NANYANG TECHNOLOGICAL UNIVERSITY

Advanced Feature Extraction and Selection for Fault Diagnosis and Prognosis

ZHOU JUNHONG

School of Mechanical and Aerospace Engineering

2012

Advanced Feature Extraction and Selection for Fault Diagnosis and Prognosis

ZHOU JUNHONG

School of Mechanical and Aerospace Engineering

A thesis submitted to the Nanyang Technological University

in fulfillment of the requirement for the degree of

Doctor of Philosophy

Abstract

Condition based maintenance (CBM) has become increasingly important over the past decade as a means to maximize the asset usage and plant operating efficiency. The CBM methodology and techniques have attracted strong research interest of the past decade. Among these, advanced feature extraction and selection are essential research issues for realizing CBM.

In this thesis, a novel Dominant Feature Identification (DFI) methodology is proposed. Singular Value Decomposition (SVD) is used to decompose the inner product matrix of collected data from the monitoring sensors. To efficiently use DFI in the CBM framework, a methodology is developed to integrate the proposed DFI with traditional feature extraction methods of time domain and frequency domain analysis, working together with Recursive Least Squares (RLS) and Extended Least Squares (ELS) modeling methods. The developed methodology and framework is evaluated based on the accuracy of prediction of the actual milling tool wear. Augmented Dominant Feature Identification (ADFI) and Decentralized Dominant Feature Identification (DDFI) are proposed for both sensors and features reduction for industrial fault detection and isolation (FDI). The proposed ADFI and DDFI are tested on a fault simulator machine to identify imbalance, misalignment and different types of bearing faults.

Further improvement has been developed to integrate DFI with advanced feature extraction techniques of wavelet based correlation modeling. Reinforced Morlet wavelet transform is proposed to overcome the difficulties in selecting Morlet wavelet transform parameters and wavelet reconstruction. Reinforced Morlet wavelet is evaluated with both a milling machine spindle and a bearing test rig, and results show that Reinforced Morlet wavelet achieves high consistency for bearing fault detection in an early stage. The research scope is also extended from single machine CBM to complex manufacturing environment. A system framework of *iDiagnosis & Prognosis* was developed to provide distributed intelligence of diagnosis and prognosis, to provide system condition and job execution capability assessment, and provide dynamic information feedback to the supervisory control when manufacturing system breakdown or performance degradation occurs.

Acknowledgements

First and foremost, I would like to express my sincere gratitude and recognition to my supervisor Professor Zhong Zhaowei for his utmost patience, invaluable instructions and encouragement in the past three years. Without his support, my research would not reach this stage.

I would also like to thank my colleagues in SIMTech and research collaborators who have lent me various support emotionally or technically, in particular Dr. Li Xiang, Dr. Luo Ming, Professor Frank Lewis and Professor Pang Chee Khiang.

Special thanks are due to my husband, Jian and my daughters, Ying and Qi for their understanding, patience and support.

Finally but most importantly, I wish to take this opportunity to thank my parents, who love me, encourage me and support me all the time, all the way.

List of Publications

The following publications have been generated from this research work:

Refereed International Journal Papers:

1. J. H. Zhou; C. K. Pang; F. L. Lewis,; Z. W. Zhong (2009); "Tool wear forecast using Singular Value Decomposition for dominant feature identification", IEEE Transactions on Industrial Informatics, Vol. 5, No. 4, pp: 454-464.

2. J. H. Zhou, Z. W. Zhong, M. Luo, and C. Shao (2009), "Wavelet-based correlation modeling for health assessment of fluid dynamic bearings in brushless DC motors", International Journal of Advanced Manufacturing Technology, Vol. 41, No. 5-6, pp. 421-429.

3. J. H. Zhou, C. K. Pang, Z. W. Zhong, and F. L. Lewis (2011), "Tool wear monitoring using acoustic emissions for dominant feature identification", IEEE Transactions on Instrumentation and Measurement, Vol. 60, No. 2, pp. 547-559.

4. J. H. Zhou, C. K. Pang, F. L. Lewis and Z. W. Zhong (2011), "Dominant feature identification for industrial fault detection and isolation applications", Expert Systems with Applications, Vol. 38, pp. 10676-10684.

5. J. H. Zhou, Z. W. Zhong (2011), "Reinforce morlet wavelet transform for bearing fault diagnosis", submitted to Mechanical Systems and Signal Processing.

6. J. H. Zhou, Z. W. Zhong (2011), "A dominant feature identification of wavelet selection on fluid dynamic bearings health assessment", submitted to Signal Processing.

Published International Conference Papers:

7. C. K. Pang, J. H. Zhou, Z. -W. Zhong, and F. L. Lewis (2011),"Industrial fault detection and isolation using dominant feature identification," ASCC 2011, TuB3.1, pp. 1018-1023. **Best Application Paper Award.**

8. J. H. Zhou, L Wee, Z. W. Zhong, (2010), “A knowledgebase system for rotary equipment fault detection and diagnosis”, IEEE International Conference on Robotics and Automation, Robotics and Vision (ICARCV), pp.1335-1340.

9. J. H Zhou, D. H. Zhang, Arogeti, S. A, M. Luo , Z. W. Zhong (2009), “*iDiagnosis & Prognosis*-An intelligent platform for complex manufacturing”, IEEE/ASME Advanced Intelligent Mechtronics, Vol. WB5.6, pp. 405 – 410.

10. C. K. Pang, J. H. Zhou, F. L. Lewis and Z. W. Zhong (2009), "Machine wear forecast using singular value decomposition for dominant feature identification", IEEE Advanced Intelligent Mechatronics, Vol. WB5.6, pp. 421-426.

11. J. H. Zhou, M. Luo, Q. Shao and Z. W. Zhong (2007), "Fluid dynamic bearing wear detection in brushless DC motor using wavelet-based correlation modeling", WCEAM 2007, Vol. 1, pp. 2249-2258

Table of Contents

List of Figures

List of Tables

List of Abbreviation

ACO	Ant colony optimization
ADFI	Augmented dominant feature identification
AE	Acoustic Emission
ANN	Artificial neutral network
ARMAX	Auto-Regressive Moving Average with exogenous
BLDC	Brushless direct current
CBM	Condition based maintenance
DDFI	Decentralized dominant feature identification
DFI	Dominant feature identification
DWT	Discrete wavelet transform
DRAM	Dynamic resource allocation management
ELS	Extended least square
FDI	Fault detection and isolation
FFT	Fast Fourier Transform
FNN	Fuzzy neutral network
FDB	Fluid dynamic bearing
GA	Genetic algorithm
HHT	Hilbert-Huang transform
JADE	Java agent development framework
LSE	Least square error
MAS	Multi-agent system
MCSA	Motor current signature analysis
MLP	Multi-layer perception
MSE	Mean square error
MRE	Mean relative error
MRM	Multiple regression model
MSSCS	Manufacturing system supervisory control system
NN	Neutral network
PCA	Principle component analysis

PFA	Principle feature analysis
RMS	Root mean square
RLS	Recursive least square
SBS	Sequential backward selection
SFS	Sequential forward selection
SOM	Support vector machine
SOA	Service-oriented architecture
SVM	Support vector machine
SVD	Singular value decomposition
TCM	Tool condition monitoring

List of Symbols

Λ	diagonal matrix
U	orthonormal matrix
Σ	Eigenvalue
V	Eigenvector
$\| \|^2$	second norm
$O( )$	O notation for computational complexity
f_0	fundamental frequency
f_{BPFO}	ball pass frequency for outer raceway
f_{BPFI}	ball pass frequency for inner raceway
f_{BSF}	ball spin frequency
f_c	cage frequency
$\Psi_i(f)$	Fourier transform of the mother wavelet
f_i	centre frequency
σ	shape of the mother wavelet
α	width of the wavelet window
Cw	wavelet coefficients matrix
F^{-1}	inverse Fourier transform
$X(f)$	Fourier transform of the original signal $x(t)$
f_s	sampling frequency
f_R	shaft rotational frequency
C_r	magnification factor
C_{rt}	threshold magnification factor
C_s	max-energy spectrum

CHAPTER 1 INTRODUCTION

1.1 Background

In an era of intensive competition, where asset usage and plant operating efficiency must be maximized, unexpected downtime due to machinery failure has become more costly than before. The earliest maintenance technique is basically run-to-failure maintenance and maintenance takes place only at breakdowns. A later maintenance technique is time-based preventive maintenance, which sets a periodic interval to perform preventive maintenance regardless of the health status of the equipment. With the rapid development of modern technology, more efficient maintenance approaches such as condition based maintenance (CBM) has been actively pursued in the manufacturing industry in recent years.

CBM [1] recommends maintenance actions based on the information collected through condition monitoring. CBM attempts to avoid unnecessary maintenance tasks by taking maintenance actions only when there is evidence of abnormal behaviors of the equipment. CBM can significantly reduce maintenance cost by reducing the unnecessary scheduled preventive maintenance operations where equipment outages are predicted and maintenance is carried out only when necessary.

Two important techniques in CBM are diagnostics and prognostics [1]. Computerised diagnostics and prognostics systems have become a valuable tool to assist maintenance person in making maintenance decisions, or possibly even replace maintenance experts in future. Today's concept of machine diagnosis comprises the automated detection and classification of faults, whereas machine prognosis is the automated estimation of how soon and likely a failure will occur.

To ensure successful condition based maintenance, it is necessary to detect, identify and classify different kinds of failure modes in the manufacturing process. Sensing technology is necessary in implementing the detecting system. There are two major approaches using sensing technology: the direct method, which measures and evaluates the fault from a single direct sensor, and the indirect method that adopts different kinds of sensors at different positions to sense different failure modes. Features are then extracted and selected to analyse the

signals from all these sensors to assess the condition of the system. In many applications, direct method is not possible for on-line realization. Indirect method may work as an on-line systematic technique as it measures the machine condition and operating parameters during the manufacturing process. In the indirect method, amount of sensors could be large and determining failure could mean taking minutes or hours to analyse data from all the sensors and make a reasonably accurate deduction.

Numerous machine parameters have been used to form the basis of machine health assessment, including motor speeds, input currents, transmitted torques, and vibration parameters such as accelerations, velocities, and displacements, etc. Each of these parameters might produce characteristic features that can be used for fault detection at the data processing phase. Also, many methods like statistical moment computation and spectral analysis of various machine parameters [2], Hilbert-Huang Transform (HHT) [3], minimum variance cepstrum, amplitude modulation and envelope analysis [4], Fast Fourier Transform (FFT), Statistical Computations, Wavelets [5–6], etc., have been used for feature calculation and extraction

In many complicated applications, it is not unusual to find problems involving hundreds of features. However, it has been observed that, beyond a certain point, the inclusion of additional features leads to a worse performance. Moreover, the choice of features affects several aspects of the recognition process such as accuracy, learning time and necessary number of samples. Most importantly, this leads to an increase in time and computational space complexity of the recognition process. The main goal of feature subset selection is to reduce the number of features used in classification without compromising on accuracy.

Feature subset selection algorithms can be classified into two categories based on whether or not feature subset selection is performed independently of the learning algorithm used to construct the verifier. The filter approach to feature subset selection is done independently of the learning algorithm. Another technique, known as wrapper feature subset selection [7], uses the method of classification itself to measure the importance of a feature or feature set. The wrapper approaches such as Genetic algorithm, ant colony are more

computationally expensive and tends to provide better results than the simpler filter methods [8]. For the filter approach, data compression schemes include Principal Component Analysis (PCA) [9-10] and Principal Feature Analysis (PFA) [11-13], which had been applied to pattern recognition to identify key features of original data. Both PCA and PFA require computationally intensive calculations of eigenvectors (and their corresponding eigenvalues) based on the correlation or covariance matrix to derive a linear transformation matrix for feature selection to compress the raw data into lower dimensions. New feature identification methods to improve the computational efficiency and provide an autonomous decision for data and sensor reduction are desired for further research.

Industry has always had an urgent need for prediction of fault progression and remaining use life. Prediction leads to improving management and hence effectiveness of equipment, and multifaceted guarantees are increasingly being given for industrial machines, products, and services [14]. Time series prediction has been extensively studied using various methods, e.g., (a) linear methods including classical estimation theory using time- or frequency-domain methodologies, (b) nonlinear estimation and function approximations as well as identification and fault diagnosis using Neural Networks (NNs) [15-16], or (c) methods using wavelet transforms [17], support vector machines [18], diagnosis model based on fault trees [19], etc. These methods normally segregate raw data into a "training" set and "validation" set — the former used to train the underlying prediction mechanism, and the latter used to test if the identified model is of satisfactory accuracy [20]. These methods are typically offline approaches and require extensive computations. It is postulated that the integration of a formal decision methodology for selecting the dominant features with the advanced feature extraction and intelligent modelling method will significantly improve the machine health prediction accuracy and efficiency.

CBM researches have been focused on the diagnosis and prognosis of individual equipment with little research regarding present day complex manufacturing systems. Complex manufacturing systems usually consist of many different types of equipment. Supervisory control of complex manufacturing systems requires the monitoring and control with a clear objective to economically maximize the Quality of Service (QoS) provided by the prevailing manufacturing

resources and to achieve near zero down time operations. Three most important elements of the control inputs are the system/equipment health condition, system capability and job execution capability of the manufacturing systems. This is because any unexpected system/equipment down time in the process chain will have a ripple effect on downstream processes and can be highly disruptive to the shop floor as equipment, labour and materials have to be rescheduled and reallocated.

Over the past ten years, extensive research and development work have been devoted to e-maintenance. E-maintenance is the gradual replacement of traditional maintenance types by more predictive / proactive types [21]. Regular periodic maintenance should be advanced and shifted to the intelligent maintenance philosophy to satisfy the manufacturer's high reliability requirements [22]. Hence, Koc and Lee [23] defined e-maintenance (system) to be a predictive maintenance (system) which provides monitoring and predictive prognostic functions. E-maintenance systems provide the equipment monitoring, diagnosis and prognostics functions, but none of previous systems provides manufacturing system condition and job execution capability assessment functions. Other related researches focused on the e-maintenance platform development based on the Internet, web, and agent technologies. Zhang et al. [24] consider that e-maintenance is a combination of web service technology and agent technology, which provides a way to realize intelligent and cooperative features for the systems in an industrial automation system. Crespo Marquez and Gupta [25] define e-maintenance as a distributed artificial intelligence environment, which includes information processing capability, decision support and communication tools, as well as the collaboration between maintenance processes and expert systems. Despite these efforts, we are still lacking a system that can provide distributed diagnosis and prognosis to assess system condition and job execution capability and provide dynamic information feedback to the supervisory control when manufacturing system breakdown or performance degradation occurs. When working with other modules in the supervisory control system, a distributed diagnosis system can greatly improve the shop floor efficiency and enhance the competency of the manufacturing companies in the global competition environment.

1.2 Objectives of the research

This research aims to address the critical issues concerning the realization of a maintenance decision support system for the CBM in a complex manufacturing environment. A system framework of *iDiagnosis & Prognosis* is proposed and the system architecture is shown in Figure 1.1. *iDiagnosis & Prognosis* system provides distributed intelligence of diagnosis and prognosis, system condition and job execution capability assessment, and dynamic information feedback to the supervisory control when manufacturing system breakdown or performance degradation occurs. Meantime, machine diagnosis and prognosis information is also provided for maintenance decision making. Inside the iDiagnosis/Prognosis frame work, sensing first collects information from different types of sensors. The collected information is fused to yield maximum useful information. Digital signal process removes the noise using wavelet decomposition. Time domain, frequency domain and time-frequency domain analyse are used to extract features. Dominant feature selection reduces the feature set to a more representative subset which is sensitive to the equipment condition. Intelligent decision making processes, classifies and associates the information of the sensory signal pattern with the monitored equipment condition and recommends efficient maintenance policies. This thesis firstly concentrates on the development of a novel Dominant Feature Identification (DFI) algorithm to provide an autonomous rule-base for data and sensor reduction which is presented in Chapter 3. An advanced feature extraction and selection methodology is then developed that integrates the proposed DFI with feature extraction methods of time domain and frequency domain analysis methods, working together with Recursive Least Squares (RLS) and Extended Least Squares (ELS) modelling and Neutral Network methods. The efficiency of the advanced feature extraction and selection methodology is evaluated based on the accuracy of health condition assessment of the actual milling tool wear, brushless DC motor, machine spindles and rotary equipments. The research works of advanced feature extraction and selection methodology are presented in Chapters 4, 5, 6 and 7. The system framework, architecture, functions, and information flow of *iDiagnosis & Prognosis* are developed to extend the research from single machine CBM to a complex manufacturing environment and are presented in Chapter 8.

