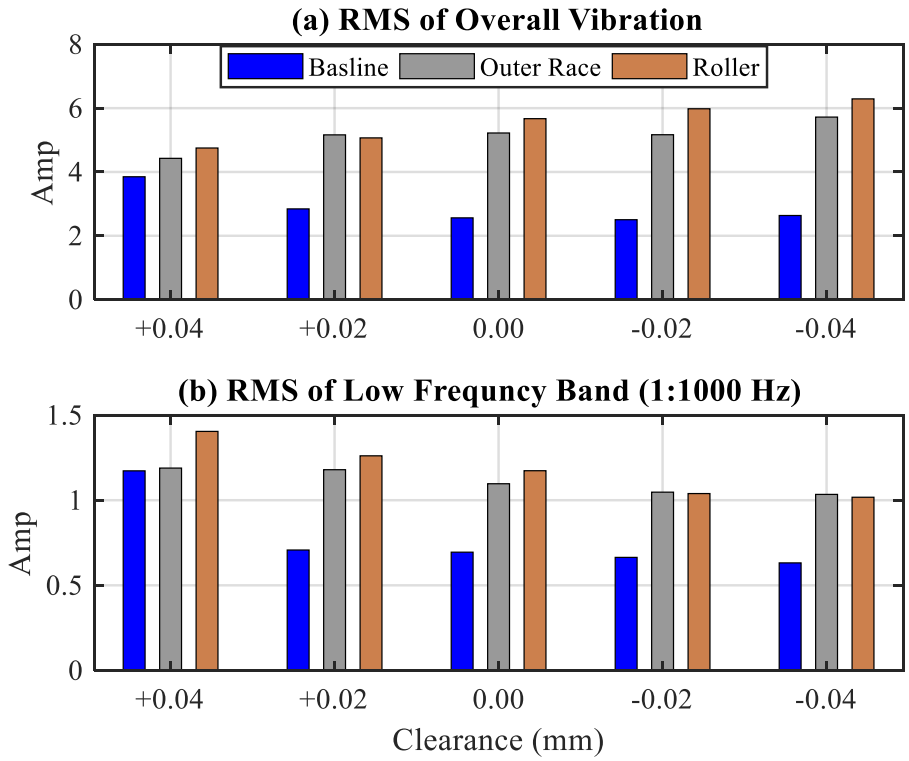

Figure 8-21 Low-Frequency Vibration for Clearance Estimation

8.7 Summary

In this chapter, for theoretical fault frequency calculation, two sets of data are used to estimate the missing key parameters: pitch diameter and contact angle. Based on measured data with artificial faults on outer race and a roller, the shaft frequency f_s, and f_{BPO}, and $2f_{BS}$ were identified from the experiment results. Moreover, alongside baseline data is used as a reference, sets with different internal clearances data for outer race, the roller with two fault severities were analysed. The analysis started by exploring the time and frequency domain analysis, then envelope spectrum analysis also adopted in this chapter. Despite the slight change of the signal amplitude in the time domain and time domain statistical analysis, it is still not adequate to identify the defects occurred in the two faulty cases. Moreover, from the results obtained using Fast Fourier Transform, it can be concluded that it is very unlikely to detect the peaks belong to the calculated faults at 156 Hz and 134 Hz with their harmonics for the outer race defects and the roller defects. It can be seen that there is no significant increase in the amplitudes at the defects frequencies compared with the no-fault case.

The use of well-known envelope analysis shows the ability of the technique in detecting defects with large severity with their harmonics and spaced with the expected sidebands, also, reasonable results for small faults with high preloaded conditions. However, when clearance become larger within the specified range provided by the manufacturer, the envelope does not provide clear results for the small fault severity especially for the internal clearance of (+0.04). Moreover, an important step of applying envelope analysis is to choose the best bandpass filtering, which still not an easy task. Despite the development of advanced techniques to determine the best band such as Fast Kurtogram, these techniques reported to fails in many cases.

Furthermore, to clarify the impact of mounting clearance settings or wear evolution on bearing fault diagnostics, an adjustable clearance mechanism method was adopted in the test to track the variation in vibration signatures. Two feature parameters are used to diagnose the internal clearance variation: the change in magnitude of characteristic frequencies which is usually used to diagnose fault severity, and the deviation of characteristic frequencies, which is suggested in this research to estimate internal clearance changes and wear evolution. The experimental results show that the defective magnitude declines remarkably with the increase of the clearance (wear evolution) for different fault cases. Moreover, the outer race fault shows a declining trend as the preload value increases, on the other hand, the deviation for roller faults exhibits an increasing trend as the preload value increases. Especially, these frequency deviations from nominal characteristic can be an

indication of wear evolution. Further investigation will be carried out to study the clearance estimation by analysing the low-frequency band; also, the influence of clearance on geometrical information of the TRB and, consequently, on the fault features will be studied.

CHAPTER NINE

AUTOMATED DATA ANALYSIS USING COMPONENTIAL CODING NEURAL NETWORK

In order to evaluate the performance of CCNN, the detail of its implementation is presented based on different types of data in this Chapter. Firstly, it is applied to simulated datasets to show its performance in characterising different types of signals met in condition monitoring. Then it is assessed based on experimental data from different fault types and severity.

9.1 Introduction

This chapter presents the evaluation of CCNN using simulation data and presents the results obtained from applying CCNN to real data collected from the test bearings with discussion, In order to efficiently and effectively use the CCNN for condition monitoring, the detailed rules for the selection of CCNN parameters is studied based on typical condition monitoring data features. Vibration signal from condition monitoring is one-dimensional, it fluctuates in both time and frequency domains, and often only small variations occur in the signal from condition to another.

9.2 CCNN Evaluation Using Simulated Data

This section focuses on the selection of parameter values when applying CCNN to data. The novelty detection is inferred when the network is trained with a baseline signal, the novelty or the changes of the faulty signal from the baseline training signal can be found through the reconstruction error of the new signal using the trained neural network. The reconstruction error can be calculated by:

$$E^u = \langle(|x - \hat{x}|^2)\rangle \tag{9.1}$$

The reconstruction error E^u is obtained from unseen data-set during the validation stage. Thus, the magnitude of the error is the differences between the new data and the training data. This averaged error then gives a general measurement of the novelty. This method is called Reconstruction Error Based Detection Model (REM). REM is computed using ADI as shown in (9.2). ADI will have a value close to zero (ADI<0.1) if the monitored data-set is healthy. However, ADI should have a much larger value than zero (ADI>0.1) if the monitored dataset is anomaly compared to the unseen dataset used in the calibration stage during training because the accuracy of reconstruction should be much less than the reconstruction accuracy of non-anomaly dataset. The reason behind using

unseen dataset R_u is to improve the generalization of the derived data model from the training process to avoid the problem of run-into overfitting to the dataset used in the training.

$$ADI = \frac{R_m - R_u}{R_u} \tag{9.2}$$

9.2.1 Test Procedure and CCNN Parameters

In order to ensure the stability of the CCNN and illustrate the capability of the method in anomaly detection, network firstly needs to be optimised and then the network will be tested in terms of performance evaluation using simulated signals. These steps will be discussed in details below.

Firstly, the network will be optimized using synthetically created signals (simulated signals) based on the features of condition monitoring data. With the aiming of learning the training experiences, the simulated signal is formed with simple structures. Knowing the structure of the data allows the characteristics of the parameters and detection performances to be considered more easily.

Secondly, the performance evaluation of CCNN using simulation data will be carried out using simulated signals. Two types of signals will be used periodic and impact signals. These simulated signals will consist of the principal frequency with modulated signal frequencies and random noise to simulate REBs condition monitoring data.
Using simulated signals, where the structure of the signals is known, gives us a better understanding in the followings:

- To determine the optimal training parameters.
- To explore the impact of the noise on learning the features.
- To explore the capability of CCNN in the detection of signal amplitude changes.
- The detection of changes in the frequency components.
- The detection of changes in the signal structure.

9.2.2 Non-adaptive Network Parameters Optimization

Non-adaptive network parameters are optimised using trial and error procedure in order to get the optimal parameters for anomaly detection. A simple periodic signal was generated to be used in the optimization process. To simulate condition monitoring data for optimizing the network, the simple simulated signal contains a principal frequency component and a random portion of noise as seen in (9.3).

$$x(t) = (A_1 \cos(2\pi f_z t)) + n(t) \quad (9.3)$$

where f_z is the principle frequency with amplitude of A_1, $n(t)$ is a random noise.

Weight Vector

Simulations were carried out to evaluate the influence of the number of the weight vectors and the size of each weight vector on the data reconstructions and anomaly detection capabilities.

Weight Vector Size

Simulations tests show that the size of the weight vector has an impact on the ADI amplitudes. In other words, when the size of each weight vectors increases to be close to the number of data points, the network gets better anomaly detection capability. Therefore, In the simulation test, the frequency is 500 Hz while the sampling rate is 15KHz, so the size of each weight vector is set to be 32 (15000/500=30) to cover one period of the principle frequency component.

Weight Vector Number

For the number of weight vectors, the optimization process shows that the more of the weight vector we create, the more details about the signal structure the network gets. Thus, in simulation, the number of weight vectors is set to be 6 to represent the expected frequency components from generated 5 principle frequencies and the random noise (500Hz, noise).

Network Threshold Value

Simulations have shown that as the value of the threshold increases, the training errors generally increase. The largest value of ADI was found around the threshold of 0.40. Thus, the threshold value is set to 0.40 in the simulation test.

Softness Value

Varying the softness value has led to more dramatic influence on the results, and stronger influence on the training error. The best results were obtained when the softness set to between 0.16 – 0.2.

After identifying the values of non-adaptive parameters, in this test two signal structures will be considered in the evaluation of the performance of the CCNN. Firstly, the sensitivity of the CCNN to the changes of the signal structure using simple periodic signal, and secondly, the discrimination capability over additive noise levels using impact signals.

9.2.3 Anomaly Detection Results for Periodic Signal

The capability of CCNN in the detection of frequency variation is tested. A simple periodic signal was generated from a simple cosine wave signal with a small portion of noise as shown in (9.3) where $A1$ is signal amplitude and set to 2, f_z represents frequency and set to 500 Hz and t is time set to be one second, $n(t)$ random noise and set to 0.1 and the sampling frequency set as 15 kHz.

The network parameters and training parameters were set to the optimal values found in the network optimization. The network parameters used in these experiments are shown in Table 9-1. This configuration allows the study of network training characteristics and detection capability of the CCNN.

Table 9-1 Network parameters

Item	Value
Weight vectors numbers	6
Weight Vector dimension	32
Threshold	0.40
learning rate	0.05
sigma	0.2
Batch number	Data length /WV dimension
Iteration	200

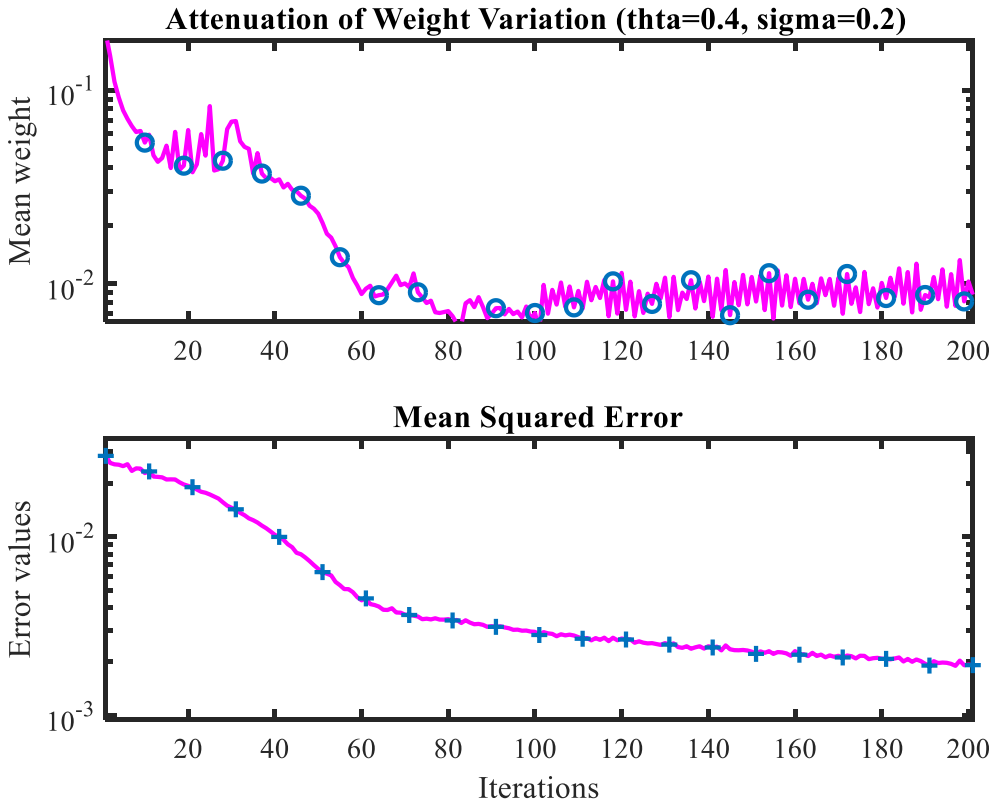

Figure 9-1 Iterations Number

Figure 9-1 shows that after 80 iterations, both the weight variation and MSE were nearly stabilised and become almost flat with regards to the remaining training iterations. The weight vectors obtained from the training stage are shown in Figure 9-2. It can be seen that the weight vectors reflect the signal components, as described in Figure 9-3, the weight vectors match both the deterministic and the random components.

From the training result, as described in Figure 9-4, the signal reconstruction test shows that the model learned the required features that were needed to be able to reconstruct the very similar to the original signal using the learnt model.

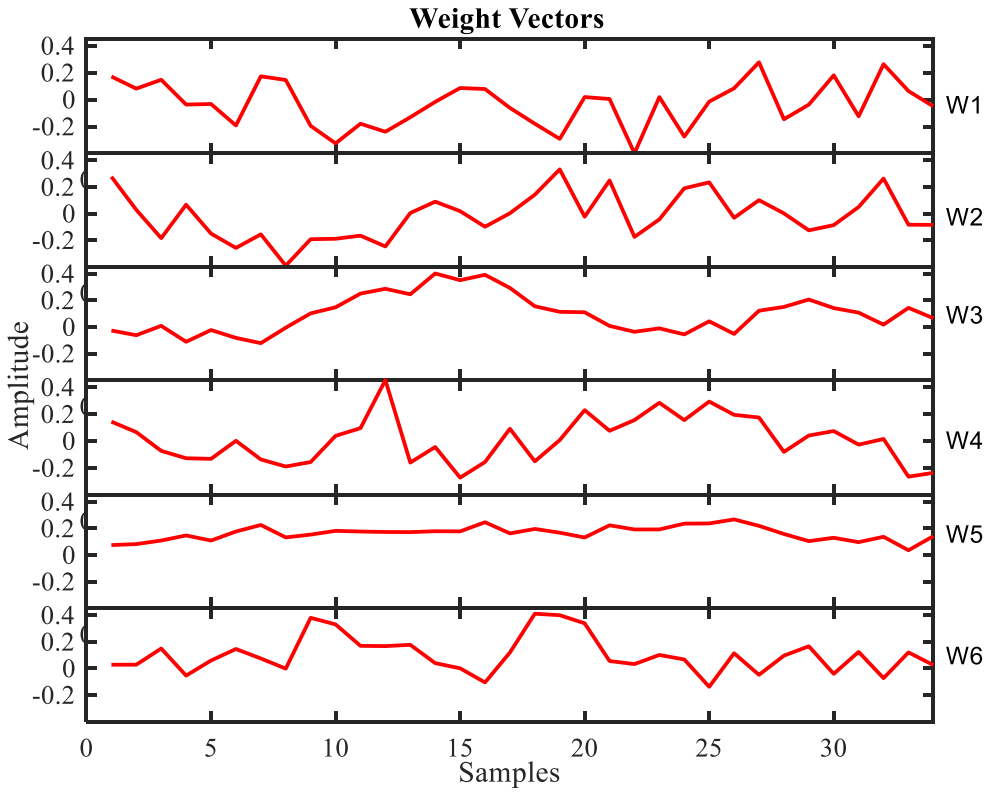

Figure 9-2 Weight Vectors Obtained from Simulated Periodic Signal

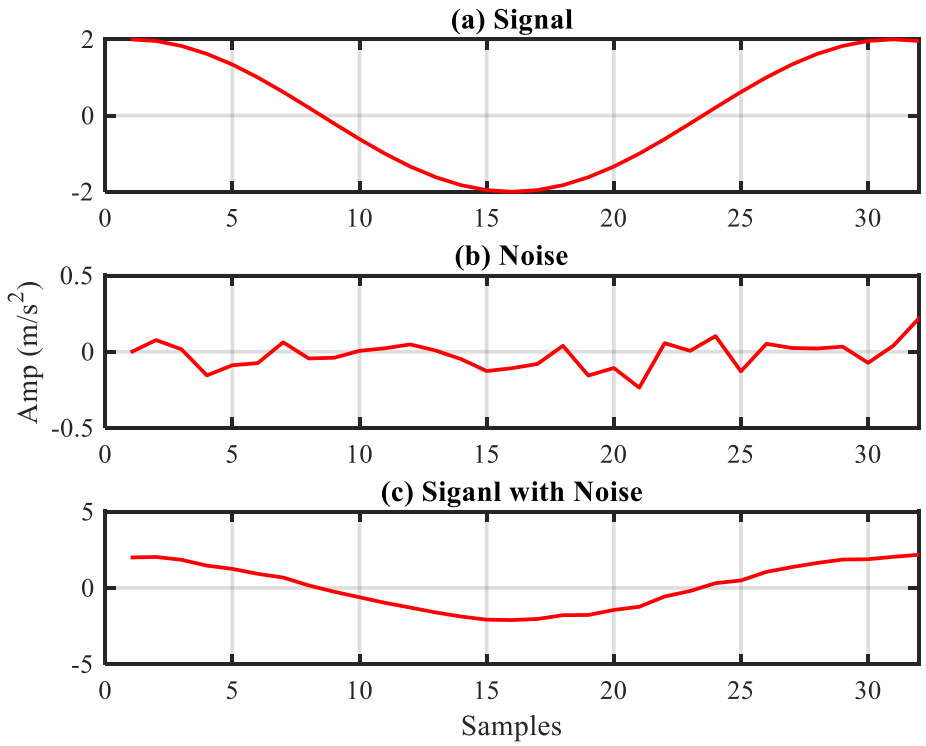

Figure 9-3 Simulated Periodic Signal

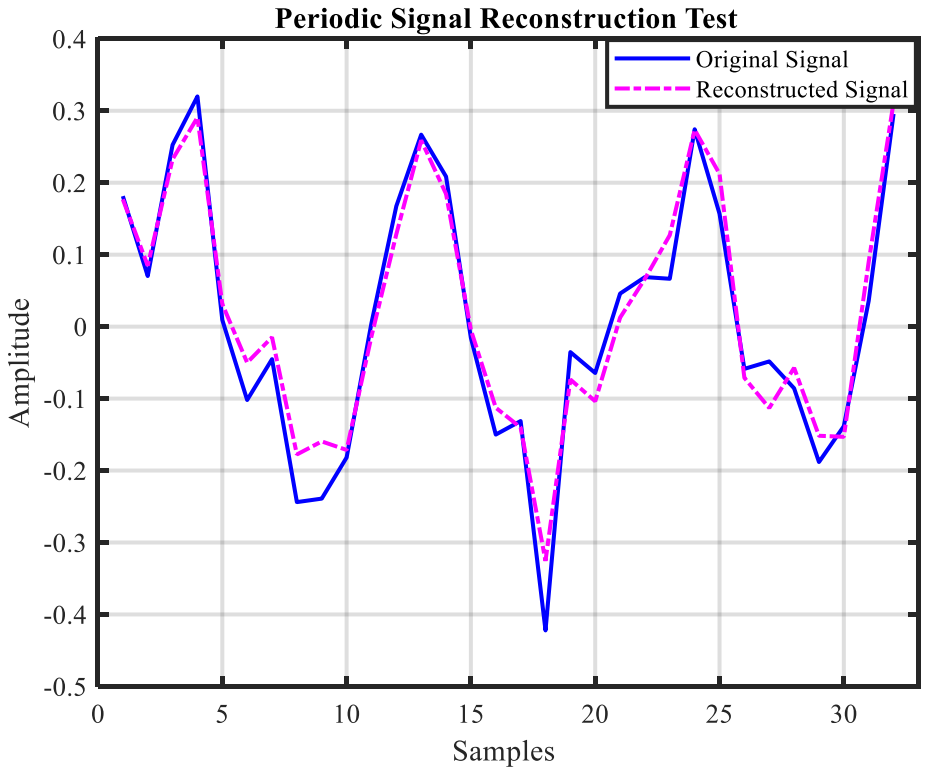

Figure 9-4 Reconstruction Test

The frequency variations are inspected. A number of 5 test signals were formed with a variation in 5 Hz steps around the frequency values of the training signal. The same noise level was used for the test signals and that of training signals.

As seen in Figure 9-5, the results show clearly that the CCNN is very sensitive to changes in signals structure (frequency). Both MSE and ADI show that the CCNN is capable of detecting the changes in the frequencies with regard to the training signal.

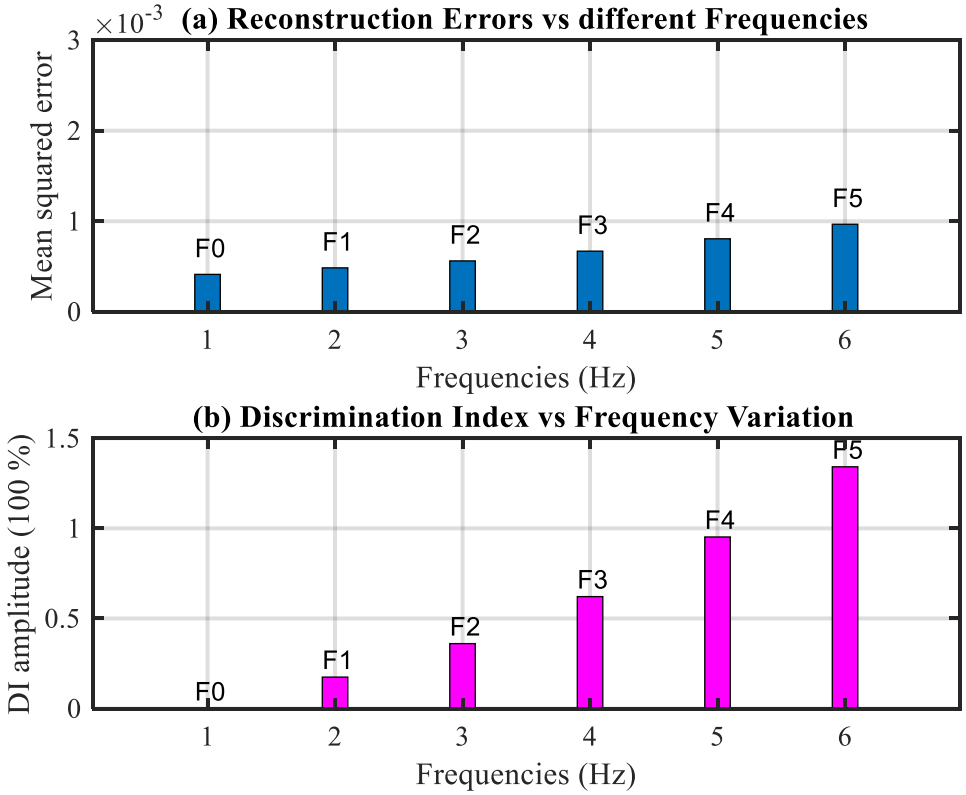

Figure 9-5 Performance of Frequency Variation Using Periodic Signal Detection

9.2.4 Anomaly Detection Results for Impact Signals

This section presents the evaluation of CCNN using simulation data, so simulation signals were generated and different levels of noise were added to the signal. The anomaly detection will be carried out using CCNN.

9.2.4.1 Bearing fault model

Several models have been developed to simulate the vibration signals since the first attempt to model a nonstationary vibration signal by McFeddan in [175]. However, in reality, the combination of impacts produced by a defect on the surface of a bearing and stochastic components have periodically time-varying ensemble statistics. This variation was found to be caused by the slippage (sliding) between the components of the bearing, as a results signals produced are not strictly periodic but considered as a cyclostationary signal. The first attempt perhaps to simulate the bearing as a cyclostationary signal was introduced by McCormick and Nandi in [92]. Since then a number of vibration models have been presented. In this research the model adopted, described in Equation (9.4), is the one that takes into account, the periodicity nature and the random variation in the spaces between adjacent pulses, also it takes into account the modulation might occur due to effects

described in chapter Eight Table 8-4. It was presented by Randall et al. in [181] as a statistical vibration model produced by a single defect as shown in (9.4).

$$x(t) = \sum_{j} A_j y\left(t - jT - \tau_j\right) + n(t) \tag{9.4}$$

A_j can be a possible modulator, T stands for average time between two adjacent impacts and it is derived by $T = 1/f_r$ where f_r is the fault frequency, jT is the jth time of single impact occurrence, τ_j is the randomness of the time lag between the impacts, $n(t)$is the additive white noise comes from other vibrations in the system. $y(t)$ is considered as the impulse response function which can be simplified as an exponential damping cosinusoidal signal. However, later on, Randall and Antoni in [288] pointed out that the randomness of fault pulse spacing variation was modelled incorrectly as shown in (9.5)

$$T_j = jT + \delta T_j \tag{9.5}$$

Where jT is the jth time of single impact occurrence and δT_j a random variable around each period, this model version identifies the spacing around each period using a jitter, which is defined usually by a percentage value. However, the model was refined and enhanced later as shown in (9.6).

$$\Delta T_j = T_{j+1} - T_j \tag{9.6}$$

The random variable ΔT_j is the space itself, the difference can be highlighted between the former and the enhanced model in the way of defining the randomness of space, clearly, in the former model is defined by the jitter δT_j, whilst, in the later model is defined ΔT_j which is the interval time itself between every two adjacent impacts. This makes the later model simulates non-stationary signals more realistically.

The enhanced model is considered in this research and the impulse response function $y(t)$ is derived as in (9.7).

$$y(t) = \begin{cases} e^{-at}\cos(2\pi f_d t)\,; & t > 0 \\ 0\,; & otherwise \end{cases} \tag{9.7}$$

Where f_d is the resonance frequency and set as [3 KHz+5 KHz+8 KHz], and a is the damping ratio and set as 0.05. The time of each impact is shorter when compared to the resonance frequency.

In order to simulate a vibration signal similar to the signal derived from experimental case studies, two forms of signals are generated. For outer race fault, spaced with shaft frequency, A_j, can be simplified as:

$$A_j = A_1 \cos(2\pi f_k (jT + t_j) \tag{9.8}$$

Where A_1 considers the amplitude of the modulator, f_k is the shaft frequency, $jT + t_j$ represents the specific time of the jth impact. Whilst, for roller fault, f_k is set as the cage frequency f_c and $f_r - f_c$ as described in Table 8-4.

Test signals were formed using the adopted bearing fault model to replicate a common form of the condition monitoring signal. Variations are made to the amplitude of transient, also the variation in transients number is set to be based on the modelled fault in the simulations. The carrier frequency was set to the same as that found usually in measured monitoring data due to the resonance of the system. The distribution of the impacts was set to random within the simulations based on the used model. The variations in signal amplitude and impact numbers can result in a change in the SNR value and a change in the structure of signals.

9.2.4.2 Training stage using simulated impact signal

The generated signal for roller defect is based on the given parameters: the amplitude of the impulses A_1 set to be 0.9, f_k is equal to 25 Hz, sampling rate F_s equal to 50 KHz, fault characteristic frequencies f_r are calculated as described in Table 8-2 and is set to be 133.7.

The generated signal, as it can be seen in Figure 9-6 (a), shows the random phase error induced by slippage for the three models. It shows clearly that error-free model has no variation in time (y-axis) whilst model 1 (red trace) caused a little fixed percentage of variation, however, model 2 (black trace) showed a clear random behave of variations.

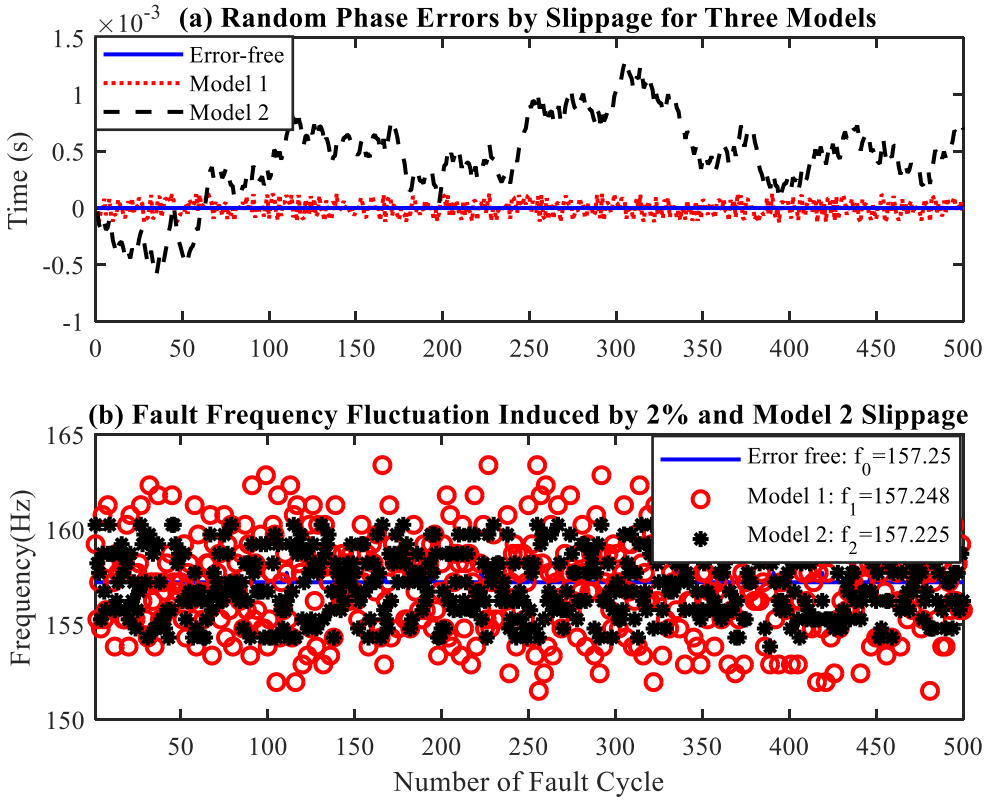

Figure 9-6 Random Phase Error and Frequency Fluctuation Induced by Slippage

The time domain of the generated signal is shown in Figure 9-7 and the envelope spectrum is shown in Figure 9-8. In both figures, (a) shows the slippage free form whilst, (b) illustrates signal generated using model 2 described above in (9.5) . In (c), the generated signal using the enhanced model 3 is depicted, the spaces between impacts are dominated by enhancement described in (9.6).

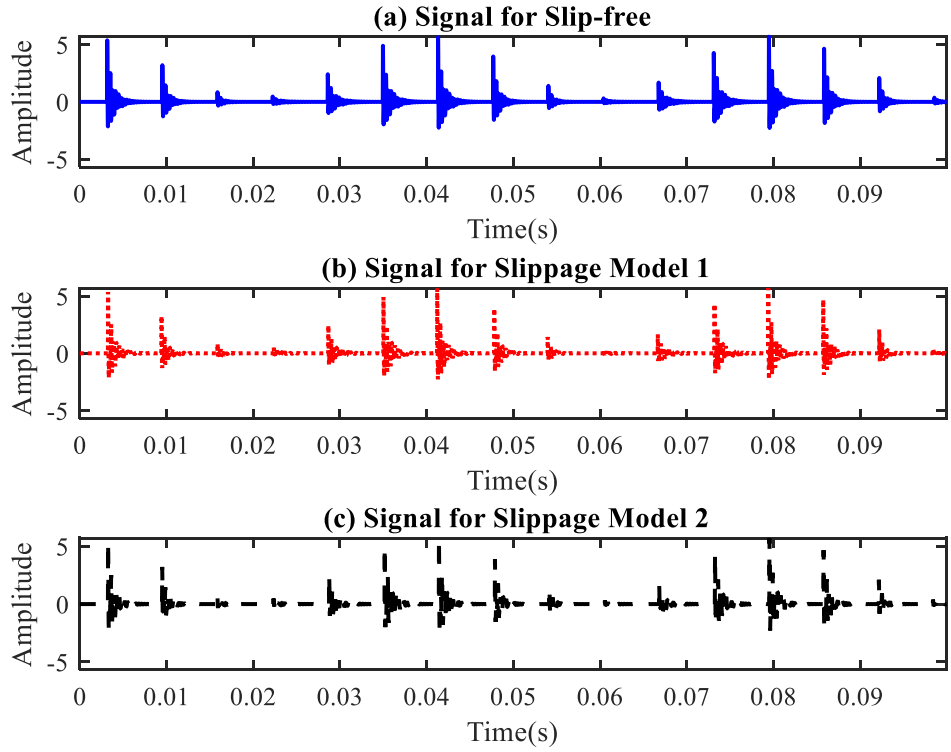

Figure 9-7 Time Domain of the Generated Signal Using Three Models

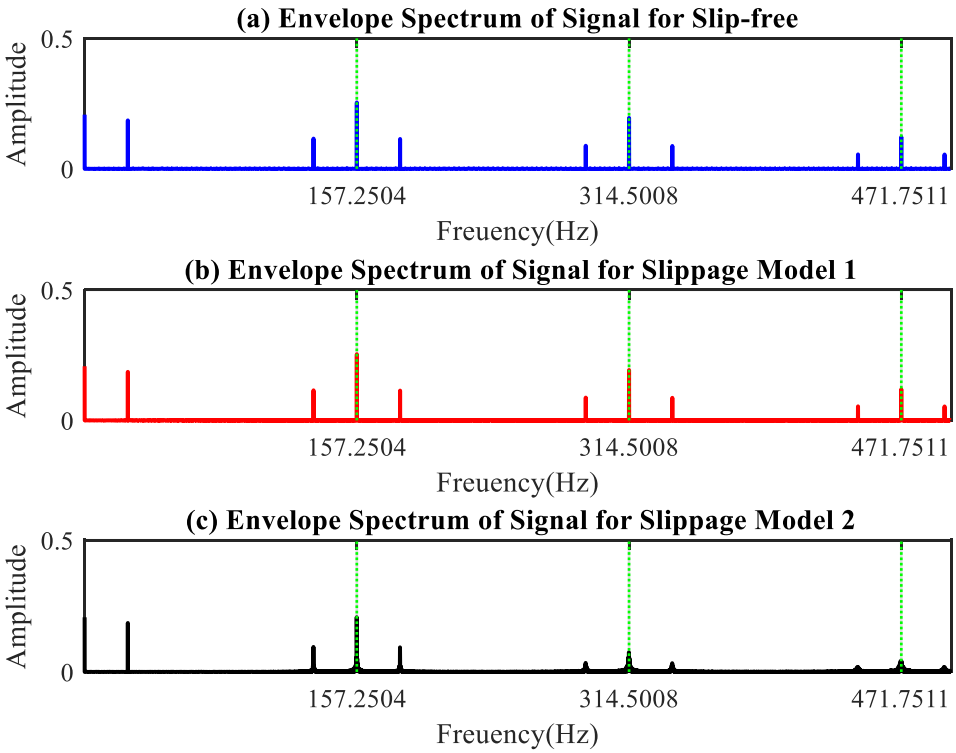

Figure 9-8 Envelope Spectrum of the Generated Signal Using Three Models

The network has been first tested using simulation impact noise-free signal as seen in Figure 9-9, (a) and (b) shows the time domain and the envelope spectrum respectively. The network parameters

shown in Table 9-2, were set to the same values found in the optimization process with an increase in the weight vector size as the signal has a more complex structure and contain more components.

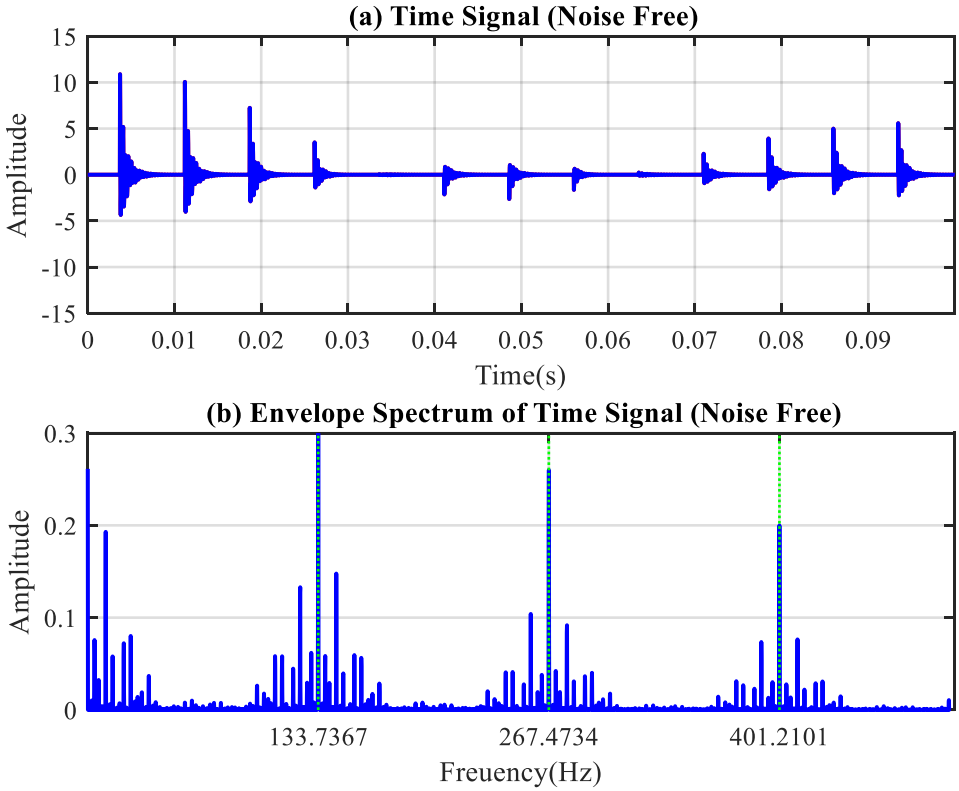

Figure 9-9 Noise Free training Signal

Table 9-2 Impact Signal Training Parameters

Parameter	Value
Weight vector number	8
Weight vector dimension	512
Threshold	0.4
Learning rate	0.05
Sigma	0.2
Batch number	Data length / WV dimension
Iterations	80

Figure 9-10 (a) shows the attenuation of weight variations and (b) shows the reconstruction error for noise-free signal, it shows that both the attenuation and the reconstruction error continue to decline till iteration number 60 and remained stable until it reached the end of the training process.

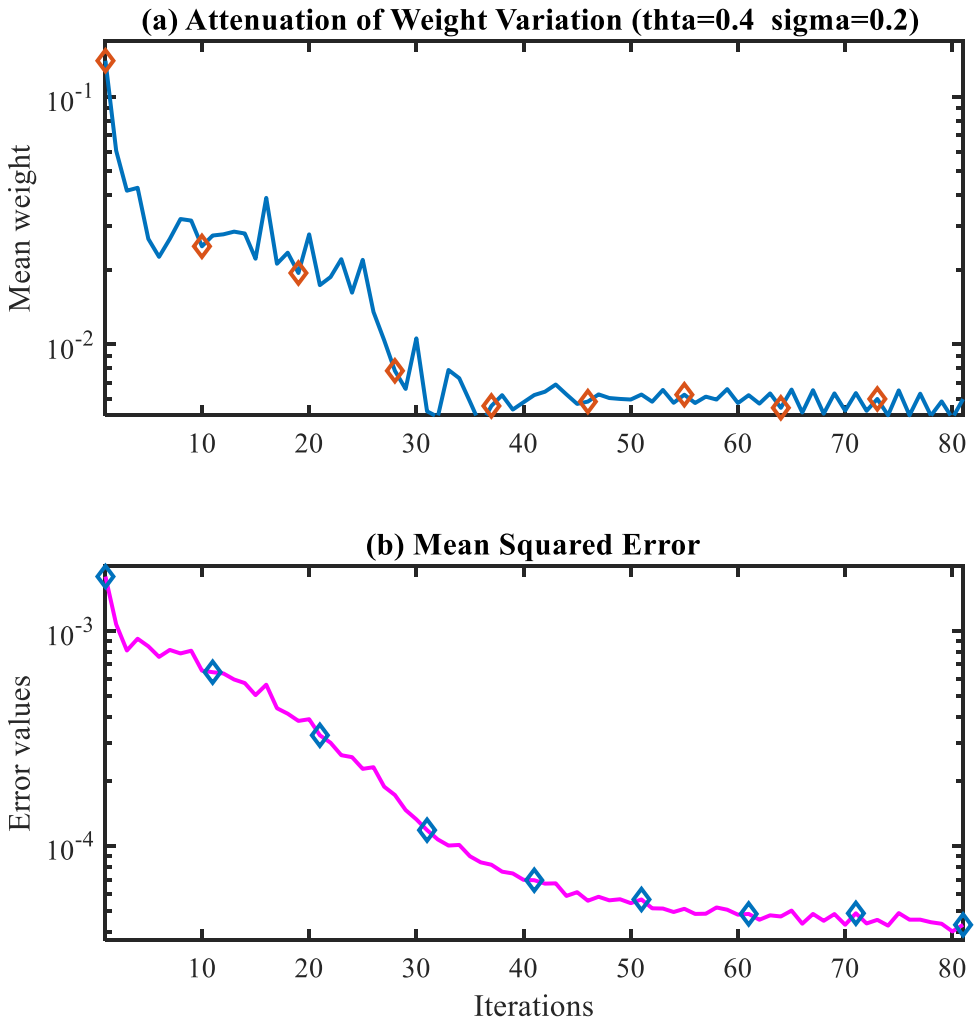

Figure 9-10 Training Error

The reconstruction of the unseen test signal is shown in Figure 9-11 alongside with the original signal used in training the CCNN. The reconstructed signal has nearly the same structure as the original signal and this indicates that the CCNN learned the dominated features in the signal structure.