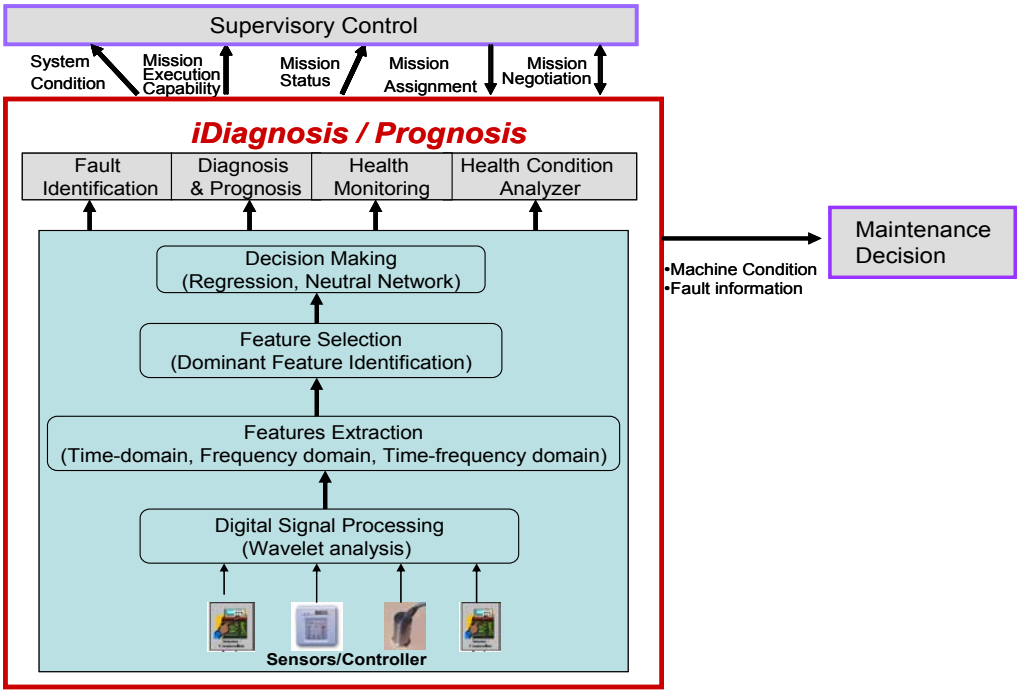


Figure 1.1 System architecture

The overall objectives of the research are:

1. To develop a methodology for selecting the dominant features that are most essential in predicting time series that provides an autonomous decision for feature reduction (Chapter 3).

2. To develop a framework to integrate the feature extraction methods with dominant feature selection method and intelligent modelling methods to provide efficient and accurate machine health condition prediction. The developed framework is evaluated with real application of tool wear prediction (Chapter 4).

3. To enhance the DFI methodology from features selection to sensors selection to provide an autonomous rule-base for both data and sensors reduction. Using Neural Networks (NNs) to classify faults based on reduced features set and extend the (DFI) method for multiple components and multiple types of faults detection and isolation (Chapter 5).

4. To enhance the DFI methodology from features selection to dominant mother wavelet selection and dominant wavelet decomposition levels selection. The developed methodology is evaluated with real application of motor health condition prediction (Chapter 6).
5. To develop a new feature extraction method to diagnose bearing fault. The proposed method needs to overcome the difficulties in implementing Morlet wavelet transform, to achieve high consistency and detect fault in early stage (Chapter 7).
6. To develop a system framework to provide distributed intelligence of diagnosis and prognosis, to provide system condition and job execution capability assessment, and provide the dynamic information feedback to the supervisory control when manufacturing system breakdown or performance degradation occurs (Chapter 8).

1.3 Scope of the research

The scope of this research is to address a few of the important problems in condition based maintenance.

Firstly, this thesis concentrates on the development of a novel Dominant Feature Identification (DFI) to provide an autonomous rule-base for data and sensor reduction. Singular Value Decomposition (SVD) is used to decompose the inner product matrix of collected data from the monitoring sensors. The principal components are optimized in a least squares sense in a certain reduced space, and the dominant features are extracted using the k-means clustering algorithm. The proposed DFI decision tool is numerically efficient, and reduces the complexity of feature selection greatly while reducing the least square error during dominant feature selection and clustering automatically.

Secondly, a CBM framework is developed that integrates the proposed DFI with traditional feature extraction methods of time domain and frequency domain analysis methods, working together with Recursive Least Squares (RLS) and Extended Least Squares (ELS) modelling methods. The developed methodology and framework are evaluated based on the accuracy of prediction of the actual milling tool.

Next, Augmented Dominant Feature Identification (ADFI) and Decentralized Dominant Feature Identification (DDFI) are proposed for industrial Fault Detection and Isolation (FDI). The proposed ADFI and DDFI are tested on a fault simulator machine for both data and sensors reduction. Neural Networks (NNs) is used for the identification of different types of faults.

Furthermore, two important research problems of using wavelet approaches in the condition based maintenance are studied: one is appropriate choosing mother wavelets and decomposition levels. A methodology using DFI to select the dominant mother wavelet and decomposition levels is proposed. The proposed methodology is evaluated based on the accuracy of health condition assessment of the brushless DC motor; another important aspect for study is to optimize the wavelet parameters. Reinforced Morlet wavelet transform for bearing fault diagnosis is proposed to optimize the wavelet parameters to achieve high consistency for bearing fault detection in an early stage.

Finally, to extend the research from single machine CBM to a complex manufacturing environment, a system framework of *iDiagnosis & Prognosis* is developed to provide distributed intelligence of diagnosis and prognosis, to provide system condition and job execution capability assessment, and provide dynamic information feedback to the supervisory control when manufacturing system breakdown or performance degradation occurs. The system framework, architecture, functions, and information flow are developed for implementing the diagnosis & prognosis in a complex manufacturing environment.

1.4 Organization of the thesis

This thesis is organized into 9 chapters. Chapter 2 presents a broad and thorough literature review on the key issues and related investigations in the CBM. These include the critical challenges for sensing, feature extraction, feature selection, intelligent modelling for decision support system. Based on the overall review of techniques for CBM, some conclusions and gaps are drawn from the literature. In Chapter 3, a rigorous mathematical framework for Dominant Feature Identification (DFI) is developed that provides an autonomous rule-base for data and sensor reduction. Singular Value Decomposition (SVD) is used to decompose the inner product matrix of collected data from the monitoring sensors. The

principal components are optimized in a least squares sense in a certain reduced space, and the dominant features are extracted using the k-means clustering algorithm. Chapter 4 proposes the framework to integrate the DFI with traditional feature extraction method and least square modelling method. The proposed framework is evaluated based on the accuracy of prediction of the actual milling machine tool. Chapter 5 proposes Augmented Dominant Feature Identification (ADFI) and Decentralized Dominant Feature Identification (DDFI) for both sensors and features reduction. Neural Networks (NNs) is used for faults identification. The proposed ADFI and DDFI are tested on a fault simulator machine for industrial Fault Detection and Isolation (FDI). Chapter 6 proposes wavelet based correlation modelling to integrate with DFI. The proposed methodology is evaluated based on the accuracy of health condition assessment for brushless DC motor. Chapter 7 proposes reinforced Morlet wavelet transform to overcome the difficulties in implementing Morlet wavelet transform and achieve high consistency for bearing fault detection in an early stage. Chapter 8 proposes a system framework of *iDiagnosis & Prognosis* to provide distributed intelligence of diagnosis and prognosis, to assess system condition and job execution capability, and provide dynamic information feedback to the supervisory control for a complex manufacture environment. The last chapter, Chapter 9 summarizes the main conclusions reached in this work and highlights the overall direction of the future work for this research.

CHAPTER 2 LITERATURE REVIEW

This chapter introduces the area of condition based maintenance and various techniques applied in this area. The first Section discusses the importance of condition based maintenance (CBM) and introduces two important application areas in CBM. The second Section explains the key steps for realizing CBM. Different fault diagnosis and prognosis technology like sensing, signal processing, features extraction, features selection and Artificial Intelligence (AI) modelling techniques in the condition based maintenance are discussed. Based on the overall review of CBM techniques, some conclusions and proposed research areas are drawn from the literature.

2.1 Introduction

2.1.1 Condition Based Maintenance (CBM)

In an era of intensive competition, where asset usage and plant operating efficiency must be maximized, unexpected downtime due to equipment failure has become more costly than before. Equipment reliability has always been an important aspect in modern industry. No matter how good the equipment designs, equipment performance degradation over time since the equipment is operating under certain stress or load in the real environment. Thus, maintenance has been introduced as an efficient way to assure a satisfactory level of reliability during the useful life of the equipment.

In general, maintenance strategies can be divided into three categories based on the underlying principles employed, i.e. run-to-failure maintenance, preventative maintenance and Condition Based Maintenance (CBM), as shown in Table 2.1.

Table 2.1 Category of maintenance strategies

Maintenance Strategy	Basic Principle
Run-to-failure maintenance	Machine runs to failure and then repair is carried out
Preventative maintenance	Periodic component replacement
Condition-based maintenance	Maintenance based on machine condition

CBM [1] is a maintenance program that recommends maintenance actions based on the information collected through condition monitoring. CBM attempts to avoid unnecessary maintenance tasks by taking maintenance actions only when there is evidence of abnormal behaviors of the equipment. CBM can significantly reduce maintenance cost by reducing the unnecessary scheduled preventive maintenance operations.

Two important researches in CBM are diagnostics and prognostics. Computerized diagnostics and prognostics systems have become a valuable tool for maintenance personnel in decisions making and might replace maintenance experts in future. Today's concept of machine diagnosis comprises the automated detection and classification of faults, whereas machine prognosis is the automated estimation of how soon and likely a failure will occur. The literature on CBM is huge and diverse. Hundreds of papers in this area appear every year in academic journals, conference proceedings and technical reports. Related reviews on CBM have been reported in the literature. Posey and Roemer [26] provided a broad overview of the development in diagnostics and prognostics technologies applicable to high-performance turbo-machines up until year 1999. Jar dine et al. [27] provided an overview of publications on data acquisition, data processing, diagnostics and prognostics of various machines up to year 2005. Vachtsevanos et al. [28] described intelligent fault diagnostics and prognostics approaches for engineering systems through examples. Ma [29] discussed the need for a new paradigm shift in CBM research for engineering asset management. Heng [30] provided an overview of the state of the art, challenges and opportunities in rotating machinery prognostics.

The CBM methodology and techniques have been extensively researched in a variety of application areas, such as rotating machinery, aerospace systems, chemical manufacturing, electronic and electrical components, etc.

In general, a CBM system usually [31] [32] consists of sensing, signal conditioning devices, signal processing algorithms, signal interpretation and decision-making procedure. This system should sense signals indicating the equipment status and its health changes, interpret incoming sensed information,

and decide on the appropriate control or repair actions. As illustrated in Figure 2.1, a CBM system should include the following five key steps:

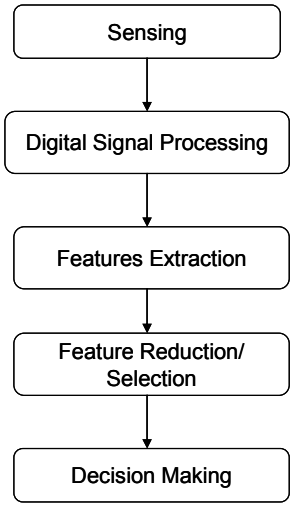


Figure 2.1: Five key steps in a CBM system

Step 1: Sensing technique: indirect sensing techniques such as force, vibration, acoustic emission are used. Different types of sensors and sensory data are fused to yield maximum useful information.

Step 2: Signal processing: the data or signals collected in step 1 are handled and analyzed for better understanding and interpretation of the data.

Step 3: Feature extraction: from the digital signals employed in step 2, different types of features are computed and represented by a compact vector of features. The feature set is sensitive to the monitored equipment condition.

Step 4: Feature reduction/selection: computed features at the signal processing stage may consist of noise, irrelevant or redundant information. Reducing the feature set to a more representative one which is sensitive to the equipment condition and insensitive to noise is a vital step in CBM.

Step 5: Decision making: the information of the sensory signal pattern is processed, classified and associated with the monitored equipment condition. Efficient maintenance policies are recommended by intelligent decision making.

The literature review of this chapter in the following Sections will review methods and techniques addressing each step of a CBM system.

2.1.2 Two important application areas in CBM

With the increasing demand for the production cost reduction and product quality improvement, tool condition monitoring and rotary equipment condition monitoring are two important areas that attract much attention of research works.

2.1.2.1 Tool condition monitoring

End milling cutters suffering from various types of progressive wear will degrade the quality of workpiece produced. If cutting continues with a worn cutting tool, it will break causing further damage to the workpiece being milled. Machine operation specialists are able to predict the amount of wear on the cutting edges by listening to the cutting, feeling the vibrations of the milling machine, and inspecting the chips produced during cutting. However, it is not feasible to dedicate a specialist to constantly monitor tool wear. The use of inexperienced workers for tool condition inspection (tool wear or tool breakage) will result in unawareness of the developing tool wear until several defective products have been produced, or they may change tools prematurely resulting in underutilization of the tool, which increases machining cost. An automatic tool condition monitoring (TCM) system is needed that is able to alert the operator when cutter replacement is warranted due to process abnormality. The TCM is defined as an integrated system consisting of sensing, signal conditioning devices, signal processing algorithms and signal interpretation and decision making [33]. Many research works have been carried out in TCM [34] [35] [36] [37]. Effective TCM systems can be used for several purposes, including the institution of tool change policy, economic optimization of machining operations, compensation of tool wear online, and avoidance of catastrophic tool failure.

2.1.2.2 Rotary equipment condition monitoring

Rotating machines have widespread applications ranging from generators in power plant, gearbox in wind turbine, aerospace engines and stamping machines in factories to motorised vehicle engines. As a result, they have gained such significance that they become one of the inseparable components in our lives.

Typical rotating machines composed of numerous components such as shafts, bearing, gearbox, coupling, electric motors and belt drives [38]. The conditions of these components contribute to the overall machine condition and are very important diagnostic objects in industry. Among the machinery parts, rolling bearings are most often used mechanical parts. There are many problems which can be related to bearings and their vibrations. Installation problems are relatively common and are often caused by improperly forcing the bearing onto the shaft or in the housing. Misalignment of bearings is a common result of defective bearing installation. The problems can cause bearing failure. Usually the failure is accelerated by overloading, over speeding or starving the bearings of lubricants.

Figure 2.2 presents the results surveyed by Brüel & Kjaer (B&K) on the faults detected in a machine. These findings show that bearing related faults are the most common machine fault accounting for about 40% of total machine failures. Belt related faults book the next 25%.