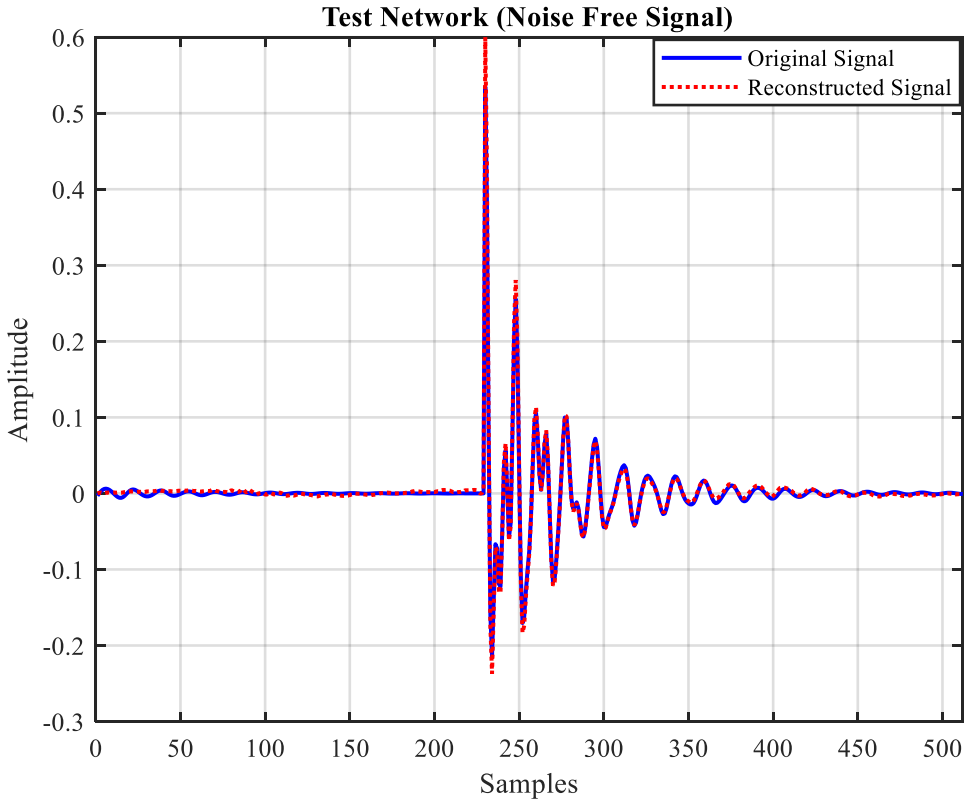

Figure 9-11 Original Signal and Reconstructed Signal during training

The weight vectors derived from the training stage are described in Figure 9-13, the pattern of the noise-free signal, illustrated in Figure 9-12, can be seen in many weight vectors. The learned features during training are used to reconstruct the signal in the test stage.

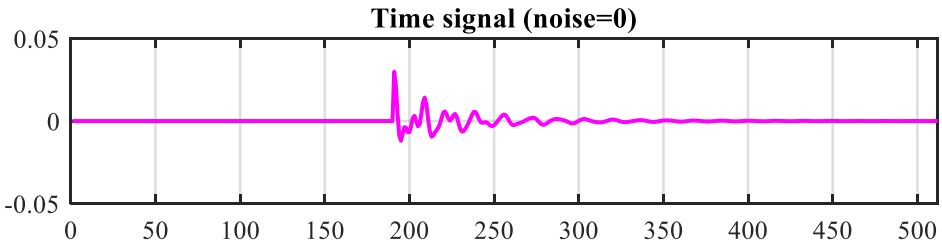

Figure 9-12 Time Signal used to Train the CCNN

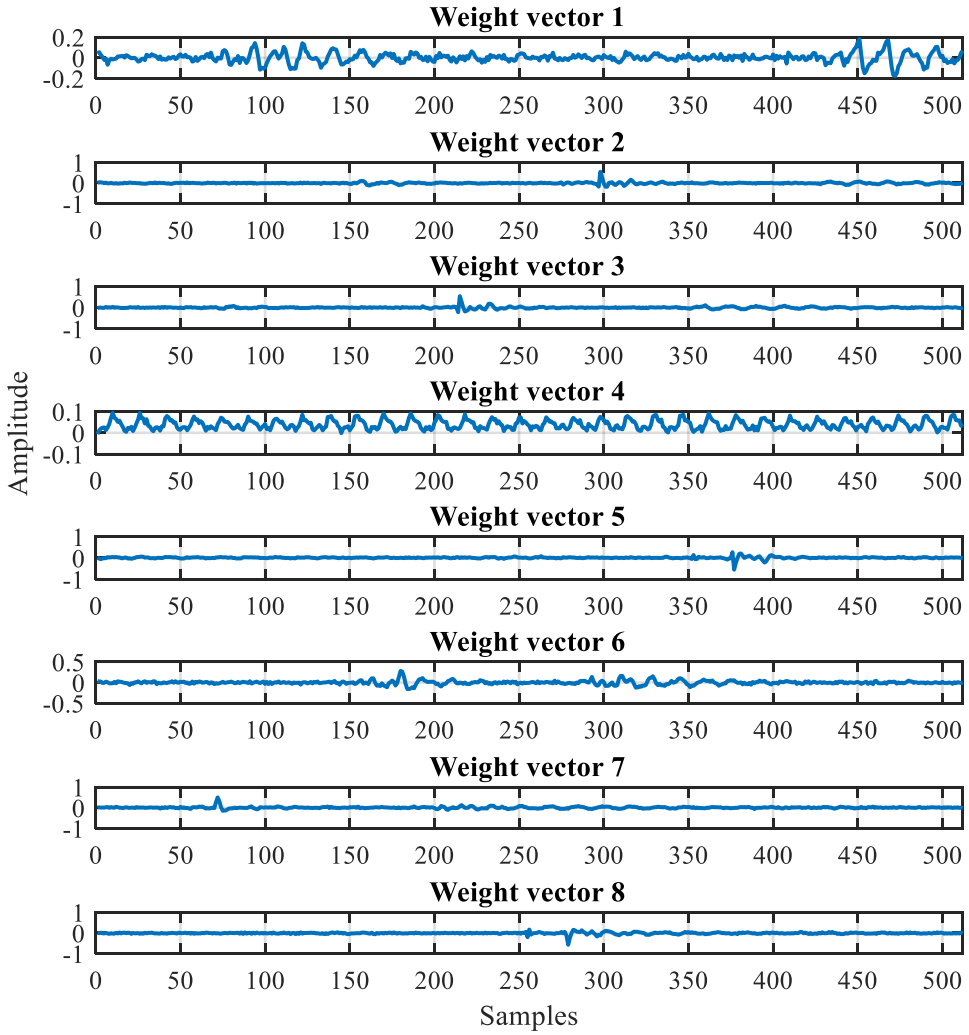

Figure 9-13 Weight Vectors of Model for Noise-Free Signal

After testing the network using noise-free signal, the CCNN is evaluated using generated signals with different noise levels to simulate a common form of bearings condition monitoring impact signal. Simulated data is generated with different levels of additional noise to test the influence of variations in noise strength on the performance of anomaly detection and how signals may be discriminated by the trained network.

Figure 9-14 and Figure 9-15 shows the signals used for training CCNN with different portions of noise start from 0.1 to 1.2 step 0.1. As shown in Figure 9-14, signal (a) impacts appear clear with regard to natural frequency and the noise. However, Signal with increased noise levels are shown in Figure 9-14 (b), (c), (d), (f) and (e), also plots in Figure 9-15 shows how SNR declined with the increase of noise levels to reach -7.3 in (e).

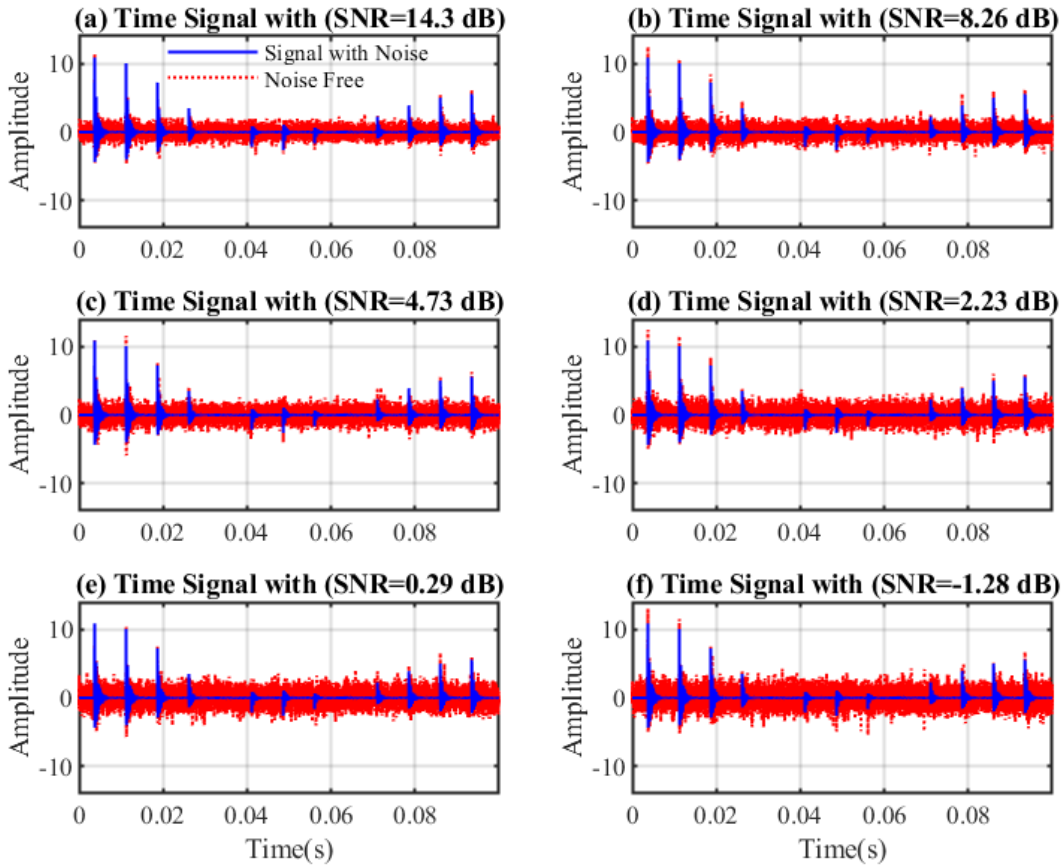

Figure 9-14 Simulated Impact Signals with Noise used for Training and Testing CCNN

The network was trained using impact signal with a different portion of noise for each training case, as seen in Figure 9-15, the signals are contaminated by the noise levels and the signal shown in (f) was nearly buried in the noise. Figure 9-16 shows the iterations of the network using signal (f) with signal to noise ratio SNR of (-7.3) and after 60 iterations of training the network, the MSE fallen to the lowest level and almost flatten.

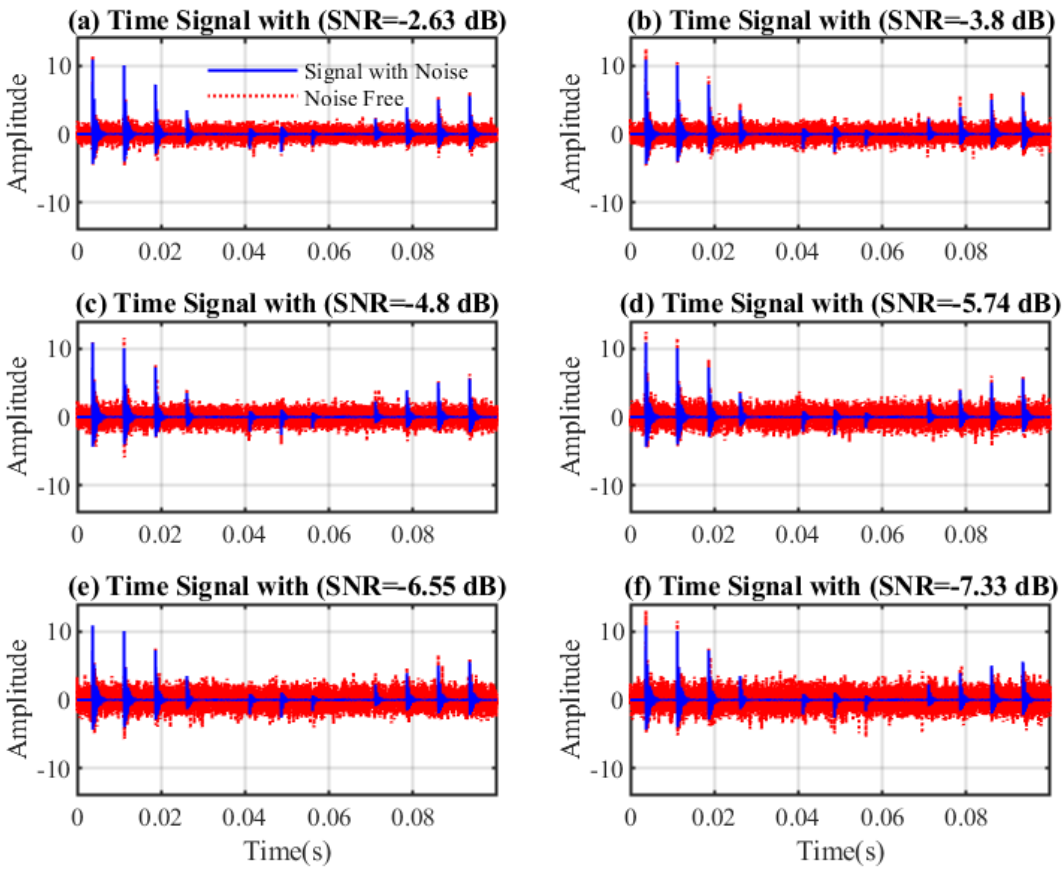

Figure 9-15 Simulated Impact Signals with Noise used for Training and Testing CCNN

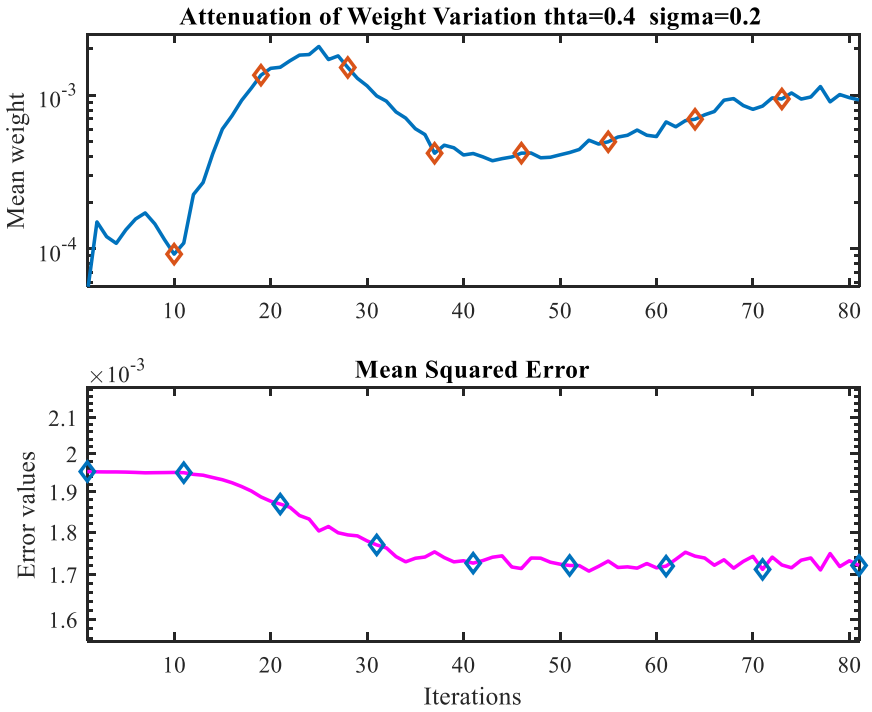

Figure 9-16 Attenuation of Weight Variation and Reconstruction error for the case with (SNR -7.3)

The features learned in the data model are shown in Figure 9-17, and it can be seen that the model derived the data pattern from the training signal shown in Figure 9-18.

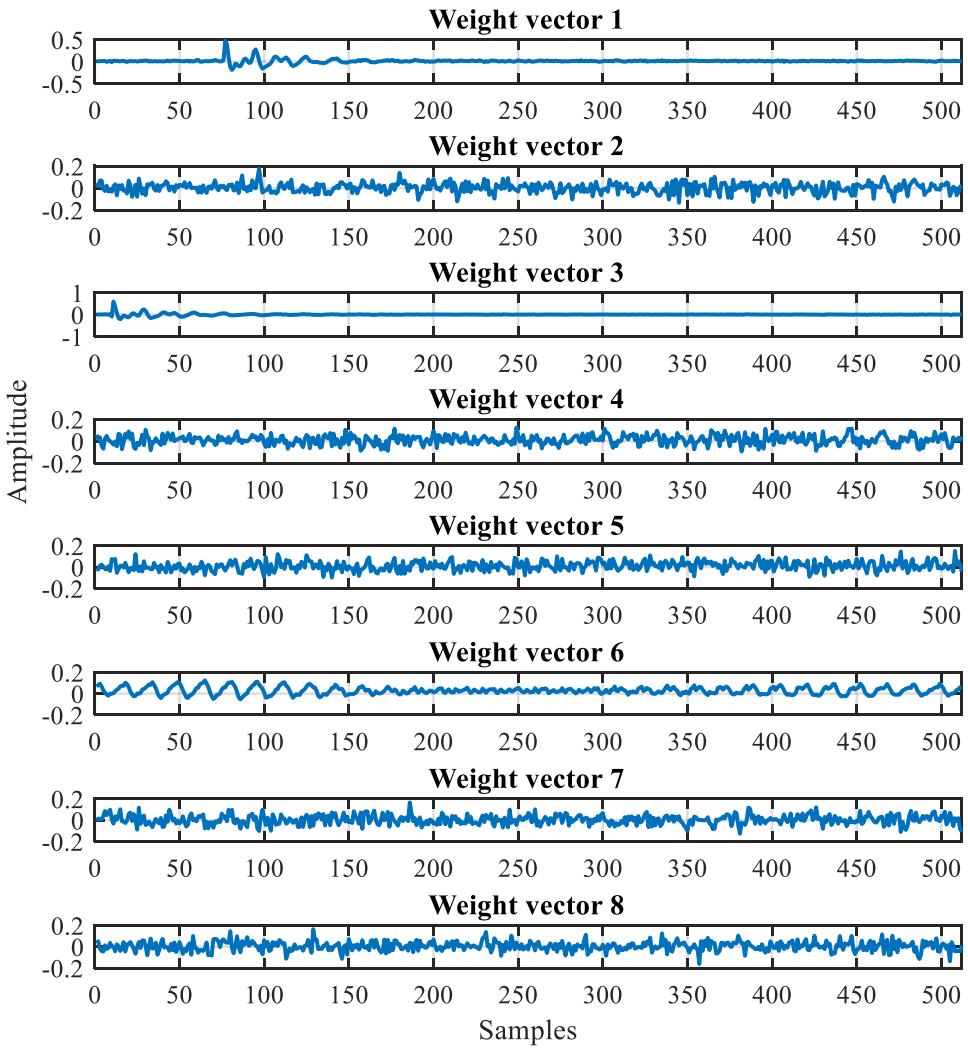

Figure 9-17 Weight Vectors for the Signal with SNR (-7.3)

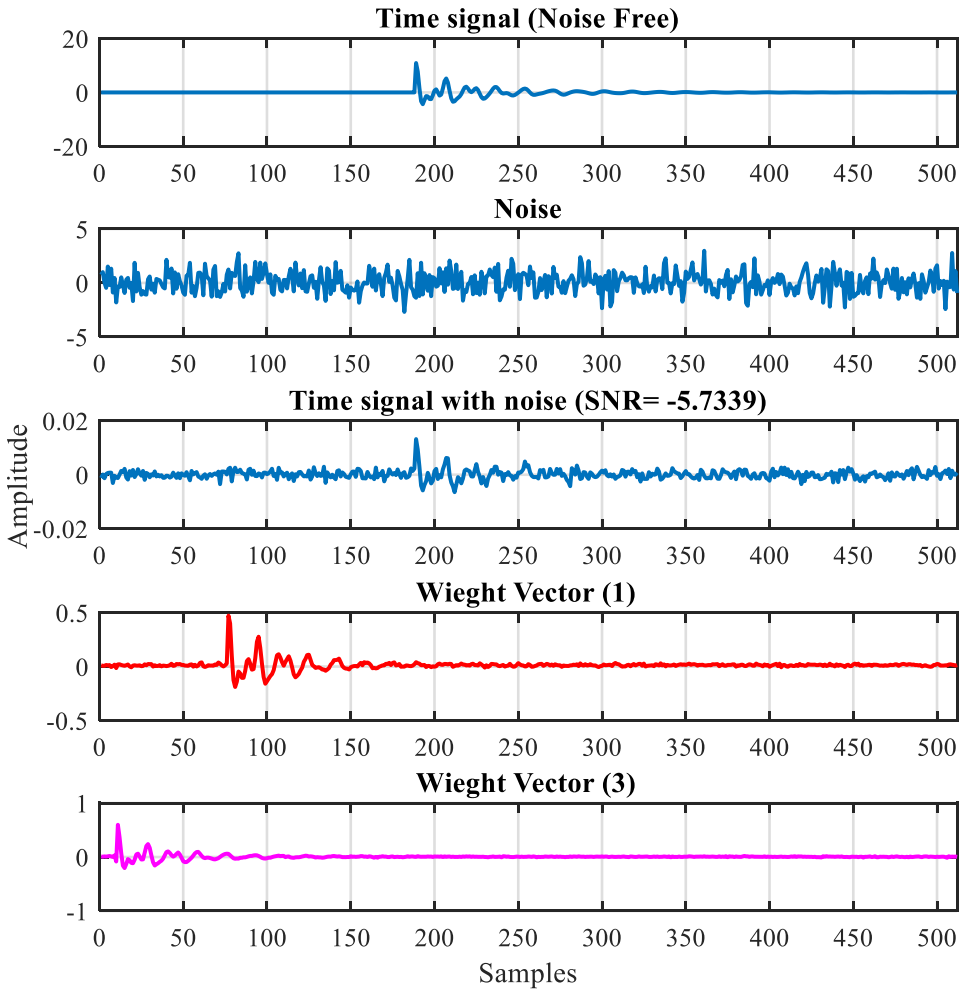

Figure 9-18 Time Signal with SNR (-7.3) and Learned Features in the Data Model

9.2.4.3 Testing Stage using Impact Signal

The performance evaluation test starts by firstly, training the network using noise-free signal which shown in Figure 9-14(a), then the trained network is tested with another unseen generated data with the same signal structure. The result obtained, shown in case number 1 (bar1) of Figure 9-19, clearly ADI showed no anomaly was detected when tested with the same noise-free signal structure which used in the training stage.

Secondly, as shown in Figure 9-20 (b), the noise levels were increased gradually by (0.1) for cases from number 2 until case number 13, so each case represents a different noise level and each case was tested with the same noise level used in the training stage. The results obtained can be seen in Figure 9-19 and in Figure 9-20 (a) with SNR information. The results proved that the network CCNN is capable to identify the data patterns even in the very low SNR values of (-6 dB). However, the network started to show an increase in the ADI level for cases with SNR of (-7 dB) and (-8 dB).

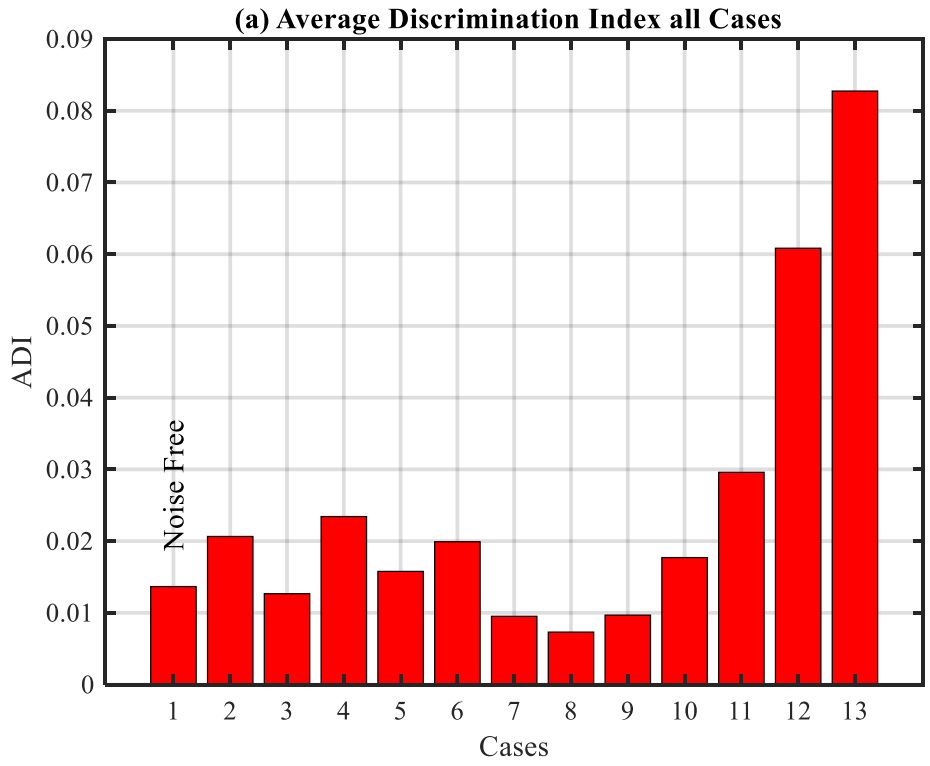

Figure 9-19 ADI of Testing Network with Same Data of Training

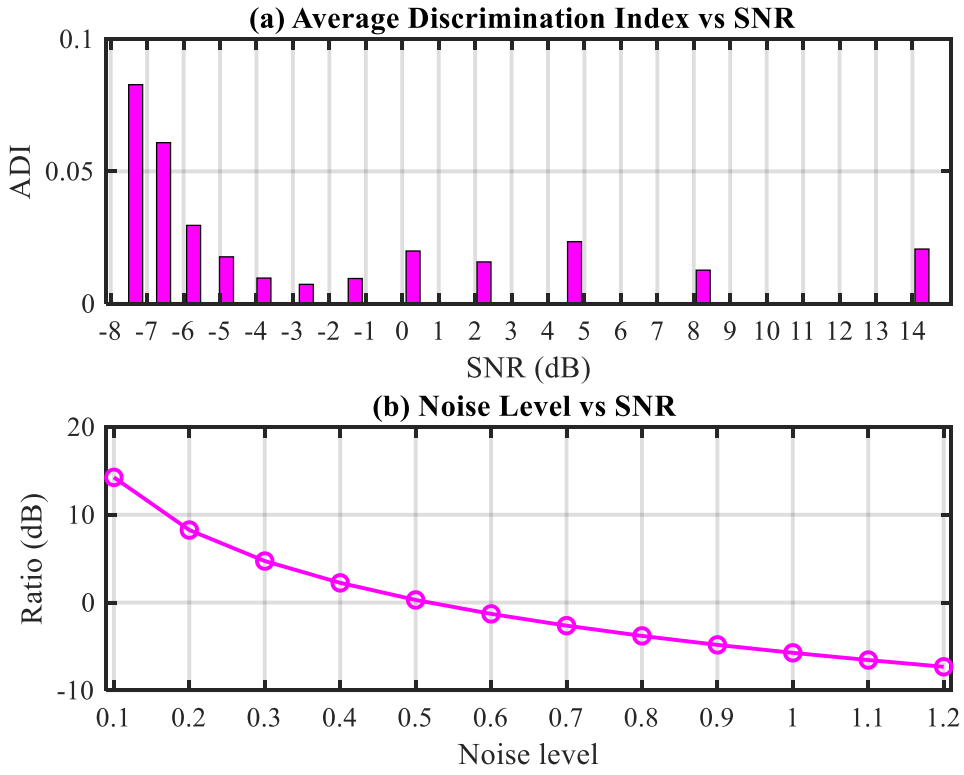

Figure 9-20 Simulated Data Used For Training

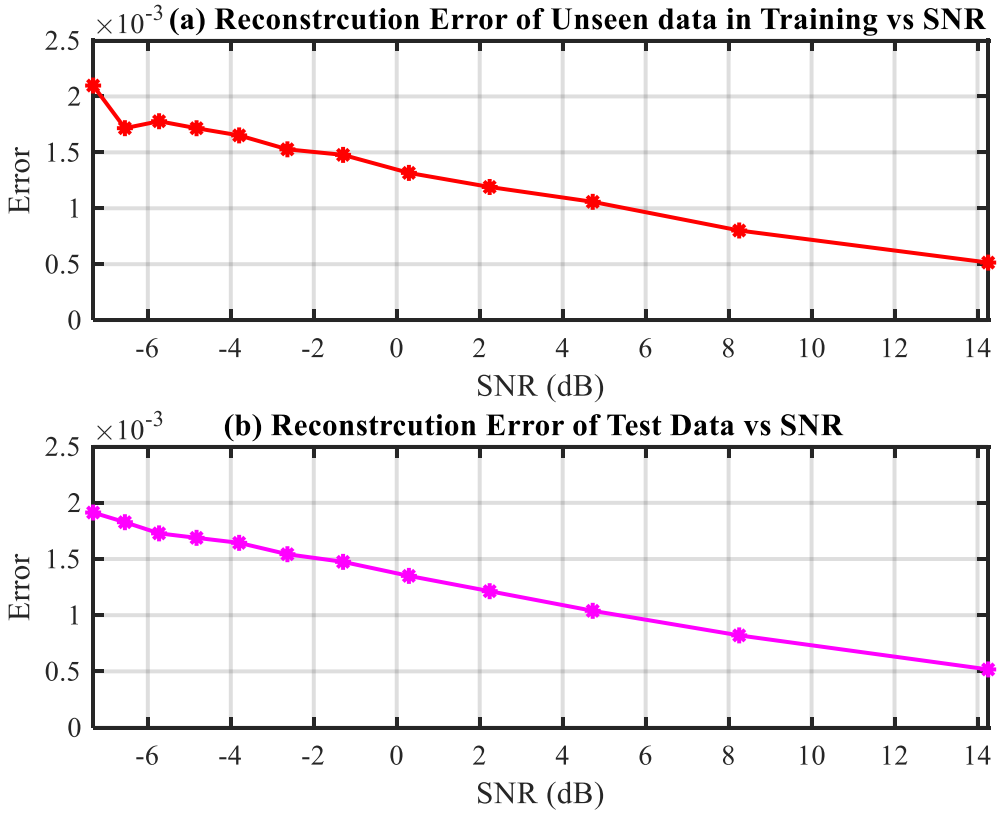

Figure 9-21 Reconstruction Error for Unseen Data Used in Training and Testing Stage

The reconstruction error for both unseen data during the training stage and for data used during the testing stage is shown in Figure 9-21 (a) and (b) respectively. The error shows an increasing trend with the increase of the noise level, this is because of the influence of noise at the learning process.

The second aspect of network evaluation considers the capability of the network in detecting the change of the structure of the impact signal similar to signals usually appears in a defective bearing data in the real world. Therefore new signal structures were generated to represent the outer race fault, inner race fault and cage fault with different noise levels.

In this test, the change of f_r is set to be 157 Hz to represent outer race frequency with the same portion of noise levels used in the training stage for each case. The same procedure in the training and testing was followed for all the signals. Figure 9-22 illustrates that the network can effectively detect the anomalies in the impact signals with outer race fault until the SNR reached (-6) by an amplitude of 0.14 ADI. Moreover, the network effectively discriminates the severity of the anomalies. The trend of the anomaly declines with the decrease of the value of SNR, however, overall the sensitivity of the network in detecting the novelty is promising and reliable.

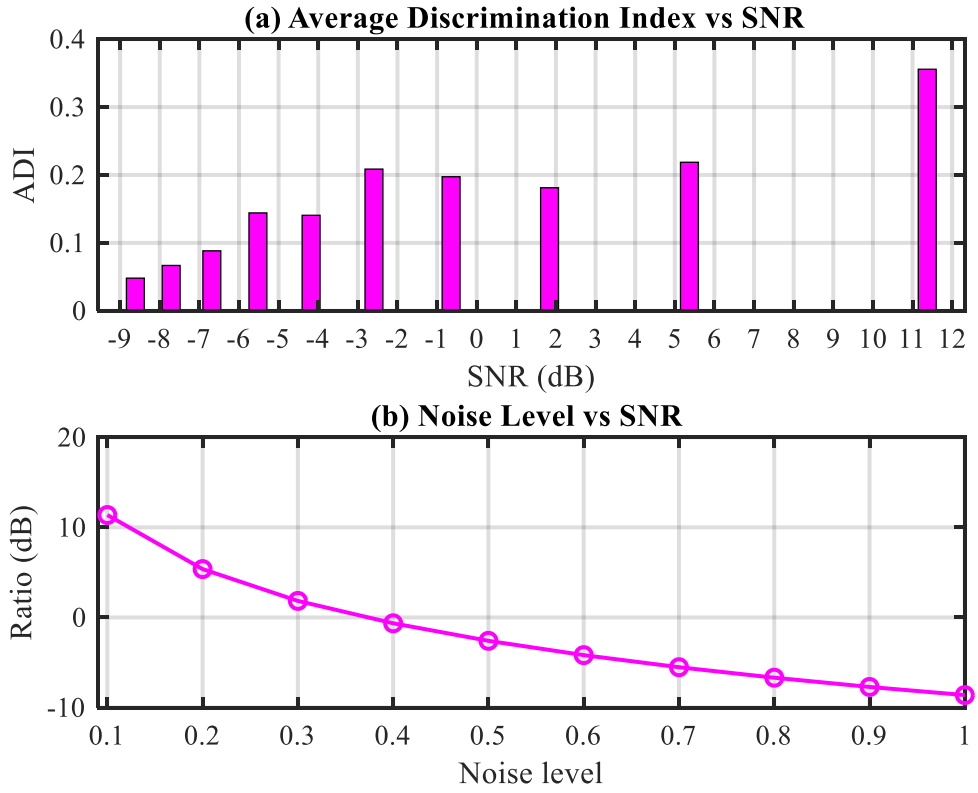

Figure 9-22 ADI Detection Performance of Simulated Signal with Outer Race Defect

The reconstruction error, shown in Figure 9-23, illustrates the trend of the reconstruction error for unseen data used during the training stage and the error for new test data, it can be seen that the error is higher for the cases till SNR reached (-6 dB) than the error obtained from testing during training stage. Although the noise level affected the results after SNR of (-6), however, results show that CCNN is a very effective method in anomaly detection even in the high noise levels.

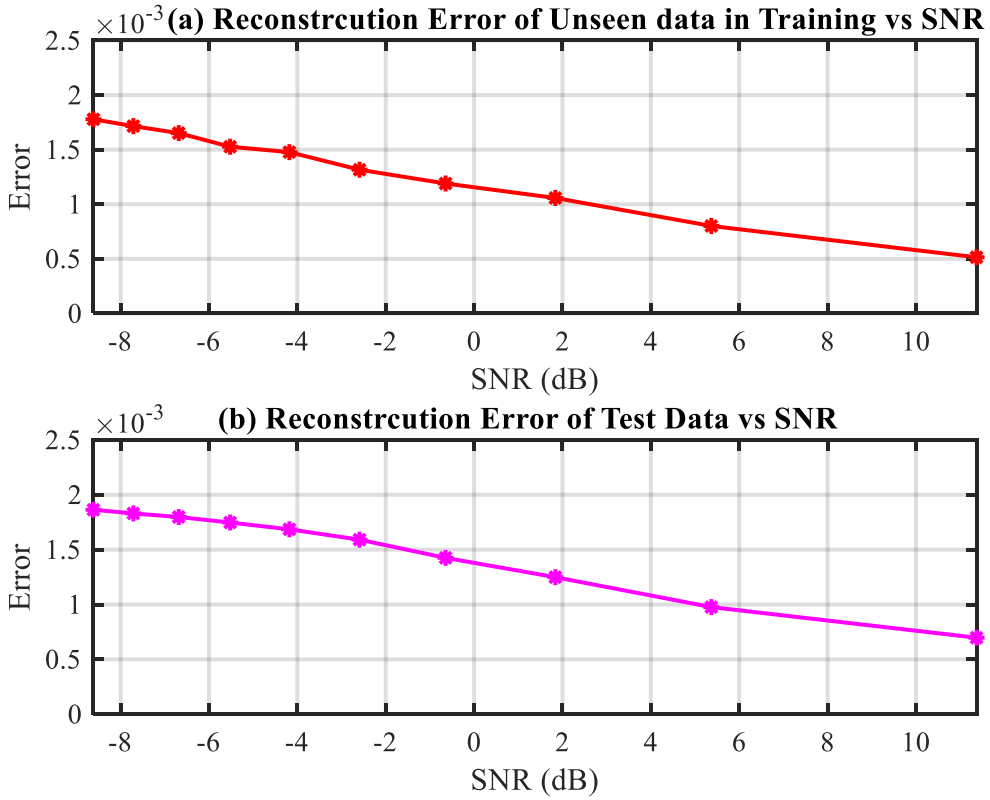

Figure 9-23 The Reconstruction Error for Unseen Training Data and Signal with Outer Race Fault

In the next test, the change of f_r is set to be 217 Hz to represent inner race fault frequency with the same portion of noise levels used in the training stage for each case. The same procedure in the training and testing was followed for all the signals. Figure 9-24 illustrates that the network can effectively detect the anomalies in the impact signals with inner race fault until the SNR reached (-6). Moreover, the network effectively discriminates the severity of the anomalies. The trend of the anomaly declines with the decrease in the value of SNR.

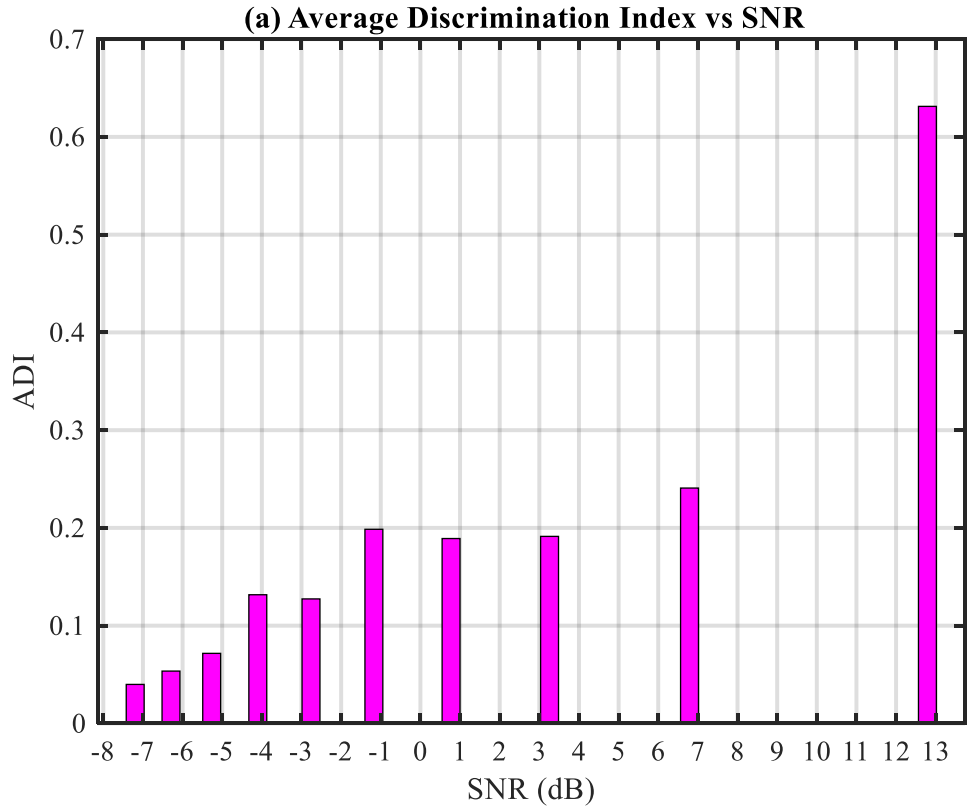

Figure 9-24 ADI Detection Performance for The Simulated Signal with Inner Race Fault

The reconstruction error, shown in Figure 9-25, illustrates the trend of the reconstruction error for unseen data used during the training stage and the error for new test data, it can be seen that the error is higher for the cases till SNR reached (-4 dB) than the error obtained from testing during training stage. Although the noise level affected the results after SNR of (-4), however, results show that CCNN is a very effective method in anomaly detection even in the high noise levels.

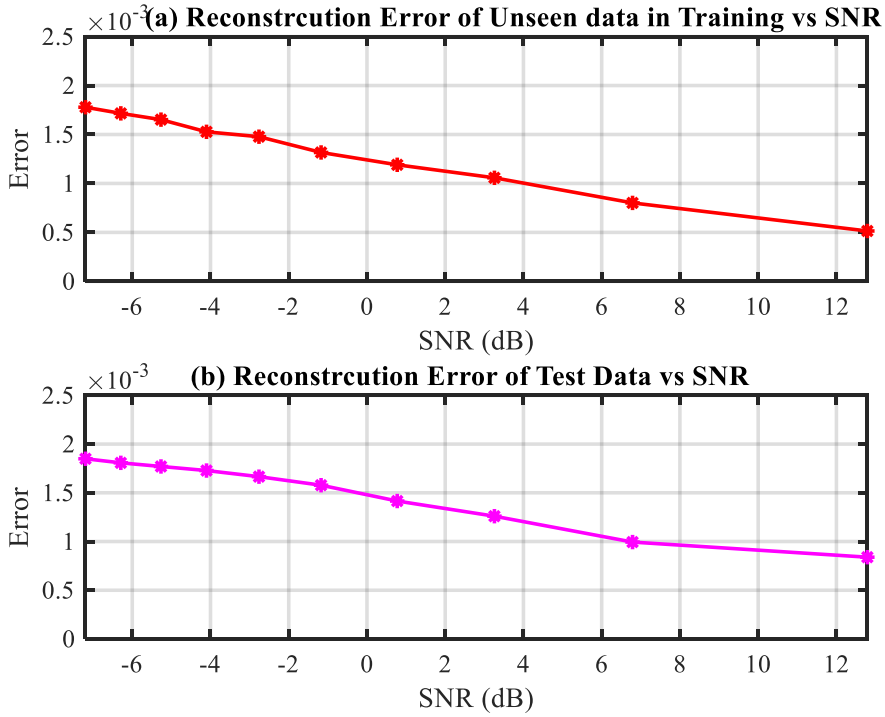

Figure 9-25 Reconstruction Error for Unseen Training Data and Signal with Inner Race Fault

In this test, the change of f_r is set to be 14.5 Hz to represent cage fault frequency with the same portion of noise levels used in the training stage for each case. The same procedure in the training and testing was followed for all the signals. Figure 9-26 illustrates that the network can effectively detect the anomalies in the impact signals with cage fault until the SNR reached (-6) by an amplitude of 0.14 ADI. Moreover, the network effectively discriminates the severity of the anomalies. The trend of the anomaly declines with the decrease of the value of SNR, however, overall the sensitivity of the network in detecting the novelty is promising and reliable.

The reconstruction error shown in Figure 9-27, (a) shows the trend of the reconstruction error for unseen data used during the training stage and in (b) the error for new test data, it can be seen that the error is higher for the cases till SNR reached (-20 dB) than the error obtained from testing during training stage. Although the noise level affected the results after SNR of (-18), however, results show that CCNN is a very effective method in anomaly detection even in the high noise levels.