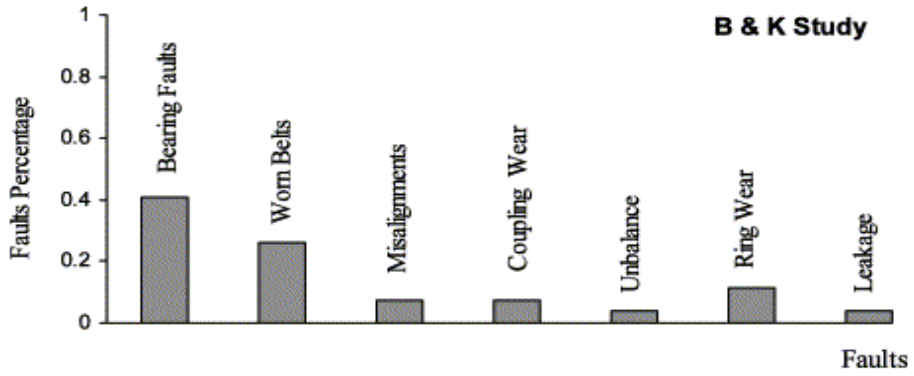


Figure 2.2: B& K survey on major faults detected in a machine [38]

Other two studies were conducted to demonstrate which fault is most probable to occur in motors. Both studies were in concordance that the leader is bearing faults accounting for about 40% of the total failures as shown in Figure 2.3.

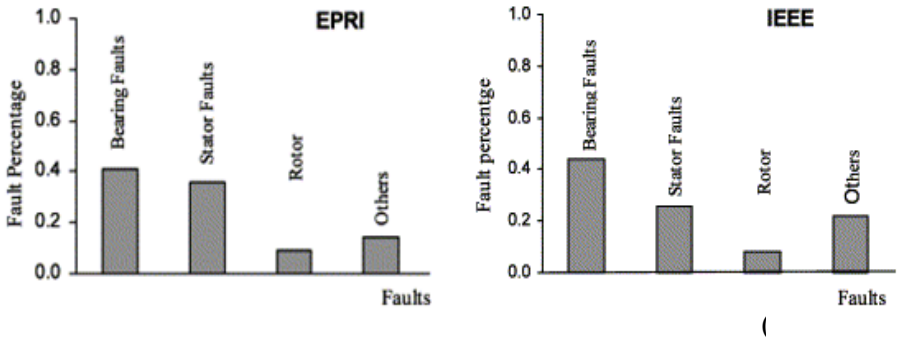


Figure 2.3: Study on induction motor failure by (a) EPRI and (b) IEEE [41]

A fault in the early stage may cause unit damage, high energy losses and efficiency reduction of the machine [39]. If not diagnosed, the fault becomes a catastrophic failure and production shutdown is a likely outcome.

Signal processing and feature extraction are important techniques for diagnosing rotating machines faults. The techniques are widespread and can range from statistical to model based techniques and compose of a variety of signal processing algorithms. Fault detection and identification is a subsequent step and recently artificial intelligence techniques are incorporated for fault diagnosis and prognosis. The advances of fault identification techniques enhance efficiency and reliability of fault diagnosis [40] [41] [42]. With the advancement of modern fault diagnosis techniques, researchers synthetically employ techniques to achieve effectiveness of fault diagnosis.

2.2 Five key steps in CBM

In this Section, CBM research and techniques are reviewed in detail. The Section is organized according to the five keys steps shown in Figure 2.1. The review covers sensing, signal processing, feature extraction, feature reduction/selection and decision making methods in CBM.

2.2.1 Sensing in condition monitoring

Data acquisition is a process of collecting and storing useful information from targeted equipment for the purpose of machinery fault diagnosis and prognosis. This process is an essential step in implementing a CBM program. The data captured for equipment condition monitoring are very versatile. They can be

vibration data, acoustic data, oil analysis data, temperature, pressure, moisture, humidity, weather or environment data, etc. Various sensors, such as accelerometer, current transducer, power sensors, acoustic emission sensors, etc., have been designed to collect different types of data [43, 44]. With the rapid development of computer and advanced sensor technologies, data acquisition technologies have become more powerful and less expensive, making data acquisition for CBM implementation more affordable and feasible.

As mentioned in Section 2.1.2, tool condition monitoring and rotary equipment condition monitoring are two important application areas. The reviewing in this Section mainly covers the sensing techniques in tool condition monitoring and rotary equipment condition monitoring.

2.2.1.1 Sensors in tool condition monitoring

Tlusty & Andrews [45], Dan et al. [46], Dutta et al. [47] and Sick [48] reviewed sensors and techniques for tool condition monitoring. They reported that most CBM systems in tool condition monitoring employ forces, acoustic emission, vibration, current/power sensors or combination of these sensors. The sensors are listed below in more detail.

- Cutting Force

Cutting force is one important characteristic variable to monitor during cutting processes [45]. Research results show that tool breakage, tool wear and workpiece surface quality are strongly related to cutting force [49] [50].

Force measurements are commonly taken by using a table-mounted dynamometer during machining. The dynamometer measures the cutting force in three directions: the X, Y and Z axis. During end milling the Z axis cutting force component contains little information, but the X and Y axis signals have been shown to contain useful information for process modeling [51].

- Acoustic Emission

During metal cutting, the workpiece undergoes considerable plastic deformation as the tool pushes through it. Within the deformation zones, strain energy is released. This released energy is commonly referred to as acoustic emission. Acoustic emission has been used to study the fundamentals of the

machining process, e.g., the chip formation mechanism (shear and plastic deformation), friction between chip, tool rake face, tool flank wear in both turning and milling operations [52].

The AE sensor is small and easy to position on the workpiece or tools holder. The use of AE sensing for tool condition monitoring is becoming an attractive sensing method due to its non-intrusiveness, simplicity of operation and fast dynamic response.

- Vibration

The vibration signal generated by the end milling operation provides a great deal of information about the state of the end mill cutter. Information about the natural frequencies of workpiece, fixture, machine and spindle [53] is contained in the vibration signals.

When the cutting edge of a tool is sharp, vibrations are generated by the formation of chips, as metal is being removed from the surface of the workpiece by the cutting tool. At the cutting edge wears, additional signal information is generated by the sliding contact between the workpiece and wear-land along the flank side [54]. However, it was found that vibration signal is sensitive to the environment noise, e.g. coolant or other machine centers that operate near by, thus causing challenges using vibration signals in TCM.

- Power of spindle motor

In the end milling process, spindle motor current/power is related to the motor shaft torque or force component in the feed direction. When the tool is sharp, less force is required in feed or axial directions. As the tool develops flank wear, forces in all these directions increase to varying degrees. This variation in forces, particularly in the feed direction, increases the power required to drive the shaft. The spindle current is usually measured through a current transducer. The measured signal is found to drop instantaneously and soon recover to the pre-drop level when tool breakage occurs. This behavior is used to detect tool breakage. However, it is found that power/spindle current is less sensitive to tool wear sensing compared to force and AE measurement [55].

In this research, our experiments are focused on the using of force and AE

sensors as they are more effective and informative in tool condition monitoring.

2.2.1.2 Sensors in rotary machine condition monitoring

Two important sensing techniques for rotary equipment condition monitoring are vibration and electrical monitoring.

- Vibration Monitoring

Vibration monitoring is based on the concept that mechanical vibrations at various frequencies can be used to provide an indication of the condition of the machine. Bearing faults, rotor eccentricities, gear faults and unbalanced rotors are the best candidates for vibration based diagnostics. The vibration energy of the machine is measured in units of one of three related quantities: displacement, velocity, or acceleration. These measurements are accomplished using displacement probes, velocity transducers, or accelerometers respectively. The vibration monitoring is accomplished through the use of broad-band, narrow-band, or spectral (signature) analysis of the measured vibration energy of the machine [56].

- Electrical Monitoring

Monitoring the stator current, the stator voltage, sensing the air-gap flux, measuring shaft voltages and currents, all fall under the category of electrical monitoring. These methods are used to detect various kinds of machine and inverter faults. The most common form of signal analysis technique used in electrical monitoring is the current spectra analysis, often termed as "Motor Current Signature Analysis (MCSA)" [57]. It has been shown that there is a relationship between the mechanical vibration of a machine and the magnitude of the stator current component at the corresponding harmonics [58]. For increased mechanical vibrations, the magnitude of the corresponding stator current harmonic components also increases. This is because the mechanical vibration modulates the air gap at that particular frequency. These frequency components then show up in the stator inductance, and finally in the stator current. For this reason, the MCSA can be used to detect rotor and bearing faults. The purpose of the MCSA is to detect changes in a machine's condition by examining the current signals [59]. The MCSA uses the current spectrum of the machine for locating characteristic fault

frequencies. The spectrum is obtained using a Fast Fourier Transformation (FFT) that is performed on the signal under analysis. The fault frequencies that occur in the motor current spectra are unique for different motor faults. Electrical monitoring holds much promise for the future as it is inexpensive and reliable.

2.2.2 Signal analysis and features extraction

The most common waveform data in condition monitoring are vibration signals and acoustic emissions. Other waveform data are cutting force, ultrasonic signals, motor current, partial discharge, etc. There are numerous signal processing techniques and algorithms in the literature for diagnostics and prognostics. In the literature, there are three main categories of waveform data analysis techniques: time-domain analysis, frequency-domain analysis and time-frequency analysis. Various signal processing and features extraction researches have been carried out for the implementation of CBM.

In rotary equipment CBM, Tandon [60] presented a review of vibration and acoustic measurement methods for detection of the defects in rolling element bearings. Vibration measurement in both time and frequency domains along with signal processing techniques such as the high-frequency resonance technique have been covered. Recent trends in research on the detection of bearings defects, such as the wavelet transform method and automated data processing, have also been included. Time domain features of Root Mean Square (RMS), kurtosis and shock pulse methods have been analyzed. In the frequency domain, Fast Fourier Transform (FFT), power spectrum, enveloping detection or high-frequency resonance technique (HFRT) are reviewed as important signal processing techniques.

In tool condition monitoring, various features have been proposed to be extracted for monitoring the force signals in milling processes. Altintas [61] proposed a first-order auto-regression (AR1) model to evaluate the difference between the actual measurement and the predicted value, which was called the residual error of the cutting force. In another attempt by Altintas [62], the first- and second-order differencing of a time-averaged resultant force was used to detect tool failures. Tarn [63] employed four parameters of maximum force level, total amplitude of the cutting force, combined incremental force changes, and the

amplitude ratio from each tooth period to monitor the tool and cutting conditions. Tarng [64] defined the frequency range between the DC component and the tooth frequency as a tool breakage zone. He extracted the tool breakage zone components using a band pass filter. The standard deviation of the force data was then calculated. Zhang [65] used the peak rate of cutting forces to detect tool breakages. It was defined as the ratio between the difference and the sum of the force peaks in adjacent tooth periods. Elbestawi [66] found that the harmonic contents of the cutting forces were sensitive to tool flank wear. He summed up the powers at the fundamental tooth frequency and its harmonics, and derived a total harmonic power. Tarng [67] used the average force and the variable force, derived by subtracting the median force from the average force, as input features to a multi-layer perceptron (MLP) neural network for sensing tool breakage. Leem [68] extracted five statistics from the cutting force: the mean, standard deviation, crest factor, skew, and kurtosis. Then, these features were fed into a customized neural network to monitor tool wear.

2.2.2.1 Time-domain analysis

Time-domain analysis [60] is directly based on the time waveform to calculate characteristic features from time waveform signals as descriptive statistics such as mean, peak, peak-to-peak interval, standard deviation, crest factor, high-order statistics: root mean square, skewness, kurtosis, etc. These features are called time-domain features. The changes in the statistics features could be due to fault, load changes or speed changes.

Instead of studying the probability density curves, it is often more informative to examine the statistical moments of the data, defined by [60] as

$$M_x = \int_{-\infty}^{+\infty} x^n P(x) \qquad n = 1,2,3,.....m \tag{2.1}$$

Where $P(x)$ is the probability density function of the instantaneous function amplitude x. The first and second moments are well known, being the mean value and the variance, respectively. The third moment normalized with respect to the cube of standard division is known as the coefficient of 'Skewness'. Kurtosis is defined as the fourth moment of the distribution and measures the relative the

peakedness or flatness of a distribution compared to a normal distribution. Kurtosis provides a measure of the size of the tails of distribution and is used as an indictor of major peaks in a set of data. The formulas for these four moments [60] are described below:

- Mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.2}$$

- Standard Deviation

The standard deviation of a set of N sample data points x_i is given by

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}} \tag{2.3}$$

where $\bar{x}$ is the sample mean.

- Skewness

Skewness is generally the measure of symmetry or lack of symmetry of the probability distribution of a data set. Two kinds of skewness: negative skewness and positive skewness are shown in Figure 2.4.

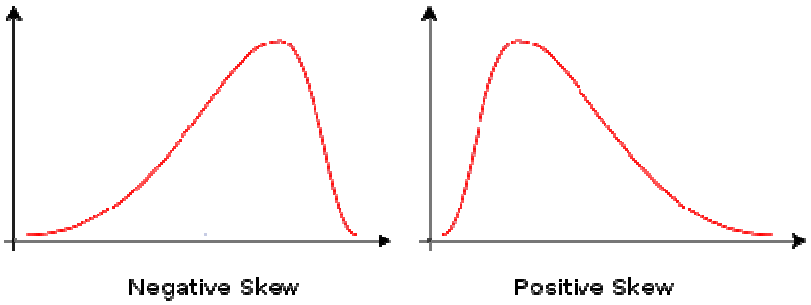


Figure 2.4: Two types of skewness [69]

Skewness is described as:

$$Skewness = \sum_{i=1}^{N} \frac{(x_i - \bar{x})^3}{(N-1)s^3} \tag{2.4}$$

where,

$\bar{x}$ = Mean of data points,

x_i = Data point at i,

s = Standard deviation,

N = Number of data points

The skewness is one of the good indicators to detect the change of the shape of the signal.

- Kurtosis

Kurtosis is one of the most popular condition indicators used in fault analysis. It is defined as the normalized fourth central moment.

$$Kurtosis = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^4 \Big/ N}{\sigma^4} \tag{2.5}$$

The kurtosis is a measurement of the number and magnitude of large excursions from the RMS value in a signal [68]. When a tool wears, this feature should signal an error due to the increased level of vibration.

- Root Mean Square (RMS)

Root mean square (RMS) is the statistical measure of the magnitude of the varying quantity. It can be calculated as:

$$RMS = \sqrt{(\frac{1}{N}\sum_{i=1}^{N} x_i^{\,2})} \tag{2.6}$$

The RMS value of a signal measures the power content in the signal. Root mean square can be used as the energy indicator of the signal but it does not characterize the shape of the signal [70]. Delta RMS is the difference between the current RMS value and previous. It measures the changes in the Root Mean Square of the signal. RMS and Delta RMS are very efficient in measuring the system overall performance changes, but are often not sensitive enough to detect incipient fault and which component is failing.

- Peak and peak to peak

Peak [63] [65] is simply the maximum value of the signal. Peak amplitude is the maximum excursion of the wave from the zero or equilibrium point. Peak to peak [63] is the range between the maximum and minimum values of the signal. It

is the difference between the maximum positive and the maximum negative amplitudes of a waveform. Figure 2.5 illustrates the ideas of peak and peak to peak.

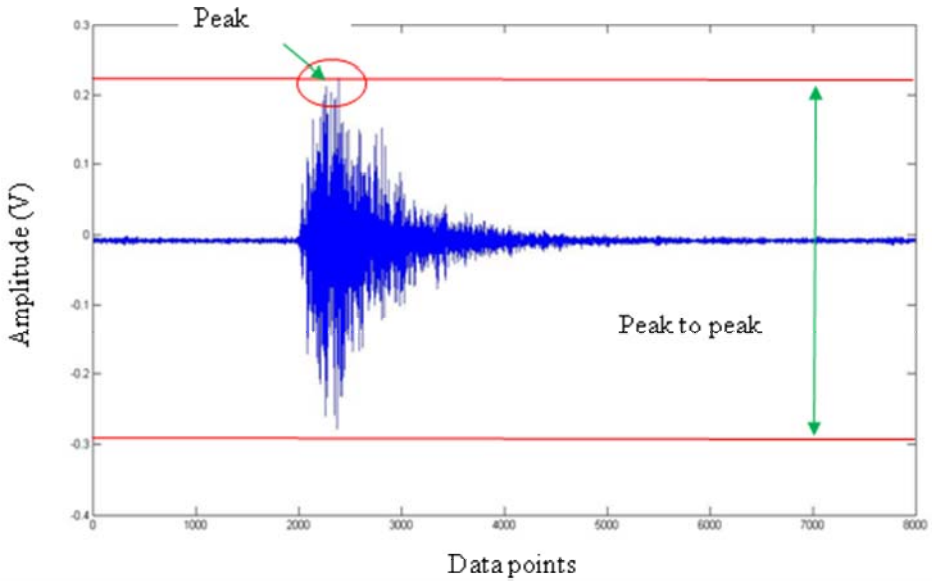


Figure 2.5: Peak and peak to peak

- Crest Factor

The RMS level may not show appreciate changes in the early stages of gear and bearing damage. A better measure is to use "Crest Factor" [63] which is defined as the ratio of the peak level of the input signal to the RMS level. The crest factor can be calculated as:

$$Crest\ Factor = \frac{Peak}{RMS} \tag{2.7}$$

where,

Peak = Maximum value of signal

RMS = Root mean square of signal

Therefore, peaks in the time series signal will result in an increase in the Crest Factor value. Crest Factor may reach between 2 and 6 in normal operations, and a value above 6 is usually associated with machinery problems. This feature is used to detect changes in the signal pattern due to impulsive vibration sources such

as tooth breakage on a gear or a defect on the outer race of a bearing. Crest factor can be used to characterize the time signal induced by defects [71].

Examples of use RMS, peak value, kurtosis and crest factor and combined with a high frequency resonance technique to detect and localize the damage in rolling bearing was reported in [71].

- Absolute deviation

The absolute deviation can be denoted as the measure of central tendency. It can be defined as:

$$D_i = \left| x_i - \bar{x} \right| \tag{2.8}$$

Where,

D_i=Absolute deviation,

x_i = Data set,

$\bar{x}$ = Mean of the data

Generally, absolute deviation is used to measure the statistical dispersion which can be denoted as variety of spread in variable or in probability distribution. It can be used to analyze changes in signal during cutting time [67] and these changes can compare with the tool wear or surface roughness.

- Count, rise time and duration

Count is the number of time signal exceeds the preset threshold value. Rise time is the duration of time interval in which signal starts to exceed the threshold value and reaches to its peak value. Duration is the time duration in which signals starts to exceed the threshold value to the time it goes down under the threshold value. Figure 2.6 shows the rise time and duration of the signal in a cycle of time.

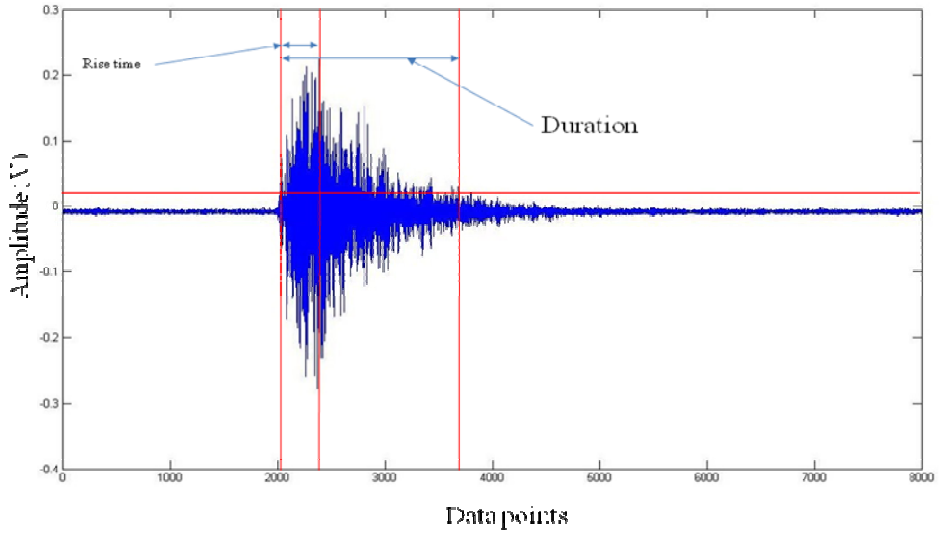


Figure 2.6: Rise time and duration

- Area under curve

  Area under curve can be calculated as:

$$Area = \int_{a}^{b} f(x)dx \tag{2.9}$$

Figure 2.7 explains the idea of area under the curve.

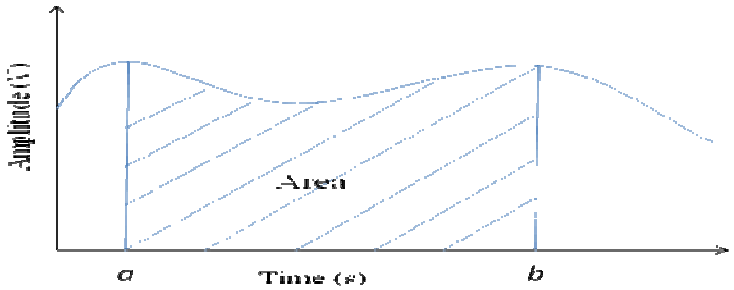


Figure 2.7: Area under curve

2.2.2.2 Frequency feature extraction techniques

Frequency-domain or spectral analysis of the vibration is the most widely used approach of rotary machine defect detection. The Discrete Fourier Transform (DFT) and Fast Fourier Transform (FFT) are the most conventional diagnosis techniques which have been widely used [72]. FFT and DFT provide complex frequency spectra that can in turn be used to calculate both the phase and

magnitude spectra of signal. Each one of the frequency bins has a resolution of f_s / N, where f_s is the sampling frequency. Fast Fourier Transform is given as [72] :

$$X(k) = \sum_{j=1}^{N} x(j)\omega_N^{(j-1)(k-1)} \tag{2.10}$$

where $X(k)$ is the Fourier Transformed data

$x(j)$ is the raw signal data

ω_n is the n^{th} root of unity

In order to conduct the FFT smoothly, the data length needs to be satisfactorily long. The number of points also needs to be adequate so that the data will have improved signal to noise ratio. In addition, the sampling rate must be at least twice the maximum frequency component as defined by the Nyquist criterion.

The power spectrum whose amplitude is square of amplitude of spectrum and it is an effective method for diagnosing machinery faults. The power spectrum calculates the average of the magnitude of the DFT, and is defined as [72]:

$$S_{xx}(k) = E\{X(k)X^*(k)\} \tag{2.11}$$

where $X^*(k)$ is the complex conjugate of $X(k)$. This is probably the most common spectral technique used. By examining the peaks, harmonics can be spotted in the spectrum, and from this, the frequencies of interest can be calculated, and related back to the faulty components.

Almost all the rotor and bearing faults cause mechanical vibrations at specific frequencies [73, 74]. A listing of possible machine faults and their related vibration frequencies is given in Table 2.2 [60] [2].

Table 2.2

Occurrence of some common faults and characteristic frequencies

Fault	Frequencies at Which High Vibration Levels are Expected (f_r: Rotational frequency = RPM/60)
Unbalance	f_r (Mainly in radial direction)
Misalignment	f_r, $2f_r$ (Often $2f_r$ is higher than f_r and axial vibration may be higher than radial vibration)
Mechanical looseness and rubbing	$0.5f_r$, f_r and a number of their higher harmonics
Oil whirl in journal bearings	40-50% of f_r
Damaged rolling bearings	At rolling element pass frequency given by: Outer race defect = $\frac{N}{2} f_r \left(1-\frac{RD}{PD}\cos\alpha\right)$ Inner race defect = $\frac{N}{2} f_r \left(1+\frac{RD}{PD}\cos\alpha\right)$ Rolling element defect = $\frac{PD}{RD} f_r \left[1-\left(\frac{RD}{PD}\cos\alpha\right)^2\right]$ Cage defect frequency = $\frac{f_r}{2}\left(1\pm\frac{RD}{PD}\cos\alpha\right)$ (+ sign if outer race is rotating, - sign if inner race is rotating) N: Number of rolling elements (balls/rollers) F_r: shaft rotational speed, Hz RD: Rolling element diameter PD: Pitch circle diameter α: Contact angle

Other useful signal processing spectra have been developed and shown their own advantages over FFT spectrum in certain cases. Cepstrum [75] has the capability to detect harmonics and sideband patterns in power spectrum. Cepstrum is calculated by determining the natural logarithm of magnitude of the Fourier transform of x (raw signal), then obtaining the inverse Fourier transform of the resulting sequence. It was used as the effective method in the diagnosis of bearing faults.

Few methods have been derived such as singular spectrum, envelope spectrum, and power spectrum. The procedures for obtaining the spectrum of the envelope signal by a high frequency resonance technique are well established and applied in bearing fault detection [76].

2.2.2.3 Time-frequency feature extraction techniques

In recent years, the time-frequency signal analysis has been studied and

applied to machinery fault diagnosis due to its capability of representing signals in both time and frequency domains. This characteristic of time-frequency analysis technique meets the requirement for analyzing non stationary signals, such as vibration, current or acoustic emission signals.

The concept of wavelet transform and fundamentals of different basis of discrete wavelet packet analysis are cited as they form the basis for features extraction in Chapter 6.

Given a signal s of length N, the Discrete Wavelet Transform (DWT) consists of at most $\log_2 N$ stages. The first step produces two sets of coefficients: approximation coefficients cA_1, and detail coefficients cD_1. These vectors are obtained by convolving s with the low-pass filter Lo_D for approximation, and with the high-pass filter Hi_D for detail, followed by dyadic decimation [77].

More precisely, the first step is shown in Figure 2.8:

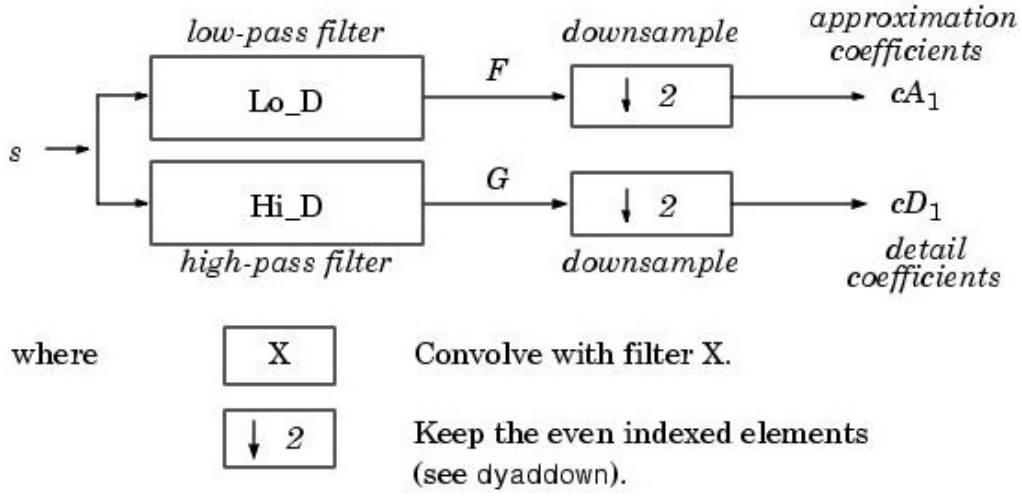


Figure 2.8: First step of DWT algorithms [77]

The length of each filter is equal to $2N$. If $n = length(s)$, the signals F and G are of length $n + 2N - 1$, and then the coefficients $cA1$ and $cD1$ are of length:

$$floor\left(\frac{n-1}{2}\right) + N \qquad (2.12)[77]$$

The $floor()$ function gets the greatest integer of $\frac{n-1}{2}$. The next step splits the approximation coefficients $cA1$ in two parts using the same scheme, replacing by cA_2 and cD_2, and so on. This is shown in Figure 2.9.

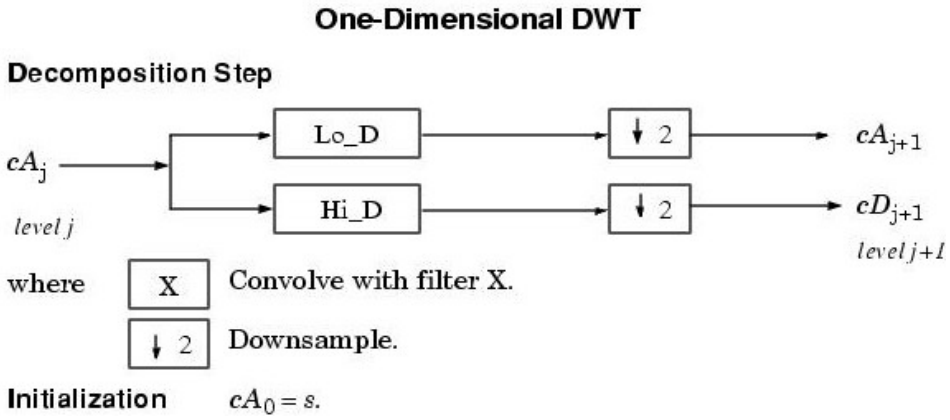


Figure 2.9: One-dimensional DWT [77]

The wavelet decomposition of the signal s analyzed at level j has the following structure: [cA_j, cD_j, ..., cD_1]. An example of this structure for $j = 3$ terminal nodes of the following tree is shown in Figure 2.10.

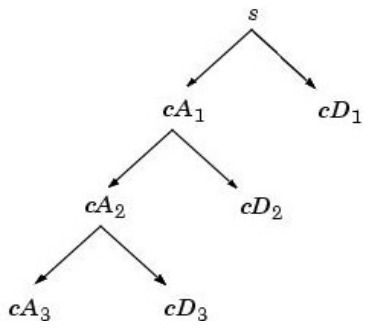


Figure 2.10: One-dimensional wavelet decomposition tree [77]

For many signals, the low-frequency content is the most important part. It is what gives the signal its identity. The high-frequency content, on the other hand, imparts flavour or nuance [77].

In wavelet analysis, it is often spoken of approximations and details. The approximations are the high-scale, low-frequency components of the signal. The details are the low-scale, high-frequency components.

The decomposition process can be iterated, with successive approximations being decomposed in turn, so that one signal is broken down into many lower resolution components.

Next, the fundamental concept of discrete wavelet packet analysis is described as it will be used in Chapter 6. A wavelet packet is composed of elementary functions called wavelet packets $2^{j/2}\omega_n(2^j x-k): j,k,n\in Z\}$, where *j, k* and *n* represent the index of scale, position and degree of oscillation, respectively. Wavelet packets can be represented by a filter bank constructed from quandrature mirror filters.

In wavelet analysis, a signal is split into an approximation and a detail. The approximation is then itself split into a second-level approximation and detail, and the process is repeated. For an *n*-level decomposition, there are $n+1$ possible ways to decompose or encode the signal [77].

In wavelet packet analysis, the details as well as the approximations can be split. This yields more than $2^{2^{n-1}}$ different ways to encode the signal. Figure 2.11 is the wavelet packet decomposition tree.