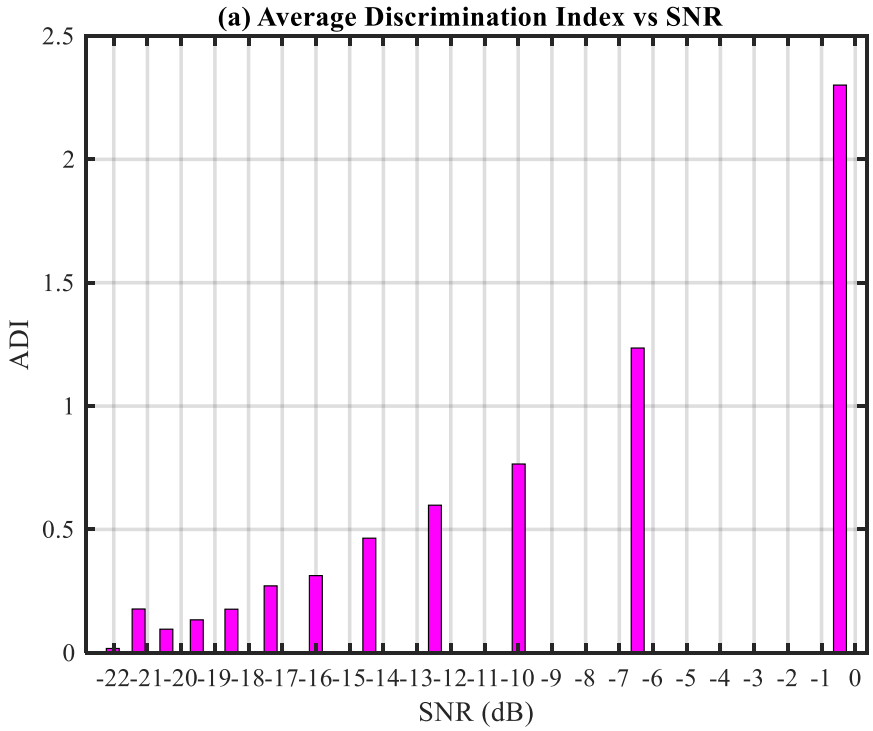

Figure 9-26 ADI Detection Performance of Simulated Signal with Cage Fault

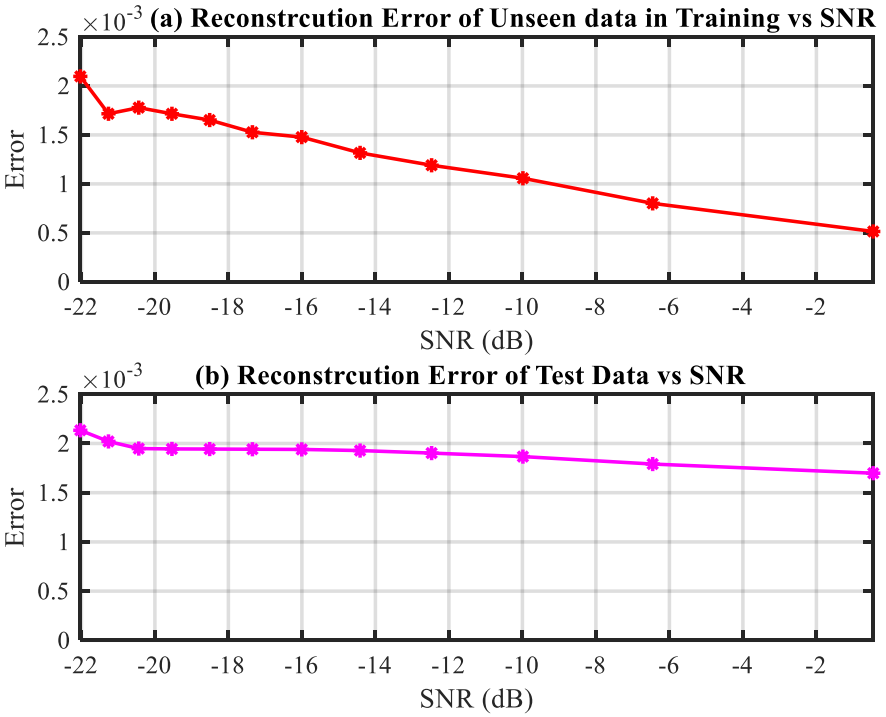

Figure 9-27 The Reconstruction Error for Unseen Training Data and Signal with Cage Fault

9.3 Implementation of the CCNN to Real Data

In order to verify the CCNN, the results analysis of the collected vibration data is presented with the discussion of the results. This chapter will assess the capability of the method for bearings fault detection. This chapter also reports the successful application of the CCNN to fault severity classification based on the ADI. The results indicate that CCNN can be a suitable method for fault detection and diagnosis of REBs.

After the successful data collections have been conducted, five sets of data have been used in this study from three bearings. Set one is taken during healthy condition operation form one bearing, and then for each of the remaining, set two from small outer race fault, set three from large outer race fault, set four from the small roller fault and set five large roller fault. Each contains vibration measurements and has 1524000 samples. The potential of these measurements for detection and discrimination has been firstly validated using time domain, spectrum and then envelop analysis. Thus, these data sets provide further validation of the performance of the CCNN networks. CCNN has been applied to the recorded real data. Firstly, two sets of the baseline data used in the training process one set used to train the network and the other set (unseen data-set) used in the calibration process, as during the calibration process the network needs to be fed with unseen data-set.

9.3.1 Training Stage

The baseline dataset used for training consists of 1524000 data points and this large dataset is covering more than 750 shaft rotations and was adopted to ensure that CCNN is trained adequately in order to build a consistent data model. The network was designed to have 16 weight vectors, which allows capturing the complex patterns hiding in the data structures. The interval between the expected faulty peaks for outer race fault is calculated as $\Delta t = \frac{1}{f_{BPO}} F_s$ which results in about (320) data points and for roller fault $\Delta t = \frac{1}{2f_{BS}} F_s$ which results in about (378) data points, F_s represents the sampling rate. Therefore, each weight vector is set to be consisting of 512 data points so as to ensure that each vector is covering more than one period of the expected fault and reflecting the basic characteristics of the periodicity of the vibration signals.

The initial non-adaptive parameter (configuration) of the network has been optimised using the simulated data and best detection results were obtained when the parameters were set as shown in Table 9-3. The adjustment of the threshold is based on the principle of non-negative ADI.

Table 9-3 Network Parameters

Item	Value
Weight vector numbers	16
Weight vector dimension	512
Threshold	0.40
Learning rate	0.05
Sigma	0.2
Batch number	Data length /WV dimension
Iteration	100

The training steps as following:

- The initialisation of the weight vectors with random values and scaling into unit length;
- Segment the training data into predefined patches with the same length of weight vectors
- The input data is fed into the network and iterate the network until the weight vectors become stationary at the lowest possible level.

For an ideal network, the variation of weight and training error should be reduced smoothly. A reduced smooth trend can be achieved if there is a defined structure exists in the dataset. The convergence of the network during training usually gets smooth with small statistical fluctuations depending on the size of the training patterns and learning rate. On the other hand, if data contains no defined structure or is entirely random, no minimum point can be found and the route of convergence fluctuates with large amplitudes.

As seen in Figure 9-28 (a), mean squared error has reached the minimum level at about iteration number 60 and remained almost flat. Also, weight vectors obtained after 80 iterations, when both the residual errors and weight variation had stabilised with regard to further training iterations. (b) shows the scale parameters derived during the training process.

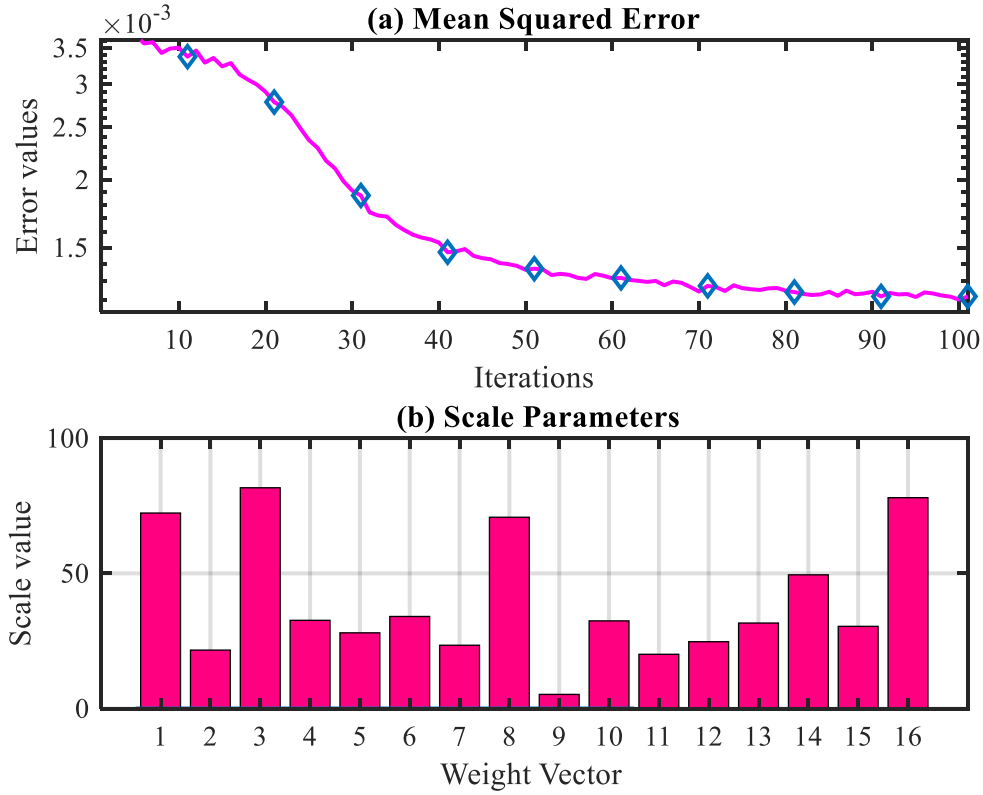

Figure 9-28 Network Convergence

9.3.2 Validation

In the validation stage, the network is tested using data model learned from training data shown in Figure 9-29. Unseen baseline data is fed into the network, the reconstructed signal obtained is shown in Figure 9-30 and Figure 9-31, it looks very similar to the original signal used in the training stage, and this shows that the network learned the most important features required to reconstruct the signal.

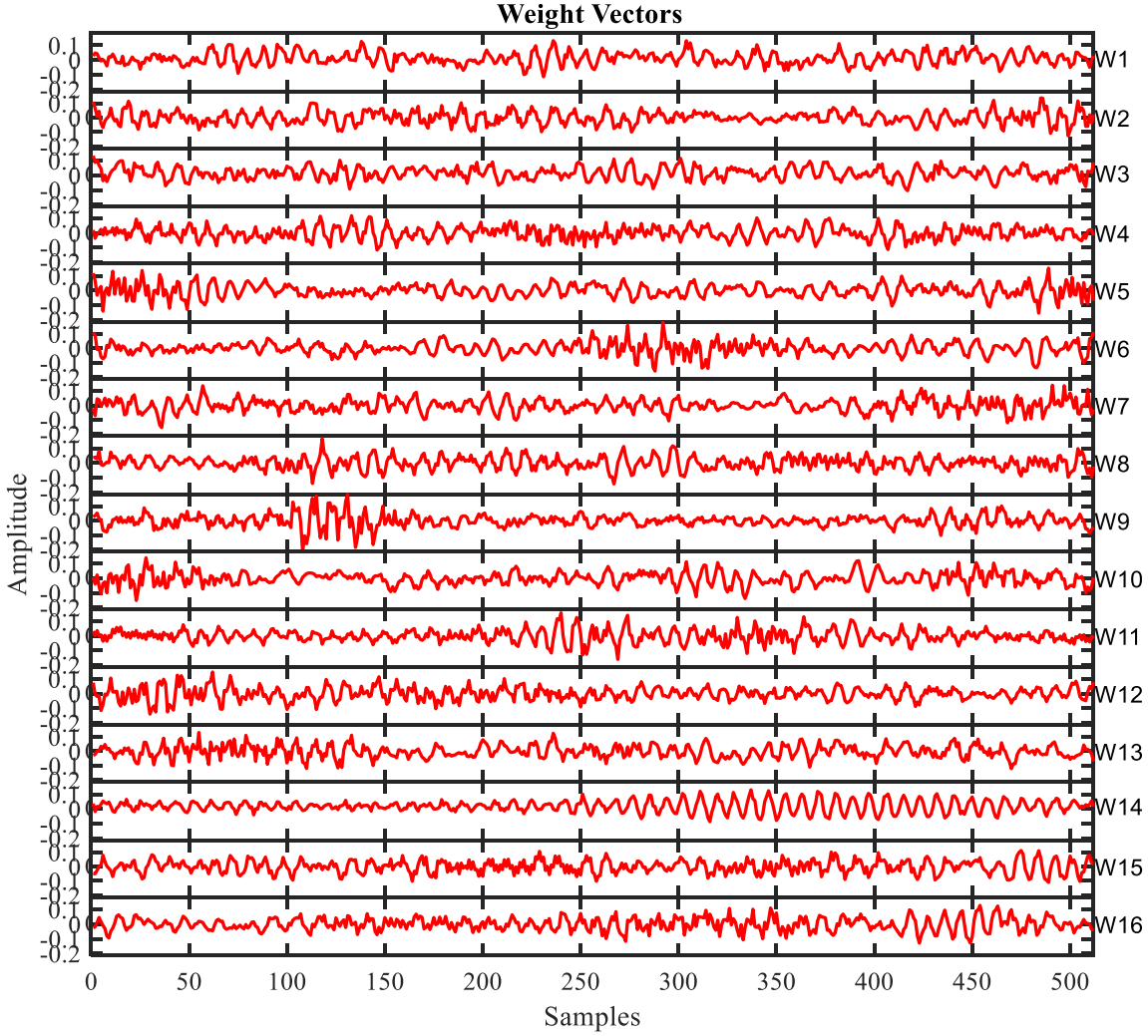

Figure 9-29 Weight Vectors Obtained during training

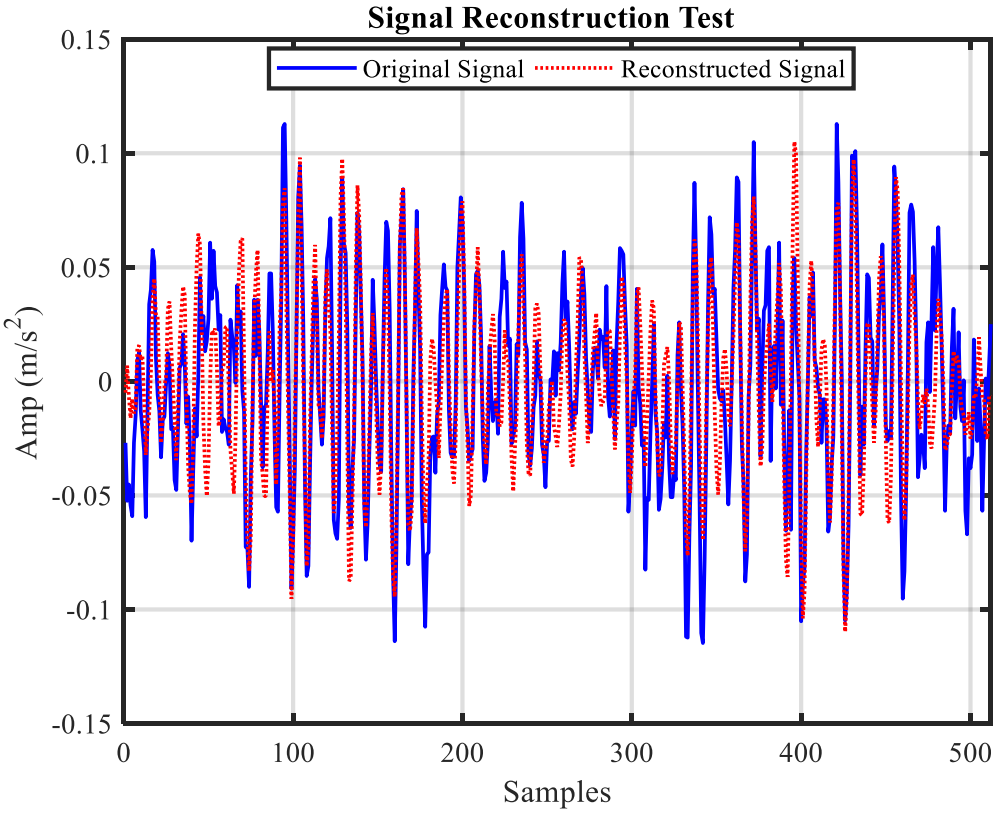

Figure 9-30 Reconstructed Signal in Validation Stage

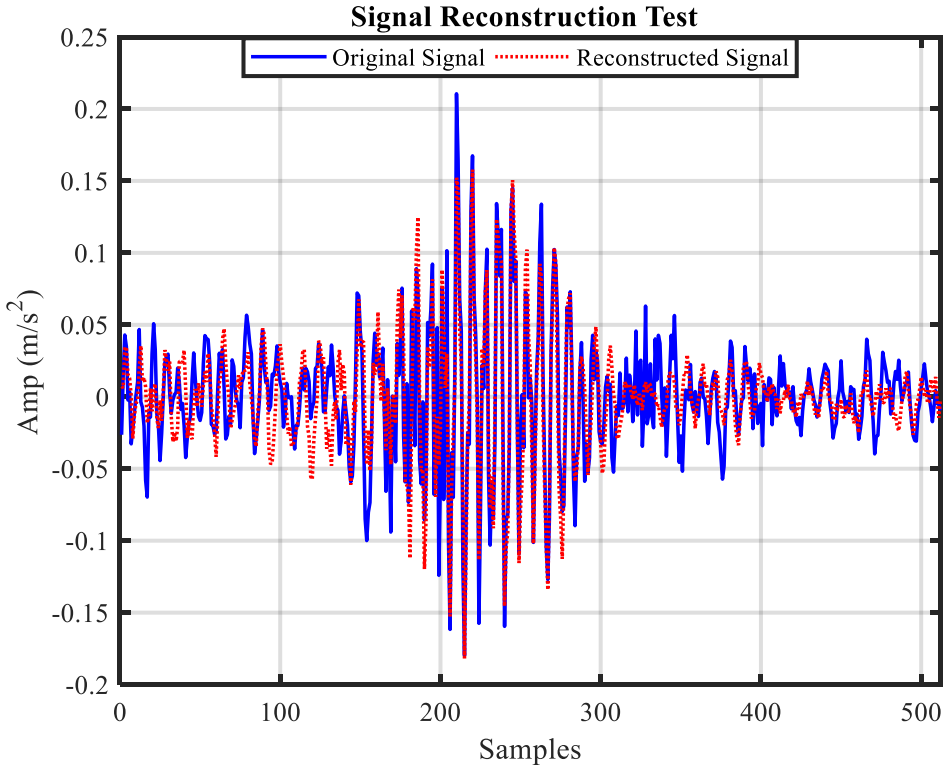

Figure 9-31 Reconstructed Signal in Validation Stage

9.3.3 Implementation Results and Discussion

Based on the data model obtained from training stage using baseline data, the data model is used with unseen baseline data with five clearances and four faulty datasets with five different clearances collected from tapered roller bearings as detailed in chapter six.

9.3.3.1 Outer race data analysis

This section presents the results when using CCNN to detect anomalies in three datasets, dataset 1 was collected from a healthy condition and dataset 2 was collected from defective bearings with small outer race fault whilst, dataset 3 was gathered from a bearing with large outer race fault.

Figure 9-32(a) and (b) presents the RMS of both raw signal and envelope spectrum respectively, both RMS values shown no clear trend or discrimination between the healthy and small outer fault conditions.

Figure 9-33 illustrates the RMS of raw signal in (a) whilst the application of CCNN to baseline data with different five clearances is shown in (b). The RMS show a decreasing trend with the decline of clearance in the baseline data sets. Signal collected under zero clearance is used to train the network and another four sets of baseline data are tested to highlight the response of CCNN to healthy data under different operating conditions. Although CCNN does not show anomalies in all baseline datasets, it can be seen the CCNN has good sensitivity to the slight change in the data.

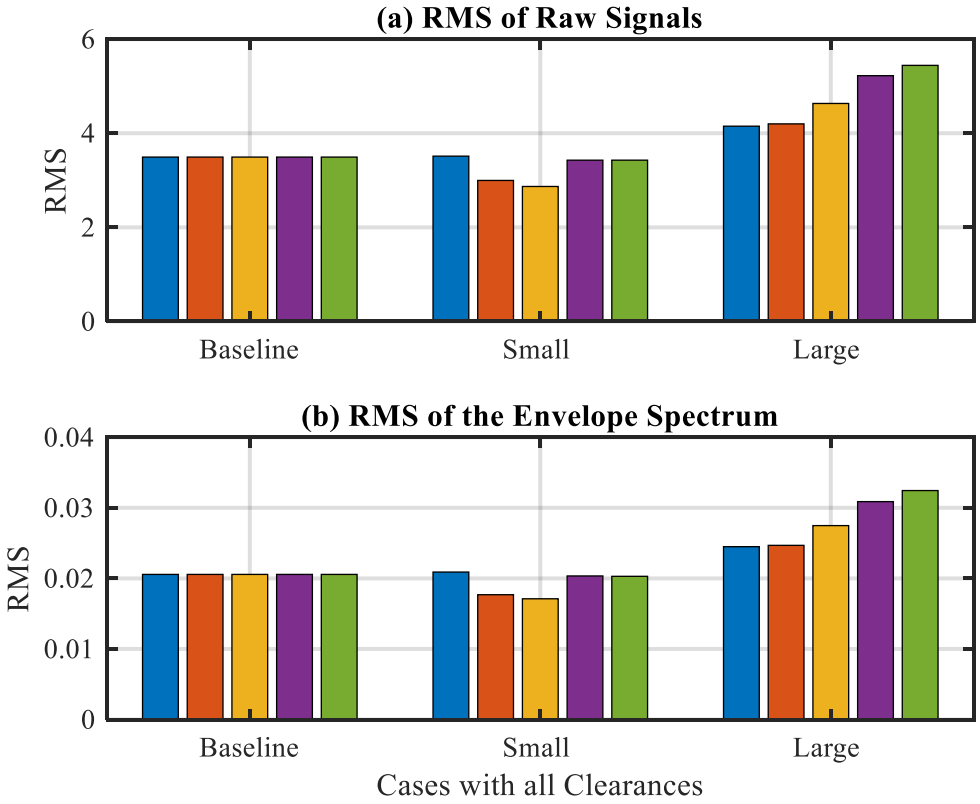

Figure 9-32 RMS of Raw Signal and Envelope Spectrum for all Outer Race Cases

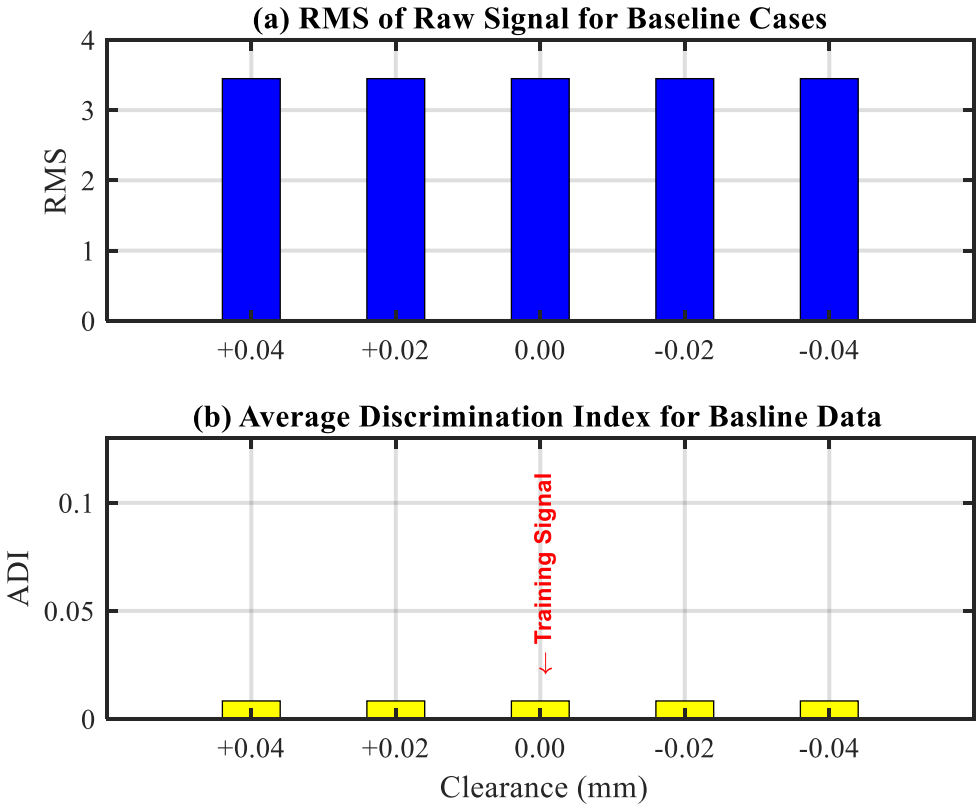

Figure 9-33 ADI Detection Results of Baseline Data with all clearances

Figure 9-34 shows the RMS of raw data and the results obtained from applying CCNN to the dataset with small outer race fault, it can be seen in (a), there is no trend exists in the RMS results. However, CCNN can clearly detect the faults with an increasing trend as the clearance declines compared to the healthy condition described above in Figure 9-33. The small fault for all clearance cases are detected by CCNN, the largest clearance case has the lowest amplitude of 0.12 ADI whilst the smallest clearance case detected with the maximum amplitude value of 0.16. This shown that CCNN effectively detects the anomalies and can discriminate the clearance variance.

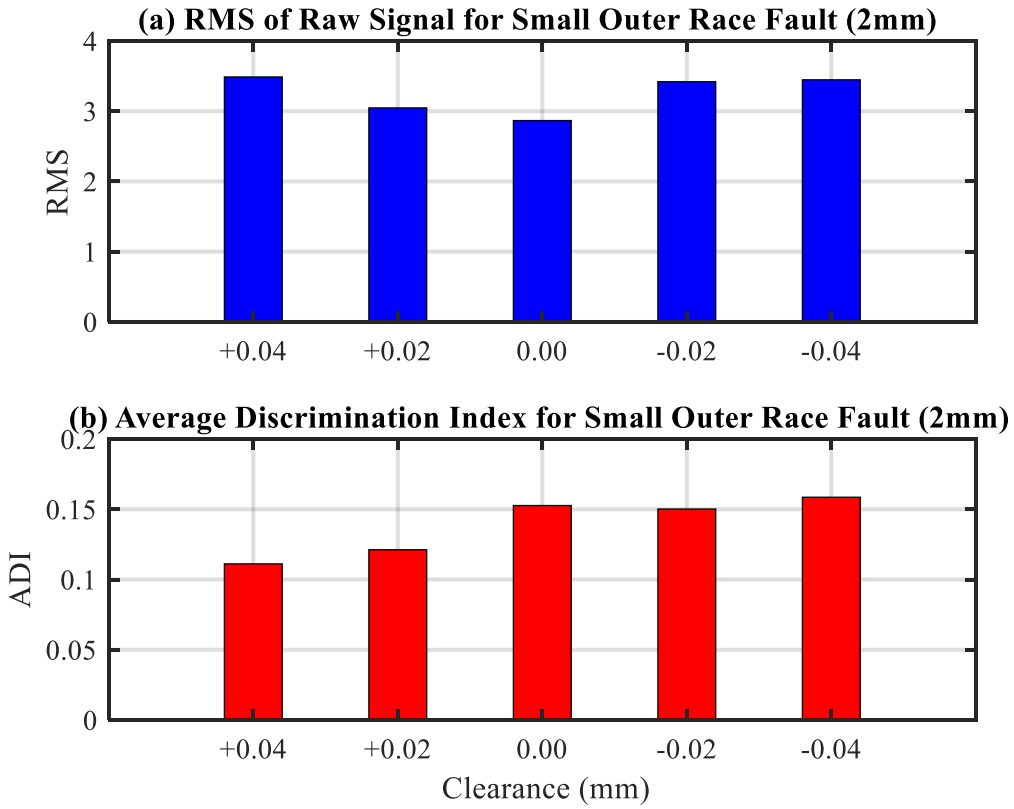

Figure 9-34 RMS of Raw Data and ADI Detection Results of Small Outer Race all clearances

Figure 9-35 (a) and (b) illustrates RMS of raw signal and ADI for large outer race fault respectively. The CCNN can effectively detect the fault and also shows the ability to estimate the severities of the fault compared to the small fault cases. Moreover, CCNN can is very sensitive to the change in signal signature due to variance in internal clearances. Thus, CCNN can clearly detect the faults with an increasing trend as the clearance declines.

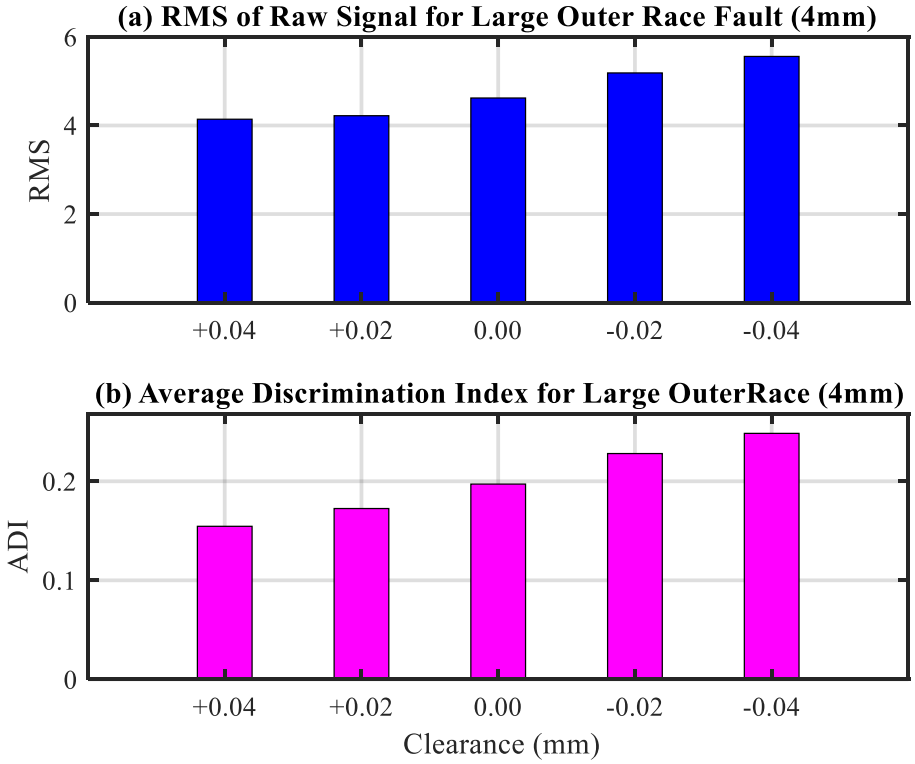

Figure 9-35 RMS of Raw Signal and ADI Detection Results of Large Outer Race all clearances

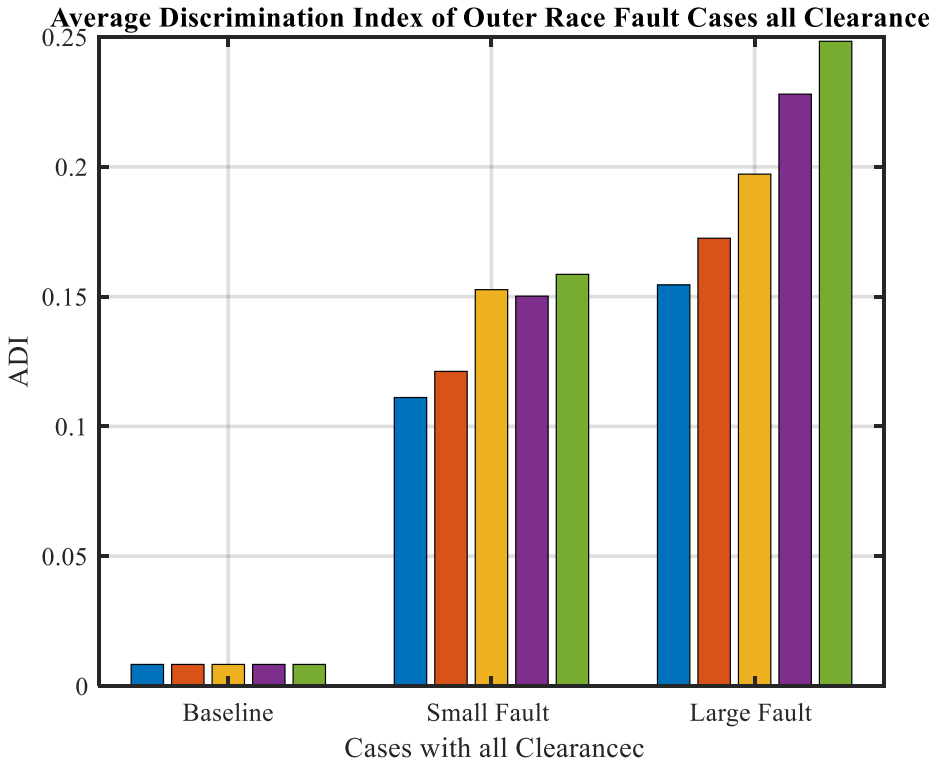

Figure 9-36 Detection Results for all Baseline, Small and Large Outer Race fault Cases

For more understanding, a comparison has been presented in Figure 9-36 and in Figure 9-37 between baseline and two faulty cases with all clearance cases. It can be clearly seen that CCNN

has a very good performance in anomaly detection and in estimating the severities. For the outer race data sets, the CCNN has shown the capability of detecting the faults with two severities. Compared to the baseline data, the two outer race faults have significant higher amplitudes as a result of the changes occurred in the signals structure due to the presence of induced faults.

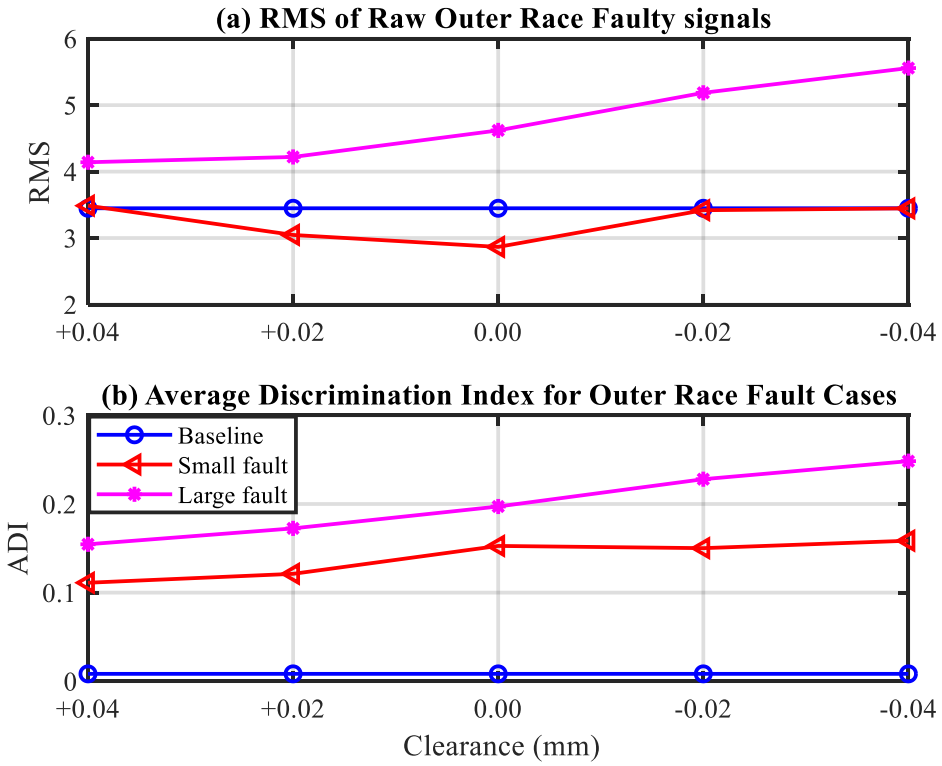

Figure 9-37 RMS of Raw Data and ADI Detection Results of Baseline and Outer Race all Cases

9.3.3.2 Rolling Element data analysis

This section presents and discusses the results obtained from implementing CCNN to datasets collected from a defective TRBs with roller fault. As it can be seen in Figure 9-38, RMS of Raw Signal in (a) and RMS of the Envelope Signal show no clear separation nor estimation of the fault severities. Figure 9-39 shows in (b) the average discrimination index of small roller fault case with all clearances. Compared to the RMS of raw data shown in (a), CCNN can effectively detect the change in the signal structure compared to the healthy condition. Moreover, it has demonstrated clearly the capability in estimating the increased severity of the small fault due to the declining of the clearances.

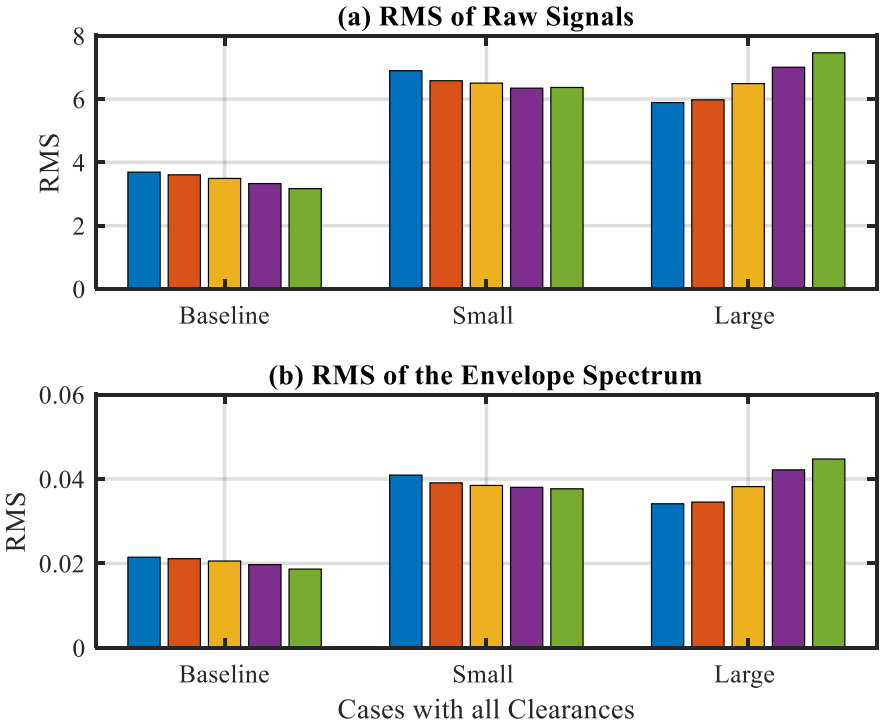

Figure 9-38 RMS of Raw Signal and Envelope Spectrum for Roller Faults

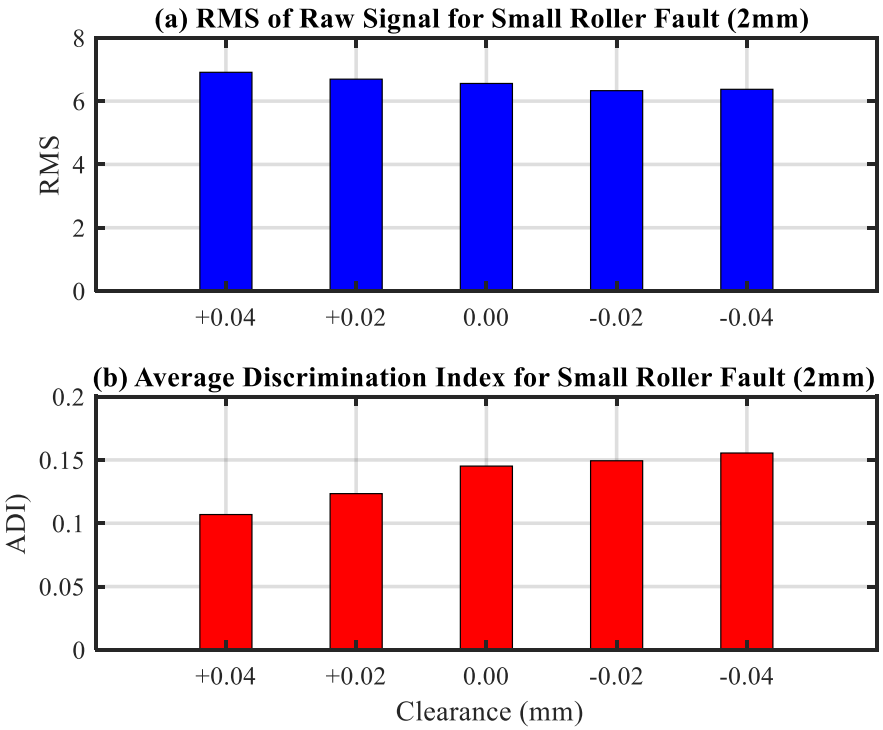

Figure 9-39 RMS and ADI Detection Results of Small Roller Fault

In order to test the capability of CCNN in separating different fault severity of the roller fault cases, Figure 9-40 (b) illustrates the results obtained from applying CCNN to large roller fault whilst, (a)

shows the RMS of Raw data. It can be seen that CCNN detected the faults and estimated the severity gradually with the increase of the preload. CCNN showed a sharp increase, almost doubled, in the amplitude of ADI for the large fault cases compared to small roller case shown in Figure 9-39.

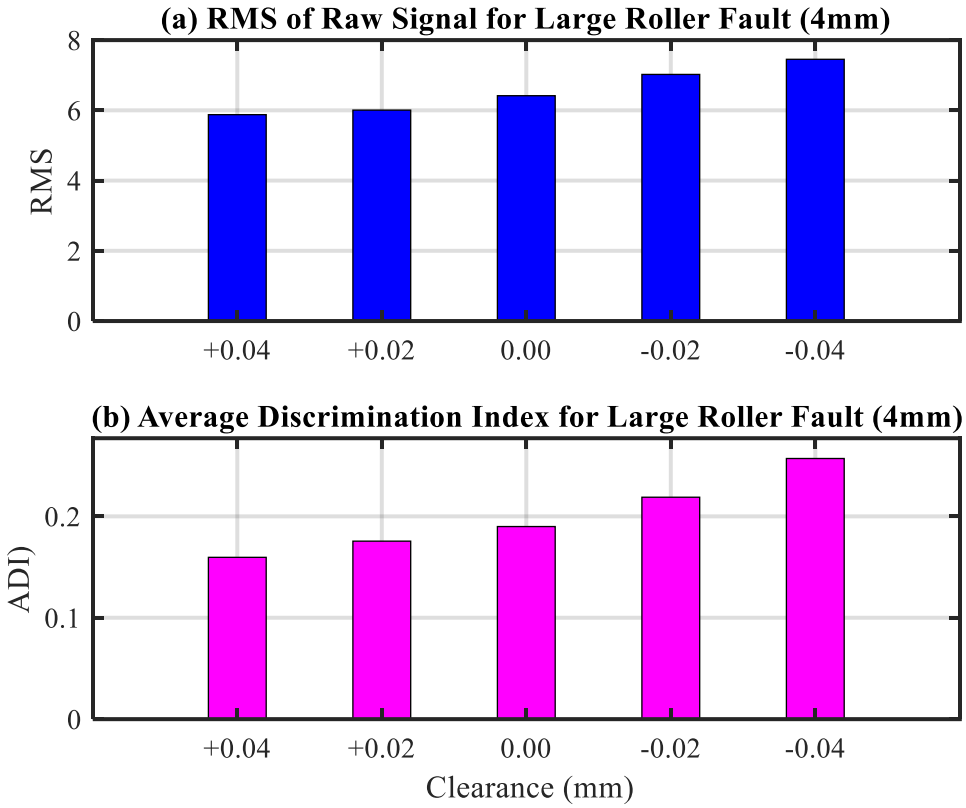

Figure 9-40 RMS and ADI Detection Results of Large Roller Fault

For more demonstrations, a comparison is made and depicted in Figure 9-41. The comparison shows the capability of CCNN in early fault detection task also in estimating the severity of the defects due to the fault size or due to variance in internal clearance. The amplitude of the small roller fault increased gradually from 0.11 to 0.16 with the increase of the preload. Figure 9-42 in (a) shows the RMS and in (b) shows the detection results obtained from CCNN, it can be seen clearly that CCNN can separate the condition and also results show the increasing trend with the increase of preload settings.

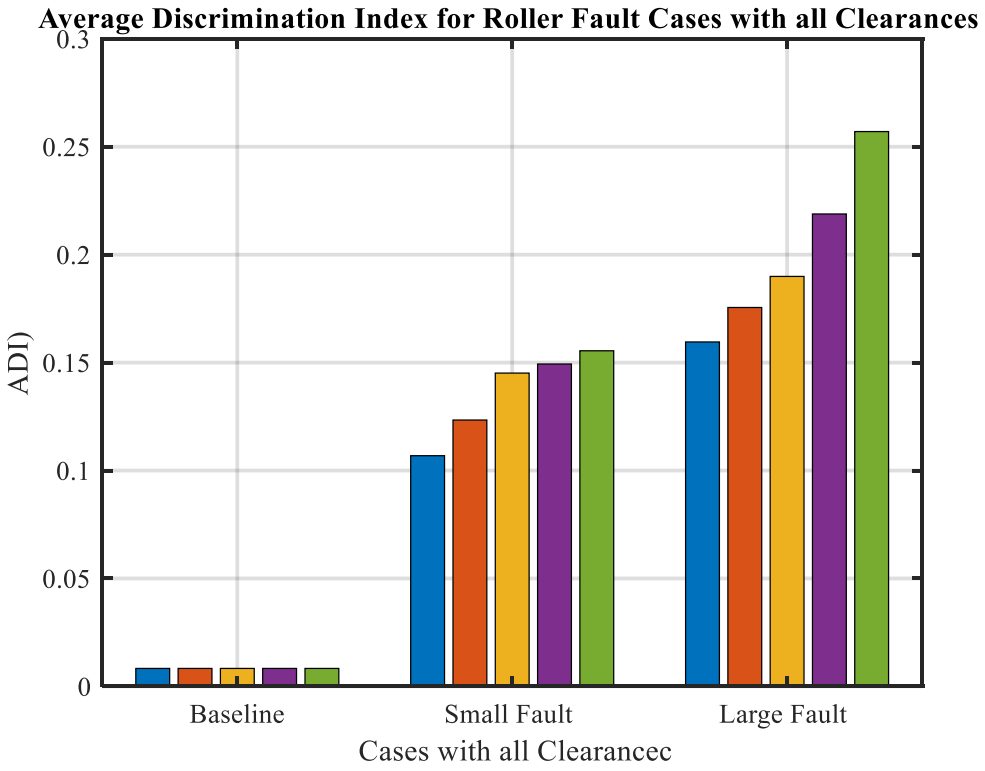

Figure 9-41 ADI Detection Results of Baseline, Small and Large Roller Fault

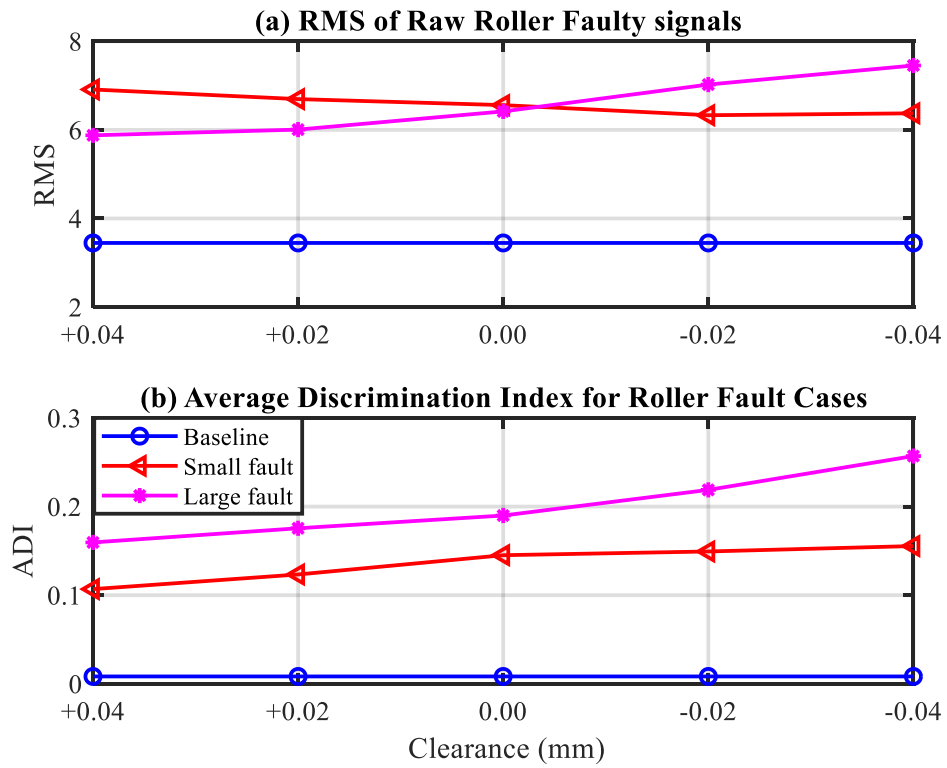

Figure 9-42 RMS of Raw Signal and ADI Detection Results for Roller Fault Cases

9.4 Summary

In this study, to accurately evaluate the performance of CCNN, based on simulated data and measured real data, several tests have been carried out. Two types of synthetic (simulated) signal were generating. The first simulation test started using a simple periodic signal to assess the capability of CCNN in detecting frequency variation. In the second simulation test, to simulate bearing fault signal (cyclo-stationary), impact signal was generated and used to train and then with another signal used with a change in frequency for testing the network. This test is carried out to assess the ability of CCNN in learning features from impact signals and detecting the anomalies when it happens. The results show that the method is effective in learning the features and detecting the anomalies even in high noise levels, the network is very effective when the noise level is low, however still effective and capable in detecting the anomalies when the signal to noise ratio SNR reached -6.

Moreover, the applications of the Componential Coding neural network to measured real data have demonstrated that the network has good performance in anomaly detection and fault discrimination. The networks perform well in fault detection and discrimination is achieved for both simple fault cases of the out race and for the rolling element of the bearings. The advantage of CCNN is that it can be used with minimum knowledge about data, also can be applied to raw vibration data directly. This will greatly improve the efficiency of implementing condition monitoring. It has been demonstrated with both simulated and real data that when a trained network is applied to the fault data, it reveals the novelty accurately. Moreover, the CCNN does not require sensors to record angular position when condition-monitoring data is collected since it incorporates a feature of translation invariance.

CHAPTER TEN

SEMI-AUTOMATED WAVELET DATA ANALYSIS BASED ON ADAPTIVE THRESHOLDING METHOD

In this chapter, in order to verify the proposed method DD-DWT alongside with the developed thresholding method (HSR), the collected vibration data is analysed with the proposed method using DD-DWT. Also the performance of both DT-ℂ𝑊𝑇 and DDD-DWT were investigated and compared to DD-DWT. The results obtained presented with the discussion. This chapter will assess the capability of the method for bearings fault diagnostics. The results indicate that DD-DWT is an effective and reliable method with harmonic to signal ratio thresholding technique and it can be a suitable method for fault detection and diagnosis of REBs.

10.1 Introduction

Vibration data from machinery systems can contain important information for fault diagnostic. However, in reality, complex systems produce non-stationary and nonlinear vibration data, also unavoidable background noise in practice makes the task of identifying fault features, not an easy task. Therefore, denoising the measured data can increase the SNR and leads to a better diagnosis. However, the limitations of both time and frequency domain analysis, which were described in chapter one, have increased the need to investigate the time-frequency domain in analysing such complex data. Wavelet transform has been for a long time an effective method for data analysis in both time and frequency domain.

The adaptive property gives wavelet the ability to obtain high time but low-frequency resolution at high frequencies, whilst, at low frequencies, it can obtain low time but high-frequency resolution. This makes WT suitable for vibration signal produced from bearings where better frequency resolution is desired at low frequencies to detect the slow time-varying peaks, while good time resolution is sought at high frequencies to track rapid changing transients[62].

Wavelet gives an excellent representation for several types of vibration signals that containing jumps and spikes (singularities) at different scales or resolutions, it provides optimal sparse representation for such signals, the sparsity comes from the fact that since wavelets oscillate locally, only wavelets overlapping a singularity will have large wavelet coefficients whilst the rest of other coefficients will have smaller coefficients [108].

One of the key issues in the application of WT for feature extraction is to select the appropriate coefficient from the decomposed levels in which the components of interest exist. One of the well-known methods to illuminate the unwanted component is thresholding. Thresholding in the transformation domain is an effective method for illuminating unwanted components based on the decomposition levels. Different families of wavelets are available for different tasks, in practice, there is no ideal way to adopt the best suitable wavelet type for a particular application or task. This is usually is conducted by matching the expected shape of peaks to be detected with the suitable wavelet which has similar shape[61]. Daubechies wavelet [289] or any wavelet analogous to Daubechies is a reasonable choice and was applied in this research, as it has the greatest matching property to vibration produced by faulty bearing with near-exponentially damped vibration responses and non-stationary features [177, 290]. Figure 10-1 shows the matching properties of Daubechies WT to spikes produced by a faulty bearing.

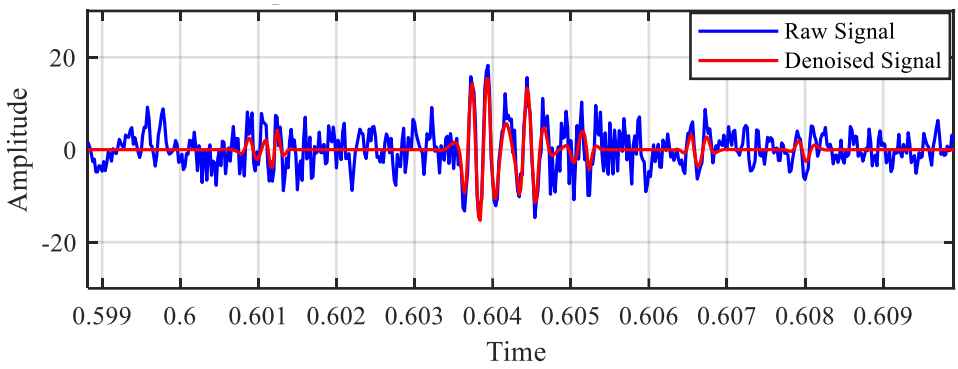

Figure 10-1 Daubechies wavelet

In this chapter, to extract the diagnostic features, the collected data will be analysed using Dual Density DWT and another two types of expensive wavelets, Dual-Tree Complex Wavelet Transform (DT-ℂWT), and Dual Density Dual-Tree Complex Wavelet Transform (DDD-ℂTW). A comparison between theses wavelets will be taking a place in this chapter to highlight the effectiveness and reliability of DD-DWT in analysing the collected vibration data with HSR thresholding method. The collected data will be decomposed into several levels. It was found that scale numbers should be calculated based on data length by which scale numbers should not be greater than as $log_2(N)$ where N is the data length[291]. Another approach can be used to calculate the scale numbers based on sampling frequency. Based on the Nyquist rule as the original signal have a bandwidth of $F_s/2$ and the first level will be $(0 - F_s/4 \,\&F_s/4 - F_s/2)$, whilst the level 2 will have a bandwidth of $(0 - F_s/8 \,\&F_s/8 - F_s/4)$ and so on until the desired lowest level is reached. In this approach, the number of decomposition levels is to be set based on the basis of desired frequency band features. In this study, the data length is 30 seconds, the sampling frequency is 50 KHz, and thus, the signal $F_s/2$ (25 KHz) will be decomposed into five levels. An estimation of the initial threshold value is calculated for each level independently and iteratively increased by a fixed amount. The implementing steps of the proposed method are described in Chapter Six section (6.4.1).

10.1.1 The implementation of the proposed method

To evaluate the effectiveness and performance DD-DWT with the developed method thresholding in data denoising, a comparison is carried out between three types of expensive wavelets using the proposed thresholding method. The experimental signal will be analysed using Wavelet Transform, whilst, the developed thresholding method will be used to denoise the signal. The envelope spectrum technique will be applied based on the whole band for all cases. The WTs decomposes

the experimental vibration signals collected from bearings and subsequently denoise the signal using the developed adaptive level-dependant threshold on the wavelet coefficients.

10.2 Enhancement of Diagnostic Features Using DD-DWT

The double density DWT was developed as a less expansive version of the undecimated DWT. Also, DD-DWT has very smooth wavelets and it is nearly shift-invariant. This property is important for extracting non-stationary periodical peaks. Another property is the reduced frequency aliasing effects which claimed to be effective for detecting harmonic features and makes the DD-DWT well suited for applications such as non-stationary signal processing that rolling bearing produces. It has more wavelets than necessary which give a narrower spacing between adjacent wavelets within the same scale and is less redundant than undecimated wavelet [113].

Therefore, DD-DWT was applied to the measured signals with five different clearance for each fault case ranged from +0.04 to -0.04. The signal decomposed into 5 levels and the optimal decomposition levels are selected based on the energy distribution as principal criterion. Thus coefficients with the largest values have been selected[191]. The proposed thresholding method used for denoising the selected wavelet coefficients. Figure 10-2 shows in (a) the time domain for both raw signal and the denoised signal for clearance -0.04. It can be seen that the shrinkage process left the faulty features nearly unchanged. Whilst in (b), both low and high-frequency components were nearly removed and the frequency band with high SNR was modified and the important information was preserved.

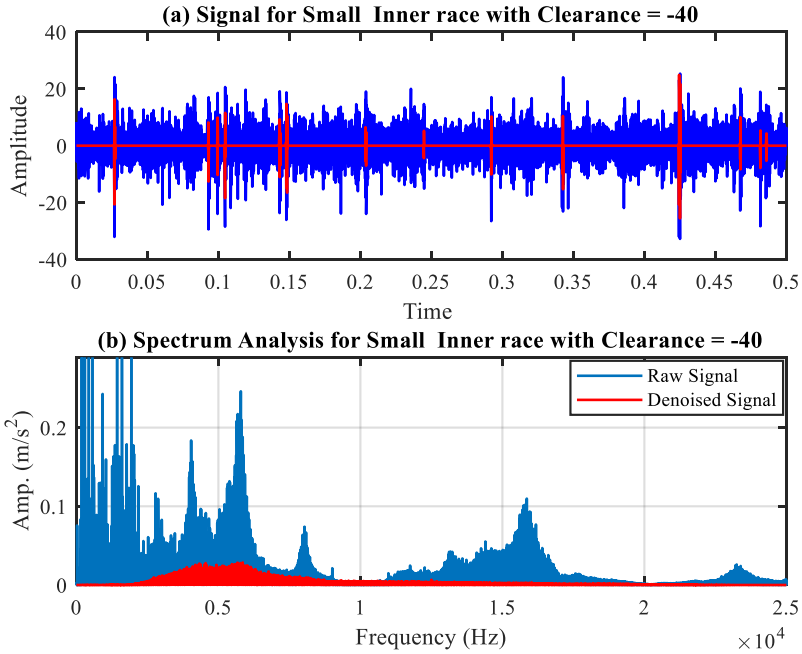

Figure 10-2 (a) Time Domain and (b) Spectrum of the Original and Denoised Signal

10.2.1 Outer race Small Fault (2mm)

Five sets of data for a defective bearing with small outer race fault are analysed using DD-DWT. The five sets belong to five clearances as detailed earlier, Figure 10-3 shows of the improvement of the first three harmonics to signal ratio compared to reconstructed coefficients. It shows the effectiveness of DD-DWD in noise suppression and enhancing the fault features, although the original signal of cases with a large internal clearance has a low harmonic ratio, the HSR threshold with DD-DWT has greatly improved the harmonic ratio.

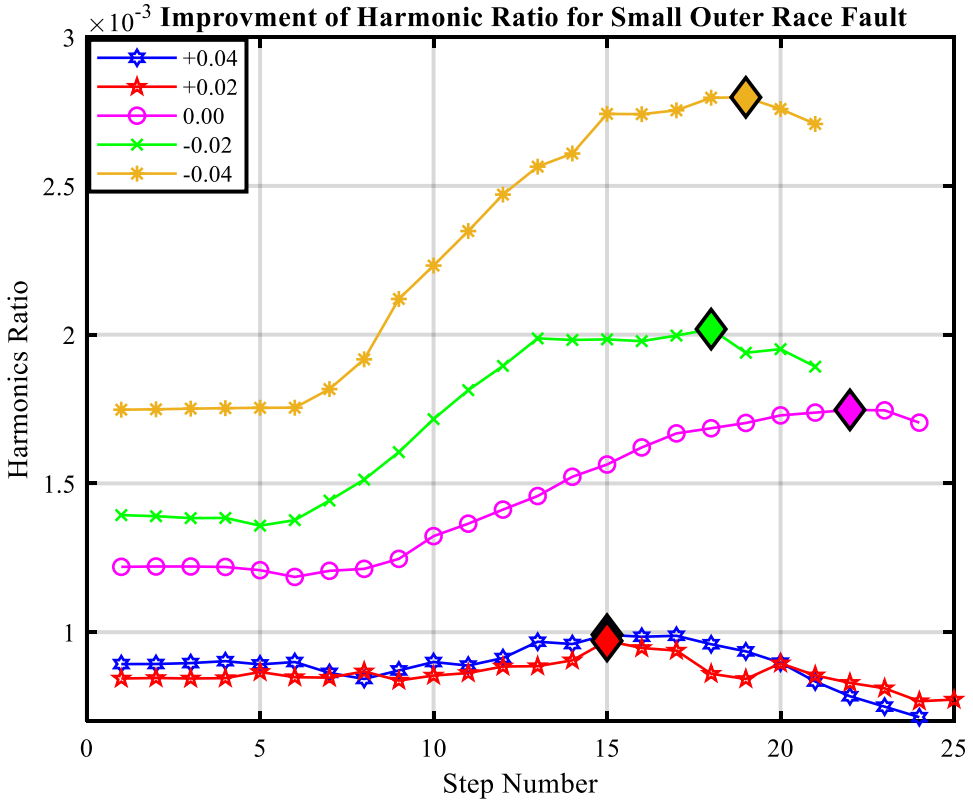

Figure 10-3 Improvement of Harmonic Ratio during Steps using DD-DWT

The improvement of fault harmonics ratio to signal ratio for all cases are illustrated in Figure 10-4, to quantify the improvement achieved, the percentage of improvement is estimated and expressed in the bars. It can be observed that the improvement has a linear relationship with the decline of the internal clearances. The case with clearance (-0.04 mm) reached the highest improvement by 62% while the case with the largest internal clearance of (+0.04mm) improved by 12%.

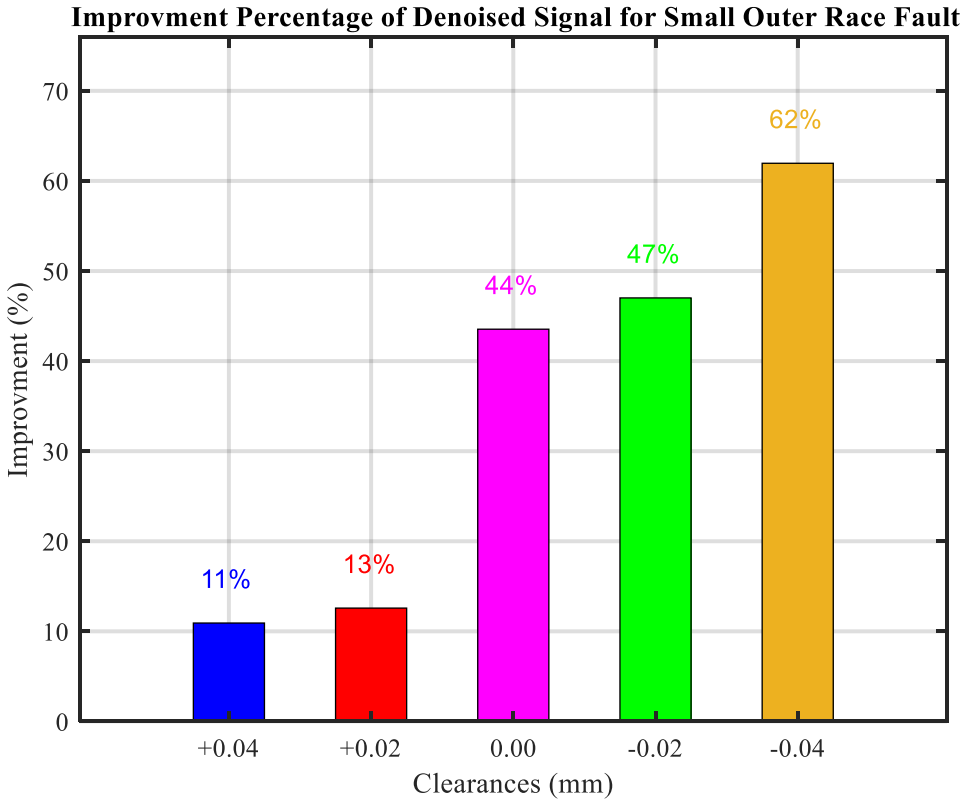

Figure 10-4 Improvement Percentage of Denoised Signal using DD-DWT

In this section, a comparison in the time domain is presented between original signals and denoised signals of small outer race faulty cases. Time domain of both the original signal and denoised signals illustrated in Figure 10-5 and Figure 10-6 respectively. Although few components left in the residual signals, the envelope spectrum shows that the diagnostic features are preserved and faults can be clearly identified from the denoised coefficients.

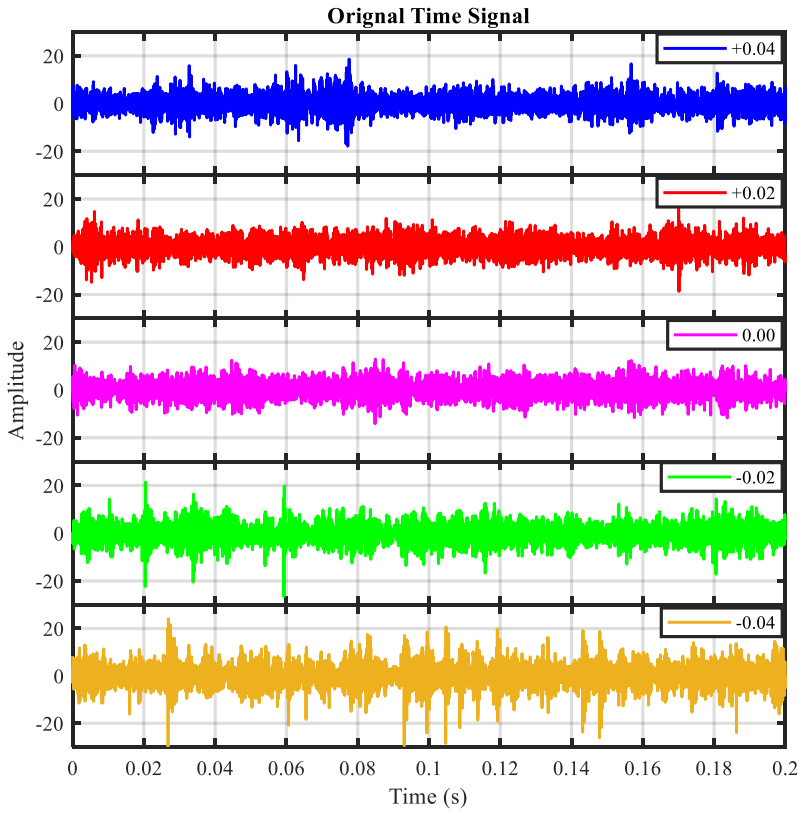

Figure 10-5 Time Domain of Small Outer Race Fault signals

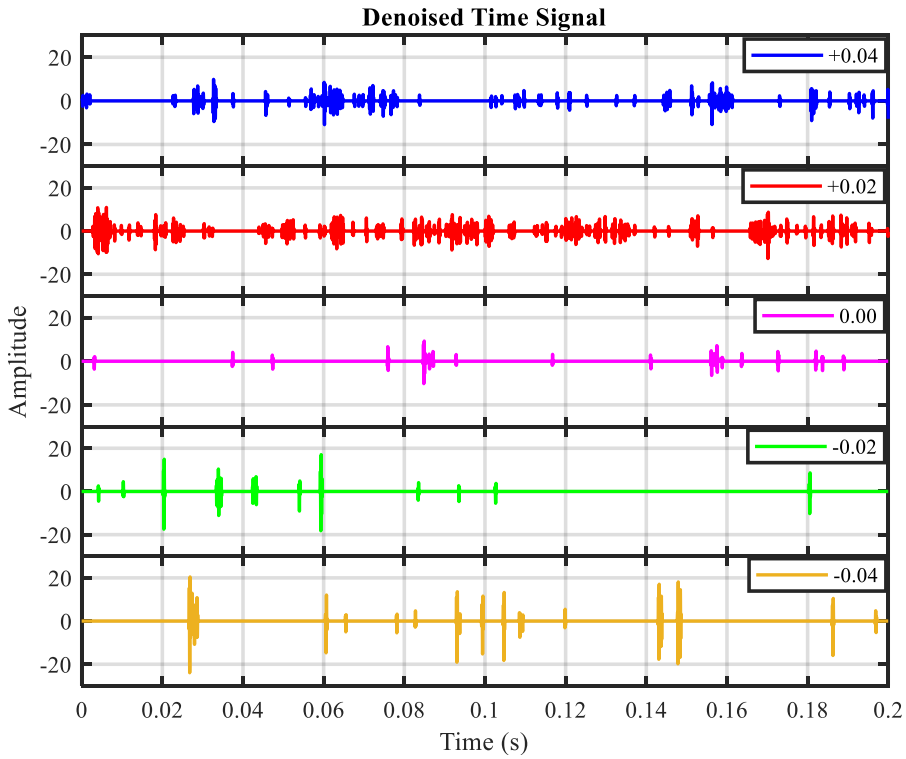

Figure 10-6 Time Domain of Denoised Small Outer Race Fault signals

For more details, the denoised signal was demodulated using envelope analysis. As it can be seen in the envelope analysis of the denoised coefficients using the proposed method that the faulty features become much clearer and the noise reduced dramatically. the results obtained using the DD-DWT are illustrated in Figure 10-7, Figure 10-8, Figure 10-9, Figure 10-10 and Figure 10-11 below for all cases. It can be observed that the sidebands due to unbalanced shaft become clearer in Figure 10-9 and Figure 10-10, furthermore, the first harmonic become clearer in the large clearance cases where the faulty peaks are very weak as seen in Figure 10-7 and Figure 10-8.

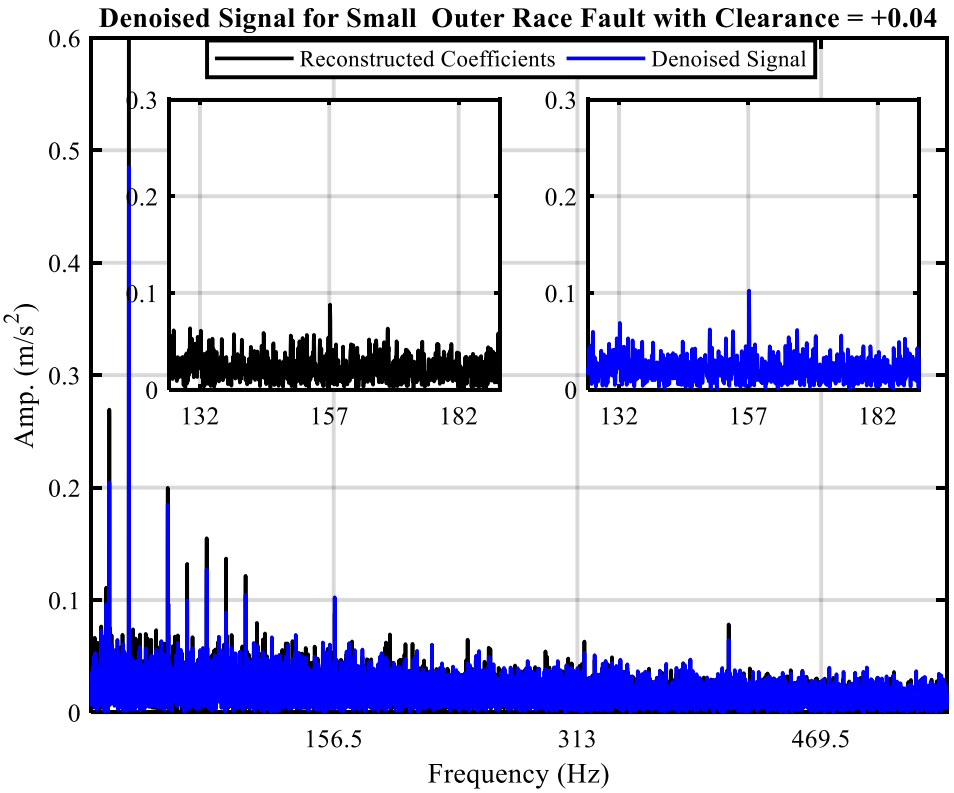

Figure 10-7 Envelope Spectrum of Reconstructed and Denoised Coefficients

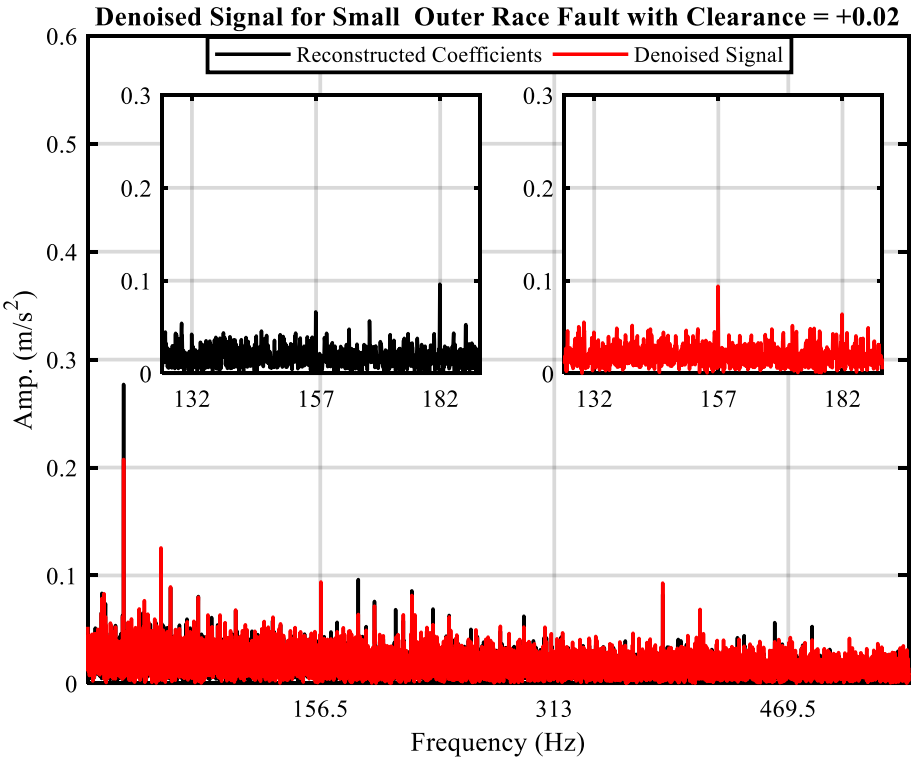

Figure 10-8 Envelope Spectrum of Reconstructed and Denoised Coefficients

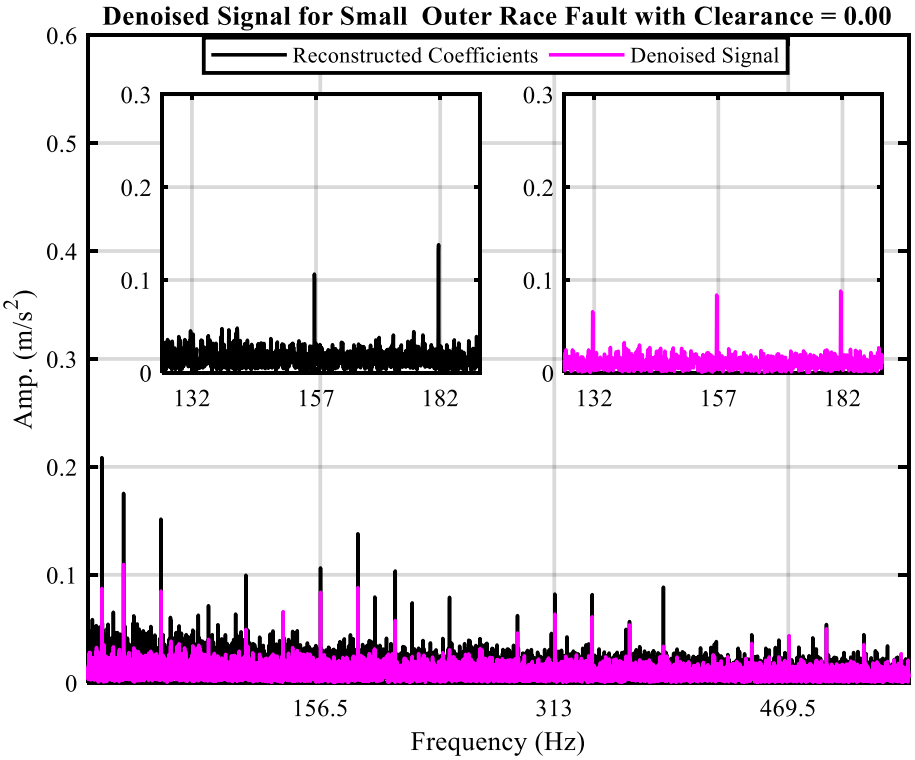

Figure 10-9 Envelope Spectrum of Reconstructed and Denoised Coefficients

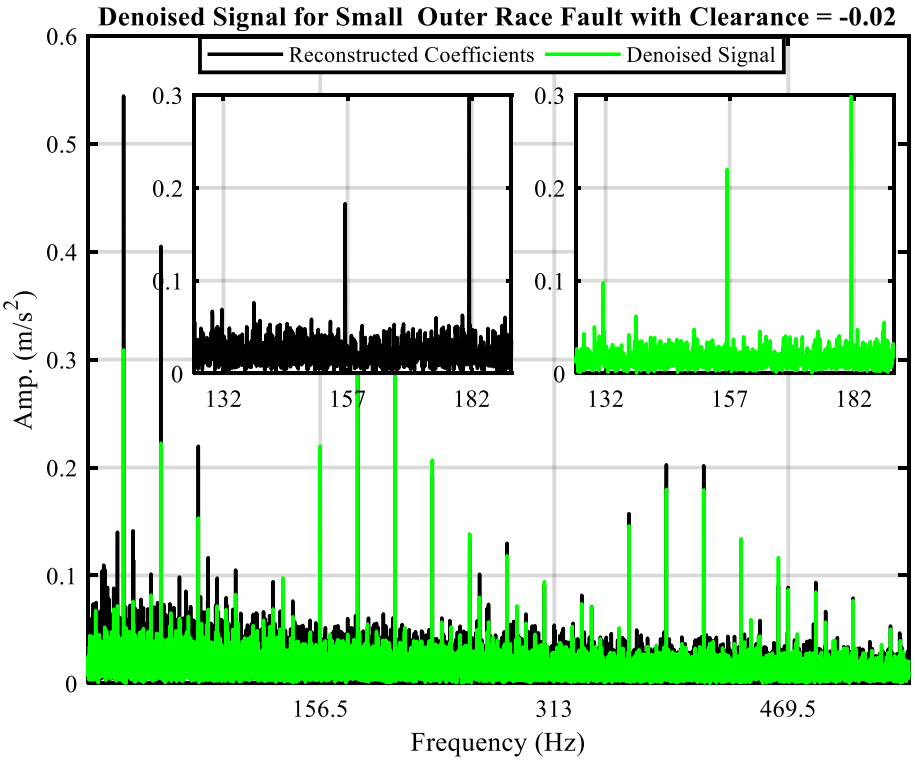

Figure 10-10 Envelope Spectrum of Reconstructed and Denoised Coefficients

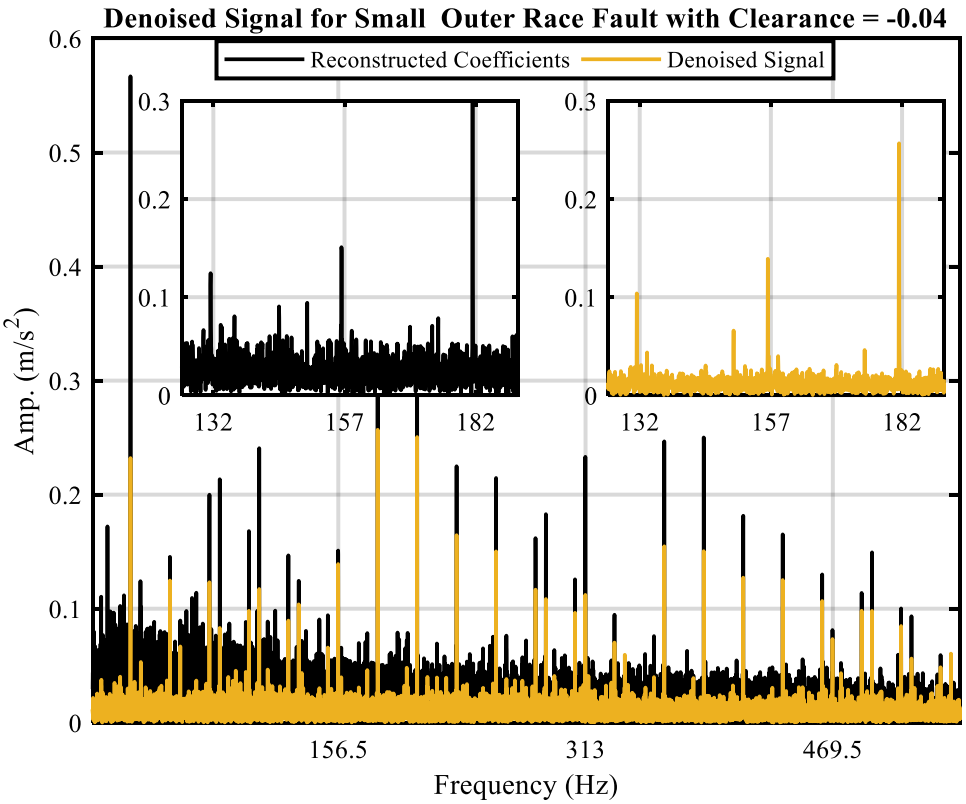

Figure 10-11 Envelope Spectrum of Reconstructed and Denoised Coefficients

Results of the envelope spectrum for Original signal, Reconstructed and denoised coefficients are compared and shown in Figure 10-12, Figure 10-13 and Figure 10-14 respectively. It can be observed that, although the diagnostic features improved in the selected coefficients when processed with DD-DWT, however, a remarkable improvement was achieved in the denoised signal. Furthermore, the noise level significantly reduced compared to both original and reconstructed coefficients.