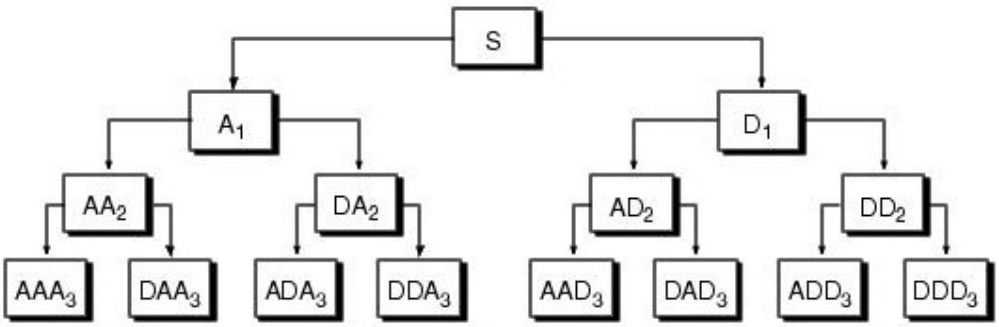


Figure 2.11: Wavelet packet decomposition tree [77]

Wavelet packets are generated by the following iterations [77]:

$$\omega_{2n}(2^{j-1}t-k)=\sqrt{2}\sum_m h_{m-2k}\omega_n(2^j t-m) \tag{2.13}$$

$$\omega_{2n+1}(2^{j-1}t-k)=\sqrt{2}\sum_m g_{m-2k}\omega_n(2^j t-m) \tag{2.14}$$

where *h* and *g* represent the respective high- and low-pass quandrature mirror

filters, and ω_0 and ω_1 correspond to the father wavelet (scaling function) and mother wavelet (analysing function).

For a given signal the wavelet packet coefficients can be iteratively computed by the following equations [77].

$$c_{2n,1}^{j-1} = \sum_k h_{k-2i} c_{n,k}^j \tag{2.15}$$

$$c_{2n+1,1}^{j-1} = \sum_k g_{k-2i} c_{n,k}^j \tag{2.16}$$

Next, different wavelet functions that will be used in this thesis are cited below:

A. Harr wavelets

Any discussion of wavelets begins with Harr wavelet [80], the first and simplest. Harr wavelet is discontinuous (5.6) (as shown in Figure 2.12), and resembles a step function.

$$\psi_H = \begin{cases} 1 & 0 \le x \le 1/2 \\ -1 & 1/2 \le x \le 1 \\ 0 & otherwise \end{cases} \tag{2.17}$$

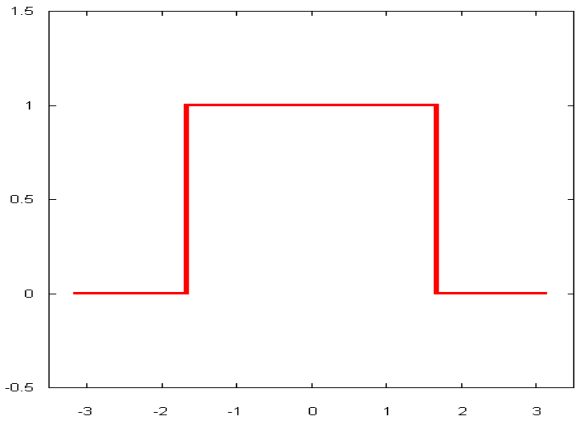


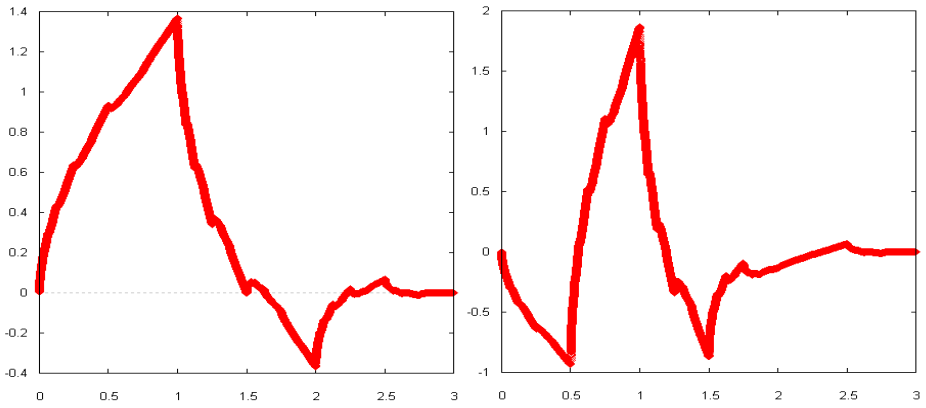
Figure 2.12: Harr wavelet [78]

B. *Daubechies wavelets*

Daubechies wavelets are a family of orthogonal wavelets defining a discrete wavelet transform and characterized by a maximal number of vanishing moments for some given support. With each wavelet type of this class, there is a scaling

function (also called father wavelet) which generates an orthogonal multi-resolution analysis. Daubechies orthogonal wavelets $db2-db20$ are commonly used. The index number refers to the number N of coefficients. Each wavelet has a number of zero moments or vanishing moments equal to half the number of coefficients. For example, $db1$ (the Haar wavelet) has one vanishing moment, $db2$ has two, etc. A vanishing moment limits the wavelet's ability to represent polynomial behaviour or information in a signal. For example, $db1$ with one moment, easily encodes polynomials of one coefficient, or constant signal components. $db2$ encodes polynomials with two coefficients, i.e. constant and linear signal components; and $db3$ encodes 3-polynomials, i.e. constant, linear and quadratic signal components. Examples of Daubechhies $db1$, Daubechhies $db2$ and Daubechhies $db3$ are shown in Figure 5.8.

Daubechies wavelets are very asymmetric because they are constructed by selecting the minimum phase square root of $Q(e^{-iw})$. Filters corresponding to a minimum phase square root have their energy optimally concentrated near the starting point of their support. They are highly non-symmetric, which yields very asymmetric wavelets.

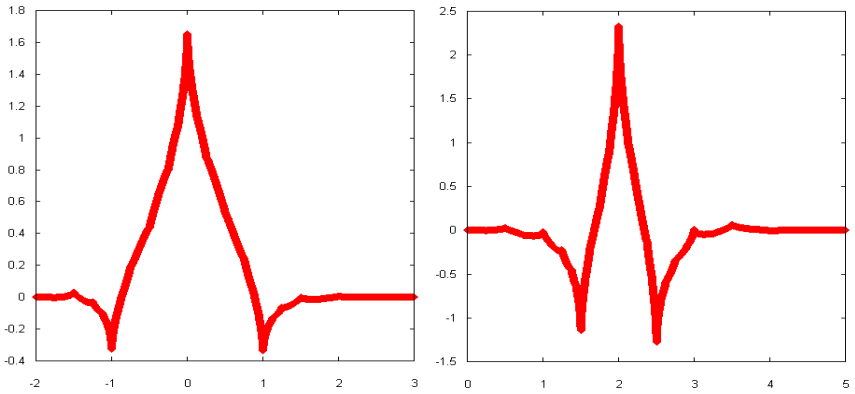


db2 (a) Father function (b) Wavelet function

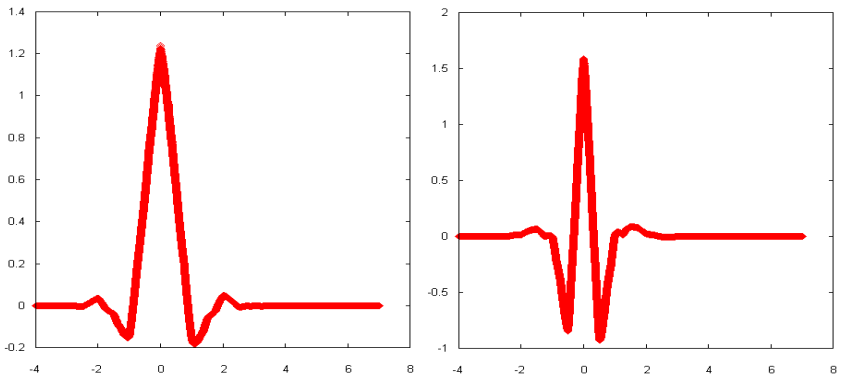
Figure 2.13: Daubechhies wavelet [78]

C. *Coiflet (CoifN)*

Coiflet wavelet is near symmetric; their wavelet functions have $N/3$ vanishing moments and scaling functions $N/3-1$, and have been used in many applications. Examples of *Coiflet K1 (coif1)* and *Coiflet K2 (coif2)* are shown in Figure 2.14.



coif1 (a) Father function (b) Wavelet function

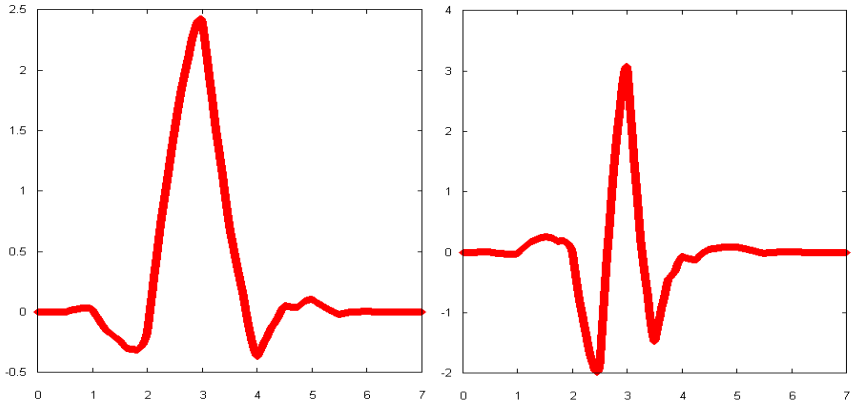


coif2 (a) Father function (b) Wavelet function

Figure 2.14: Coiflet wavelet [78]

D. *Symmlets*

To obtain a symmetric or antisymmetric wavelet, the filter h must be symmetric or antisymmetric with respect to the centre of its support. The Symmelet filters (as shown in Figure 2.15) of Daubechies are obtained by optimizing the choice of the square root $R(e^{-iw})$ of $Q(e^{iw})$ to obtain an almost linear phase.



(a) Father function (b) Wavelet function

Figure 2.15: Symmlet wavelet [78]

There is tradeoff between symmetry and computation complexity of a wavelet function. Among the orthogonal wavelet, Harr wavelet has computational advantages due to its symmetry property. However, Harr wavelet can hardly be used in the analysis because it has straight waveforms, which has nothing in common with current signal waveforms. Symmlet family and coiflet family have most properties such as compactly supposed, near symmetry, orthogonal analysis. Daubechhies family, symmlet family and coiflet family will be used in this study.

Wavelet transform has been successfully applied to data analysis in fault diagnostics of gears [79, 80], bearings [81, 82] and other mechanical systems [83, 84]. Baydar and Ball [85] applied the wavelet transform to both acoustic signals and vibration. The wavelet transform combined with Fourier transform to enhance feature extraction capability was proposed in [86]. A more advanced transform, known as wavelet packet transform, was studied in [87 – 88].

With the development for about 20 years, the wavelet transform has been widely used in fault diagnostics. But, compared with the FFT, the applications of the wavelets have still not achieved a standard status [89]. Many reasons have caused the current status. For example, many functions can be used as the wavelet basis, but there is no a standard or a general method to select the wavelet function for different tasks. It is an obstacle for the popularization of the wavelet transform. Additionally, an ignored problem is how to determine the range scales used in the wavelet transform. The solution to this problem is important. Wavelet transforms with scales out of this range would bring some meaningless information, which

will mislead the signal analysis. Unlike the FFT, the results of the wavelet transform have no straightforward physical implications, and therefore it is difficult to obtain useful information directly from the results of the wavelet transform. Furthermore, different wavelet functions may result in different analysis results, which will make the analysis results become confused. In Chapter 6, the author will have more detail discussion about this problem and demonstrate our achievements in this research area.

In conclusion, statistical analysis has been applied to analyze time domain signals. The FFT and high order spectrum analyses of signals have been applied in the frequency domain. Wavelet transforms can show features of signals in the frequency as well as the time domains. Generally, frequency domain analysis is more consistent in the fault diagnosis than time domain analysis. However, sufficient evidence shows that it would be unreliable to depend exclusively on any one technique to detect bearing damage [90]. In many cases it can be difficult to distinguish harmonic peaks within a spectrum, dependent upon the degree of noise present in the raw data, and the number of other vibration sources in the location where the data was sampled. In this research, all these three methods are used as important signal processing techniques for extracting useful features related to machine health conditions.

2.2.3 Feature selection

2.2.3.1 Feature subset selection

Feature subset selection (FSS) is a method for enhancing the performance of learning algorithms, reducing the hypothesis search space and storage requirement, through pre-processing data [91]. It looks at the issue of dimensionality reduction from a different perspective by selecting an optimum subset of features that are most responsible for a given outcome, and has a unique set of methodologies.

The objectives of FSS are to improve the prediction performance with faster and more cost-effective predictions, and to provide a better understanding of the underlying process that generates the data [92]. FSS is able to choose a subset of the original features to be used for subsequent process, thus reducing the number of features used in classification without compromising on accuracy. The reduction

of features is crucial in decreasing the training time, computational (and sensor) costs, and improving the prediction accuracy. In addition to improving the prediction accuracy, feature subset selection also helps to provide a better understanding of the underlying concept that generates the data.

In the feature subset selection problem, a learning algorithm is faced with the problem of selecting a relevant subset of features, and ignoring the rest of the features, so that an induction algorithm that runs on data containing only these features generates a classifier with the highest possible accuracy [93]. Thus in order to achieve the best possible performance with a particular learning algorithm on a particular training set, a feature subset selection method should consider how the algorithm and the training set interact.

As described in their paper, Blum and Langley [93] argued that feature selection algorithms, as shown in Figure 2.16, consist of the following four main components:

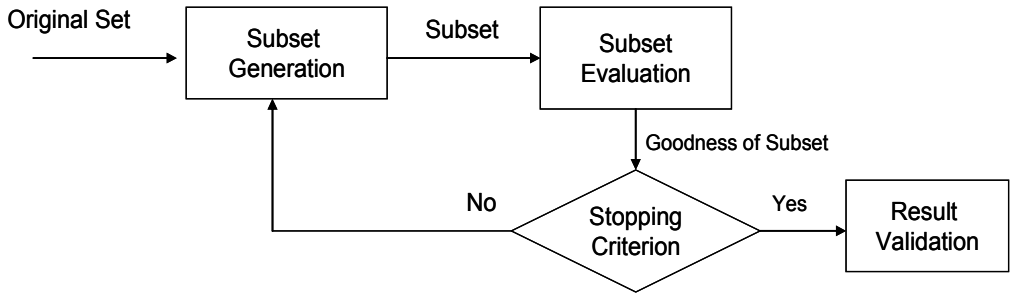


Figure 2.16: General procedures of feature subset selection

i. The first component is the initialization in feature space, which could include no features, all features, or random subset of features. In the first case, the search algorithm adds features successively, while in the second case, features are successively removed. In the last case of starting with a random subset, features could be successively added / removed, or reproduced by a certain procedure.

ii. The second component is the search procedure of the feature selection algorithm. Ideally, the best subset could be found by searching exhaustively and evaluating all the possible subsets. However, this becomes prohibitive as the number of features increases, where there are 2^d possible feature subsets for d features. Although several search procedures that have been developed, they do

not guarantee to find the optimal subset of features. In addition, search procedures differ in their computational cost and the optimality of the subsets they find.

iii. The third component is the evaluation function, which is an important component of any feature selection method. Evaluation functions measure how good a specific subset can be in discriminating between classes, and the aim is to reduce the dimensionality of the data by removing the irrelevant and redundant information to allow the machine learning program to operate more effectively. Evaluation functions can be divided into two main groups, namely filters (e.g. Principle Component Analysis (PCA), Principle Feature Analysis (PFA)) and wrapper (e.g. genetic algorithm) approaches, as shown in Figure 2.17. Filter approach operates independently of any learning algorithm, where undesirable features are filtered out of the data before learning begins. In contrast, wrapper approach operates dependently of the learning algorithms, which are used to select features.

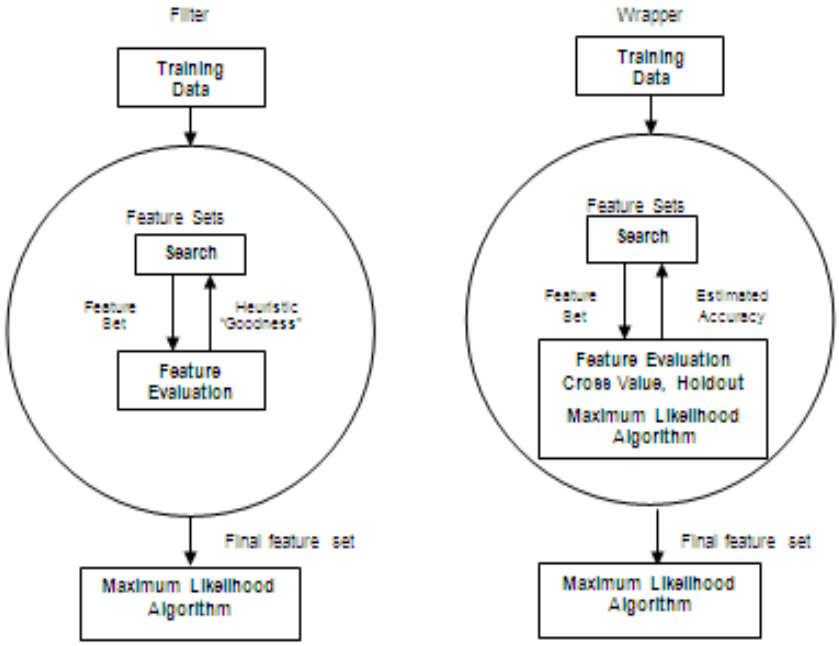


Figure 2.17: Filter and Wrapper approaches

iv. The last component is the criterion for stopping the search of feature selection algorithm. Feature selection methods need to decide when to stop searching through the space of feature subsets. Some methods require user to predefine the number of selected features, while others are based on the evaluation function.

2.2.3.2 Filter approach

The filter approach [94], as shown in Figure 2.18, evaluates subsets by their information content, and this is done independently of the induction algorithm (used to run on dataset, usually partitioned into internal training and holdout sets, with different sets of features removed from the data). The filter approach pre-processes data and selects feature subsets based on general criteria (properties that good features are presumed to have).

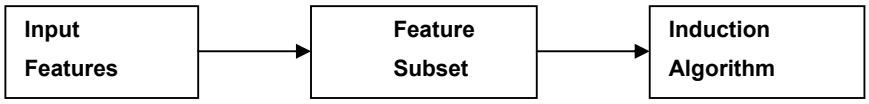


Figure 2.18: Filter approach

This approach can be relatively fast and computationally efficient. However, it may produce disappointing redundancy results in the selected features as the features are considered independently and not measured on the basis of whether they contribute new information. Lastly, interactions among features generally cannot be explicitly incorporated.

Therefore to solve the problem of redundancy in the selected features, rather than retaining a subset of e.g. "*s*" features, dimension reduction can be performed by projecting features onto "*s*" principal components of variation. Thus, the resultant will be a linear combination of features, rather than separate features. Examples of filter approaches [13] that performed dimensional reduction to solve the problem of redundancy in the selected features are PCA, PFA, etc.

PCA performs a linear mapping of the data to a lower dimensional space in such a way that the variance of the data in the low-dimensional representation is maximized. The original space has been reduced to the space spanned by a few

eigenvectors. PFA selects a subset from the set of the original features, is computationally efficient, and preserves the optimality properties as PCA does. The difference is that instead of performing linear combination of features, PFA clusters the weight vectors by similarity and picks a representative of the vector from each cluster. Further details of PCA and PFA will be discussed in Chapters 3 and 4 respectively.

2.2.3.3 Wrapper approach

The wrapper approach of feature subset selection [95], as shown is Figure 2.19, uses the method of classification itself to measure the importance of a feature or feature set. This approach tailors feature subsets to suit the inductive bias of a given learning algorithm. This is done by treating the feature selection as a search problem in the space of all possible feature subsets, and performing an exhaustive search through all possible feature subsets to select the optimal features.

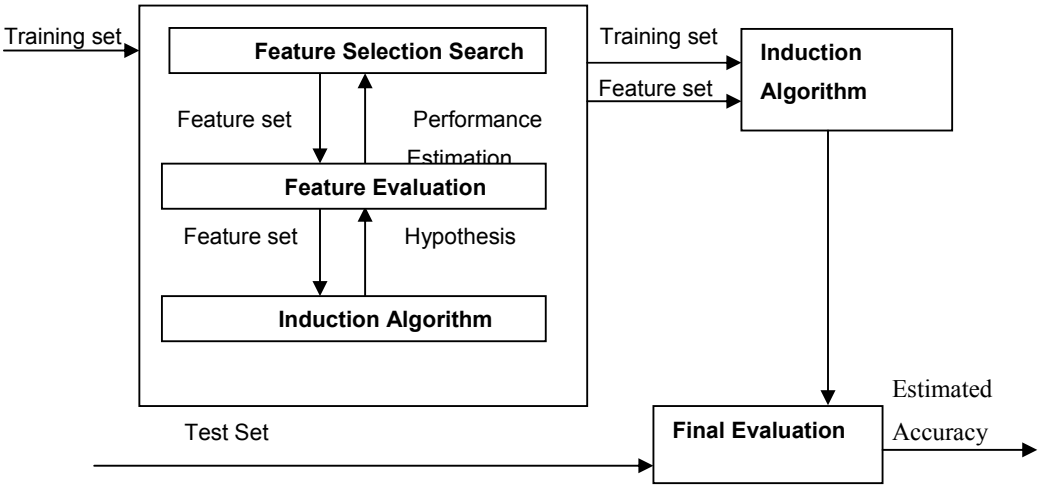


Figure 2.19: Wrapper approach

The feature subset selection algorithm exists as a wrapper around the induction algorithm. The feature subset selection algorithm conducts a search for a good feature subset using the induction algorithm as part of the evaluation of the feature subsets. An induction algorithm is typically presented with a set of training instances, where each instance is described by a vector of feature (or attribute) values and a class label. The task of the induction algorithm is to induce a classifier that will be useful in classifying future cases. The classifier is a mapping from the

space of feature values to the set of class values.

The wrapper approach makes use of the heuristic search algorithm to evaluate each candidate feature subset. The feature subset with the highest evaluation is chosen as the final set on which to run the induction algorithm. The resulting classifier is then evaluated on an independent test set that was not used during the search.

In contrast with the filter approach, the wrapper approach may discover which feature deteriorates performance and avoid to be selected. Furthermore, it also selects feature subsets that achieve the lowest possible prediction error. However, a major disadvantage of the wrapper approach is that it requires much more computational effort. These disadvantages make the wrapper approach computational more expensive than the filter approach.

Examples of search strategies that use wrapper approach to search for the optimal attribute subset are sequential forward selection (SFS) [96], sequential backward elimination (SBE) [95], Genetic Algorithm (GA) [99], etc.

SFS [96] is an algorithm that starts with an empty set of attributes, and greedily adds attributes that has the smallest *P* value (most significant, typically less than 0.05 or 0.10) one at a time. SFS is computationally intensive since many possible enumerations of the feature combinations exist. SBE [97] starts with all the attributes and greedily removes the largest *P* value (least significant) until the remaining attributes have very small *P* value, typically about 0.05 or 0.10.

The Genetic algorithm [98] is an optimization technique that mimics the evolutionary process of "survival of the fittest". It starts off with an initial random population of solutions, and evolves new populations by mating (crossover) pairs of solutions and mutating solutions according to their fitness (objective function). The better the fitness, the more likely it is to be selected for the mating and mutation operations and therefore carry their "genetic code" from generation to generation. Although these approaches improve performance, they still require *m* training passes per wrapper cycle.

Research using genetic algorithms for feature selection in fault diagnosis has been active. Li et al. [99] presents a novel two stages feature selection method for gear fault diagnosis based on ReliefF feature filtering method and genetic

algorithm. In the first stage, the ReliefF was employed to evaluate the quality of every individual feature and a sequential feature sets were obtained according to the marks evaluated by ReliefF. Then the cross validation technique was used to get the candidate feature set from the sequential sets. Dong [100], proposes an integration of Genetic Algorithm (GA) and Support Vector Machines (SVM) for efficient optimization of both features and parameters for detection models. Nguyen [101] [102] presents feature selection by decision tree and genetic algorithm (GA) to remove irrelevant information in the feature set. New data with the selected features is used to train an expert system for classification. Testing results show that systems with selected features can reliably diagnose different conditions of induction motor, which has better performance compared to original one without feature selection. Zhou [103] presents a method to determine optimum feature subset selection with a modified wrapper-based multi-criteria approach using genetic algorithms. The method used to predict machinery tool wear demonstrates the viability of the feature subset selection method for diagnosis applications.

Other research has been carried out for feature selection in fault diagnosis. Ant colony optimization (ACO) is one of the approaches for feature selection. Wang et al. [104] combine Supporting Vector Machines (SVMs) with a novel quantum ant colony optimization (QACO) algorithm to select fault features for improving the performance of fault diagnosis in a complex chemical process. Liao [105] studied acoustic emission signals collected during grinding operations, processed by autoregressive modeling or discrete wavelet decomposition for feature extraction, and then using three different feature selection methods, including two proposed ant colony optimization (ACO)-based methods and the famous sequential forward floating selection method to select the best feature subsets. The methods can find the essential fault variables exactly and improve the fault diagnosis performance of SVMs. Zhou [106] determines optimum feature subset selection with ant colony optimization and mutual information hybrid algorithms. The best compound features found by ant colony algorithms are verified by multiple regression models and are used to construct fault prediction models.

PCA has also been used for feature subset selection in fault diagnosis [107]

[108] [109]. Chen [110] used PCA to select features and eliminate the redundancy features in process of rolling bearing fault monitoring. A new algorithm of P-PCA (Parts Principal Component Analysis) was proposed. PPCA reserves more useful information but has a bigger dimension of final feature space than PCA. Principal Feature Analysis [11] involves k-means formulation on significant principal components and is used for reduction of feature sets in tool wear prediction.

Other methods of Self Organizing Maps (SOM) and decision tree are also used in the research. Liao et al. [111] presents a study that uses self-organizing maps to realize feature selection and reduce dimensionality of the raw feature space for machine faults classification. By means of evaluating the responses of every dimensional feature in SOM networks neurons weights to the input data, the feature sets having the main responses and being sensitive to pattern recognition are selected. Industrial gearbox vibration signals measured under different operating conditions are analyzed using the method. Zheng et al. [112] present a case study of feature selection for high-dimensional data in astronomy comparing the performance and computational cost of different feature selection methods. Applied on the sample, the result indicates that from the viewpoints of computational cost, the filter method is superior to the wrapper method. Sugumaran [113] illustrates the use of a decision tree of J48 algorithm to examine the normalized information gain (difference in entropy) for feature selection in fault diagnosis of bearings. The set of features identified through the algorithm are differing from the set of features reported in the literature in some respects. The limitation of algorithm is that it needs a large number of data points in the data set. As the number of data points becomes small, the results of applying the algorithm become less meaningful as it works based on the statistical measure. A comparison summary of different feature selection methods is given in Table 2.3.

After the dominant features have been selected, the features will be used for machine condition assessment using intelligent modeling methods and this will be discussed in Section 2.2.4.

Table 2.3 List of feature selection methods and their merits and limitations

Approach	Merits	Limitations
Sequential forward selection (SFS) [99]	• Simple • Interacts with the classifier • Models feature dependencies	• Classifier dependent selection • Risk of over fitting • More computationally intensive than SBS, more computationally intensive than GA. AC
Sequential backward selection (SBS) [100]	• Simple • Interacts with the classifier • Models feature dependencies • Less computationally intensive	• Classifier dependent selection • Risk of over fitting
GA [101] [102][104 [105] [106]	• Less prone to local optimation • Models feature dependencies • Less prone to local optima ion • Models feature dependencies	• Computationally intensive • Higher risk of over fitting • Performance depends on initial parameters setting
Ant Colony Optimization (ACO) [107] [108]		
PCA [110] [111] [112] [113]	• Better computational complexity than GA and ACO	• Perform linear combination of features for feature transformation instead of feature selection.
PFA [11]	• Perform feature selection by clustering and pick a representative of the vector from each cluster. • Better computational complexity than GA and ACO	• Computationally intensive calculations of eigenvectors and their corresponding eigen values when data sets are large
Weight vector of SVM [115]	• Better computational complexity than GA and ACO • Models feature dependencies	• Classifier dependent selection
Decision trees [116] [117]		• Needs a large number of data points in the data set

2.2.4 Intelligent modeling with AI techniques

In the literature, popular AI techniques for machine diagnosis are artificial neural networks (ANNs), fuzzy logic systems, fuzzy-neural networks (FNNs), neural-fuzzy systems and evolutionary algorithms (EAs). Related reviews on AI techniques in CBM have been reported in the literature. Fault diagnosis of electronic systems using intelligent techniques was reviewed by William G. [114] in 2001. Applications of AI techniques for induction machine stator fault

diagnostics were given by Siddique et al. [115] in 2003. A review of Intelligent Diagnostic Methods for power transformer was given by [116] in 2008.

Artificial neural network (ANN) is currently the most commonly found data-driven technique in the prognostics literature. An ANN consists of a layer of input nodes, one or more layers of hidden nodes, one layer of output nodes and connecting weights. The network learns the unknown function by adjusting its weights with repetitive observations of inputs and outputs.

Numerous studies across various disciplines have demonstrated the merits of ANNs, including the abilities to

(a) Perform faster than system identification techniques in multivariate prognosis [117];

(b) Capture complex phenomenon without a priori knowledge.

The most simple ANN-based machinery prognostics approach was time series prediction models. Tse and Atherton [118] and Yam et al. [119] used recurrent neural networks (RNNs) to trend condition indices and forecast successive condition index value at the next time step. Wang and Vachtsevanos [120] developed recurrent wavelet neural network (RWNN) to predict rolling element bearing crack propagation. Wang et al. [121] used a Neuro-Fuzzy (NF) network to predict spur gear condition value. The fuzzy interference structure is determined by experts. The fuzzy membership functions are trained by the neural network. The NF system performed much better than RNN when there was sufficient training data. However, it could not predict well when the train-set was small or when there were fast dynamic fluctuations, such as during the chipping of gear tooth surface material or just prior to gear failure. An adaptive training technique was later proposed by Wang [122] to improve the NF model.

Another technique is the extraction of knowledge from existing databases or casebases. Using existing information to automatically generate a knowledge-based system can greatly speed development time and greatly reduces the knowledge acquisition bottleneck. However, it is only suitable where large databases of domain information are available. Therefore, it is inappropriate for new systems where actual data is not yet available. In [123], knowledge base generation from General Motor's diagnostic database is described. This database

contains 300,000 cases of vehicle symptoms and repair information. An extended form of the decision tree induction algorithm ID3 [124] is used to extract general diagnostic rules from the database. ID3 uses database examples to generate decision trees, which are then used to classify the examples into suitable diagnostic rules. The extensions deal with the presence of inconclusive data sets, that is, when the set of examples used is not enough to specify a single conclusive outcome.