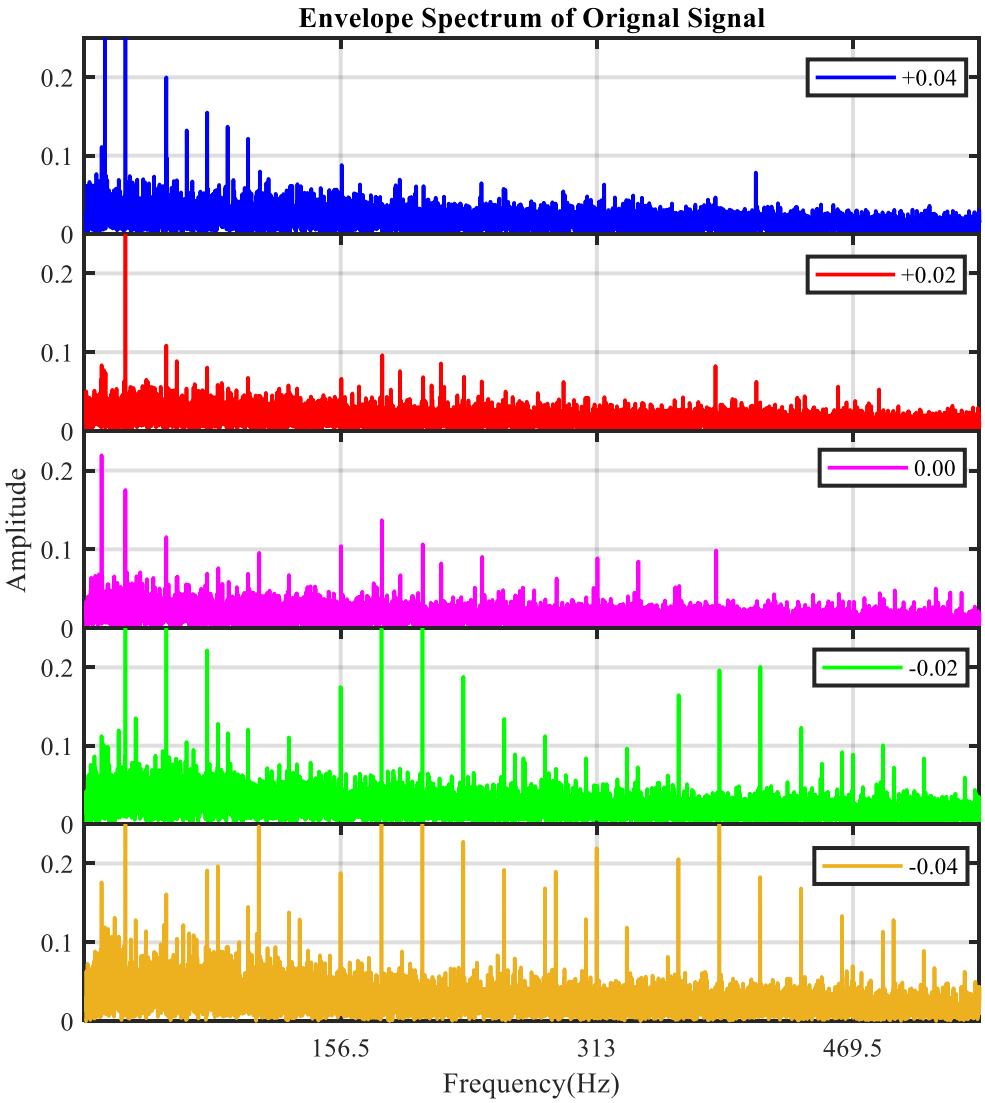

Figure 10-12 Envelope Spectrum of Original Signals for Small Outer Race Fault

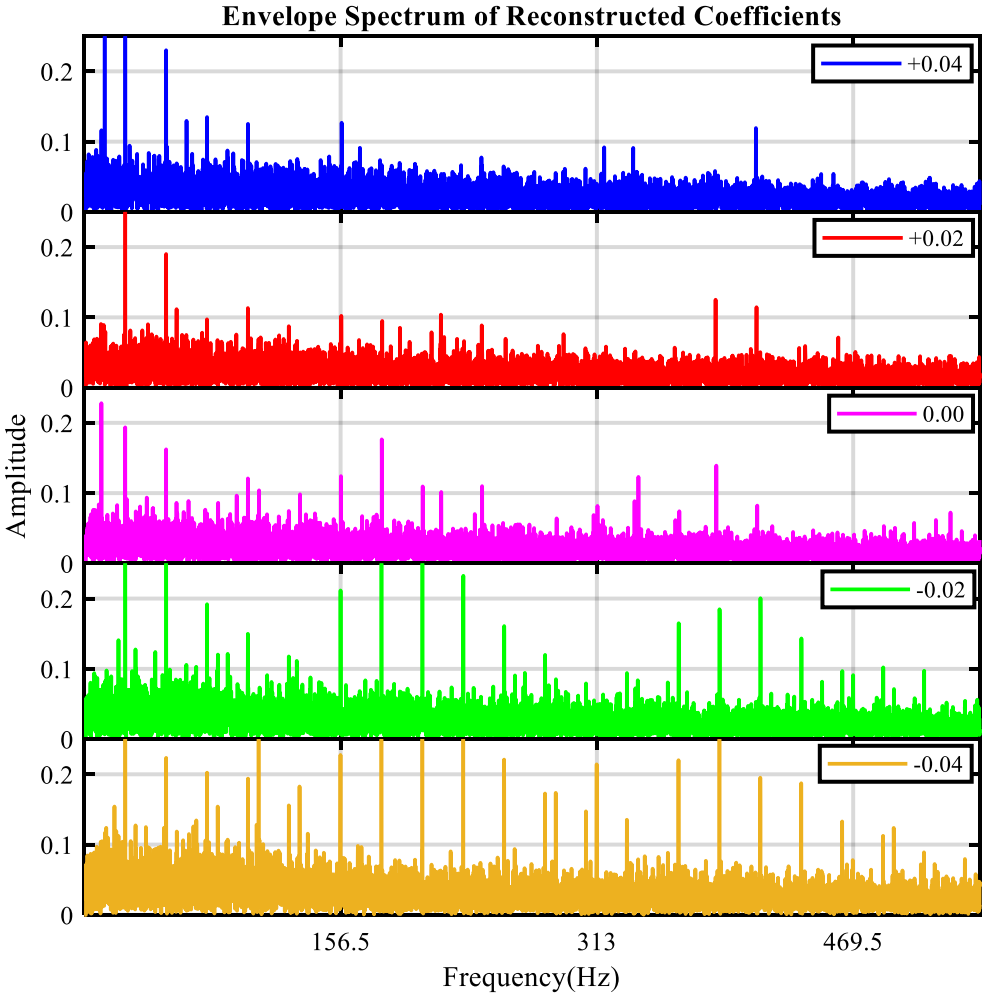

Figure 10-13 Envelope Spectrum of Reconstructed Coefficients for Small Outer Race Fault

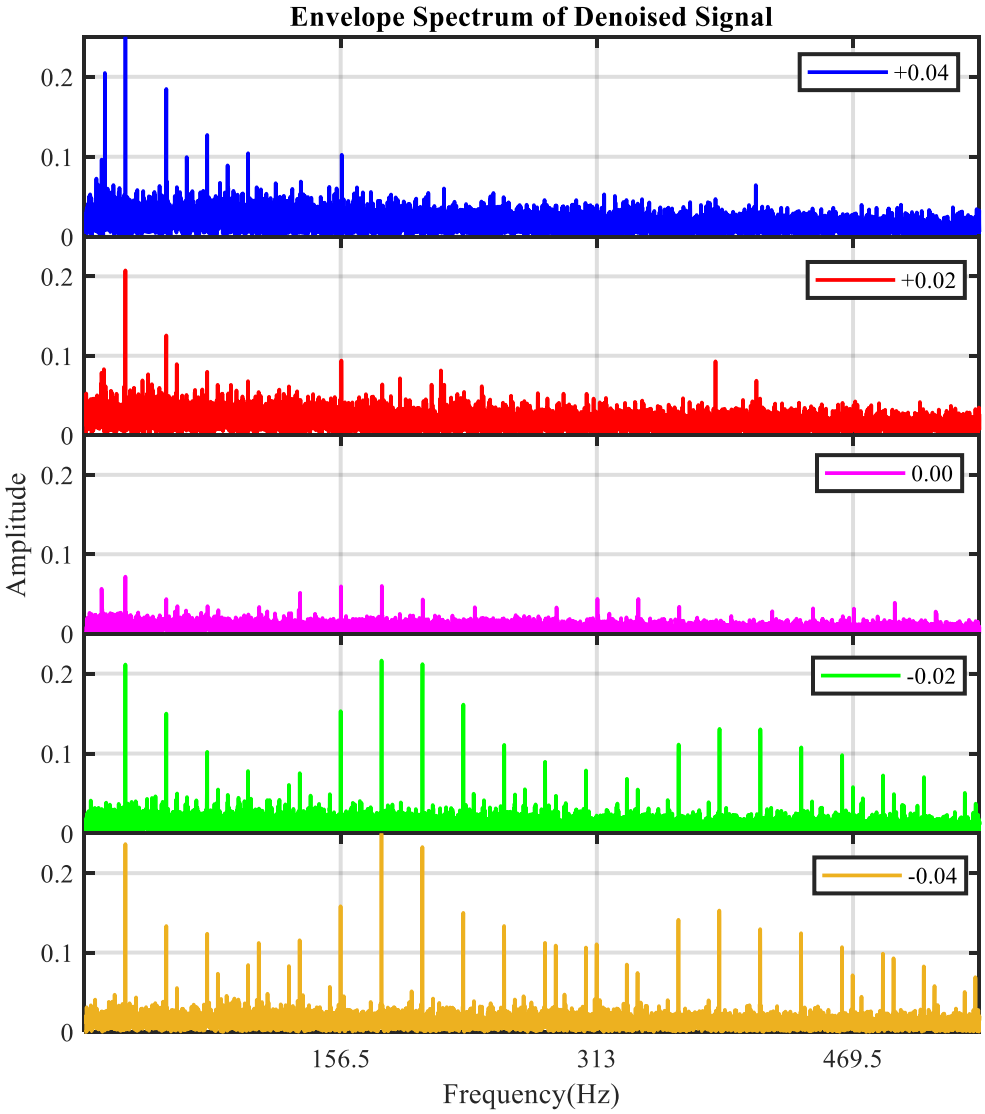

Figure 10-14 Envelope Spectrum of Denoised Signals for Small Outer Race Fault

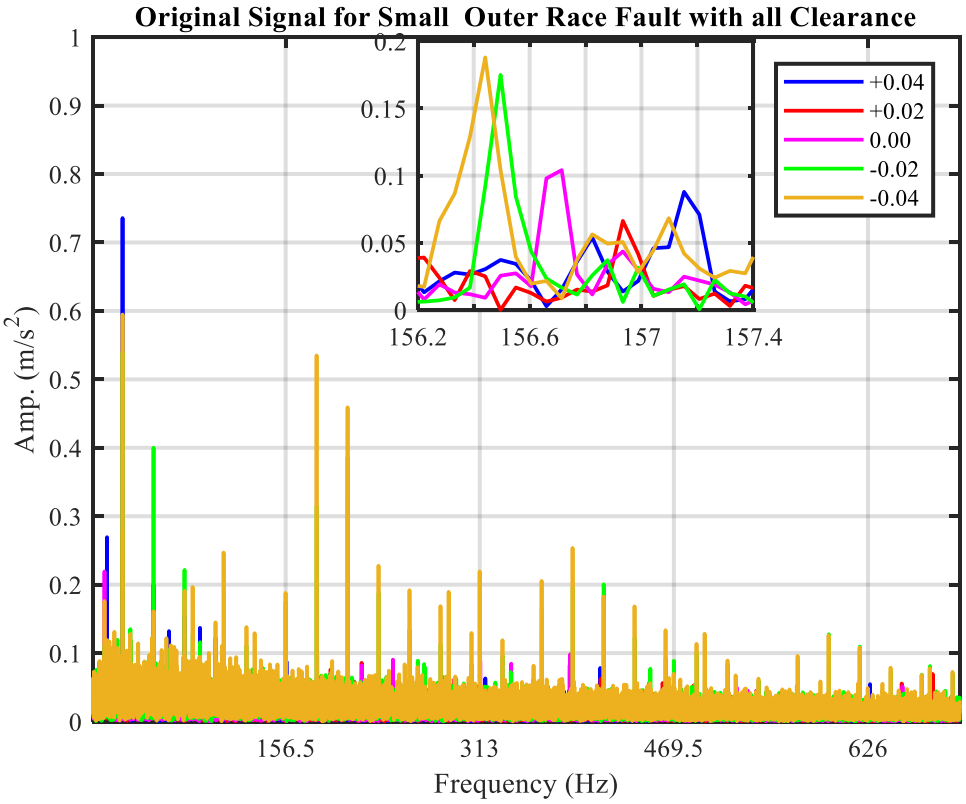

Figure 10-15 Envelope Spectrum of Original Signal for all Clearance Cases

Original signals for small outer race fault with all clearance cases are sown in Figure 10-15 and the selected coefficients after the reconstruction process are shown in Figure 10-16. The improvement can be observed, however, in the denoised version of the selected coefficients using HSR technique, as shown in Figure 10-17, a remarkable improvement can be seen for all clearance cases.

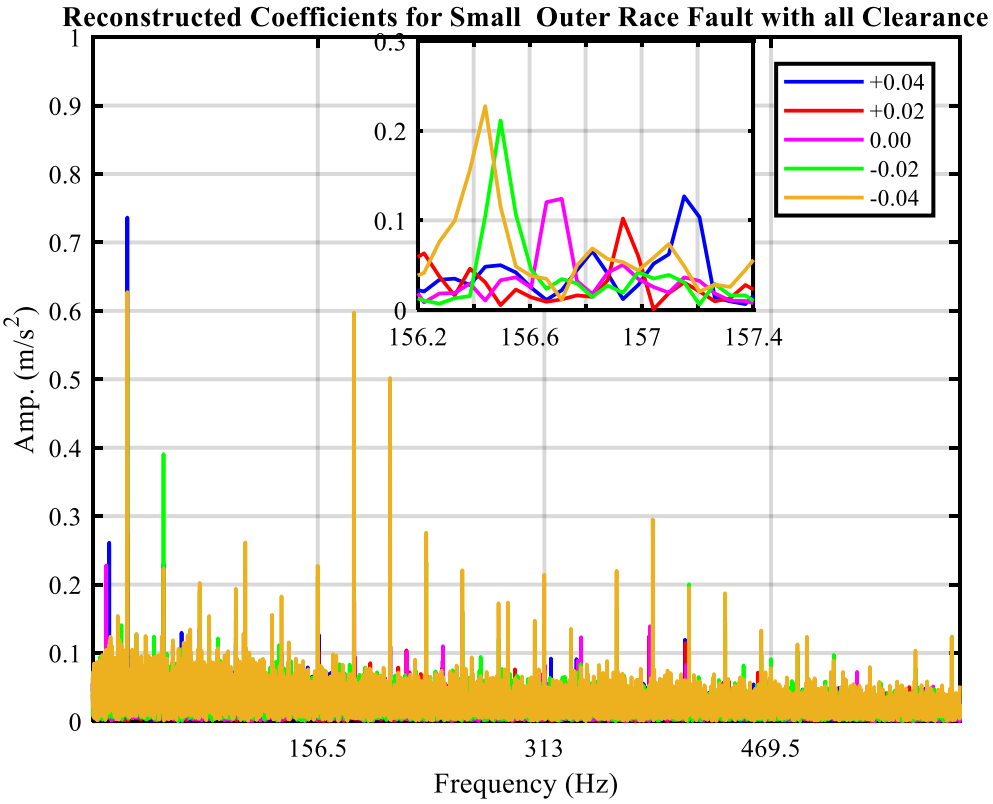

Figure 10-16 Envelope Spectrum of Reconstructed Coefficients for Small Outer Race Fault

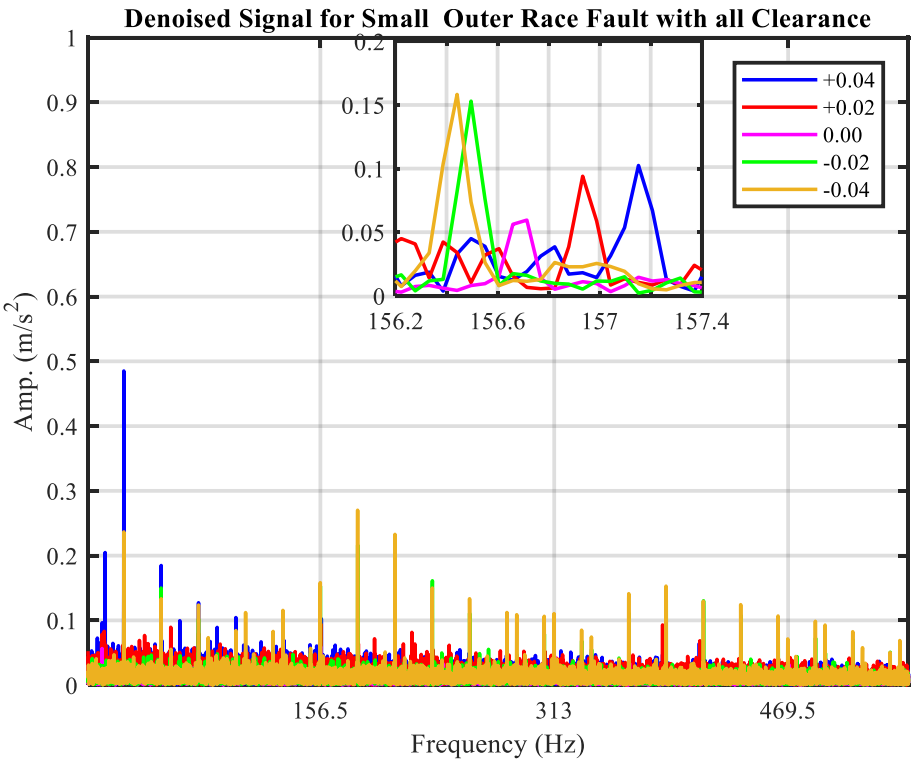

Figure 10-17 Envelope Spectrum of Denoised Signal for Small Outer Race

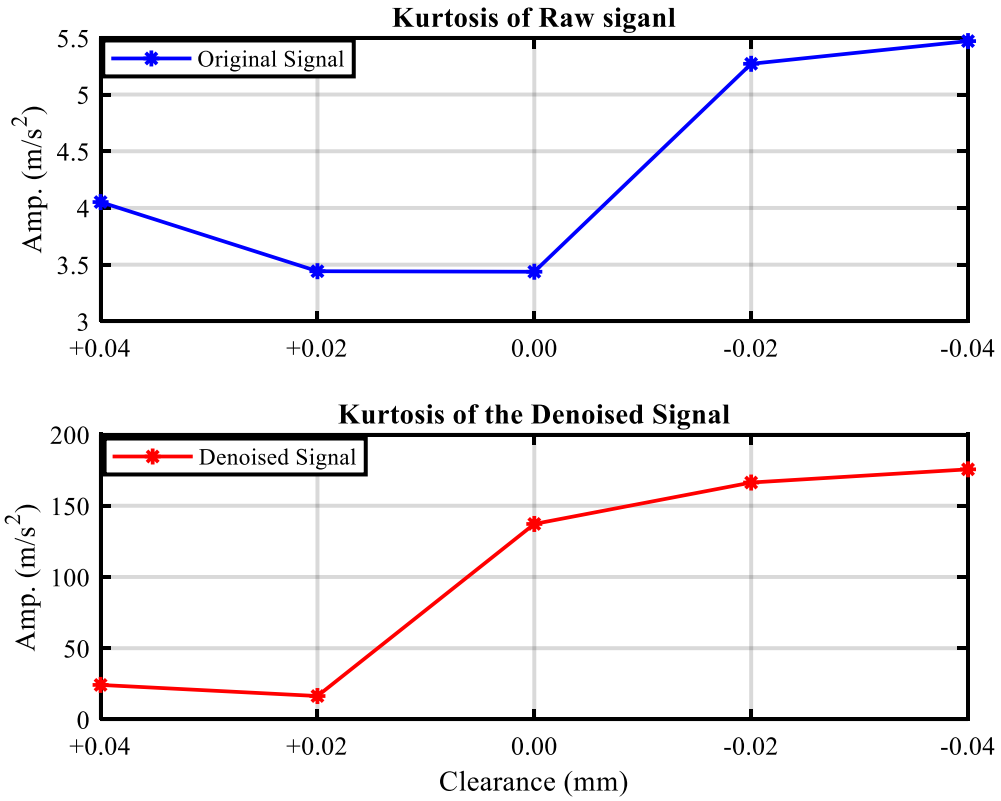

Figure 10-18 RMS of Kurtosis and Denoised Outer Race Signals with all Clearance Cases

The impulsiveness of the denoised signal is measured and compared with the original signal using Kurtosis, the results obtained shown in Figure 10-18. The kurtosis value has significantly increased to reach nearly n amplitude of 180 compared to the original signal.

10.2.2 Small Roller Fault

In this section, five datasets with different internal clearances gathered from a defective bearing with small roller fault are analysed using HSR with DD-DWT. The improvement achieved of the first three harmonics to signal ratio is depicted in Figure 10-19. It shows the effectiveness of HSR with DD-DWD in noise suppression and enhancing the diagnostic fault features. The original signal of cases with a large internal clearance has low harmonic ratio compared to preloaded cases (-0.02) and (0.04), however, the HSR threshold with DD-DWT has improved the ratio.

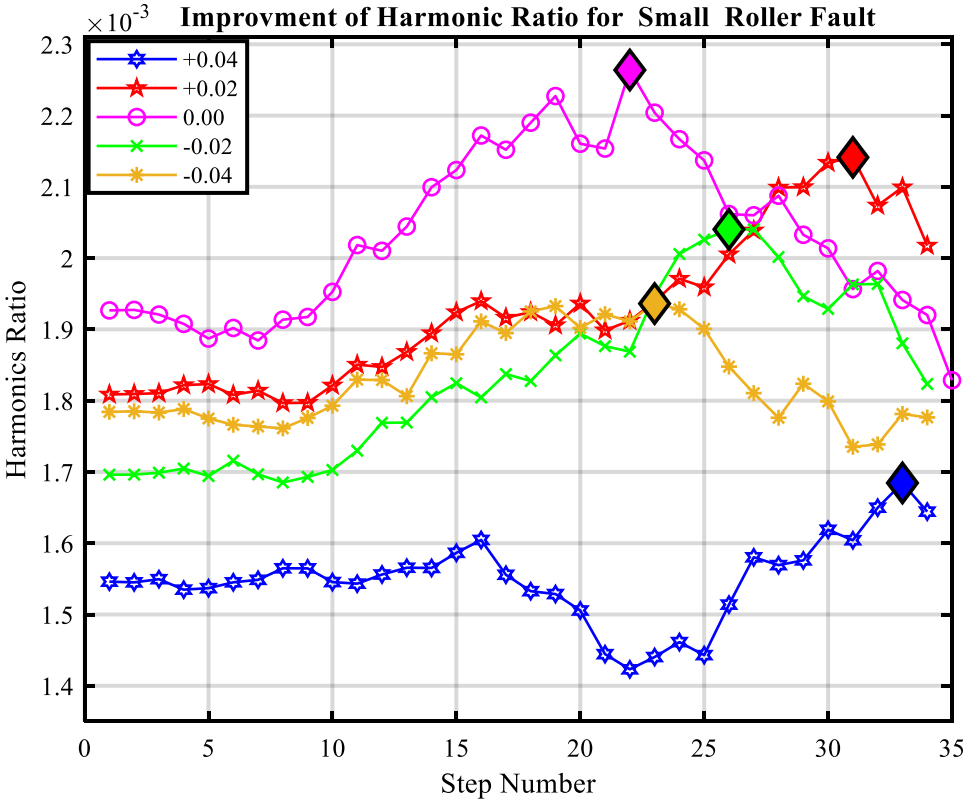

Figure 10-19 Improvement of the Harmonic to Signal Ratio for Small Roller Fault (2mm)

The percentage of improvement for all cases are illustrated in Figure 10-20, cases (+0.04) and (-0.04) has the lowest improvement by 9% while the other cases improved by around 20%.

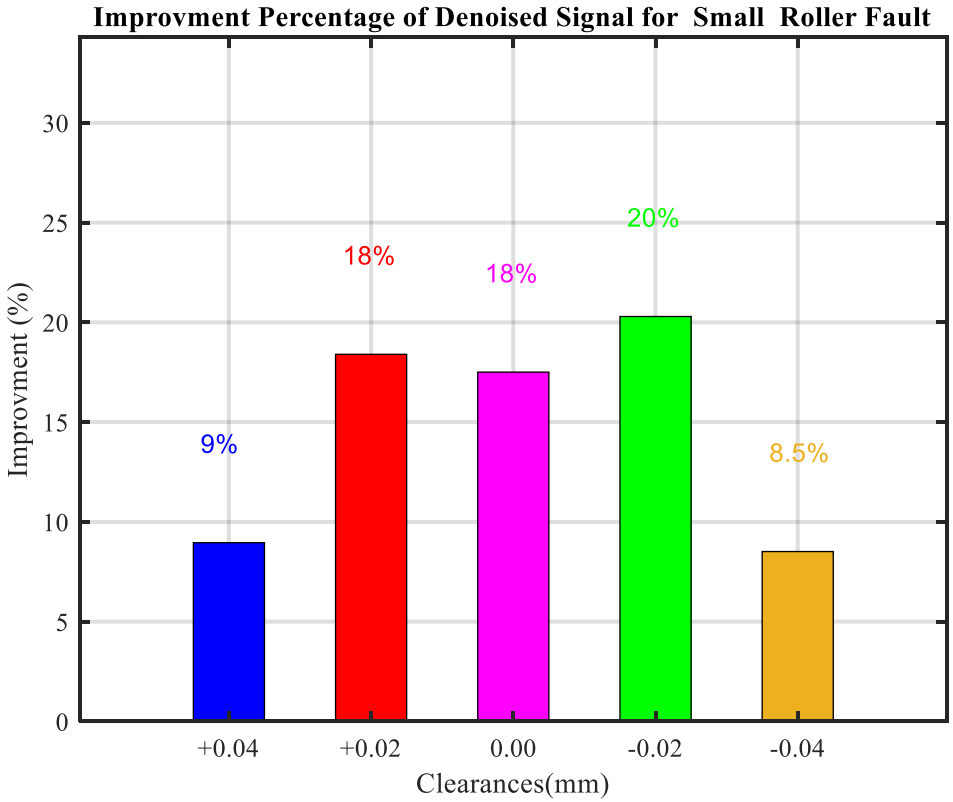

Figure 10-20 Percentage of improvement for Small Roller Fault (2mm)

Both time domain of the original signal and denoised signals are illustrated in Figure 10-21and Figure 10-22 respectively. Although few components left in the denoised signals, the envelope spectrum shows that the diagnostic features are preserved and faults can be clearly identified from the denoised coefficients.

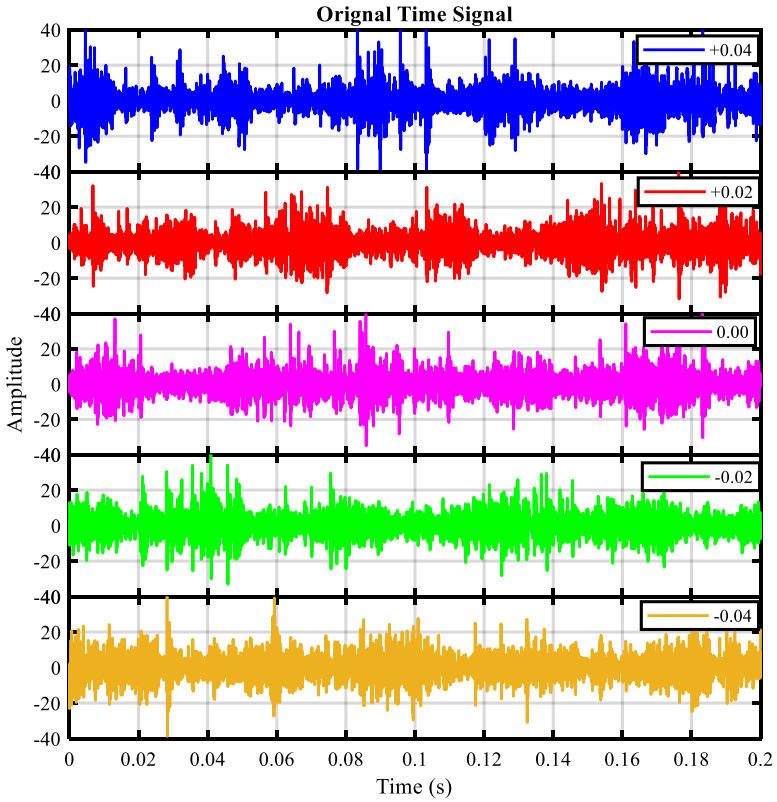

Figure 10-21 Time Domain of Raw Data for Small Roller Fault

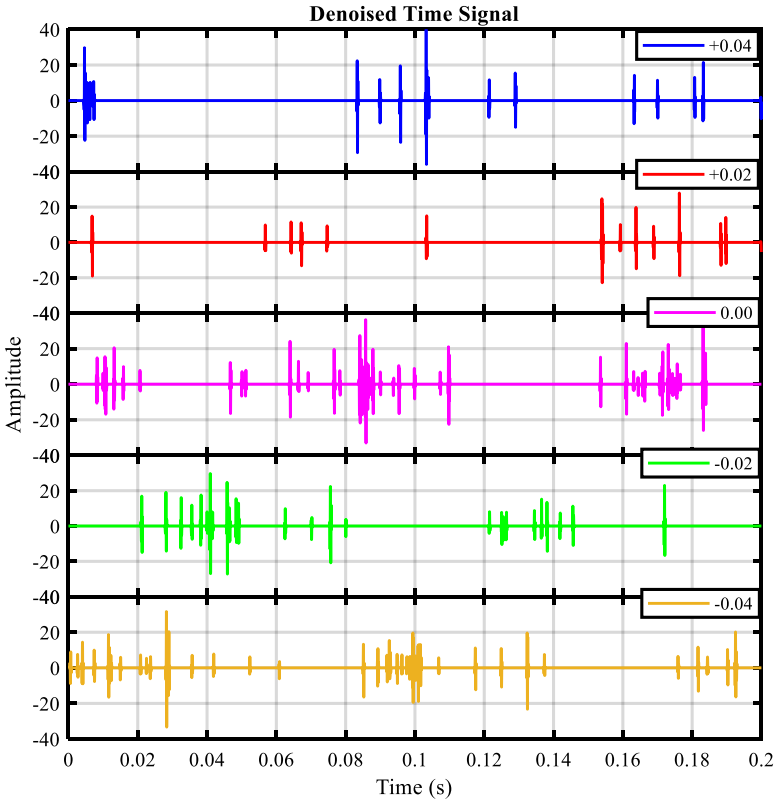

Figure 10-22 Time Domain of Denoised Signals for Small Roller Fault

Selected coefficient compared with the denoised coefficients for each case, Figure 10-23, Figure 10-24, Figure 10-25, Figure 10-26 and Figure 10-27 show the selected and the denoised coefficients for all clearance cases. The diagnostic features improved significantly in almost all cases particularly the noise reduced remarkably while diagnostic periodic features improved.

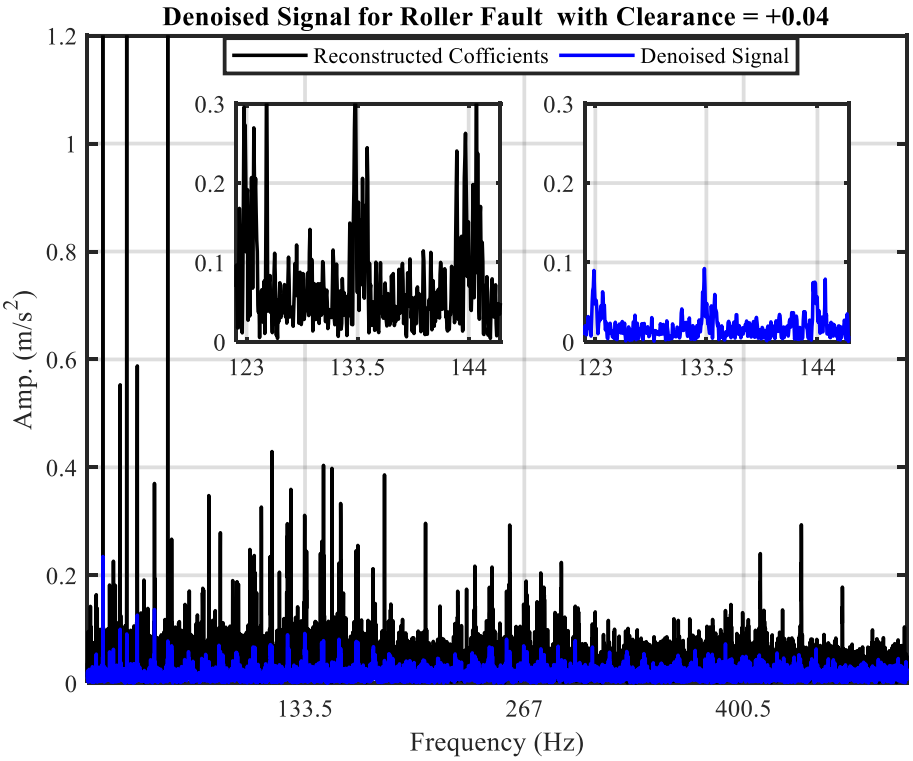

Figure 10-23 Envelope Spectrum of Reconstructed and Denoised Coefficients

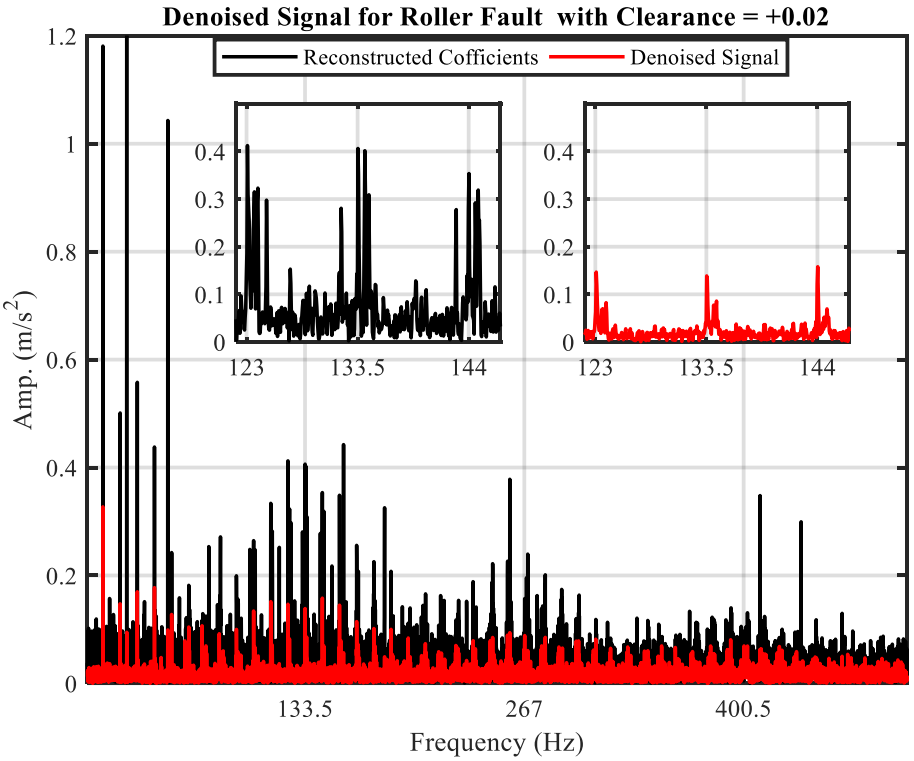

Figure 10-24 Envelope Spectrum of Reconstructed and Denoised Coefficients

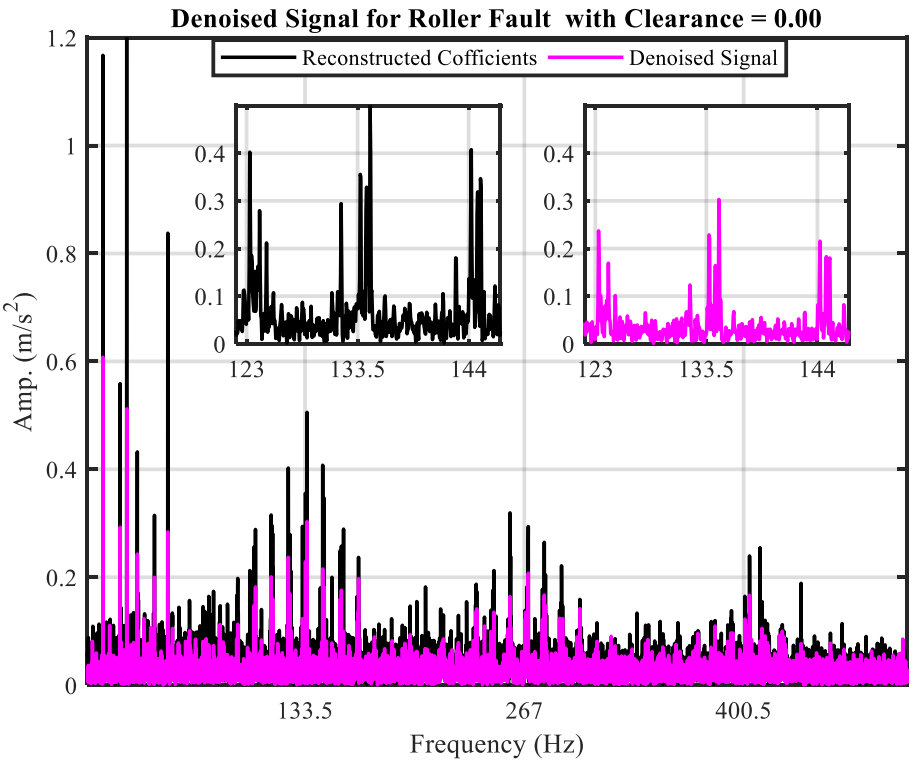

Figure 10-25 Envelope Spectrum of Reconstructed and Denoised Coefficients

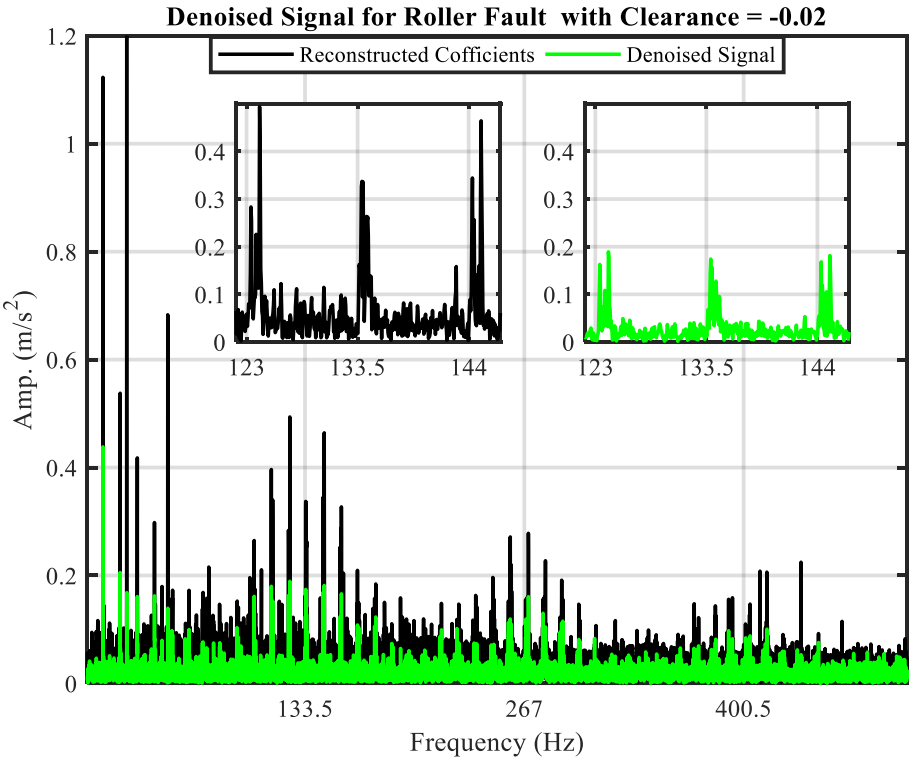

Figure 10-26 Envelope Spectrum of Reconstructed and Denoised Coefficients

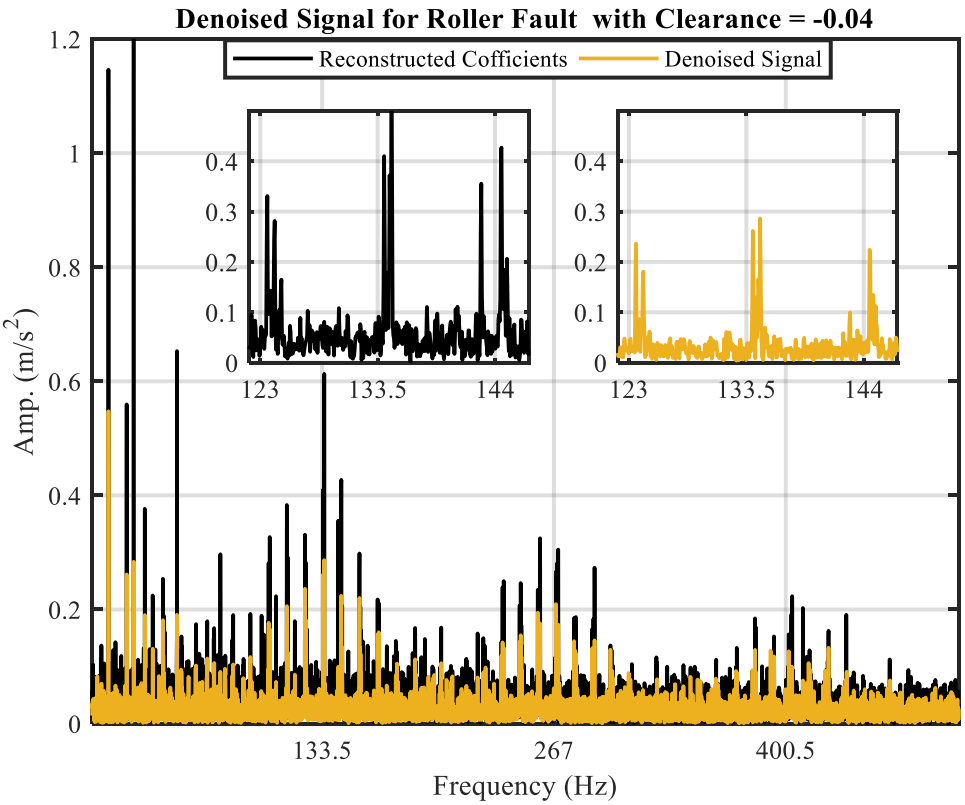

Figure 10-27 Envelope Spectrum of Reconstructed and Denoised Coefficients

Figure 10-28, describes the original signal of small roller fault with all clearance cases, Figure 10-29 shows the selected coefficients after reconstructions and Figure 10-30 shows the denoised version of the selected coefficients.

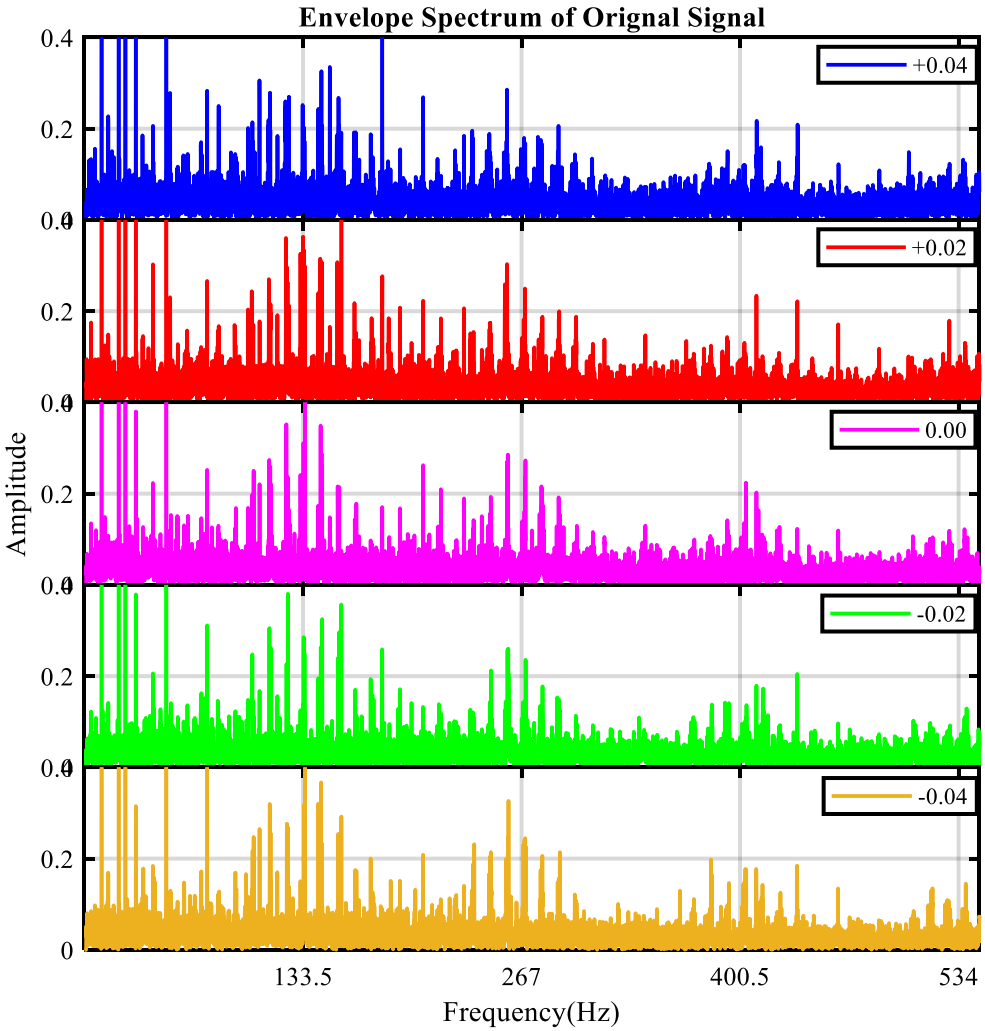

Figure 10-28 Envelope Spectrum of Original Signal for Roller Fault

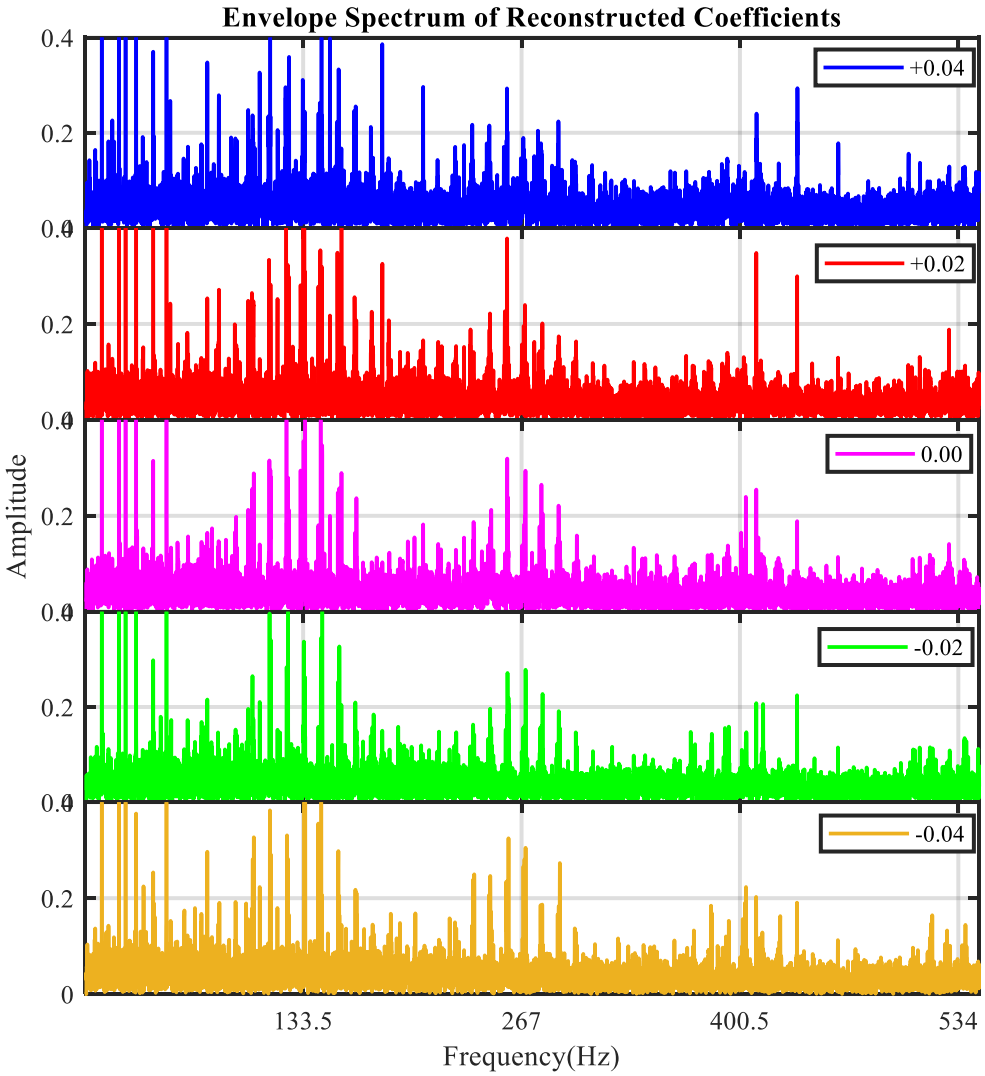

Figure 10-29 Envelope Spectrum of Reconstructed Coefficients

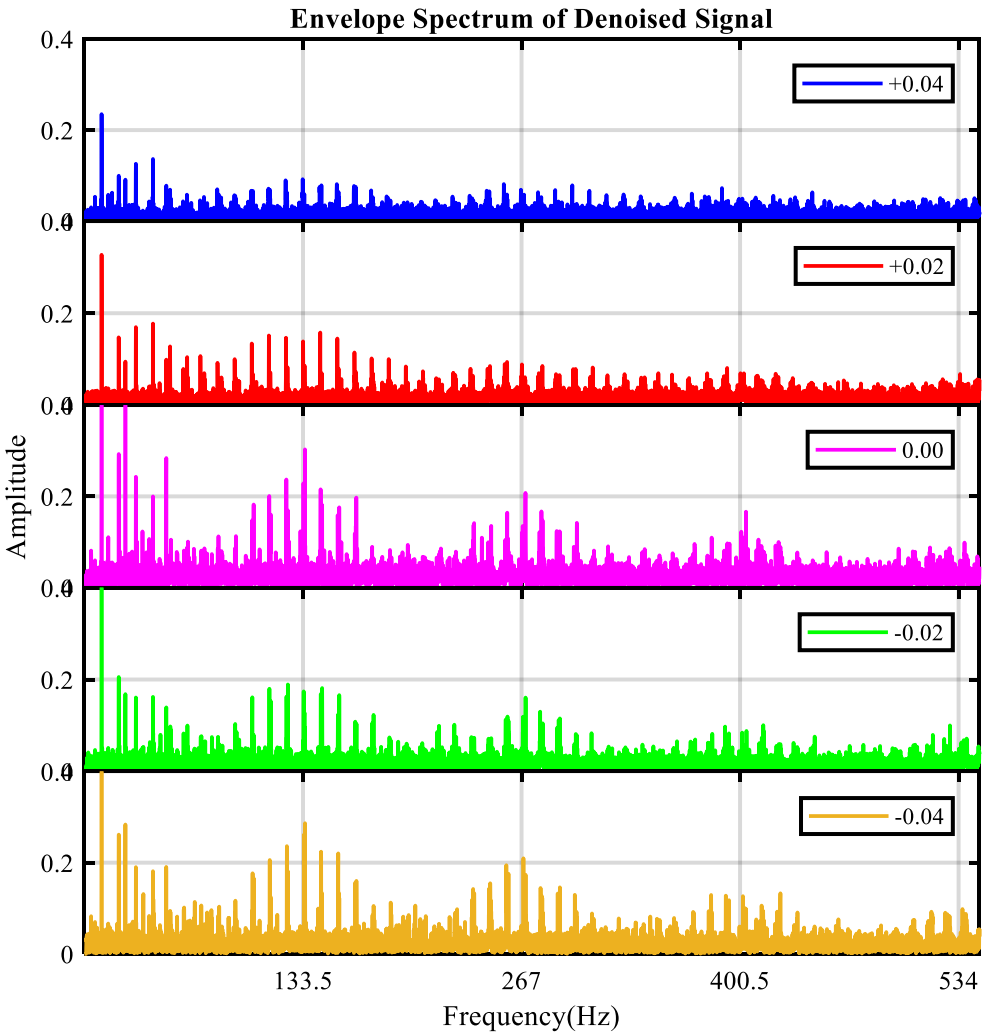

Figure 10-30 Envelope Spectrum of Denoised Coefficients for Small Roller Fault Using DD-DWT

The kurtosis of both original and denoised signals are calculated and as it can be seen in Figure 10-31, the impulsiveness of the denoised signal is increased is significantly increased compared to the original signal.