Fuzzy Logic is another approach used in fault diagnosis. Fuzzy logic provides mechanisms to represent and manipulate linguistic concepts. It deals with approximates rather than exact measurements and is based on fuzzy set theory [125]. In traditional sets, membership is either true '1' or false '0', and there is no concept of partial membership. In fuzzy sets, partial membership is allowed and membership is represented by a value between 0 (definitely not a member) and 1 (definitely a member). In fuzzy set theory, a series of operators is defined for manipulating sets. Many of these are analogous to those used in conventional sets, such as, union (OR), interSection (AND), and complement (NOT). Fuzzy reasoning consists of manipulating a series of unconditional and conditional fuzzy propositions or rules using fuzzy rules of inference. With its concept of partial set membership, fuzzy logic provides a good alternative for reasoning with uncertain and inaccurate data. Most of the research work relating to fuzzy logic and diagnosis has occurred in the area of dynamic industrial processes. In this domain, fuzzy logic has been applied primarily to the following tasks [126 -127].

a) Fault Detection: Industrial processes are characterized by dynamic continuous variables (symptoms). Such variables are prone to measurement errors, noise, and operating conditions. Therefore, reliable measurement thresholds are difficult to define. Fuzzy logic provides a good solution to this problem, by representing signal values using overlapping linguistic variables.

b) Fault Diagnosis: Fault diagnosis in dynamic processes is always approximate, as measured signal values are only known to a certain degree of accuracy. A fuzzy inference system based on fuzzy IF–THEN rules can provide a solution to this problem, and is proposed and reported by many researchers.

Because of using of linguistic variables, fuzzy logic provides a very human-like and intuitive way of representing and reasoning with incomplete and inaccurate information. It is typically combined with other approaches such as

rules, models or cases, and provides a good alternative for reasoning under uncertainty.

Besides supervised learning, Salles and Filippetti [129] suggest an unsupervised self-organization mapping (SOM) neural network to distinguish among the signatures introduced by different load faults and those introduced by other machine troubles such as rotor faults. This technique utilizes a selected set of current spectral components as inputs to the system to perform fault classification. The data sets constructed from computer simulation results are used to train the neural network offline. Therefore, this method may not be suitable for online condition monitoring due to the relatively high computational cost.

2.3 Summary and proposed research directions

In an era of intensive competition, where asset usage and plant operating efficiency must be maximized, unexpected downtime due to machinery failure has become more costly than before. Therefore condition based maintenance has been actively pursued in the manufacturing industry in recent years, where equipment outages are predicted and maintenance is carried out only when necessary. To ensure successful condition based maintenance, it is necessary to detect, identify and classify different kinds of failure modes in the manufacturing process. Sensing technology is necessary in implementing the detecting system. The amount of sensors could be large and determining failure could mean taking minutes or hours to analyze data from all the sensors and make a reasonably accurate decision. Advanced feature extraction and selection are essential for analyzing signals from sensors, extracting useful information and providing better prediction accuracy and less learning time. Many researches have been developed to address this challenge. However, existing methodologies for advanced feature extraction and selection usually require computationally intensive calculations and lack of standard to select the wavelet function which causes the wrong decision about equipment health. Therefore it is highly desirable to develop a new feature identification method to improve the computational efficiency and provide an autonomous decision for data and sensor reduction. It is envisaged that such a feature identification method will play an important role in improving the efficiency and accuracy of a CBM system, by addressing the following issues:

1. Dominant feature selection (DFI)

First, our research focuses on the feature subset selection method to determine which features are important in equipment health monitoring. As discussed in Section 2.2.3, the current drawbacks of feature selection methods include:

- Current method requires computationally intensive calculations and will result in the delay of the fault diagnosis response time.
- The number of features that needs to retain is artificially decided. It is lacking of a systematic method to decide the retained numbers of features.

As such, a new feature selection method is desired for achieving numerically efficient performance, reducing the complexity of feature selection while reducing the least square error during dominant feature selection. In this thesis, the author will propose a new dominant feature selection method using Singular Value Decomposition (SVD) to decompose the inner product matrix of collected data from the monitoring sensors. The principal components are optimized in a least squares sense in a certain reduced space, and the dominant features are extracted using the k-means clustering algorithm.

2. Framework for integrating the feature extraction and intelligent modelling methods

In order to realize an efficient CBM system, it is desirable to develop a framework for integrating the feature extraction methods with DFI method and intelligent modelling methods to provide efficient and accurate machine health condition prediction.

In this framework, time domain and frequency domain analysis feature extraction methods are used to extract useful features from the sensor signals. DFI is then used to select the most dominant features that are related to machine health condition. The resulting dominant features are used with Recursive Least Squares (RLS) or Extended Least Squares (ELS) modelling methods or neutral networks to provide efficient and accurate machine health diagnosis and prognosis.

3. Advanced feature extraction method using wavelet decomposition

Next, our research focuses on the essential step of signal processing and feature extraction. The current obstacles for using wavelet decomposition for feature extraction in CBM include:

- Many functions can be used as the wavelet basis, but there is no a standard or a general method to select the wavelet function for different tasks. It is an obstacle for the popularization of the wavelet transform.
- The results of the wavelet transform have no straightforward physical implications, and therefore it is difficult to obtain useful information directly from the results of the wavelet transform.

The author proposes a new approach to use DFI for dominant wavelets selection. The new approach integrates DFI, discrete wavelet transformation and correlation modelling together to achieve robust machine performance degradation prediction and failure prevention.

4. Sensors and features reduction for industrial fault detection and isolation (FDI)

Industrial fault detection and isolation (FDI) is important for CBM. It is desirable to reduce both hardware and computational time from signal processing without compromising the FDI accuracy. Augmented Dominant Feature Identification (ADFI) and Decentralized Dominant Feature Identification (DDFI) are proposed for both sensors and features reduction for FDI. The earlier proposed Dominant Feature Identification (DFI) method is extended for FDI using Neural Networks (NNs) to classify faults based on a reduced set of features. It is shown how to apply DFI to fault detection by two methods that seek to identify the important features in a given set of faults. Then based on the determined reduced feature set, an NN is used for online fault classification and multiple fault prediction in such a proposed two-stage framework. The proposed ADFI and DDFI are tested on a fault simulator machine to identify imbalance, misalignment and different types of bearing faults.

5. *i*Diagnosis and prognosis

Current researches have been carried out on the diagnosis and prognosis of individual equipment health condition with little concern about the complex manufacturing environment which consists of many different types of equipments.

These complex manufacturing systems usually face many different types of disruptions which arise from shop floor dynamics such as material shortages, machine breakdown, machine performance degradation, shortage of operators, etc. This entails time-driven, event-driven and demand-driven monitoring of the operation of complex manufacturing systems, as well as their working conditions, tendency of performance degradations, minor faults, major failures, slow response, and so on. Detection and identification of various faults in complex manufacturing systems require many tools and different methods, and some of the methods will be discussed in Chapters 3 to 7. To achieve desirable results, a framework is required to enable the developed algorithms and their integration in an efficient manner. As such, it is desirable to develop a system framework to provide distributed intelligence of diagnosis and prognosis, system condition and job execution capability assessment, and dynamic information feedback to the supervisory control when manufacturing system breakdown or performance degradation occurs. The system framework, architecture, functions, and information flow are needed for implementing the diagnosis and prognosis in a complex manufacturing environment.

In summary, major findings from the literature review have identified the necessity of developing a new advanced feature extraction and selection framework to increase the efficiency and accuracy of the CBM system. It is postulated that the realization of such an advanced feature extraction and selection framework focusing on the four requirements listed above could help improving the CBM system with more accurate decision and enhance the competitive advantages of a manufacturing company.

CHAPTER 3 DOMINANT FEATURE SELECTION

As per discussion in Chapter two, feature selection is an essential step in CBM. Feature subset selection (FSS) is a method for enhancing the performance of learning algorithms, reducing the hypothesis search space, and storage requirement, through pre-processing data [13]. It looks at the issue of dimensionality reduction from a different perspective by selecting an optimum subset of features that are most responsible for the outcome.

The objectives of FSS are to reduce the features redundancy, provide faster and more cost-effective performance [95]. Data compression schemes include Principal Component Analysis (PCA) [9-10] and Principal Feature Analysis (PFA) [11-13], which had been applied to pattern recognition to identify key features of original data.

We use a real ($R^{m\times n}$) matrix X to represent the data collected from the sensors. m is the number of the measuring data points from sensors while n is the number of features and in real applications $m >> n$. We represent the data into two spaces: feature space and range space. In Figure 3.1, R^n can be identified as the feature space and R^m as the range space of X which is termed as the data space.

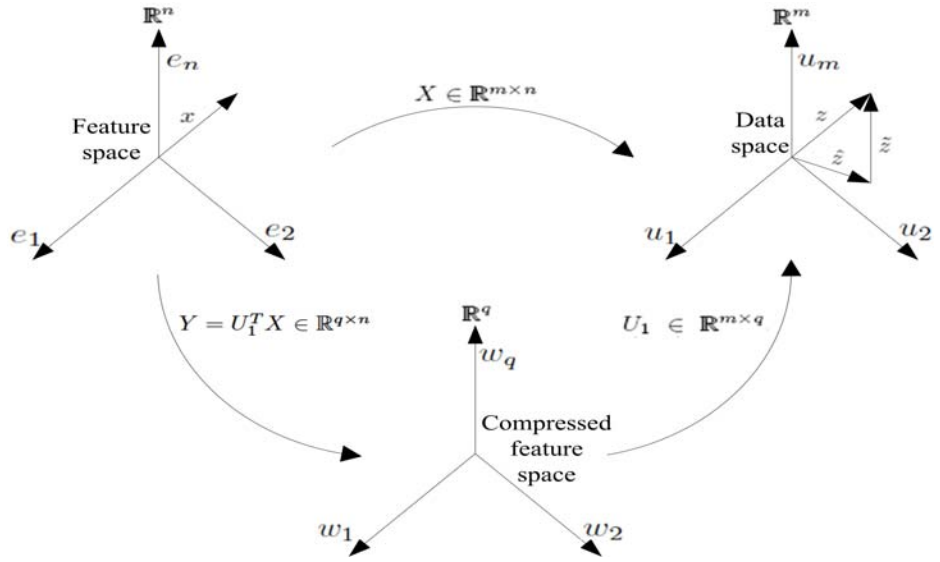


Figure 3.1 Three spaces (feature space R^n, compressed feature space R^q, and data (singular value)) space R^m involved in the computation

Note that traditional PCA performed approximation with respect to data space R^m, but the features however reside in R^n. PCA performs approximation to outer product $XX^T = U\Sigma^2 U^T \in R^{m\times m}$ and the results are given in terms of the q retained singular values. PFA performs the data compression and dimension reduction to isolate the principal components based on the same criteria as PCA. It proceeds to derive principal feature components to arrive at a subset of the original feature vector using the k-means clustering algorithm after singular value decomposition. Further details of PCA and PFA will be discussed in Sections 3.1 and 3.2 respectively. Both PCA and PFA require computationally intensive as they perform the data compression and dimension reduction with respect to data space R^m. This is computationally expensive in general, since $m >> n$ in most practical applications. We propose DFI to conduct data compression on the inner matrix $X^T X \in R^{n\times n}$ and obtain the compressed feature space R^q. The compressed features space keeps all the original features and it is the best approximation of the original data space. The dominant features are then selected in the compressed feature space R^q. The retained features could obtain the best approximation to minimize the total least-squares estimation error. The computation is highly simplified as the approximation is working on the inner matrix $X^T X \in R^{n\times n}$ with $n << m$. The concept of these three spaces can be found in Figure 3.1.

In this chapter, a rigorous mathematical framework for Dominant Feature Identification (DFI) is developed that provides an autonomous rule-base for data and sensor reduction. Singular Value Decomposition (SVD) is used to decompose the inner product matrix of the collected data. The principal components are optimized in a least squares sense in a certain reduced space, and the dominant features are extracted using the k-means clustering algorithm [130]. This DFI framework uses formal mathematical analysis to select dominant features. A numerically efficient scheme for implementation is presented that is based on the inner product matrix of the collected data, not the correlation (outer product) matrix. The proposed DFI decision tool is numerically efficient, and reduces the complexity of feature selection greatly while reducing the least square error during dominant feature selection and clustering automatically.

3.1 Principal Component Analysis (PCA)

In this Section, the theoretical fundamentals of PCA using SVD are reviewed in a rigorous manner, which is essential for selecting dominant features in Section 3.3.

PCA performs a linear mapping of the data to a lower dimensional space in such a way that the variance of the data in the low-dimensional representation is maximized. The original space has been reduced to the space spanned by a few eigenvectors. The concept of PCA [8] with singular value decomposition is introduced below:

3.1.1 Spectral decomposition of a square matrix

Any real symmetric R^m $m \times m$ matrix A has a spectral decomposition of the form,

$$A = U\Lambda U^T \tag{3.1}$$

where U is an orthonormal matrix (matrix of orthogonal unit vectors: $U^T U = I$ or $\sum_{k=1}^{m} U_{ki} U_{kj} = \delta_{ij}$) and Λ is a diagonal matrix.

The columns of U are the eigenvectors of matrix A and the diagonal elements of Λ are the eigenvalues.

Multiplying with U on both side of equation (3.1):

$$AU = U\Lambda U^T U = U\Lambda \tag{3.2}$$

This can be written as a normal eigenvalue equation by defining the i^{th} column of U as

u_i , and the eigenvalues as $\lambda_i = \Lambda_{ii}$:

$$A\mu_i = \lambda_i \mu_i \tag{3.3}$$

3.1.2 Singular value decomposition (SVD)

A real ($R^{m \times n}$) matrix X , where $m > n$ has the decomposition:

$$X = U\Sigma V^T \tag{3.4}$$

where U is a $R^{m \times n}$ matrix with orthonormal columns ($U^T U = \mathrm{I}_n$), where V is a R^n

orthonormal matrix ($V^T V = I_n$), with I_n being an identity matrix of dimension *n*.

$\Sigma \in R^{n \times n}$ is a diagonal matrix whose elements are corresponding singular values (principal gains) arranged in descending order, *i.e.,* with $\Sigma = diag(\sigma_1, \sigma_2, \cdots, \sigma_n)$, and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$.

3.1.3 Properties of a data matrix - first and second moments

Let x (with components x_j $(j = 1,....,n)$) be a stochastic vector with probability distribution $P(x)$. Let $\{x^\alpha \mid \alpha = 1,...,m\}$ be a sample from $P(x)$. A convention for the data matrix X will be chosen, where the rows denote the features $(j = 1,....,n)$ and the columns of the samples $\alpha = 1,...,m$: in other words the components are $X_{j,\alpha} = x_j{}^\alpha$.

Principal component analysis is based on the two first empirical moments of the sample data matrix. The mean vector,

$$\langle x \rangle \equiv \frac{1}{m} \sum_{\alpha=1}^{m} x^\alpha \tag{3.5}$$

and the empirical covariance matrix,

$$C \equiv \frac{1}{m} \sum_{\alpha=1}^{m} (x^\alpha - \langle x \rangle)(x^\alpha - \langle x \rangle)^T \tag{3.6}$$

Using the matrix formulation,

$$C \equiv \frac{1}{m} X X^T \tag{3.7}$$

where the mean of the data has been removed: $X_{j,\alpha} := X_{j,\alpha} - \langle x_j \rangle$.

3.1.4 Principal Component Analysis (PCA)

In principal component analysis, eigenvectors are corresponding to the largest eigenvalues of the covariance matrix, and project the data onto these directions. The motivation for doing this is that the most second order information is in these directions. If we denote the matrix of eigenvectors sorted according to

eigenvalue by $\tilde{U}$, then PCA transformation of the data is $Y = \tilde{U}^T X$. The eigenvectors are called the principle components. By selecting only the first d rows of Y, the data have been projected from *n* down to *d* dimensions.