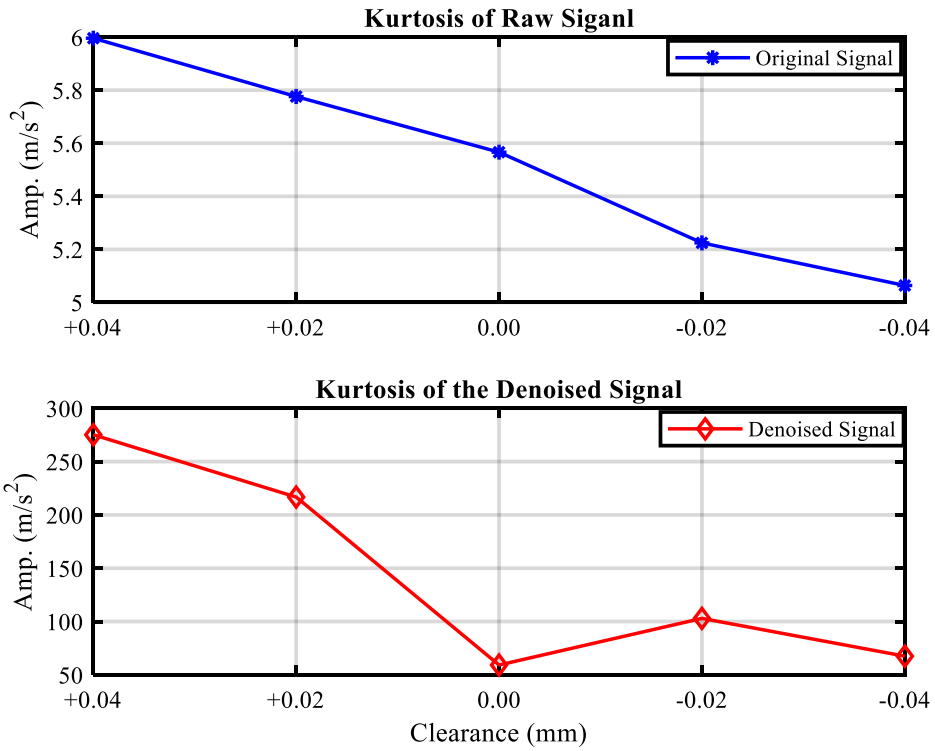

Figure 10-31 Kurtosis of Raw & Denoised Signal Using DD-DWT

10.3 Enhancement of Diagnostic Features Using DT-DWT

Following the same procedure adopted in DD-DWT, DT-ℂWT was applied to the measured signals with five different clearance for each fault case ranged from +0.04 to -0.04. The signal decomposed into 5 levels and the HSR thresholding method used for denoising the selected wavelet coefficients in the transformation domain. The analysis starts with outer race cases.

10.3.1 Small Outer Race fault

Figure 10-32 shows of the improvement of the first three harmonics to signal ratio, blue trace represents the largest clearance case and red trace represents the positive clearance with +0.02. Both have a low harmonic ratio, however, the case with clearance (+0.04) has improved more than the second case. The other three cases (0.00 to -0.04) have significantly improved. To quantify the improvement, Figure 10-33 illustrates the calculated percentage of improvement achieved and it can be seen that the all signals enhanced with 14% and 4.2% for the first two cases and with high different values of improvement for the rest of cases ranges from 42% to 60%.

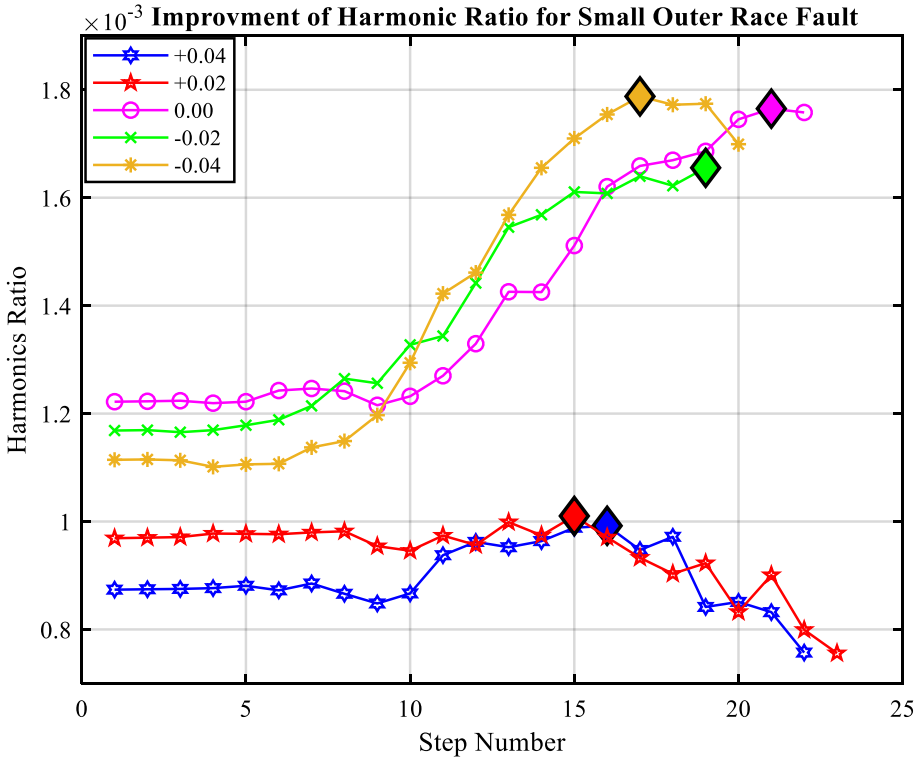

Figure 10-32 Improvement of Harmonic Ratio during Steps using HSR with DT-DWT

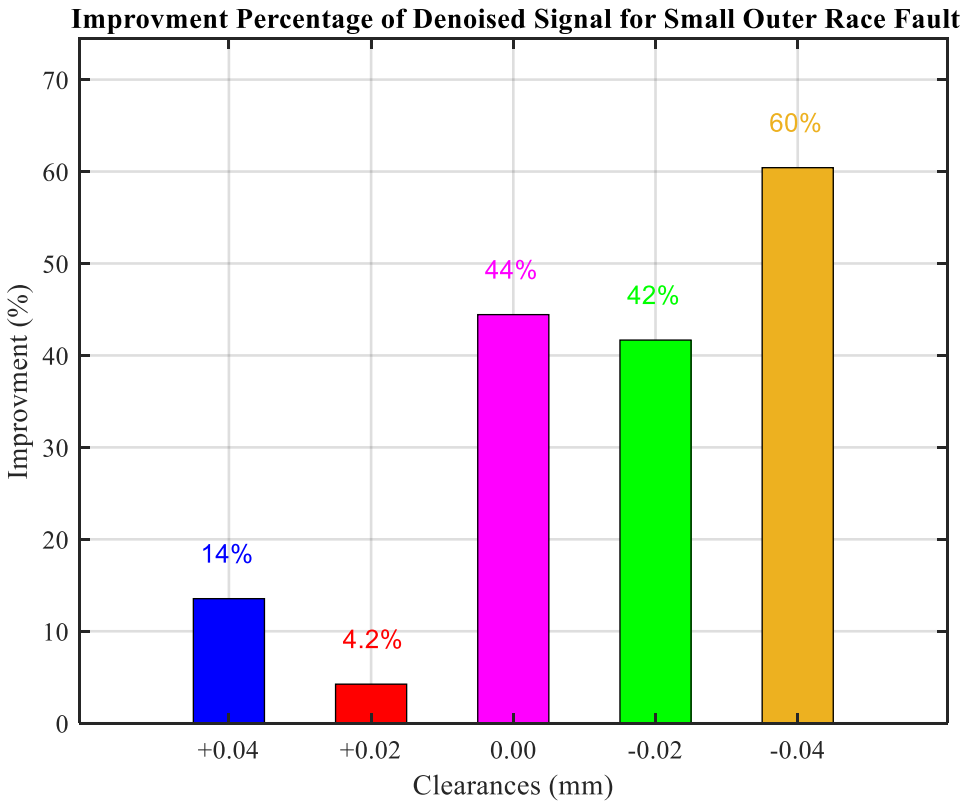

Figure 10-33 Improvement Percentage of Harmonic Ratio during Steps using HSR with DT-DWT

The denoised signal was demodulated using envelope analysis and results shown in Figure 10-34. The envelope analysis of the denoised coefficients using the proposed method shows that the faulty features become much clearer and the noise reduced dramatically. Furthermore, it can be seen that the shrinkage process left the faulty features nearly unchanged and overall the signal enhanced.

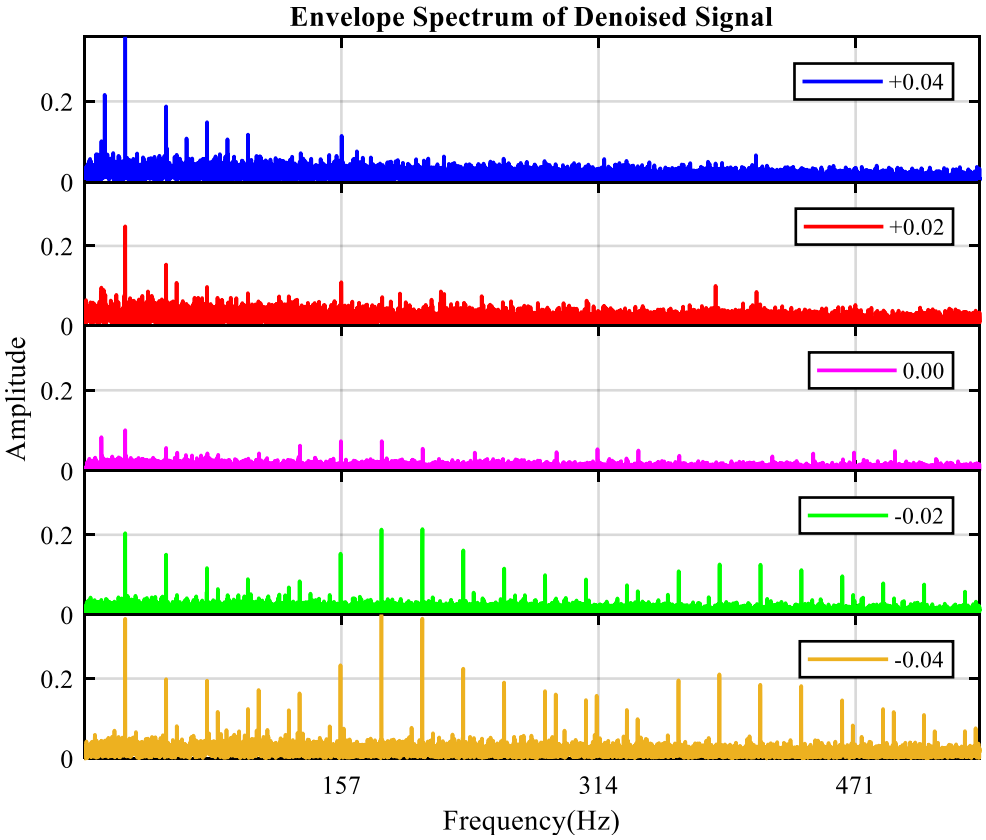

Figure 10-34 Envelope Spectrum of Denoised Signal for Small Outer Race using HSR with DT-DWT

10.3.2 Small Roller Fault

The improvement in the denoised harmonic ratio showed in Figure 10-35, all clearance cases have been enhanced with different improvement ratio levels. Clearance +0.02, 0.00 and -0.02 have higher improvement compared to the improvement of cases with clearance +0.04 and -0.04. However, the improvement achieved in all clearance cases. For more understanding, Figure 10-36 describes the improvement achieved in percentage values.

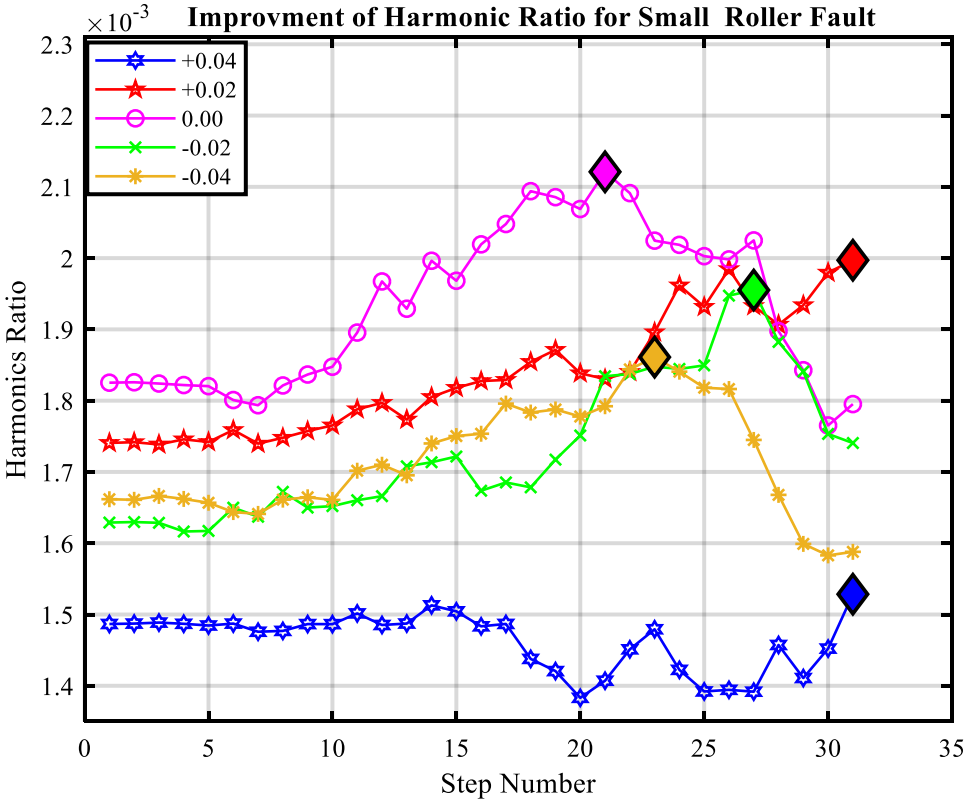

Figure 10-35 Improvement of Harmonic Ratio during Steps using HSR with DT-DWT

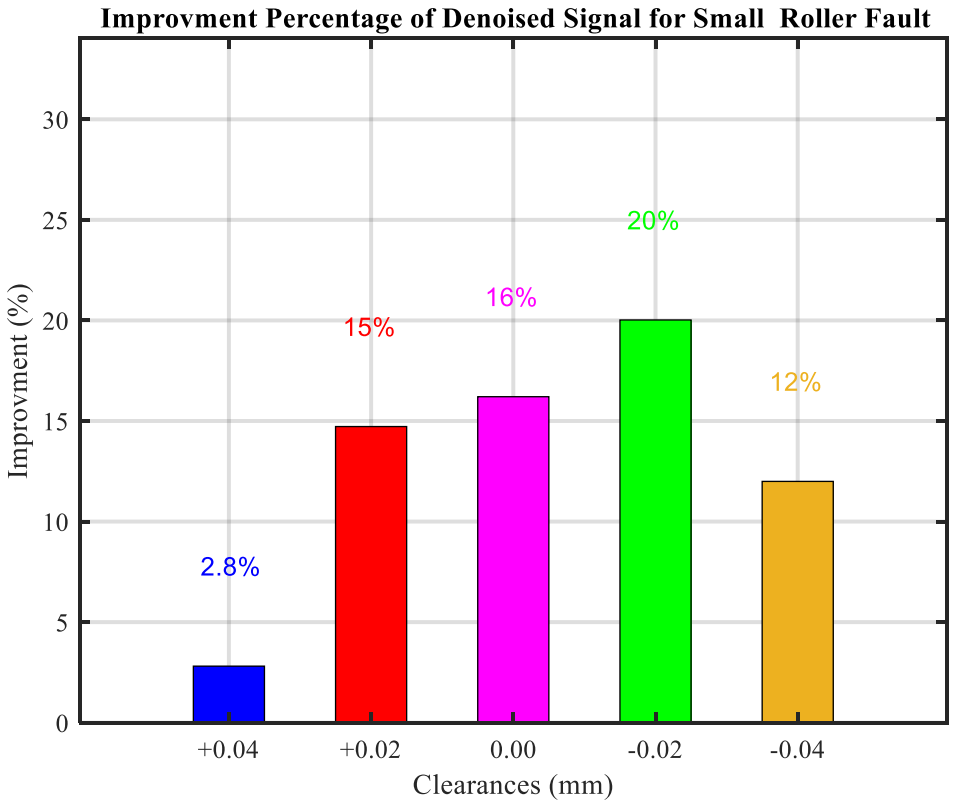

Figure 10-36 Improvement Percentage of Harmonic Ratio during Steps using HSR with DT-DWT

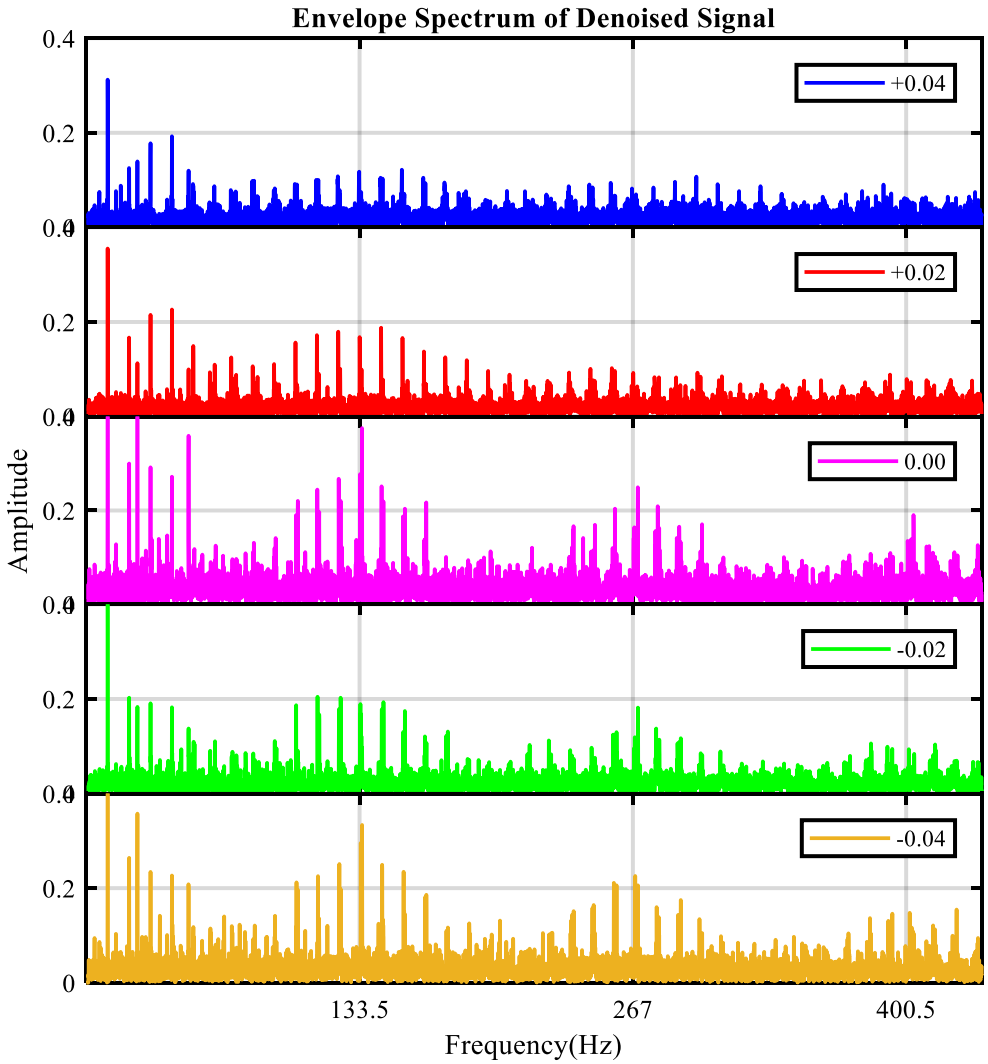

Figure 10-37 Envelope Spectrum of Denoised Signal for Small Outer Race using HSR with DT-DWT

The denoised version of the selected coefficients of small roller fault for all cases are illustrated in Figure 10-37 and it can be said that the noise reduced dramatically and the faulty peaks can easily be identified.

10.4 Enhancement of Diagnostic Features Using DDD-DWT

Double-density dual-tree complex wavelet transform is used in conjunction with envelope transform to evaluate the devolved technique. Following the same procedure adopted in DD-DWT, DT-ℂWT was applied to the measured signals with five different clearance for each fault case ranged from +0.04mm to -0.04mm. The signal decomposed into 5 levels and the HSR thresholding method used for denoising the selected wavelet coefficients in the transformation domain. The analysis starts with outer race cases.

10.4.1 Small Outer Race Fault

The diagnostic capability of the DDD-DWT based HSR threshold been validated using the data sets. Figure 10-38 shows of the improvement of the first three harmonics to signal ratio using HSR with DDD-DWT.

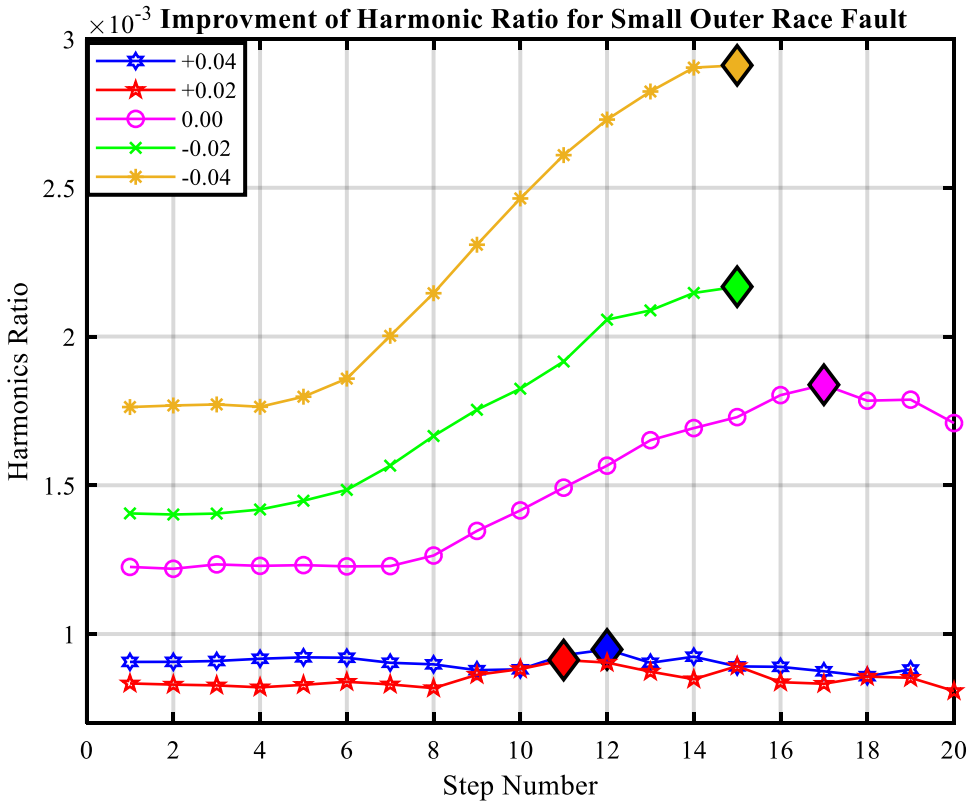

Figure 10-38 Improvement of Harmonic Ratio during Steps using HSR with DDD-DWT

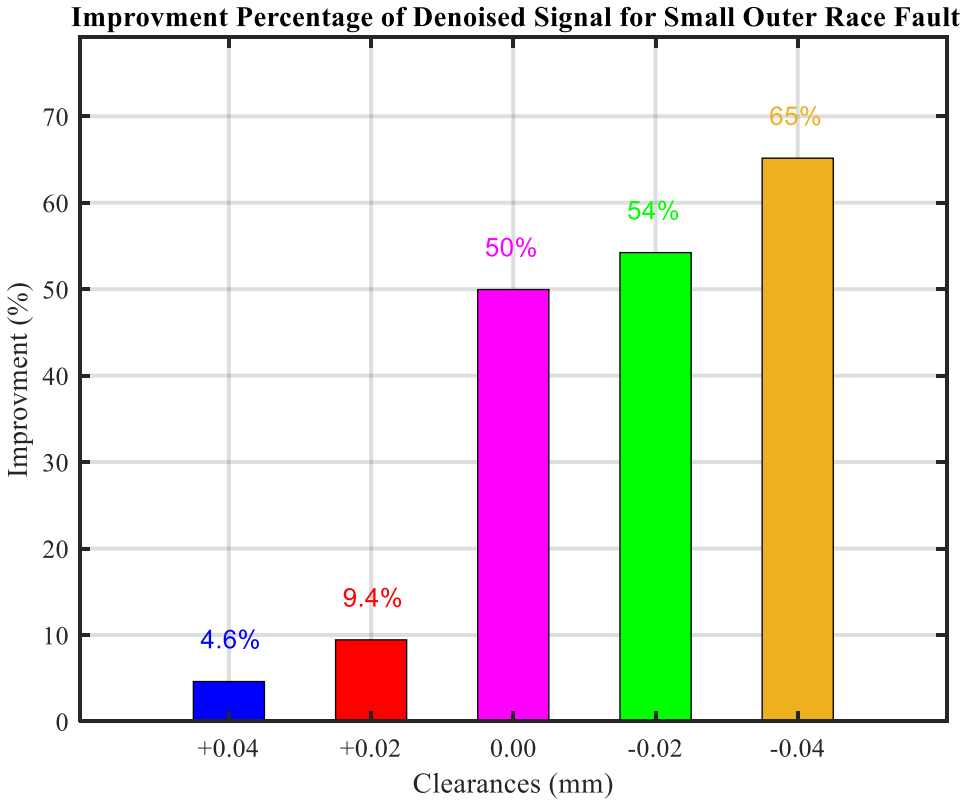

Figure 10-39 Improvement Percentage of Harmonic Ratio during Steps using HSR with DDD-DWT

Performance examination of DDD-DWT based HSR threshold is shown in percentage in Figure 10-39. A roughly exponential trend exists between the clearance variation and the improvement percentage showing that the method is capable of discrimination of the fault in different clearance cases.

Reductions in the noise level can be seen in Figure 10-40 which leads to clearer faulty peaks obtained from applying DDD_DWT based HSR threshold.

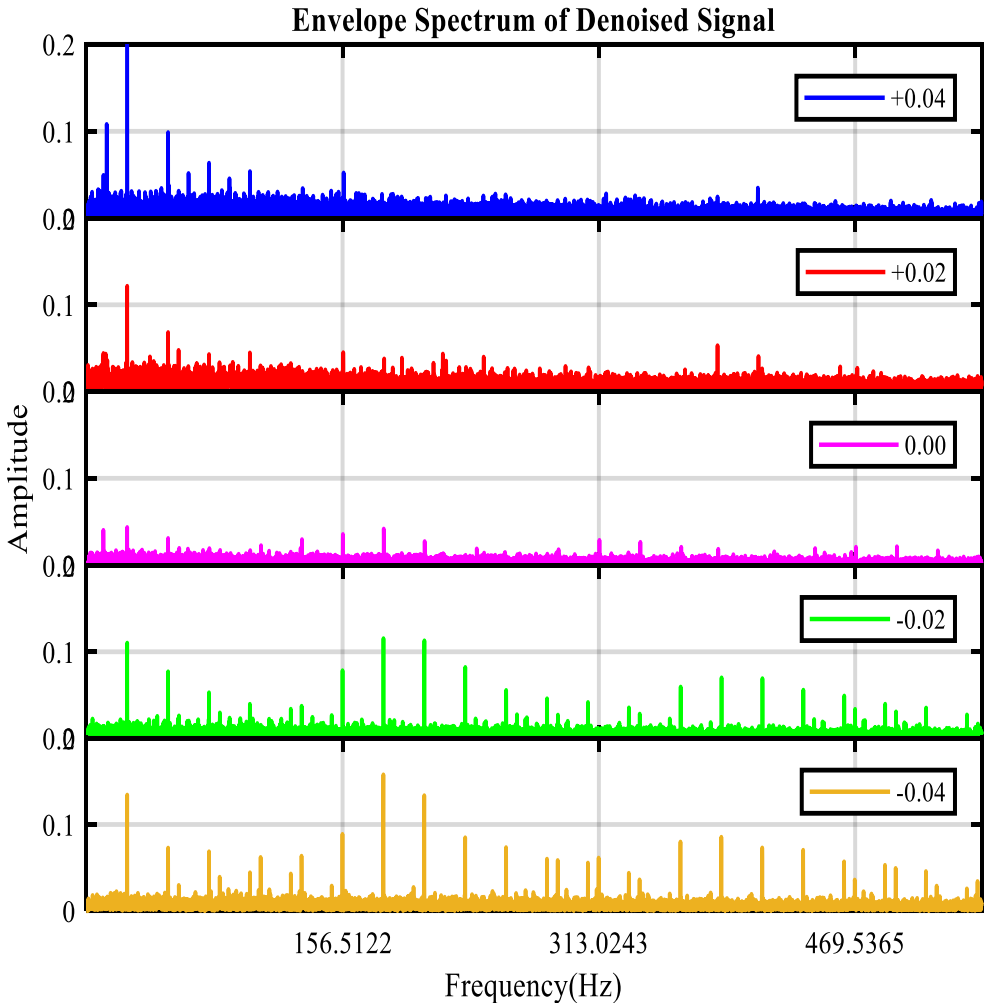

Figure 10-40 Envelope Spectrum of Denoised Signal for Small Outer Race using HSR with DD-DWT

10.4.2 Small Roller Fault

The potential of these measurements for detection and discrimination has been validated using DDD-DWT based HSR threshold and Figure 10-41 shows of the improvement of the harmonic ratio for the first three harmonics. The improvement percentage is shown in Figure 10-42, it is observed that the DDD-DWT produce good detection results and discrimination. However compared to DD-DWT, the former has shrunk the diagnostic feature more the later, thus further investigation still required to obtain better potential results.

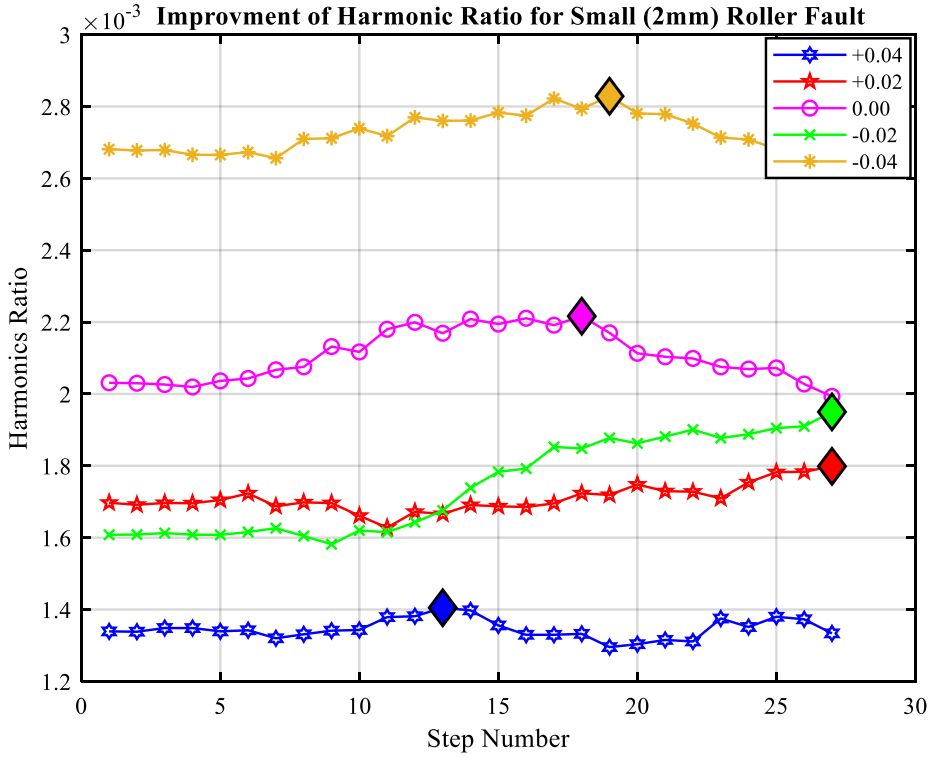

Figure 10-41 Improvement of Harmonic Ratio during Steps using HSR with DDD-DWT

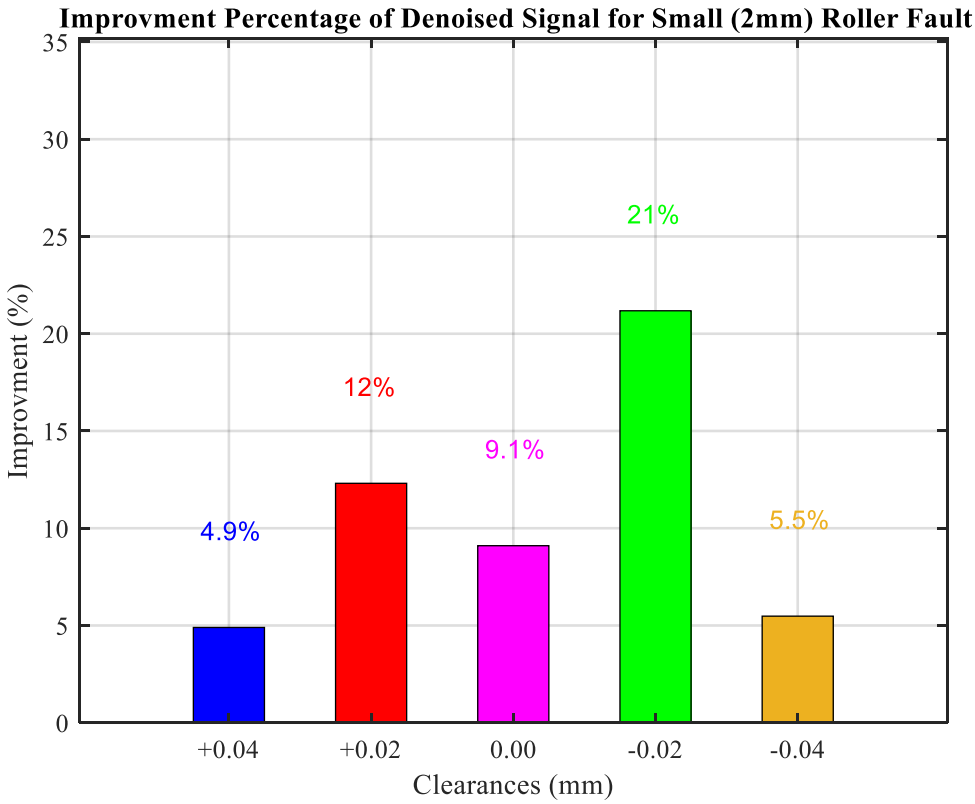

Figure 10-42 Improvement of Harmonic Ratio during Steps using HSR with DDD-DWT

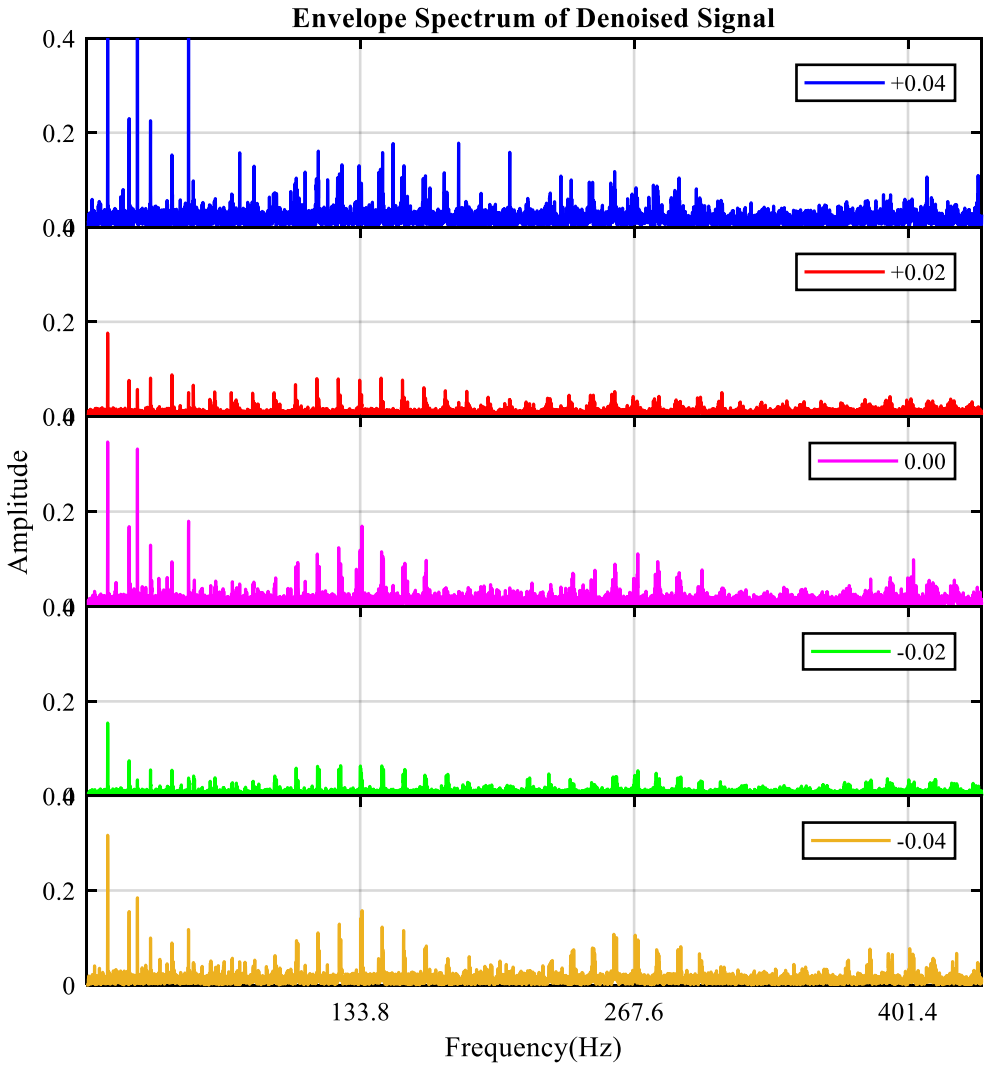

Figure 10-43 Envelope Spectrum of Denoised Signal for Small Roller Fault using HSR with DDD-DWT

As seen in Figure 10-43, overall results can be said good and more random noise reduction is achieved with the implementation of DDD-DWT for all cases, however, it does also shrink the diagnostic features more than the DD-DWT does.

10.5 The Evaluation of HSR

To evaluate the performance of HSR methods, results obtained from using HSR was compared to results obtained from using four of the best thresholding benchmark also known as state of the art methods. The diagnostic capability of the HSR and the DD_DWT has been validated using real datasets collected with zero clearance condition. The well-known thresholding methods namely as; Rigrsure (SureShrink), sqtwolog (VisuShrink), HeurSure and Minimaxi are used to denoise two datasets, one dataset is for a defective bearing with small outer race fault and the second dataset is for a defective bearing with roller fault. Both datasets were gathered with zero (0.00) clearance condition.

Figure 10-44 (a) and (b) shows the envelope spectrum of the original signal and the reconstructed coefficients respectively from datasets one, HSR was applied and results shown in Figure 10-44 (c). It can be seen that diagnostic features sharply enhanced, moreover, the noise level was reduced compared to original and reconstructed coefficients.

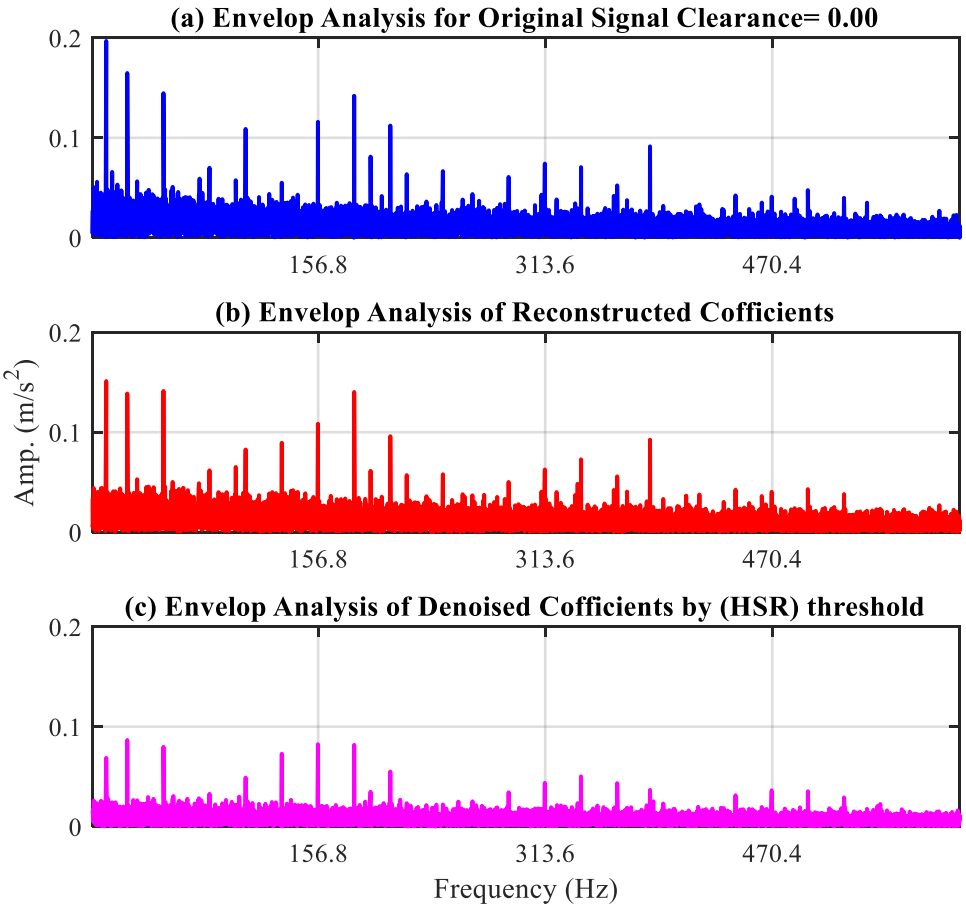

Figure 10-44 Envelope Spectrum of Small Outer Race Fault Signal Using DD-DWT

All results obtained are shown in Figure 10-45 (a) Illustrates the results obtained from applying Rigrsure, (b) shows results of applying HeurSure, (c) depicts HeurSure results, (d) illustrates results of Minimaxi method and (e) shows HSR thresholding method. Comparing all results show that HSR thresholding is effective in enhancing the diagnostic features and reducing the noise from the signal.

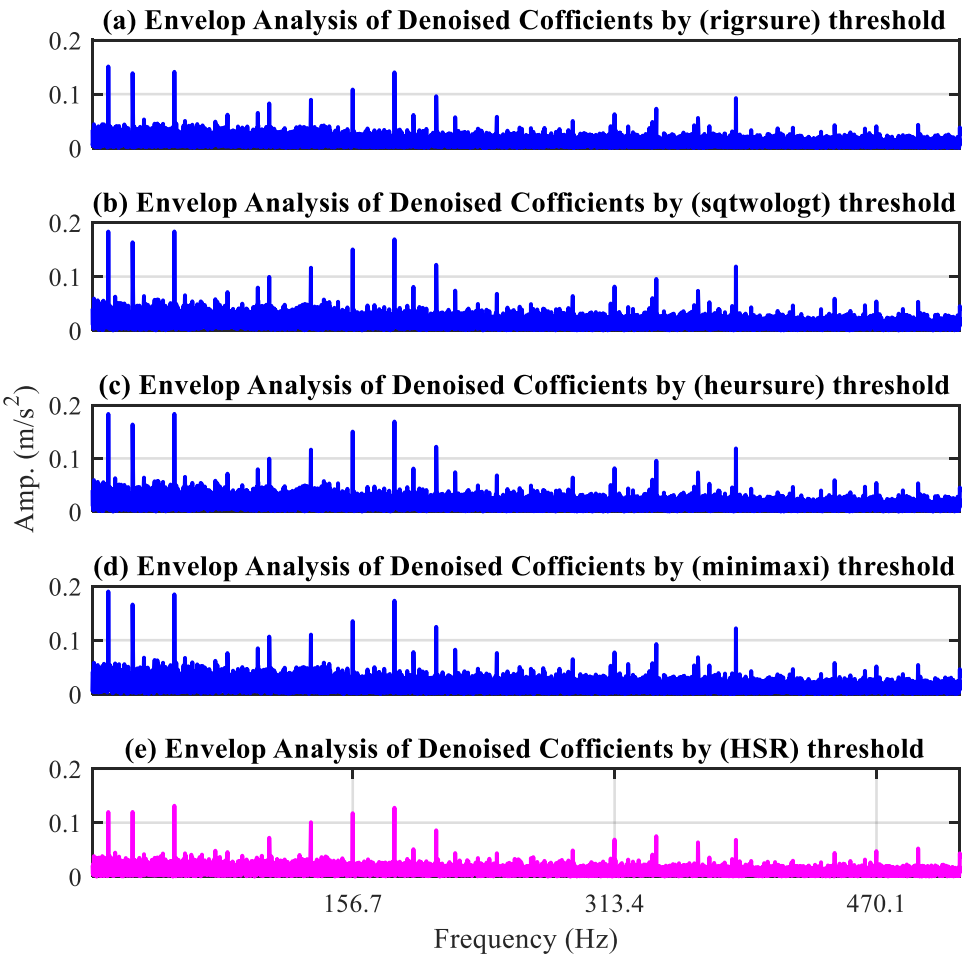

Figure 10-45 Comparison between four benchmark thresholding methods and HSR method for Outer Race Fault Signal Using DD-DWT

In the second part of the evaluation, dataset gathered from a defective bearing with roller fault, is used to evaluate HSR thresholding method. Figure 10-46 (a) and (b) illustrates the envelope spectrum of original and reconstructed coefficients respectively whilst (c) shows HSR results. Moreover, the same procedure was adopted with the second dataset and results are compared and described in Figure 10-47. As seen in (e), HSR has shrunk the noise and enhanced the diagnostic features.

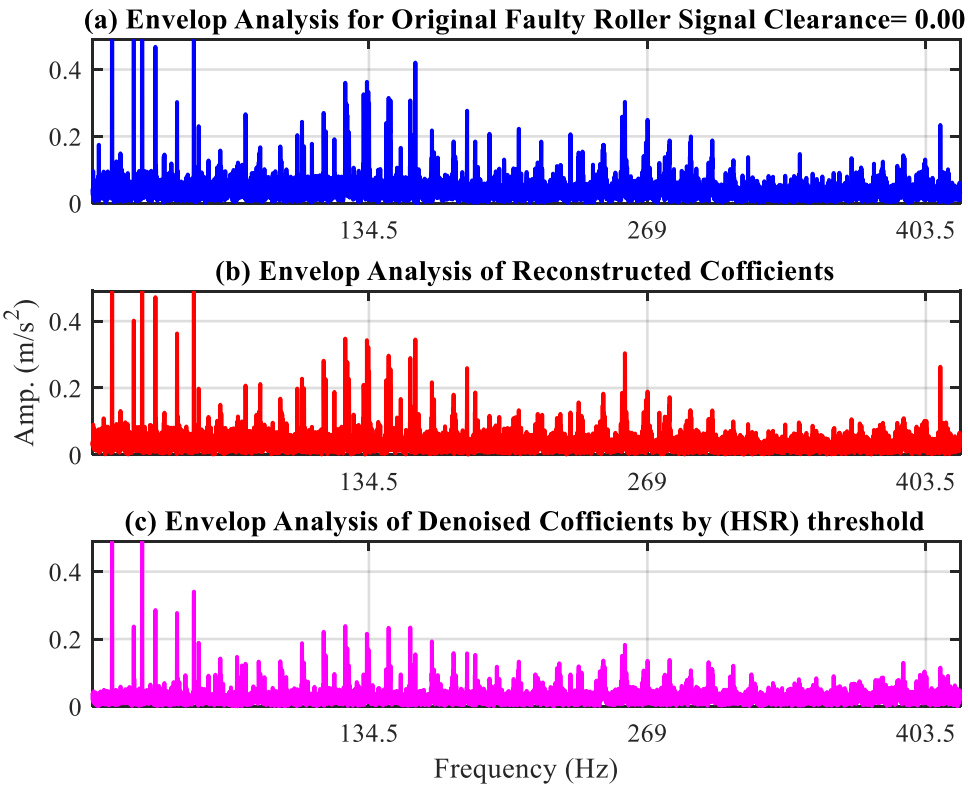

Figure 10-46 Comparison between Raw and Reconstructed Coefficients and Denoised Signal Using HSR Method with DD-DWT

It was observed from the comparison that using the HSR thresholding method yields excellent performance although benchmark thresholding techniques improved the signals with different levels of enhancement, HSR greatly decreased the noise level whilst the faulty peaks left unchanged.

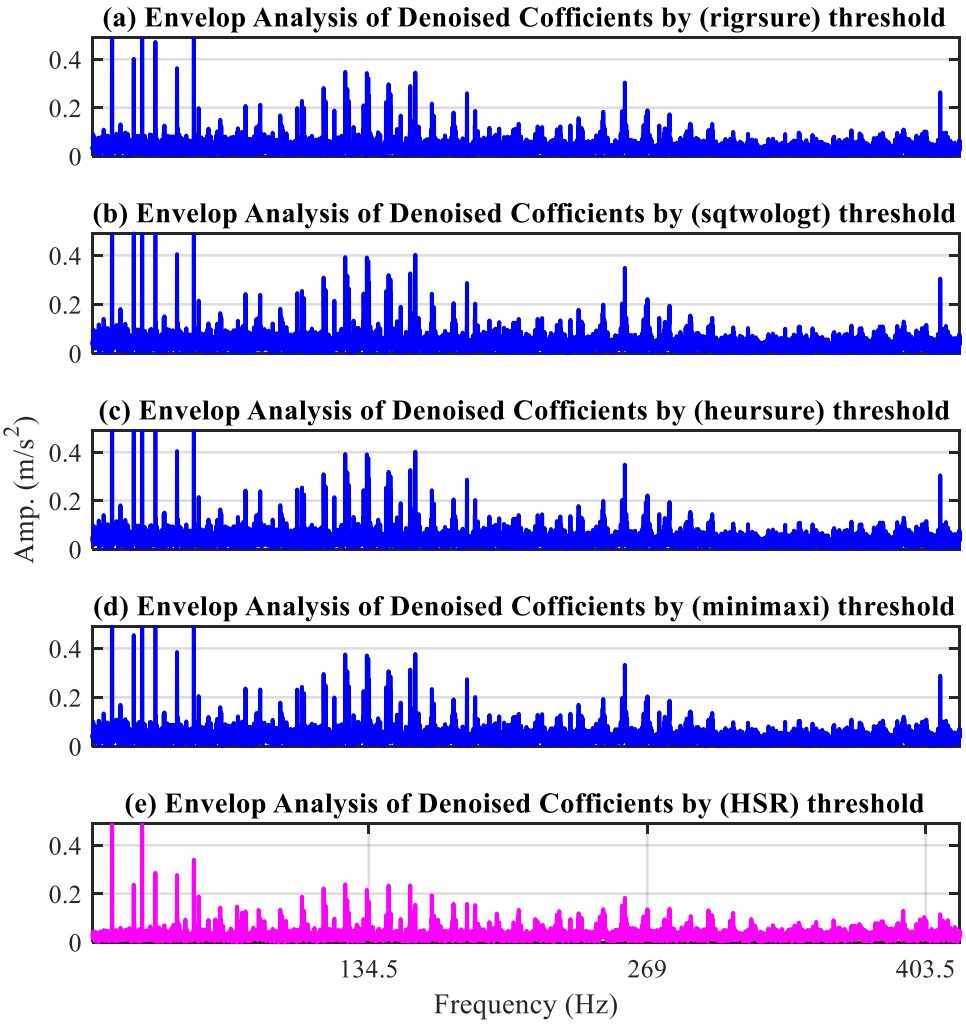

Figure 10-47 Comparison between Four Benchmark Thresholding Methods and HSR Method for Roller Fault Signal using DD-DWT

10.6 Summary

The unified method DD_DWT with HSR technique showed its effectiveness and performance in extracting the diagnostic features and in suppressing the noise, thus, enhancing the signal through experimental case studies. The proposed thresholding technique confirmed its improvement of fault harmonic ratio to signal ratio over the thresholding benchmark techniques when applied alone. The applications of the HSR with Wavelet Transform to measured data have demonstrated that the HSR has good performance in diagnostic feature enhancement and as a result in fault diagnosis. The HRS with all wavelet transform investigated in this research perform well in signal denoising and fault diagnosis, however, as it can be seen in Table 10-1, DD-DWT clearly outperforms DT-ℂWT and DDD-DWT.

Table 10-1 comparison of improvement achieved using the proposed thresholding for small outer race and roller fault (2 mm)

Outer Race						
Wavelet Type		**Clearance**				
		+40	**+20**	**0.00**	**-20**	**-40**
1	Double Density (DWT)	12%	12%	44%	48%	62%
2	Dual Tree Complex WT (DT-CWT)	14%	4.2%	44%	42%	60%
3	Double Density Dual Tree Complex (DDD-WT)	4.6%	9.4%	50%	54%	65%
Roller Fault						
1	Double Density (DWT)	9%	18%	18%	20%	8.5%
2	Dual Tree Complex WT (DT-CWT)	2.8%	15%	16%	20%	12%
3	Double Density Dual Tree Complex (DDD-WT)	5.5%	21%	9.1%	6%	5.5%

Particularly when applied to cases with large internal clearances. The enhancement is achieved in both small outer race and roller fault cases of the bearings. The advantage of DD-DWT is that it can be used to analyse data with very low SNR, also can be applied to raw vibration data directly. This will greatly improve the efficiency of implementing condition monitoring. It has been demonstrated with experimental data that HRS outperform the benchmark thresholding methods. It reveals the novelty more clearly. Moreover, the proposed method does not require any skilled labours or any advanced techniques to choose a band when applying envelope spectrum and the whole frequency band can be used to identify the faults occurred to the monitored bearing. It incorporates a feature of translation invariance thus it can be applied to non-stationary condition-monitoring data. The HRS threshold can also be viewed as an approximate optimal Hard-threshold.

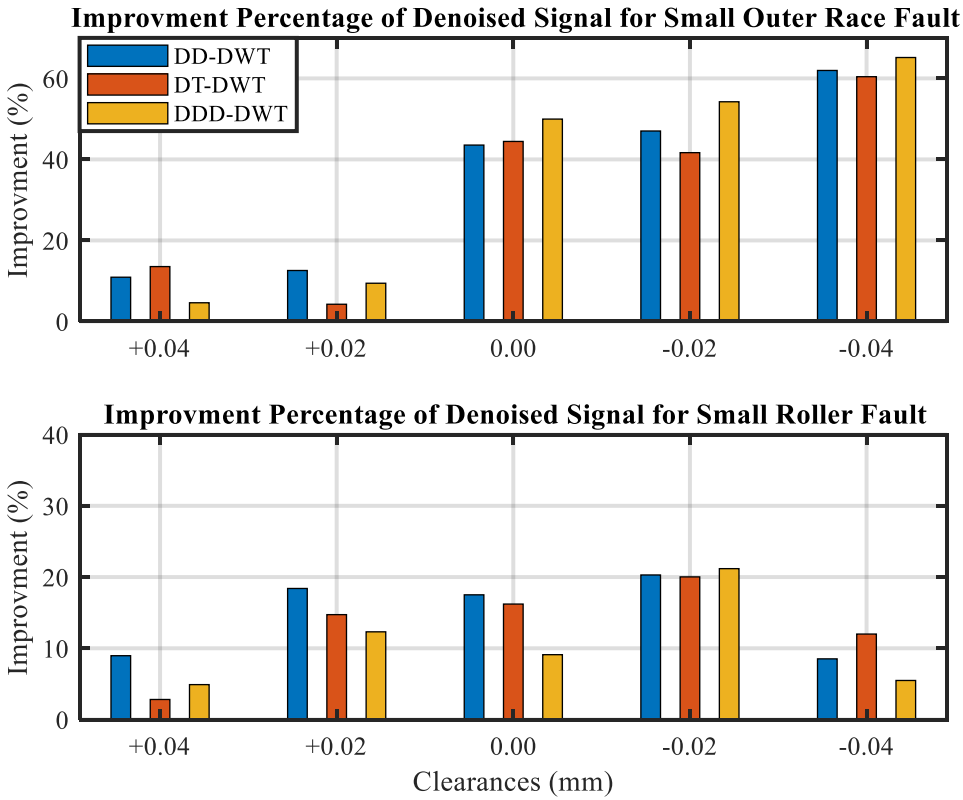

Figure 10-48 Comparison between Wavelet Transform Methods in Features Enhancement

Figure 10 50 compares the improvement percentage obtained from applying DD-DWT, DT-ℂWT and DDD-DW. It can be observed that an improvement achieved from using all wavelet types for all clearance cases, however, DD-DWT achieved higher overall improvement particularly for large clearance cases in both fault types (+0.04, +0.02).

Kurtosis is used to measure the impulsiveness of both the original signal and denoised signal. Kurtosis of Outer Race fault cases are shown in Figure 10-49 (a) whilst, Kurtosis of roller fault cases are shown in (b).

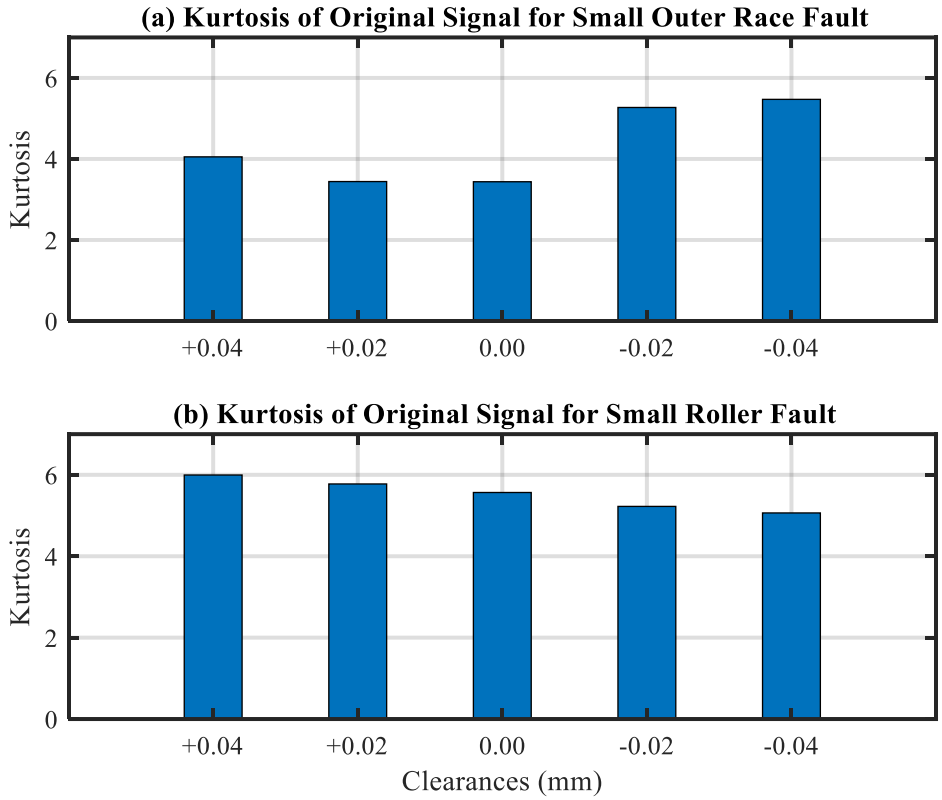

Figure 10-49 Kurtosis of Original Signal for Small Outer Race and Roller Faults

Figure 10-50 (a) illustrates the kurtosis value of denoised outer race fault signals for cases, (b) shows the kurtosis of denoised roller fault signals for all cases, it can be confirmed that kurtosis values of all denoised signals cases have increased sharply compared to baseline data. However, denoised signals using DD-DWT has overall higher kurtosis compared to other results obtained from DT-ℂWT and DDD-DWT.

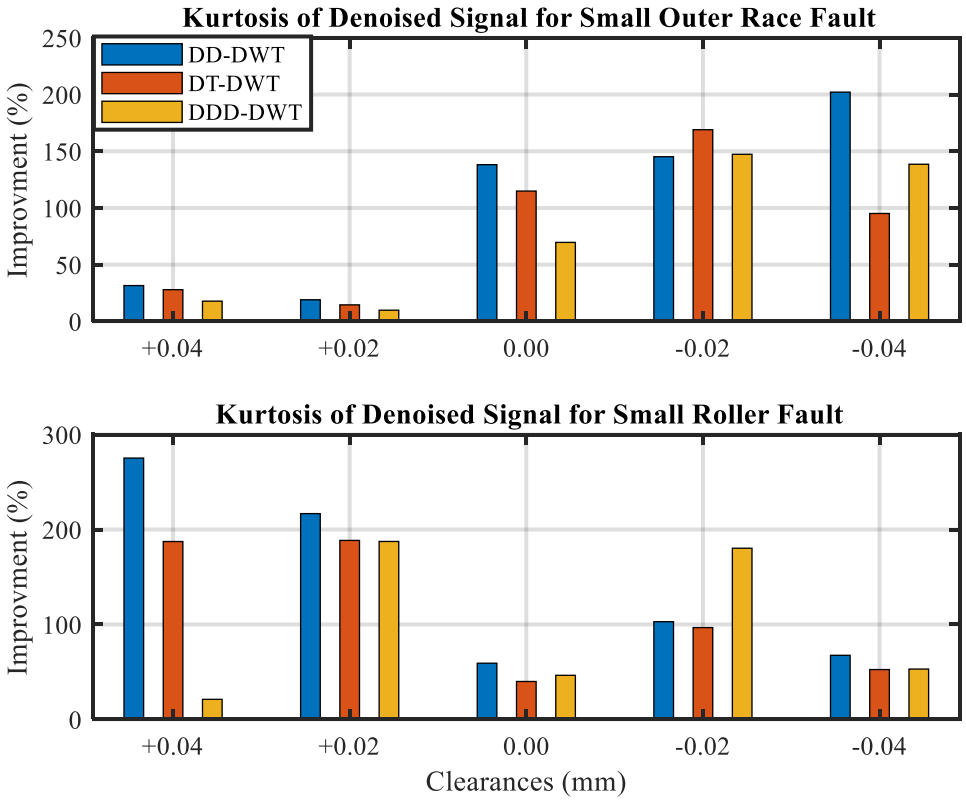

Figure 10-50 Comparison between wavelet transform methods

CHAPTER ELEVEN

CONCLUSION AND FUTURE WORK

In this chapter, the achievements of the research work are summarised and a demonstration of how all objectives of this research were achieved is presented. Also, Contributions to knowledge is described. Finally, recommendations for the future work on this research is presented.

11.1 Introduction

Generally, studying CM is carried out through three main key focuses, namely as data processing focus, data gathering technology and instrumentation focus and application focus. In this research, data processing focus is considered. A variety of vibration analysis techniques are implemented to investigate vibration datasets collected from a defective tapered roller bearing. For early fault detection, an automated approach based unsupervised machine learning algorithm is developed namely componential coding neural network. For fault isolation and identification, a semi-automated approach based wavelet transform with a novel adaptive thresholding technique is developed. The aim of this research was met throughout the achievements of the objectives as described below:

11.2 Objectives and Achievements

In this section, the objectives are reviewed alongside the achievements, the aim of this research has been met and immense progress towards the objectives has been made. The objectives and achievements were as following:

Objective one: To explore and gain insight into the methods of current CM and their applications.
Achievement one: In chapter one, machinery condition monitoring is defined and discussed in terms of CM importance, steps, data gathering techniques, CM approaches, analysis techniques of vibration data have been explored and classified according to the analysis domain.

Objective two: To review the current analysis techniques of the experimental vibration signals, as well as, artificial intelligence based techniques for anomaly detection in vibration data. Furthermore, to carry out a critical review of the wavelet-based data analysis techniques and their existing benchmark thresholding methods used for experimental data denoising and feature enhancement.
Achievement two: A comprehensive review is achieved covering the conventional analysis techniques of experimental vibration signals, also, advanced data analysis techniques time-frequency domain and adaptive data analysis techniques. Moreover, automated techniques based on Artificial intelligence for anomaly detection in CM have been reviewed.

Objective three: To automate the task of early fault detection and severity estimation by implementing an automated technique based AI approach.

Achievement three: An intelligent monitoring approach based on the mechanisms of CCNN algorithms has been successfully implemented to automate the task of early fault detection and severity estimation, which is capable of anomaly detection through an automated procedure to learn the data characteristics and to optimise CCNN configuration.

Objective four: To implement DD-DWT for vibration data analysis in the field of CM and develop a thresholding algorithm for data denoising and feature enhancing.
Achievement four: DD-DWT has been successfully implemented to extract the diagnostic features and a novel adaptive thresholding technique is developed based harmonic to signal ratio HSR and the theoretical development of HSR is presented.

Objective five: To investigate the impact of changes in internal clearance, due to inevitable wear, on the richness of diagnostic signal information and fault detection and diagnosis. Moreover, to design a test rig and develop an adjustable clearance mechanism in which the radial clearances can be controlled and the defects can be seeded into tapered bearings.
Achievement five: a test rig was developed on the basis of adjustable clearance mechanism, the effect of internal clearance on diagnostic features was highlighted and it was concluded that internal clearance affects both fault frequency and fault amplitude.

Objective six: To evaluate the performance and capability of the CCNN using simulated data in order to explore its reliability and effectiveness on bearing fault detection in comparison with conventional techniques.
Achievement six: several tests have been carried out by generating and using two types of synthetic (simulated) signal. The first test started using a simple periodic signal to assess the capability of CCNN in detecting frequency variation. In the second test, to simulate bearing fault signal (cyclo-stationary), impact signal was generated and used to train and then with another signal used with a change in frequency for testing the network. The results show that the method is effective in learning the features and detecting the anomalies even in a high noise level.

Objective seven: To evaluate the performance and capability of the CCNN using the vibration data gathered from baseline and defective bearings.
Achievement seven: The applications of the Componential Coding Neural Network to measured data showed that the networks perform well in fault detection and discrimination is achieved for both simple fault cases of the out race and for the rolling element of the bearings. The advantage of CCNN is that it can be used with minimum knowledge about data, also can be applied to raw

vibration data directly. This will greatly improve the efficiency of implementing condition monitoring.

Objective eight: To evaluate the performance DD-DWT with comparison to both DT-CWT and DDD-DWT in features extraction with the developed thresholding technique.

Achievement eight: The unified technique DD-DWT with HSR showed its effectiveness and performance in extracting the diagnostic features and in suppressing the noise, thus, enhancing the signal through experimental case studies. The proposed method DD_DWT confirmed its improvement in extracting the diagnostic features over the DT-ℂWT and DDD-DWT methods. Moreover**,** has a reliable performance and effective technique in diagnostic feature enhancement and as a result in fault diagnosis.

Objective nine: To evaluate the performance of the developed thresholding technique (HSR) in features enhancement and data denoising against the benchmark thresholding techniques.

Achievement nine: The applications of the HSR with DD-DWT to measured data have demonstrated that the HSR has a reliable performance and effective technique in the enhancement of diagnostic features and, as a result, in fault diagnosis. The developed thresholding HSR outperforms the thresholding benchmark methods.

11.3 Conclusion

The results show that the proposed framework is a robust and effective in both anomaly detection based CCNN and in the diagnosis task based on expensive wavelet transform with HSR thresholding technique. The CCNN gives excellent results in detection and discrimination of different faults whilst wavelet gives excellent results in signal denoising, feature enhancement and eventually fault diagnosis. A number of conclusions have been outlined in this research and can be concluding in this thesis as following:

Conclusion 1: Several condition monitoring (CM) techniques have been investigated for early fault detection and diagnosis in order to avoid machinery failures. However, manual techniques require well-skilled labours which will increase the cost of the monitoring process and may not always be available at the site. Thus, automate the monitoring process using artificial intelligence (AI) based techniques and semi-automatic (adaptive) techniques have promising solutions to facilitate the task of machinery CM and hence, prevent the machine failures and the potential catastrophic consequences.

Conclusion 2: the experimental investigation of the impact of wear evolution on diagnostic features in bearings shown that diagnostic features can be greatly affected by the variance of internal clearances due to wear evolution in rolling element bearing. Clearance variation can affect two feature parameters: magnitude changes of characteristic frequencies which are commonly used to assess fault severity, and shown its impact on the deviation of characteristic frequencies, which is newly suggested to show wear evolution. The experimental results show that the magnitude declines remarkably with the increase of the clearance (wear evolution) for different fault cases. However, the deviation for roller faults exhibits an increasing trend as the preload value grows, whereas the outer race fault shows a decreasing trend as the preload value increases. Therefore, these frequency deviations from nominal characteristic can be an indication of wear evolution.

Conclusion 3: In order to increase the accuracy of fault diagnosis for rolling element bearing, internal clearance increase caused by inevitable wear has to be taken into account in analysing vibration signals. Moreover, clearance variations can be estimated by analysing vibration change in the low-frequency band. The experimental results show that the magnitude of vibrations in the low-frequency band from 1Hz to 1000Hz increases remarkably with the increase of the clearance for both the baseline and different fault cases. These findings will be critical to achieving accurate diagnostics of bearing faults.

Conclusion 4: The application of automated CM approach directly to raw data can improve the task of early fault detection and reduce the cost of the monitoring process. The proposed method, CCNN, perform well when applied to measured data in fault detection and discrimination is achieved for both simple fault cases of the out race and for the rolling element of the bearings. The advantage of CCNN is that it can be used with minimum knowledge about data, also can be applied to raw vibration data directly. This will greatly improve the efficiency of implementing condition monitoring. It has been demonstrated with both simulated and real data that when a trained network is applied to the fault data, it reveals the novelty accurately. Moreover, the CCNN does not require sensors to record angular position when condition-monitoring data is collected since it incorporates a feature of translation invariance.

Conclusion 5: wavelet gives an excellent representation for several types of vibration signals containing jumps and spikes (singularities), it provides optimal sparse representation for such signals, the sparsity comes from the fact that since wavelets oscillate locally, only wavelets overlapping a singularity will have large wavelet coefficients whilst the rest of other coefficients will have smaller coefficients. Unlike Conventional DWT, DD-DWT as an oversampled wavelet

transforms that are analogous to Daubechies suit the application of analysing vibration bearing signals as they provide the best match to vibration signal produced from a defected bearing, it is time invariant. Also, it has a given number of vanishing moments and it supports FIR filters and allows the use of the fast algorithm.

Conclusion 6: for signal denoising and enhancement, unlike filtering-based methods, the wavelet de-noising method does not corrupt the important components of the signal, because the wavelet shrinks the noise using simultaneous re-scaling in both domains frequency and time. Using the nonlinear shrinkage in the transformation domain makes this method distinctive from other linear denoising methods.

Conclusion 7: Optimised wavelet analysis allows the impulsive components of faulty signals to be extracted at selected scales for fault detection and diagnosis.

Conclusion 8: Shrinkage denoising in the transformation domain allows to remove the noise from a large number of wavelet coefficients. The use of the shrinkage method has proven its ability as an effective way to suppress the noise of noisy signals with low computational complexity.

Conclusion 9: Semi-automated approach (adaptive) can achieve more accurate and consistent diagnostic results. The reason is that the algorithm can adapt itself based on the input data and treat each segment of the signal components independently, thus yields more robust diagnostic results. Therefore, the adaptive threshold has shown its effectiveness in denoising and enhancing the noisy data collected from a defective bearing.

Conclusion 10: the proposed new methods HSR adaptive thresholding is based on new findings that harmonics based thresholding is an effective method in noise suppression and diagnostic features enhancement for REBs.

Conclusion 11: DD-DWT is a robust and reliable feature extraction method for vibration data analysis and outperform both DT-ℂWT and DDD-DWT methods.

11.4 Contribution to Knowledge

In this thesis, a number of contributions are obtained and brought to the knowledge that the author believes they are novel and have not been previously implemented. These contributions are given below:

Contribution one: a novel test bearing rig was developed, which allows radial clearances to be adjusted accurately to simulate different degrees of wear during wear lifetime. In addition, it was developed to facilitate the adjusting procedure of the internal clearance using the precise tools to measure the internal clearance.

Contribution two: it was found that change in internal clearances leads to the deviation of characteristic frequencies, which is suggested in this research alongside with analysing the vibration of the low-frequency band to estimate internal clearance changes and wear evolution.

Contribution three: it is believed that a signature decrease with regard to wear evolution, suggesting that accurate severity diagnosis needs to take into account both the bearing wear conditions and the signature magnitudes. Moreover, it has been highlighted that the deviation for roller fault peaks exhibits an increasing trend as the internal clearance declines, whereas the outer race fault peaks show a decreasing trend.

Contribution four: The unsupervised algorithm CCNN has been found to be a promising novel method for monitoring REBs. Furthermore, it may be applied to a variety of rotating machinery. This thesis highlights the benefits of the network in dealing with REBs vibration signatures and its ability is particularly important when it is considered that there are no prior-knowledge found about the fault conditions of the application. Thus, the network can be valuable to many kinds of previously unseen machines.

Contribution five: DD-DWT was found to be the most effective amongst three types of wavelet transforms. It has more wavelets than necessary which give a narrower spacing between adjacent wavelets within the same scale and is less redundant than undecimated wavelet. It was developed for image processing and has been modified and successfully applied in this research. This finding has not been reported to be previously applied in condition monitoring before.

Contribution six: a novel adaptive level-dependant thresholding method (HSR) is developed to find an optimal threshed for denoising the signals and enhancing the faulty features simultaneously, which is the first use of such a denoising approach.

11.5 Recommendations for Future Work

I. The effect of variance in internal clearance on the diagnostic information can be extended to include inner ring frequency and cage frequency of TRBs. Moreover, it can be applied to investigate other types of bearings such as the spherical roller bearing.

II. The diagnostic capability of the CCNN in detecting initial faults has been validated using simulation data and experimental datasets. Further understanding of the weight vectors of the network is needed in association with the monitored data sources. This is essential for the full validation of the diagnosis capacity.

III. Although CCNN is adaptive, however an algorithm to optimize non-parametric parameters for network configuration still need to be developed further.

IV. The proposed Optimised DD-DWT with HSR threshold has been applied successfully to denoise experimental signals of REBs, however, it is essential to extend the method to be used to analyse several types or rotary machines. Moreover, it may be beneficial to apply the proposed method to acoustic data as it is well-known to be noisy data.

Reference

1. Dao, P.B., et al., *Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA data.* Renewable Energy, 2018. **116**: p. 107-122.

2. Waziralilah, N.F., et al. *A Review on Convolutional Neural Network in Bearing Fault Diagnosis*. in *MATEC Web of Conferences*. 2019. EDP Sciences.

3. Spacek, J. *Maintenance strategies of power equipments with a brief view to condition monitoring of power transformers*. in *2008 International Conference-Modern Technique and Technologies*. 2008.

4. Huang, L., et al. *Application of RCM analysis based predictive maintenance in nuclear power plants*. in *Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR2MSE), 2012 International Conference on*. 2012. IEEE.

5. Jardine, A.K., D. Lin, and D. Banjevic, *A review on machinery diagnostics and prognostics implementing condition-based maintenance.* Mechanical systems and signal processing, 2006. **20**(7): p. 1483-1510.

6. Goyal, D., et al., *Optimization of condition-based maintenance using soft computing.* Neural Computing and Applications, 2017. **28**(1): p. 829-844.

7. Lei, Y., et al., *Machinery health prognostics: A systematic review from data acquisition to RUL prediction.* Mechanical Systems and Signal Processing, 2018. **104**: p. 799-834.

8. Heyns, T., *Low cost condition monitoring under time-varying operating conditions*. 2013, University of Pretoria.

9. Randall, R.B., *Vibration-based condition monitoring: industrial, aerospace and automotive applications*. 2011: John Wiley & Sons.

10. Dai, X. and Z. Gao, *From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis.* Industrial Informatics, IEEE Transactions on, 2013. **9**(4): p. 2226-2238.

11. El-Thalji, I. and E. Jantunen, *A summary of fault modelling and predictive health monitoring of rolling element bearings.* Mechanical Systems and Signal Processing, 2015. **60**: p. 252-272.

12. Benali, J., et al. *Importance of the fourth and fifth intrinsic mode functions for bearing fault diagnosis*. in *Sciences and Techniques of Automatic Control and Computer Engineering (STA), 2013 14th International Conference on*. 2013. IEEE.

13. Shakya, P., M.S. Kulkarni, and A.K. Darpe, *Bearing diagnosis based on Mahalanobis–Taguchi–Gram–Schmidt method.* Journal of Sound and Vibration, 2015. **337**: p. 342-362.

14. Bentley, D., *Predictive maintenance through the monitoring and diagnostics of rolling element bearings, Bentley Nevala Co*. 1989, Application Note 44.

15. Rubini, R. and U. Meneghetti, *Application of the envelope and wavelet transform analyses for the diagnosis of incipient faults in ball bearings*. Mechanical systems and signal processing, 2001. **15**(2): p. 287-302.

16. Prieto, M.D., et al., *Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks.* IEEE Transactions on Industrial Electronics, 2012. **60**(8): p. 3398-3407.

17. Yu, J., *Local and nonlocal preserving projection for bearing defect classification and performance assessment.* Industrial Electronics, IEEE Transactions on, 2012. **59**(5): p. 2363-2376.

18. Zarei, J., M.A. Tajeddini, and H.R. Karimi, *Vibration analysis for bearing fault detection and classification using an intelligent filter.* Mechatronics, 2014. **24**(2): p. 151-157.

19. Duan, F., I. Nze, and D. Mba, *Low speed bearing condition monitoring: a case study*, in *Engineering Asset Management 2016*. 2018, Springer. p. 39-48.

20. Rai, A. and S. Upadhyay, *A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings.* Tribology International, 2016. **96**: p. 289-306.

21. Doguer, T. and J. Strackeljan. *Vibration analysis using time domain methods for the detection of small roller bearing defects*. in *8th International Conference on Vibrations in Rotating Machines, Vienna, Austria*. 2009.

22. Al-Bugharbee, H. and I. Trendafilova, *A fault diagnosis methodology for rolling element bearings based on advanced signal pretreatment and autoregressive modelling.* Journal of Sound and Vibration, 2016. **369**: p. 246-265.

23. Chandra, N.H. and A. Sekhar, *Fault detection in rotor bearing systems using time frequency techniques.* Mechanical Systems and Signal Processing, 2016. **72**: p. 105-133.

24. Sreejith, E., A. Verma, and A. Srividya. *Fault diagnosis of rolling element bearing using time-domain features and neural networks*. in *Industrial and Information Systems, 2008. ICIIS 2008. IEEE Region 10 and the Third international Conference on*. 2008. IEEE.

25. Mathew, J. and R. Alfredson, *The condition monitoring of rolling element bearings using vibration analysis.* Journal of vibration, acoustics, stress, and reliability in design, 1984. **106**(3): p. 447-453.

26. Wang, T., G. Lu, and P. Yan, *A novel statistical time-frequency analysis for rotating machine condition monitoring.* IEEE Transactions on Industrial Electronics, 2019.

27. Karczub, D. and M. Norton, *Fundamentals of noise and vibration analysis for engineers*. 2003, Cambridge university press.

28. Tyagi, C.S., *A comparative study of SVM classifiers and artificial neural networks application for rolling element bearing fault diagnosis using wavelet transform preprocessing.* Neuron, 2008. **1**: p. 309-317.

29. Vijay, G., et al., *Radial basis function neural network based comparison of dimensionality reduction techniques for effective bearing diagnostics.* Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2012.

30. Sheng, S., *Wind turbine gearbox condition monitoring round robin study-vibration analysis*. 2012, National Renewable Energy Lab.(NREL), Golden, CO (United States).

31. Sait, A.S. and Y.I. Sharaf-Eldeen, *A review of gearbox condition monitoring based on vibration analysis techniques diagnostics and prognostics*, in *Rotating Machinery, Structural Health Monitoring, Shock and Vibration, Volume 5*. 2011, Springer. p. 307-324.

32. Zhu, J., et al. *Survey of condition indicators for condition monitoring systems*. in *Annu. Conf. Progn. Heal. Manag. Soc*. 2014.

33. Freitas, C., et al. *Comparison of vibration and acoustic measurements for detection of bearing defects*. in *ISMA*. 2016.

34. Rehab, I.A., *The optimization of vibration data analysis for the detection and diagnosis of incipient faults in roller bearings*. 2016, University of Huddersfield.

35. Liu, H., J. Wang, and C. Lu, *Rolling bearing fault detection based on the teager energy operator and elman neural network.* Mathematical problems in engineering, 2013.

36. Xiao-Ling, Z., *The Vibrating Diagnosis Method for Rolling Bearing Fault [J]*. Journal of chongqing university of science and technology (Natural Sciences Edition), 2007. **1**: p. 012.

37. Saidi, L., J.B. Ali, and F. Fnaiech, *Application of higher order spectral features and support vector machines for bearing faults classification.* ISA transactions, 2015. **54**: p. 193-206.

38. McLaughlin, S., A. Stogioglou, and J. Fackrell, *Introducing higher order statistics (HOS) for the detection of nonlinearities.* UK Nonlinear News, 1995. **15**.

39. Chaturved, G. and D. Thomas, *Adaptive noise cancelling and condition monitoring.* Journal of Sound and Vibration, 1981. **76**(3): p. 391-405.

40. Hinich, M.J., *Testing for Gaussianity and linearity of a stationary time series.* Journal of time series analysis, 1982. **3**(3): p. 169-176.

41. Gelle, G., M. Colas, and G. Delaunay. *Higher order statistics for detection and classification of faulty fanbelts using acoustical analysis*. in *Higher-Order Statistics, 1997., Proceedings of the IEEE Signal Processing Workshop on*. 1997. IEEE.

42. Arthur, N. and J. Penman, *Induction machine condition monitoring with higher order spectra.* Industrial Electronics, IEEE Transactions on, 2000. **47**(5): p. 1031-1041.

43. Collis, W., P. White, and J. Hammond, *Higher-order spectra: the bispectrum and trispectrum.* Mechanical systems and signal processing, 1998. **12**(3): p. 375-394.

44. McCormick, A. and A.K. Nandi. *Bispectral and trispectral features for machine condition diagnosis*. in *Vision, Image and Signal Processing, IEE Proceedings-*. 1999. IET.

45. Liang, B., S. Iwnicki, and Y. Zhao, *Application of power spectrum, cepstrum, higher order spectrum and neural network analyses for induction motor fault diagnosis.* Mechanical Systems and Signal Processing, 2013. **39**(1): p. 342-360.

46. Pineyro, J., A. Klempnow, and V. Lescano, *Effectiveness of new spectral tools in the anomaly detection of rolling element bearings.* Journal of alloys and compounds, 2000. **310**(1): p. 276-279.

47. Gu, F., et al., *Electrical motor current signal analysis using a modified bispectrum for fault diagnosis of downstream mechanical equipment.* Mechanical Systems and Signal Processing, 2011. **25**(1): p. 360-372.

48. Bangalore, P. and M. Patriksson, *Analysis of SCADA data for early fault detection, with application to the maintenance management of wind turbines.* Renewable Energy, 2018. **115**: p. 521-532.

49. Pennacchi, P., N. Bachschmid, and A. Vania, *A model-based identification method of transverse cracks in rotating shafts suitable for industrial machines.* Mechanical Systems and Signal Processing, 2006. **20**(8): p. 2112-2147.

50. Pennacchi, P. and A. Vania, *Diagnosis and model based identification of a coupling misalignment.* Shock and Vibration, 2005. **12**(4): p. 293-308.

51. Howard, I., *A review of rolling element bearing vibration'detection, diagnosis and prognosis'*. 1994, DTIC Document.

52. Talebi, H. and K. Khorasani, *A neural network-based multiplicative actuator fault detection and isolation of nonlinear systems.* Control Systems Technology, IEEE Transactions on, 2013. **21**(3): p. 842-851.

53. Jiang, R., et al., *A novel method of fault diagnosis for rolling element bearings based on the accumulated envelope spectrum of the wavelet packet.* Journal of Vibration and Control, 2015. **21**(8): p. 1580-1593.

54. Ali, J.B., et al., *Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network.* Mechanical Systems and Signal Processing, 2015. **56**: p. 150-172.

55. Prieto, M.D., et al., *Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks.* Industrial Electronics, IEEE Transactions on, 2013. **60**(8): p. 3398-3407.

56. Lei, Y., et al., *An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data.* 2015.

57. Webber, C.J., *Emergent componential coding of a handwritten image database by neural self-organisation.* Network: Computation in Neural Systems, 1998. **9**(4): p. 433-447.

58. Pimentel, M.A., et al., *A review of novelty detection.* Signal Processing, 2014. **99**: p. 215-249.

59. Chandola, V., *Anomaly detection: A survey* 2007, ed.

60. Zhong, G., et al., *An overview on data representation learning: From traditional feature learning to recent deep learning.* The Journal of Finance and Data Science, 2016. **2**(4): p. 265-278.

61. Gómez, M., C. Castejón, and J. García-Prada, *Review of recent advances in the application of the wavelet transform to diagnose cracked rotors.* Algorithms, 2016. **9**(1): p. 19.

62. Jaffery, Z., K. Ahmad, and P. Sharma, *Selection of optimal decomposition level based on entropy for speech denoising using wavelet packet.* Journal of Bioinformatics and Intelligent control, 2012. **1**(2): p. 196-202.

63. Hoang, D.-T. and H.-J. Kang, *Rolling element bearing fault diagnosis using convolutional neural network and vibration Image.* Cognitive Systems Research, 2019. **53**: p. 42-50.

64. Geropp, B., *Envelope analysis-a signal analysis technique for early detection and isolation of machine faults.* IFAC Proceedings Volumes, 1997. **30**(18): p. 977-981.

65. White, G., *Amplitude demodulation-a new tool for predictive maintenance.* Sound and vibration, 1991. **25**(9): p. 14-19.

66. McMahon, S., *Condition monitoring of bearing using ESP*. Condition Monitoring and Diagnostic Technology, 1991. **2**(1): p. 21-25.

67. Ratcliffe, G. *Condition monitoring of rolling element bearings using the envelope technique*. in *IMechE Paper, Solid Mechanics and Machine Systems Group Seminar, IMechE, London*. 1990.

68. McInerny, S.A. and Y. Dai, *Basic vibration signal processing for bearing fault detection.* IEEE Transactions on education, 2003. **46**(1): p. 149-156.

69. Feng, G., et al. *The real-time implementation of envelope analysis for bearing fault diagnosis based on wireless sensor network*. in *2013 19th International Conference on Automation and Computing*. 2013. IEEE.

70. Abboud, D., et al., *Envelope analysis of rotating machine vibrations in variable speed conditions: A comprehensive treatment.* Mechanical Systems and Signal Processing, 2017. **84**: p. 200-226.

71. Bilošová, A. and J. Biloš, *Vibration diagnostics.* Investments in education development, Technical University of Ostrava, 2012.

72. Korpel, A., *Gabor: frequency, time, and memory.* Applied optics, 1982. **21**(20): p. 3624-3632.

73. Gabor, D., *Theory of communication. Part 1: The analysis of information.* Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering, 1946. **93**(26): p. 429-441.

74. Randall, R.B. and J. Antoni, *Rolling element bearing diagnostics—a tutorial.* Mechanical Systems and Signal Processing, 2011. **25**(2): p. 485-520.

75. Dwyer, R.F. *Detection of non-Gaussian signals by frequency domain kurtosis estimation*. in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'83*. 1983. IEEE.

76. Sawalhi, N., R. Randall, and H. Endo, *The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis.* Mechanical Systems and Signal Processing, 2007. **21**(6): p. 2616-2633.

77. Dwyer, R. *Detection of non-Gaussian signals by frequency domain kurtosis estimation*. in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'83*. 1983. IEEE.

78. Antoni, J., *The spectral kurtosis: a useful tool for characterising non-stationary signals.* Mechanical Systems and Signal Processing, 2006. **20**(2): p. 282-307.

79. Antoni, J., *Fast computation of the kurtogram for the detection of transient faults.* Mechanical Systems and Signal Processing, 2007. **21**(1): p. 108-124.

80. Srivastava, M., C.L. Anderson, and J.H. Freed, *A new wavelet denoising method for selecting decomposition levels and noise thresholds.* IEEE Access, 2016. **4**: p. 3862-3877.

81. McFadden, P., *Interpolation techniques for time domain averaging of gear vibration.* Mechanical systems and signal processing, 1989. **3**(1): p. 87-97.

82. Ernst, R., *Sensitivity enhancement in magnetic resonance. I. Analysis of the method of time averaging.* Review of Scientific Instruments, 1965. **36**(12): p. 1689-1695.

83. McFadden, P. and J. Smith, *A 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: Journal of Mechanical Engineering Science, 1985. **199**(4): p. 287-292.

84. Halim, E.B., et al. *Fault detection of gearbox from vibration signals using time-frequency domain averaging*. in *American Control Conference, 2006*. 2006. IEEE.

85. McFadden, P. and M. Toozhy, *Application of synchronous averaging to vibration monitoring of rolling element bearings.* Mechanical Systems and Signal Processing, 2000. **14**(6): p. 891-906.

86. Widrow, B. and S.D. Stearns, *Adaptive signal processing*. Vol. 15. 1985: Prentice-hall Englewood Cliffs, NJ.

87. Chaturvedi, G. and D. Thomas, *Bearing fault detection using adaptive noise cancelling.* Journal of Mechanical Design, 1982. **104**(2): p. 280-289.

88. Chaturvedi, G. and D. Thomas, *Bearing fault detection using adaptive noise cancelling.* ASME, Transactions, Journal of Mechanical Design, 1982. **104**: p. 280-289.

89. Tan, C. *An adaptive noise cancellation approach for condition monitoring of gear box bearings*. in *International Tribology Conference 1987, Melbourne, 2-4 December 1987: Preprints of Papers*. 1987. Institution of Engineers, Australia.

90. Ruiz-Cárcel, C., et al. *Application of linear prediction, self-adaptive noise cancellation, and spectral kurtosis in identifying natural damage of rolling element bearing in a gearbox*. in *Proceedings of the 7th World Congress on Engineering Asset Management (WCEAM 2012)*. 2015. Springer.

91. Antoni, J. and R. Randall, *Unsupervised noise cancellation for vibration signals: part I—evaluation of adaptive algorithms.* Mechanical Systems and Signal Processing, 2004. **18**(1): p. 89-101.

92. McCormick, A. and A. Nandi, *Cyclostationarity in rotating machine vibrations.* Mechanical systems and signal processing, 1998. **12**(2): p. 225-242.

93. Christian, K., et al. *On the use of time synchronous averaging, independent component analysis and support vector machines for bearing fault diagnosis*. in *First international conference on industrial risk engineering*. 2007.

94. Ismail, M.A., N. Sawalhi, and T.-H. Pham. *Quantifying bearing fault severity using time synchronous averaging jerk energy*. in *22nd InternationalCongress on Sound and Vibration*. 2015.

95. Huang, N.E., et al. *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*. in *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*. 1998. The Royal Society.

96. Wu, Z. and N.E. Huang, *Ensemble empirical mode decomposition: a noise-assisted data analysis method.* Advances in adaptive data analysis, 2009. **1**(01): p. 1-41.

97. Yu, Y. and C. Junsheng, *A roller bearing fault diagnosis method based on EMD energy entropy and ANN.* Journal of Sound and Vibration, 2006. **294**(1): p. 269-277.

98. Wiggins, R.A., *Minimum entropy deconvolution.* Geoexploration, 1978. **16**(1-2): p. 21-35.

99. Barszcz, T. and N. Sawalhi, *Fault detection enhancement in rolling element bearings using the minimum entropy deconvolution.* Archives of acoustics, 2012. **37**(2): p. 131-141.

100. Abboud, D., et al., *Advanced bearing diagnostics: A comparative study of two powerful approaches.* Mechanical Systems and Signal Processing, 2019. **114**: p. 604-627.

101. Jiang, R., et al., *The weak fault diagnosis and condition monitoring of rolling element bearing using minimum entropy deconvolution and envelop spectrum.* Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2013. **227**(5): p. 1116-1129.

102. Sacchi, M.D., D.R. Velis, and A.H. Cominguez, *Minimum entropy deconvolution with frequency-domain constraints.* Geophysics, 1994. **59**(6): p. 938-945.

103. Krishna, B.V. and M. Vishwakarma, *A review on vibration-based fault diagnosis in rolling element bearings.* International Journal of Applied Engineering Research, 2018. **13**(8): p. 6188-6192.

104. Peng, Z. and F. Chu, *Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography.* Mechanical Systems and Signal Processing, 2004. **18**(2): p. 199-221.

105. Ghazali, K.H., et al. *Feature extraction technique using discrete wavelet transform for image classification*. in *Research and Development, 2007. SCOReD 2007. 5th Student Conference on*. 2007. IEEE.

106. Chen, B., et al., *Application of wavelets and neural networks to diagnostic system development, 1, feature extraction.* Computers & chemical engineering, 1999. **23**(7): p. 899-906.

107. Tian, X., *Enhanced information extraction from noisy vibration data for machinery fault detection and diagnosis*. 2017, University of Huddersfield.

108. Selesnick, I.W., R.G. Baraniuk, and N.C. Kingsbury, *The dual-tree complex wavelet transform.* IEEE signal processing magazine, 2005. **22**(6): p. 123-151.

109. San Emeterio, J., E. Pardo, and A. Ramos, *Ultrasonic grain noise reduction using wavelet processing. An analysis of threshold selection rules.* 2006.

110. Misiti, M., et al., *Wavelets and their applications*. 2013: John Wiley & Sons.

111. Zhu, K., Y. San Wong, and G.S. Hong, *Wavelet analysis of sensor signals for tool condition monitoring: A review and some new results.* International Journal of Machine Tools and Manufacture, 2009. **49**(7-8): p. 537-553.

112. Selesnick, I.W., *The double density DWT*, in *Wavelets in Signal and Image Analysis*. 2001, Springer. p. 39-66.

113. Selesnick, I.W. *The double-density dual-tree discrete wavelet transform*. in *IEEE Trans. on Signal proc*. 2001.

114. Ahn, J.-H., D.-H. Kwak, and B.-H. Koh, *Fault detection of a roller-bearing system through the EMD of a wavelet denoised signal.* Sensors, 2014. **14**(8): p. 15022-15038.

115. Jena, D. and S. Panigrahi, *Precise measurement of defect width in tapered roller bearing using vibration signal.* Measurement, 2014. **55**: p. 39-50.

116. Yan, R., R.X. Gao, and X. Chen, *Wavelets for fault diagnosis of rotary machines: A review with applications.* Signal processing, 2014. **96**: p. 1-15.

117. Nibhanupudi, S., *Signal denoising using wavelets*. 2003, University of Cincinnati.

118. Mallat, S.G., *A theory for multiresolution signal decomposition: the wavelet representation.* IEEE transactions on pattern analysis and machine intelligence, 1989. **11**(7): p. 674-693.

119. Qiu, H., et al., *Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics.* Journal of sound and vibration, 2006. **289**(4-5): p. 1066-1090.

120. Selesnick, I.W., *A higher density discrete wavelet transform.* IEEE Transactions on Signal Processing, 2006. **54**(8): p. 3039-3048.

121. Chiementin, X., et al., *Early detection of rolling bearing defect by demodulation of vibration signal using adapted wavelet.* Journal of Vibration and Control, 2008. **14**(11): p. 1675-1690.

122. Taswell, C., *The what, how, and why of wavelet shrinkage denoising.* Computing in science & engineering, 2000. **2**(3): p. 12-19.

123. Donoho, D.L., *De-noising by soft-thresholding.* IEEE transactions on information theory, 1995. **41**(3): p. 613-627.

124. Donoho, D.L. and I.M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage.* biometrika, 1994: p. 425-455.

125. Zheng, Y., D.B. Tay, and L. Li, *Signal extraction and power spectrum estimation using wavelet transform scale space filtering and Bayes shrinkage.* Signal processing, 2000. **80**(8): p. 1535-1549.

126. Altmann, J. and J. Mathew, *Multiple band-pass autoregressive demodulation for rolling-element bearing fault diagnosis.* Mechanical Systems and Signal Processing, 2001. **15**(5): p. 963-977.

127. Shao, R., et al., *Wavelets and non-linear principal components analysis for process monitoring.* Control Engineering Practice, 1999. **7**(7): p. 865-879.

128. Daubechies, I. and W. Sweldens, *Factoring wavelet transforms into lifting steps.* Journal of Fourier analysis and applications, 1998. **4**(3): p. 247-269.

129. Li, B., et al., *An adaptive morphological gradient lifting wavelet for detecting bearing defects.* Mechanical Systems and Signal Processing, 2012. **29**: p. 415-427.

130. Lu, N., et al., *Signal denoising method based on adaptive redundant second-generation wavelet for rotating machinery fault diagnosis.* Mathematical problems in engineering, 2016. **2016**.

131. Bao, W., et al., *Anti-aliasing lifting scheme for mechanical vibration fault feature extraction.* Mechanical Systems and Signal Processing, 2009. **23**(5): p. 1458-1473.

132. Feng, K., et al., *Rolling element bearing fault detection based on optimal antisymmetric real Laplace wavelet.* Measurement, 2011. **44**(9): p. 1582-1591.

133. Su, W., et al., *Rolling element bearing faults diagnosis based on optimal Morlet wavelet filter and autocorrelation enhancement.* Mechanical Systems and Signal Processing, 2010. **24**(5): p. 1458-1472.

134. Gnanadurai, D., et al., *Undecimated double density wavelet transform based speckle reduction in SAR images.* Computers & Electrical Engineering, 2009. **35**(1): p. 209-217.

135. Celik, T. and K.-K. Ma, *Multitemporal image change detection using undecimated discrete wavelet transform and active contours.* IEEE Transactions on Geoscience and Remote Sensing, 2011. **49**(2): p. 706-716.

136. Wang, X.-Y., H.-Y. Yang, and Z.-K. Fu, *A new wavelet-based image denoising using undecimated discrete wavelet transform and least squares support vector machine.* Expert Systems with Applications, 2010. **37**(10): p. 7040-7049.

137. Hao, R. and F. Chu, *Morphological undecimated wavelet decomposition for fault diagnostics of rolling element bearings.* Journal of Sound and Vibration, 2009. **320**(4): p. 1164-1177.

138. Fucheng, Z. *Fault diagnosis method of gear of wind turbine gearbox based on undecimated wavelet transformation*. in *Computer Design and Applications (ICCDA), 2010 International Conference on*. 2010. IEEE.

139. Jena, D., S. Panigrahi, and R. Kumar, *Gear fault identification and localization using analytic wavelet transform of vibration signal.* Measurement, 2013. **46**(3): p. 1115-1124.

140. ZHANG, L.-J., et al., *Morphological undecimated wavelet and its application to feature extraction of impulsive signal [j]*. Journal of vibration and shock, 2007. **10**: p. 014.

141. Fu-cheng, Z. *Research on online monitoring and diagnosis of bearing fault of wind turbine gearbox based on undecimated wavelet transformation*. in *Information Computing and Telecommunications (YC-ICT), 2010 IEEE Youth Conference on*. 2010. IEEE.

142. Qin, Y., B. Tang, and J. Wang, *Higher-density dyadic wavelet transform and its application.* Mechanical Systems and Signal Processing, 2010. **24**(3): p. 823-834.

143. Kingsbury, N., *Complex wavelets for shift invariant analysis and filtering of signals.* Applied and computational harmonic analysis, 2001. **10**(3): p. 234-253.

144. Kingsbury, N.G. *The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters*. in *IEEE Digital Signal Processing Workshop*. 1998. Citeseer.

145. DEWANGAN, M. and D. Bhonsle, *Image denoising using complex double density dual tree wavelet transform.* International Journal of Engineering Research & Technology, 2013. **2**: p. 906-916.

146. Sveinsson, J.R. and J.A. Benediktsson. *Double density wavelet transformation for speckle reduction of SAR images*. in *Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International*. 2002. IEEE.

147. Qiao, Y.-L., C.-Y. Song, and C.-H. Zhao. *Double-density discrete wavelet transform based texture classification*. in *Intelligent Information Hiding and Multimedia Signal Processing, 2007. IIHMSP 2007. Third International Conference on*. 2007. IEEE.

148. Arfia, F.B., M.B. Messaoud, and M. Abid. *A new image denoising technique combining the empirical mode decomposition with a wavelet transform technique*. in *17th International Conference on Systems, Signals and Image Processing*. 2010.

149. Baili, J., et al. *Application of the discrete wavelet transform to denoise GPR signals*. in *Proceedings from the Second International Symposium on Communications, Control and Signal Processing (ISCCSP), Marrakech, Morocco*. 2006.

150. Vimala, C. and P.A. Priya. *Noise reduction based on double density discrete wavelet transform*. in *Smart Structures and Systems (ICSSS), 2014 International Conference on*. 2014. IEEE.

151. Li, H. *Bearing fault diagnosis based on dual-tree complex wavelet transform*. in *Advanced Materials Research*. 2012. Trans Tech Publ.

152. Qu, J., Z. Zhang, and T. Gong, *A novel intelligent method for mechanical fault diagnosis based on dual-tree complex wavelet packet transform and multiple classifier fusion.* Neurocomputing, 2016. **171**: p. 837-853.

153. Wang, Y., Z. He, and Y. Zi, *Enhancement of signal denoising and multiple fault signatures detecting in rotating machinery using dual-tree complex wavelet transform.* Mechanical Systems and Signal Processing, 2010. **24**(1): p. 119-137.

154. Yu, R. and H. Ozkaramanli, *Hilbert transform pairs of orthogonal wavelet bases: Necessary and sufficient conditions.* IEEE Transactions on Signal Processing, 2005. **53**(12): p. 4723-4725.

155. Selesnick, I.W., *The double-density dual-tree DWT.* IEEE Transactions on signal processing, 2004. **52**(5): p. 1304-1314.

156. Ma, P., et al., *Novel bearing fault diagnosis model integrated with dual-tree complex wavelet transform, permutation entropy and optimized FCM.* Journal of Vibroengineering, 2018. **20**(2).

157. Liu, X. and Z. Chen. *Signal denoising based on DT-CWT and its extraction to gear weak fault information*. in *2006 6th World Congress on Intelligent Control and Automation*. 2006. IEEE.

158. Yasin, A., A. Pavlov, and A. Hramov, *Application of the dual-tree wavelet transform for digital filtering of noisy audio signals.* Journal of Communications Technology and Electronics, 2017. **62**(3): p. 236-240.

159. Sarawale, R.K. and S. Chougule. *Noise removal using double-density dual-tree complex DWT*. in *Image Information Processing (ICIIP), 2013 IEEE Second International Conference on*. 2013. IEEE.

160. Bhonsle, D. and S. Dewangan, *Comparative Study of dual-tree complex wavelet transform and double density complex wavelet transform for Image Denoising Using Wavelet-Domain.* International Journal of Scientific and Research Publications, 2012. **2**(7): p. 1-5.

161. Cohen, R., *Signal denoising using wavelets.* Project Report, Department of Electrical Engineering Technion, Israel Institute of Technology, Haifa, 2012.

162. Chang, F., et al. *Research on wavelet denoising for pulse signal based on improved wavelet thresholding*. in *Pervasive Computing Signal Processing and Applications (PCSPA), 2010 First International Conference on*. 2010. IEEE.

163. Zhang, X., B. Wang, and X. Chen, *Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine.* Knowledge-Based Systems, 2015. **89**: p. 56-85.

164. Yuwono, M., et al., *Automatic bearing fault diagnosis using particle swarm clustering and Hidden Markov Model.* Engineering Applications of Artificial Intelligence, 2016. **47**: p. 88-100.

165. Sharma, A., M. Amarnath, and P. Kankar, *Feature extraction and fault severity classification in ball bearings.* Journal of Vibration and Control, 2016. **22**(1): p. 176-192.

166. Samanta, B. and K. Al-Balushi, *Artificial neural network based fault diagnostics of rolling element bearings using time-domain features.* Mechanical systems and signal processing, 2003. **17**(2): p. 317-328.

167. Lei, Y., et al., *Fault diagnosis of rotating machinery based on a new hybrid clustering algorithm.* The International Journal of Advanced Manufacturing Technology, 2008. **35**(9-10): p. 968-977.

168. Li, B., G. Goddu, and M.-Y. Chow. *Detection of common motor bearing faults using frequency-domain vibration signals and a neural network based approach*. in *American Control Conference, 1998. Proceedings of the 1998*. 1998. IEEE.

169. Yang, D.-M., et al., *Third-order spectral techniques for the diagnosis of motor bearing condition using artificial neural networks.* Mechanical systems and signal processing, 2002. **16**(2): p. 391-411.

170. Yang, B.-S., T. Han, and J.L. An, *ART–KOHONEN neural network for fault diagnosis of rotating machinery.* Mechanical Systems and Signal Processing, 2004. **18**(3): p. 645-657.

171. Su, H., K.T. Chong, and R.R. Kumar, *Vibration signal analysis for electrical fault detection of induction machine using neural networks.* Neural Computing and Applications, 2011. **20**(2): p. 183-194.

172. Markou, M. and S. Singh, *Novelty detection: a review—part 1: statistical approaches.* Signal processing, 2003. **83**(12): p. 2481-2497.

173. Marseguerra, M. and A. Zoia, *The autoassociative neural network in signal analysis: I. The data dimensionality reduction and its geometric interpretation.* Annals of Nuclear Energy, 2005. **32**(11): p. 1191-1206.

174. Markou, M. and S. Singh, *Novelty detection: a review—part 2:: neural network based approaches.* Signal processing, 2003. **83**(12): p. 2499-2521.

175. McFadden, P. and J. Smith, *Model for the vibration produced by a single point defect in a rolling element bearing.* Journal of sound and vibration, 1984. **96**(1): p. 69-82.

176. Kang, M., et al., *Reliable fault diagnosis for low-speed bearings using individually trained support vector machines with kernel discriminative feature analysis.* Power Electronics, IEEE Transactions on, 2015. **30**(5): p. 2786-2797.

177. Abdusslam, S.A., *Detection and diagnosis of rolling element bearing faults using time encoded signal processing and recognition*. 2012, University of Huddersfield.

178. Stergiou, C. and D. Siganos, *Neural Networks. 1996*. 2010.

179. Hagan, M.T., et al., *Neural network design*. Vol. 20. 1996: PWS publishing company Boston.

180. Zhang, R., et al., *Transfer learning with neural networks for bearing fault diagnosis in changing working conditions.* IEEE Access, 2017. **5**: p. 14347-14357.

181. Randall, R.B., 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, 2001. **15**(5): p. 945-962.

182. Sun, Q., et al., *Pattern recognition for automatic machinery fault diagnosis.* Journal of vibration and acoustics, 2004. **126**(2): p. 307-316.

183. Unal, M., et al., *Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network.* Measurement, 2014. **58**: p. 187-196.

184. Chen, X., et al., *Fault diagnosis based on dependent feature vector and probability neural network for rolling element bearings.* Applied Mathematics and Computation, 2014. **247**: p. 835-847.

185. Mahamad, A.K., S. Saon, and T. Hiyama, *Predicting remaining useful life of rotating machinery based artificial neural network.* Computers & Mathematics with Applications, 2010. **60**(4): p. 1078-1087.

186. Tian, Z., *An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring.* Journal of Intelligent Manufacturing, 2012. **23**(2): p. 227-237.

187. Webber, C.J., *Self-organisation of transformation-invariant detectors for constituents of perceptual patterns.* Network: Computation in Neural Systems, 1994. **5**(4): p. 471-496.

188. Khemili, I. and M. Chouchane, *Detection of rolling element bearing defects by adaptive filtering.* European Journal of Mechanics-A/Solids, 2005. **24**(2): p. 293-303.

189. Gan, M. and C. Wang, *Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings.* Mechanical Systems and Signal Processing, 2016. **72**: p. 92-104.

190. Chen, Z. and W. Li, *Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network.* IEEE Transactions on Instrumentation and Measurement, 2017. **66**(7): p. 1693-1702.

191. Bendjama, H., et al. *Selection of wavelet decomposition levels for vibration monitoring of rotating machinery*. in *The Ninth International Conference on Advanced Engineering Computing and Applications in Sciences*. 2015.

192. Pan, J., et al., *LiftingNet: A novel deep learning network with layerwise feature learning from noisy mechanical data for fault classification.* IEEE Transactions on Industrial Electronics, 2018. **65**(6): p. 4973-4982.

193. Stack, J.R., R.G. Harley, and T.G. Habetler, *An amplitude modulation detector for fault diagnosis in rolling element bearings.* Industrial Electronics, IEEE Transactions on, 2004. **51**(5): p. 1097-1102.

194. Wu, J., et al., *Degradation data-driven time-to-failure prognostics approach for rolling element bearings in electrical machines.* IEEE Transactions on Industrial Electronics, 2019. **66**(1): p. 529-539.

195. Hamrock, B.J. and W.J. Anderson, *Rolling-element bearings.* 1983.

196. Tenconi, A., S. Vaschetto, and A. Vigliani, *Electrical machines for high-speed applications: Design considerations and tradeoffs.* IEEE Transactions on Industrial Electronics, 2014. **61**(6): p. 3022-3029.

197. Harris, T.A., *Rolling bearing analysis*. 2001: John Wiley and sons.

198. Harris, T. and R. Barnsby, *Life ratings for ball and roller bearings.* Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2001. **215**(6): p. 577-595.

199. Bonnett, A.H. *Cause and analysis of anti-friction bearing failures in AC induction motors*. in *Pulp and Paper Industry Technical Conference, 1993., Conference Record of 1993 Annual*. 1993. IEEE.

200. Lewis, M. and B. Tomkins, *A fracture mechanics interpretation of rolling bearing fatigue.* Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2012. **226**(5): p. 389-405.

201. Gupta, P. and M. Pradhan, *Fault detection analysis in rolling element bearing: A review.* Materials Today: Proceedings, 2017. **4**(2): p. 2085-2094.

202. Choudhury, A. and N. Tandon, *A theoretical model to predict vibration response of rolling bearings to distributed defects under radial load.* Journal of vibration and acoustics, 1998. **120**(1): p. 214-220.

203. Patil, M., J. Mathew, and P. RajendraKumar, *Bearing signature analysis as a medium for fault detection: A review.* Journal of Tribology, 2008. **130**(1): p. 014001.

204. Nakhaeinejad, M., *Fault detection and model-based diagnostics in nonlinear dynamic systems*. 2010.

205. Chen, S., et al., *Wear detection of rolling element bearings using multiple-sensing technologies and mixture-model-based clustering method.* Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 2008. **222**(2): p. 207-218.

206. ISO. *Rolling bearings — Damage and failures — Terms, characteristics and causes*. Online Browsing Platform (OBP) 2017 [cited 2018 15 Oct]; Available from: https://www.iso.org/obp/ui/#iso:std:iso:15243:ed-2:v1:en.

207. SKF, C., Insocoat, Sensormount,, *Bearing damage and failure analysis*. 2017, SKF.

208. Manes, L., J.-M. De Monicault, and R. Gras, *Monitoring damage by acoustic emission in bearing steels in cryogenic environment.* Tribology international, 2001. **34**(4): p. 247-253.

209. Fernandes, P., *Contact fatigue in rolling-element bearings.* Engineering failure analysis, 1997. **4**(2): p. 155-160.

210. NSK EUROPE LTD. *Bearing & Linear Components Troubleshooting, Damage by Type*. 2018 [cited 2018 11 Oct]; Available from: https://www.nskeurope.com/damage-by-type-129.htm.

211. Williams, J.A., *Wear and wear particles—some fundamentals.* Tribology International, 2005. **38**(10): p. 863-870.

212. Bloch, H.P. and F.K. Geitner, *Machinery failure analysis and troubleshooting: practical machinery management for process plants*. 2012: Butterworth-Heinemann.

213. Engel, P.A., *Failure models for mechanical wear modes and mechanisms.* IEEE Transactions on Reliability, 1993. **42**(2): p. 262-267.

214. El-Thalji, I. and E. Jantunen, *Dynamic modelling of wear evolution in rolling bearings.* Tribology International, 2015. **84**: p. 90-99.

215. Vencl, A., V. Gašić, and B. Stojanović. *Fault tree analysis of most common rolling bearing tribological failures*. in *IOP Conference Series: Materials Science and Engineering*. 2017. IOP Publishing.

216. Kenred Stadler. *Cost reduction in wind turbines thanks to black oxidation,*. 2013 [cited 2018 12 Oct]; Available from: http://evolution.skf.com/it/riduzione-costi-nelle-turbine-eoliche-grazie-alla-black-oxidation/.

217. Tallian, T., *Rolling contact failure control through lubrication*. 1966, SKF Industries INC King Of Prussia PA Research LAB.

218. Evans, M.-H., et al., *Effect of hydrogen on butterfly and white etching crack (WEC) formation under rolling contact fatigue (RCF).* Wear, 2013. **306**(1-2): p. 226-241.

219. Loos, J., et al., *Influences on generation of white etching crack networks in rolling bearings.* Journal of Mechanics Engineering and Automation, 2016. **6**: p. 85-94.

220. Mauntz, M., U. Kuipers, and J. Gegner, *Early detection of critical operating conditions for vibration induced premature bearing failures in industrial gearboxes by sensor monitoring of oil aging.* 2012.

221. SKF. *Bearing investigation*. Extract from the railway technical handbook, volume 1, chapter 6, p. 122-135, 2012, [cited 2018 15 Oct]; Available from: http://www.skf.com/binary/86-62751/RTB-1-06-Bearing-investigation.pdf.

222. Trampe Broch, J., *Mechanical vibration and shock measurements.* Naerum: Bruel & Kjaer, 1984, 2nd ed., 1984.

223. Tandon, N. and A. Choudhury, *A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings.* Tribology International, 1999. **32**(8): p. 469-480.

224. Farrar, C. and S. Doebling, *Damage detection and evaluation II*, in *Modal analysis and testing*. 1999, Springer. p. 345-378.

225. Carden, E.P. and P. Fanning, *Vibration based condition monitoring: a review.* Structural health monitoring, 2004. **3**(4): p. 355-377.

226. Hossain, M., A. Abu-Siada, and S. Muyeen, *Methods for advanced wind turbine condition monitoring and early diagnosis: A literature review.* Energies, 2018. **11**(5): p. 1309.

227. Robert, B.H. and P.-S. Coating-Principles, *Vibration Measurement and Applications*. 1996, VCH Publishers Inc., New York, NY, USA.

228. Zhou, Y. and W. Xue, *Review of tool condition monitoring methods in milling processes.* The International Journal of Advanced Manufacturing Technology, 2018: p. 1-15.

229. Moschioni, G., B. Saggin, and M. Tarabini. *Uncertainty in hand arm vibration measurements due to the fixation method*. in *Proceedings of the 11th International Conference on Hand-Arm Vibration*. 2007. Bologna (Italy).

230. Brändlein, J., *Ball and roller bearings: theory, design and application*. 1999: Wiley.

231. Meng, L., S. Ding, and Y. Xue, *Research on denoising sparse autoencoder.* International Journal of Machine Learning and Cybernetics, 2016: p. 1-11.

232. Deng, L., et al. *Binary coding of speech spectrograms using a deep auto-encoder*. in *Interspeech*. 2010. Citeseer.

233. Rumelhart, D.E., G.E. Hinton, and R.J. Williams, *Learning internal representations by error propagation*. 1985, DTIC Document.

234. Olshausen, B.A., *Sparse codes and spikes.* Probabilistic models of the brain: Perception and neural function, 2002: p. 257-272.

235. Teh, Y.W., et al., *Energy-based models for sparse overcomplete representations.* Journal of Machine Learning Research, 2003. **4**(Dec): p. 1235-1260.

236. Poultney, C., S. Chopra, and Y.L. Cun. *Efficient learning of sparse representations with an energy-based model*. in *Advances in neural information processing systems*. 2006.

237. Lennie, P., *The cost of cortical computation.* Current biology, 2003. **13**(6): p. 493-497.

238. Simoncelli, E.P., *Statistical modeling of photographic images.* Handbook of Video and Image Processing, 2005.

239. Webber, C.J., *Generalisation and discrimination emerge from a self-organising componential network: a speech example.* Network: Computation in Neural Systems, 1997. **8**(4): p. 425-440.

240. Webber, C., et al., *Componential coding in the condition monitoring of electrical machines Part 1: principles and illustrations using simulated typical faults.* Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2003. **217**(8): p. 883-899.

241. Payne, B., et al., *Componential coding in the condition monitoring of electrical machines Part 2: application to a conventional machine and a novel machine.* Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2003. **217**(8): p. 901-915.

242. Kumar, M.S. and K.S. Babu, *Non-Linear Denoising of Images using Wavelet Transform.* International Journal of Computer Applications, 2016. **148**(10).

243. Nason, G., *Choice of the threshold parameter in wavelet function estimation*, in *Wavelets and statistics*. 1995, Springer. p. 261-280.

244. Naimi, H., A.B.H. Adamou-Mitiche, and L. Mitiche, *Medical image denoising using dual tree complex thresholding wavelet transform and Wiener filter.* Journal of King Saud University-Computer and Information Sciences, 2015. **27**(1): p. 40-45.

245. He, K., K.K. Lai, and J. Yen, *A hybrid slantlet denoising least squares support vector regression model for exchange rate prediction.* Procedia Computer Science, 2010. **1**(1): p. 2397-2405.

246. Gradolewski, D. and G. Redlarski, *Wavelet-based denoising method for real phonocardiography signal recorded by mobile devices in noisy environment.* Computers in biology and medicine, 2014. **52**: p. 119-129.

247. Fodor, I.K. and C. Kamath, *Denoising through wavelet shrinkage: an empirical study.* Journal of Electronic Imaging, 2003. **12**(1): p. 151-161.

248. Ergen, B., *Signal and image denoising using wavelet transform*, in *Advances in Wavelet Theory and Their Applications in Engineering, Physics and Technology*. 2012, InTech.

249. Tania, S. and R. Rowaida, *A comparative study of various image filtering techniques for removing various noisy pixels in aerial image.* International Journal of Signal Processing, Image Processing and Pattern Recognition, 2016. **9**(3): p. 113-124.

250. Krim, H., et al., *On denoising and best signal representation.* IEEE transactions on information theory, 1999. **45**(7): p. 2225-2238.

251. Donoho, D.L., *Denoising by soft-thresholding," dept. of statistics*. 1992, Stanford University.

252. Valencia, D., et al. *Comparison analysis between rigrsure, sqtwolog, heursure and minimaxi techniques using hard and soft thresholding methods*. in *Signal Processing, Images and Artificial Vision (STSIVA), 2016 XXI Symposium on*. 2016. IEEE.

253. Johnstone, I.M. and B.W. Silverman, *Wavelet threshold estimators for data with correlated noise.* Journal of the royal statistical society: series B (statistical methodology), 1997. **59**(2): p. 319-351.

254. Biswas, M. and H. Om, *A new soft-thresholding image denoising method.* Procedia Technology, 2012. **6**: p. 10-15.

255. Donoho, D.L. and I.M. Johnstone. *Threshold selection for wavelet shrinkage of noisy data*. in *Engineering in Medicine and Biology Society, 1994. Engineering Advances: New Opportunities for Biomedical Engineers. Proceedings of the 16th Annual International Conference of the IEEE*. 1994. IEEE.

256. Chang, S.G., B. Yu, and M. Vetterli, *Adaptive wavelet thresholding for image denoising and compression.* IEEE transactions on image processing, 2000. **9**(9): p. 1532-1546.

257. Donoho, D.L. and I.M. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage.* Journal of the american statistical association, 1995. **90**(432): p. 1200-1224.

258. Chang, H., *Research on noise reduction of structural dynamic response based on wavelet thresholds.* Revista de la Facultad de Ingeniería, 2017. **31**(9).

259. Wang, Y., et al., *Investigation and Analysis of All-Day Atmospheric Water Vapor Content over Xi'an Using Raman Lidar and Sunphotometer Measurements.* Remote Sensing, 2018. **10**(6).

260. Joy, J., S. Peter, and N. John, *Denoising using soft thresholding.* International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 2013. **2**(3): p. 1027-1032.

261. Hashemi, S. and S. Beheshti. *Adaptive image denoising by rigorous Bayesshrink thresholding*. in *Statistical Signal Processing Workshop (SSP), 2011 IEEE*. 2011. IEEE.

262. Cai, T.T. and B.W. Silverman, *Incorporating information on neighbouring coefficients into wavelet estimation.* Sankhyā: The Indian Journal of Statistics, Series B, 2001: p. 127-148.

263. AlMahamdy, M. and H.B. Riley, *Performance study of different denoising methods for ECG signals.* Procedia Computer Science, 2014. **37**: p. 325-332.

264. Roulias, D., T. Loutas, and V. Kostopoulos, *A statistical feature utilising wavelet denoising and neighblock method for improved condition monitoring of rolling bearings.* Chemical Engineering, 2013. **33**.

265. Donoho, D.L., et al., *Wavelet shrinkage: asymptopia?* Journal of the Royal Statistical Society. Series B (Methodological), 1995: p. 301-369.

266. Priya, K.D., G.S. Rao, and P.S. Rao, *Comparative analysis of wavelet thresholding techniques with wavelet-Wiener filter on ECG signal.* Procedia Computer Science, 2016. **87**: p. 178-183.

267. Han, G. and Z. Xu, *Electrocardiogram signal denoising based on a new improved wavelet thresholding.* Review of Scientific Instruments, 2016. **87**(8): p. 084303.

268. Sendur, L. and I.W. Selesnick, *Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency.* IEEE Transactions on signal processing, 2002. **50**(11): p. 2744-2756.

269. Clarcke. *Clarke 5.5 Hp, 400V Electric Motor*. 2018 [cited 2018 31 Oct]; Available from: https://www.machinemart.co.uk/AspSiteAssets/download_files/ClarkeElectricMotors.pdf.

270. National Instruments Corp, *SCB-68 User Guide*. 2009.

271. National Instruments Corporation, *SCB-68 user manual for advanced functions*. 2009.

272. RS Components Ltd. *Incremental Shaft Encoders*. 2017 [cited 2018 5 Nov]; Available from: https://uk.rs-online.com/web/p/rotary-encoders/2603724/.

273. The-Crankshaft Publishing. *Slip Gauges (Metrology)*. 2017 [cited 2017 1 Dec]; Available from: http://what-when-how.com/metrology/slip-gauges-metrology/.

274. Group, B. *Dial Indicators*. 2016 [cited 2018 5 Nov]; Available from: https://www.bowersgroup.co.uk/dip223m-dial-indicators.html.

275. Craig, M., et al., *Advanced condition monitoring of tapered roller bearings, Part 1*. Tribology International, 2009. **42**(11-12): p. 1846-1856.

276. Linet, P., *Methods of setting tapered roller bearings.* SAE transactions, 1982: p. 3275-3292.

277. TIMKEN, *Setting Techniques for Tapered Roller Bearings*. 2015, TMKEN Company: USA. p. 2-7.

278. Walker, D.C.E., PA, *Method for setting the axial end play of tapered roller bearings*, G.E. Company, Editor. 1992, USA: USA.

279. TIMKEN, *Timken Tapered Roller Bearing Catalog*, TIMKEM, Editor. 2016, TIMKEN COMPANY. p. 29.

280. Gilbert, G.G. and W.E. Harbottle, *UNIPAC-An independently sprung driving wheel bearing*. 1979, SAE Technical Paper.

281. Halme, J. and P. Andersson, *Rolling contact fatigue and wear fundamentals for rolling bearing diagnostics-state of the art.* Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2010. **224**(4): p. 377-393.

282. Nguyen-Schäfer, H., *Computational design of rolling bearings*. 2016: Springer.

283. Rehab, I., et al., *A Study of the Diagnostic Amplitude of Rolling Bearing under increasing Radial Clearance using Modulation Signal Bispectrum.* 2016.

284. Daniel Goerke, N.M.a.A.G., *Effects of Radial Clearance Changes on Vibration Frequencies in Double-Row Self-Aligning Ball Bearings: An Experimental Study.* Transactions of the Institute of Measurement and Control, 2017. **v1**(10): p. 1-14.

285. Fitzsimmons, B. and H.J.A.T. Clevenger, *Contaminated lubricants and tapered roller bearing wear.* 1977. **20**(2): p. 97-107.

286. Su, Y. and Y. Sheen, *On the detectability of roller bearing damage by frequency analysis.* Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 1993. **207**(1): p. 23-32.

287. Su, Y.-T. and S.-J. Lin, *On initial fault detection of a tapered roller bearing: frequency domain analysis.* Journal of Sound and Vibration, 1992. **155**(1): p. 75-84.

288. Antoni, J. and R. Randall, *Differential diagnosis of gear and bearing faults.* Journal of Vibration and Acoustics, 2002. **124**(2): p. 165-171.

289. Daubechies, I., *Ten lectures on wavelets*. Vol. 61. 1992: Siam.

290. Narendiranath, B.T., et al., *Journal bearing fault detection based on daubechies wavelet.* Archives of Acoustics, 2017. **42**(3): p. 401-414.

291. Raj, V.N.P. and T. Venkateswarlu. *ECG signal denoising using undecimated wavelet transform*. in *2011 3rd International Conference on Electronics Computer Technology*. 2011. IEEE.