3.1.5 PCA by SVD

SVD can be used to perform PCA. We decompose X using SVD, i.e.

$$X = U\Sigma V^T \tag{3.8}$$

and the covariance matrix can be written as

$$C = \frac{1}{n} XX^T = \frac{1}{n} U\Sigma^2 U^T \tag{3.9}$$

In this case U is a R^m matrix. SVD routine arranges the singular values in descending order. The transformed data can thus be written as

$$Y = \tilde{U}^T X = \tilde{U}^T U\Sigma^2 U^T \tag{3.10}$$

where $\tilde{U}^T U$ is a R^m matrix which is one on the diagonal and zero elsewhere. To conclude, the transformed data can be written in terms of the SVD decomposition of X.

3.1.6 Approximation of linear transformation X

X is regarded as a transformation from feature space R^n into data space R^m. Note that $X = U\Sigma V^T = \sum_{i-1}^{n} \sigma_i u_i v_i^T$, where u_i are the column vectors of U and v_i^T are the row vectors of V^T, respectively. Partition the SVD of X according to

$$X = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = U_1 \Sigma_1 V_1^T + U_2 \Sigma_2 V_2^T \tag{3.11}$$

with $q < m$ as the desired number of singular values to be retained in Σ_1 for data space of dimension *m*. As such, Σ_2 contains the $n - q$ discarded singular values. Obviously, $U_1 \in R^{m \times q}$, $U_2 \in R^{m \times (n-q)}$, $\Sigma_1 \in R^{q \times q}$, $\Sigma_2 \in R^{(n-q) \times (n-q)}$, $V_1^T \in R^{q \times n}$, and $V_2^T \in R^{(n-q) \times n}$.

Now the approximation $\hat{X}$ to X is

$$\hat{X} = U_1 \Sigma_1 V_1^T \tag{3.12}$$

Then $\hat{X} = \sum_{i=1}^{q} \sigma_i u_i v_i^T$ contains the columns u_i of U_1 and the rows v_i^T of V_1^T. The dominant singular values, *i.e.*, the q retained singular values, and their associated columns of U are called principal components in PCA.

The error induced $\tilde{X}$ by the approximation $\hat{X}$ of the linear transformation X is given by

$$\tilde{X} = X - \hat{X} = U_2 \Sigma_2 V_2^T = \sum_{i=q+1}^{n} \sigma_i u_i v_i^T . \tag{3.13}$$

The covariance matrix of the approximation error is

$$\begin{aligned} P_{\tilde{X}} &= (X - \hat{X})(X - \hat{X})^T \\ &= U_2 \Sigma_2 V_2^T V_2 \Sigma_2 U_2^T = U_2 \Sigma_2^2 U_2^T . \end{aligned} \tag{3.14}$$

The 2-norm of the approximation error is given by

$$\mathrm{tr}\{P_{\tilde{X}}\} = \mathrm{tr}\{U_2 \Sigma_2^2 U_2^T\} = \mathrm{tr}\{\Sigma_2^2 U_2^T U_2\} = \mathrm{tr}\{\Sigma_2^2\} = \sum_{i=q+1}^{n} \sigma_i^2 \tag{3.15}$$

3.1.7 Approximation in range space by principal components

Now, R^n can be identified as the feature space and R^m as the range space of X which is termed as the data space. In real applications $m >> n$ as m is the number of the measuring data points from sensors while n is the number of features.

Consider an arbitrary vector $x \in R^n$ being mapped onto a vector $z \in R^m$ by X according to $z=Xx$. As such, z in the singular value space (range space) of R^m can also be represented according to the partitioned singular value matrix in (3.11) as

$$z = Xx = U_1 \Sigma_1 V_1^T x + U_2 \Sigma_2 V_2^T x \tag{3.16}$$

with Σ_1 containing the retained q Singular Values of X.

An approximation to z is $\hat{z}$ given in terms of the q retained singular values as $\hat{z} = U_1 \Sigma_1 V_1^T x = \hat{X} x$ (3.17)

with $\hat{X}$ being the approximation of X. Note that $\hat{z} = \sum_{i=1}^{q} \mu_i (\sigma_i v_i{}^T x)$ which expresses $\hat{z}$ as a linear combination of principal components u_i with coefficients $(\sigma_i v_i{}^T x)$.

The approximation error is given by

$$\tilde{z} = z - \hat{z} = (X - \hat{X})x = U_2 \Sigma_2 V_2^T x \tag{3.18}$$

and the approximation error 2-norm is given by

$$\begin{aligned} \tilde{z}^T \tilde{z} &= x^T V_2 \Sigma_2^2 V_2^T x \\ \|\tilde{z}\|^2 &= \mathrm{tr}\{\Sigma_2^2 V_2^T x x^T V_2\} \\ &\le \sigma_{q+1}^2 x^T V_2 V_2^T x = \sigma_{q+1}^2 \|x\|^2 \\ &\le \mathrm{tr}\{\Sigma_2^2\} \|x\|^2 . \end{aligned} \tag{3.19}$$

3.1.8 Number of principal components

The number of principal components d to use could be determined from the characteristics of the singular values. When the singular values stabilize, the remaining components are usually contaminated with much noise and therefore not useful. In Figure 3.2, an example of singular values is shown. From component number 3 and up, the singular values are almost constant, indicating that d should be 3.

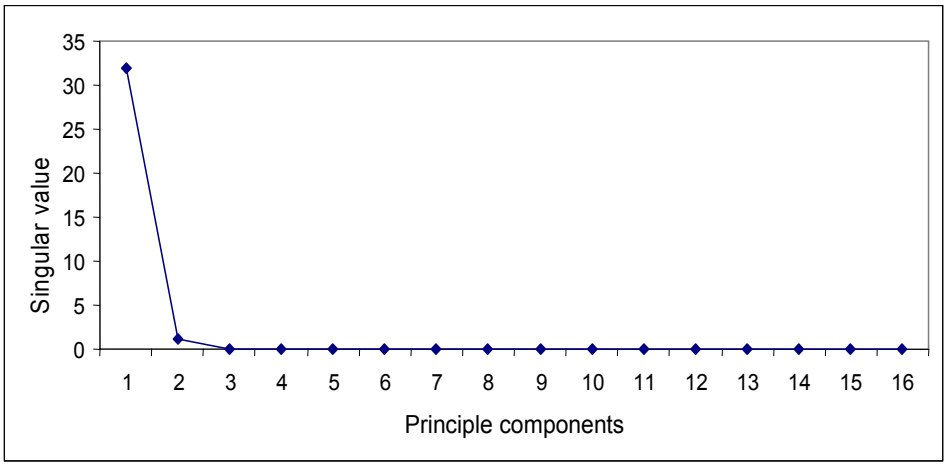


Figure 3.2: Plot of principal components vs. singular values

3.2 Principle Feature Analysis

Principal Feature Analysis, PFA [11-13], builds upon PCA to retain some of its optimality properties. Although algorithms such as PCA are useful in dimensional reduction through creation of relevant features, they have the main disadvantage that original features are transformed as projected new features into the lower dimension space. The projected new features are the combination of original features and the projected new features can not be directly used. The motivation for PFA is that using subset of the original features for analysis is sometimes more desired as compared using a linear combination of features that are transformed into the lower dimension using PCA.

For this reason, PFA first performs the dimension reduction and finds the principal components based on the same criteria as PCA. But instead of projecting the original features to the subspace, it exploits the structure of the obtained principal components to find a subset of the original feature vector through clustering of the principal components [9-10]. Therefore, the resulting subset of the features retains the measurement of original forms.

Clustering [11] involves dividing a set of data points into non-overlapping groups, or clusters of points. Points in a cluster are closer by some measure of proximity to one another than to points in other clusters. When a dataset is clustered, every point is assigned to a cluster, and every cluster can be characterized by a single reference point, usually an average of the points in the cluster.

The main purpose of clustering is to reduce the size and complexity of the dataset. Data reduction is accomplished by replacing the coordinates of each point in a cluster with the cluster's reference point. Clustered data requires considerably less storage space and can be manipulated more quickly than the original data.

In the case of PFA algorithm, it uses k-means algorithms to perform clustering of the principal components. k reference points are first chosen and all the data points are assigned to k clusters. The cluster centroids are used as reference points in subsequent partitioning. The centroids are adjusted both during and after each partitioning.

For data point x in cluster i, if the centroid z_i is the nearest reference point, no adjustments are made and the algorithm proceeds to the next data point. However, if the centroid z_j of the cluster j is the closest reference point to data point x, then x is reassigned to cluster j, the centroids of the "losing" cluster i (removal of point x) and the "gaining" cluster j (addition of point x) are recomputed, and set the new reference points z_i and z_j according to their new centroids. After each step**,** each reference points is the centroid, or mean of the cluster.

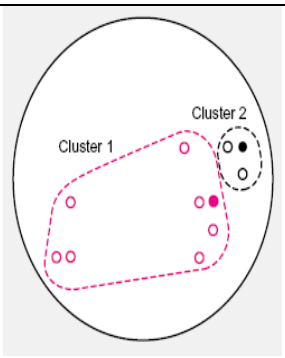
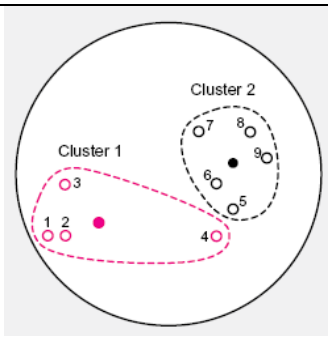
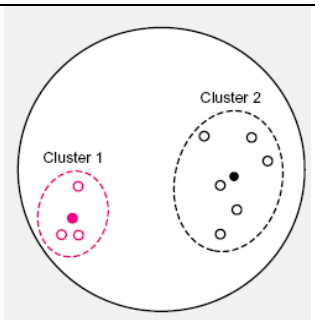
<table>
<tr><td> 
</td><td>(a) Setup:
Reference point 1 (filled red circle) and reference point 2 (filled black circle) are chosen arbitrarily. All 9 data points (open circles) are then partitioned into two clusters: each data point is assigned to cluster 1 or cluster 2, depending on whether the data point is closer to reference point 1 or 2, respectively.</td></tr>
<tr><td> 
</td><td>(b) Results of first iteration:
In the next iteration, the reference point is calculated according to the centroid of its cluster. If the data point is close to the reference point in other cluster, the data point is reassigned to that other cluster, and both cluster centroids are recomputed.</td></tr>
<tr><td> 
</td><td>(c) Results of second iteration:
During the second iteration, the process in part (b) is performed again for every data point. When partition is stable, there will not be change for any further iteration and k-means process stops.</td></tr>
</table>

Figure 3.3: Clustering by the k-means algorithm.

The diagram in Figure 3.3 illustrates how k-means performs clustering on the data set [134]. The diagrams show results during two iterations in the partitioning

data points into two clusters.

PFA performs the data compression and dimension reduction to isolate the principal components based on the same criteria as PCA, and proceeds to derive principal feature components to arrive at a subset of the original feature vector using the k-means clustering algorithm on columns of U_1^T in (3.12) after singular value decomposition of the outer product matrix XX^T [11–13]. This is computationally expensive in general, since $m >> n$ in most practical applications.

3.3 Dominant Feature Identification

In this Section, the proposed Dominant Feature Identification (DFI) methodology of using SVD to identify the dominant features is detailed. First step in DFI is to conduct data compression to obtain the compressed feature space. The retained q singular values are the best approximations to original data space. Second step is to select the dominant features in the compressed feature space using k-means clustering. The retained features should obtain the best approximation to minimize the total least-squares estimation error.

3.3.1 Data compression

Recall in Figure 3.1, X is regarded as a transformation from feature space R^n into data space R^m. Note that $X = U\Sigma V^T = \sum_{i-1}^{n} \sigma_i u_i v_i^T$, where u_i are the column vectors of U and v_i^T are the row vectors of V^T, respectively.

Partition the SVD of X as (3.11) according to

$$X = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = U_1 \Sigma_1 V_1^T + U_2 \Sigma_2 V_2^T$$

with $q < m$ as the desired number of singular values to be retained in Σ_1.

Lemma 1: For any $y = U_1^T Xx \in R^q$, the vector $\hat{z} = U_1 y$ is the best least-square approximation to data vector $z \in R^m$.

Proof 1:

Select

$$Y = U_1^T X \in R^{q \times n} \tag{3.20}$$

so that $x \in R^n$ is mapped to $y = Yx = U_1^T Xx \in R^q$. Then vector $z = Xx \in R^m$ may be approximated in terms of vector $y \in R^q$ according to

$$\hat{z} = U_1 y \tag{3.21}$$

An approximation of $z \in R^m$ can be obtained in terms of vectors in the reduced space R^q depicted by (3.22).

$$\begin{aligned} \hat{z} &= U_1 y = U_1 U_1^T Xx \\ &= (U_1 U_1^T)(U_1 \Sigma_1 V_1^T + U_2 \Sigma_2 V_2^T) x \\ &= U_1 \Sigma_1 V_1^T x = \hat{X} x \end{aligned} \tag{3.22}$$

with $\hat{X}$ depicted in $\hat{X} = U_1 \Sigma_1 V_1^T \in R^q$ with the selection of the first q singular values of X.

The approximation error $\tilde{z}$ is given by

$$\tilde{z} = (X - \hat{X})x = \tilde{X}x = U_2 \Sigma_2 V_2^T x \tag{3.23}$$

Note that

$$\begin{aligned} \hat{z}^T \tilde{z} &= (U_1 \Sigma_1 V_1^T x)^T U_2 \Sigma_2 V_2^T x \\ &= x^T V_1 \, \Sigma_1 U_1^T U_2 \Sigma_2 V_2^T x = 0, \end{aligned} \tag{3.24}$$

i.e., $\hat{z}$ is orthogonal to $\tilde{z}$ which implies that $\hat{z} = U_1 y$ best approximates $z = Xx$ in a least square sense. These prove the Lemma 1. Note that approximation of $z \in R^m$ using $\hat{z} = U_1 y$ and $y \in R^q$, the reduced space is equivalent to the approximation by principal components in $\hat{z} = \hat{X} x$. This allows us to use the reduced vector $y \in R^q$ to compute approximations to data vector $z \in R^m$, instead of using the full feature vector $x \in R^n$.

3.3.2 Selection of dominant features

Now further approximation is realized by selecting the dominant features in R^n. That is, it is desired to select the most important basis vectors from R^n to approximate the data vector z.

Lemma 2: The i^{th} feature in R^n maps into the reduced space R^q as the i^{th} column of matrix $\Sigma_1 V_1^T$.

Proof 2:

It is instrumental to note the little-realized fact that

$$Y = U_1^T X = U_1^T (U_1 \Sigma_1 V_1^T + U_2 \Sigma_2 V_2^T) = \Sigma_1 V_1^T . \tag{3.25}$$

The original basis vectors in feature space R^n are known as features, with the i^{th} basis vector corresponding to the i^{th} feature. The original basis vectors are denoted in R^n by $\{e_1, e_2, \cdots e_n,\}$ with e_i being the i^{th} column of I_n, i.e., e_i is an *n*-vector consisting of zeros except for one in the i^{th} position.

In terms of these notions, the vector generated in R^q by the i^{th} feature $e_i \in R^n$ is given by

$$Ye_i = U_1^T X e_i = \Sigma_1 V_1^T e_i . \tag{3.26}$$

Recall that the rows of V_1^T are denoted by row vectors v_i^T, i.e., the columns of V_1 are denoted by column vectors v_i. By contrast, denote now the columns of $\Sigma_1 V_1^T$ as column vectors $w_i \in R^q$. Then

$$Ye_i = \Sigma_1 V_1^T e_i = \begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix} e_i = w_i \tag{3.27}$$

Therefore, the i^{th} feature in R^n maps into the reduced space R^q as the i^{th} column of matrix $\Sigma_1 V_1^T$. There are therefore *n* vectors w_i in R^q corresponding to the *n* basis axes e_i, i.e., features, in R^n. This proves Lemma 2.

The above-mentioned notions of the proposed DFI algorithm are summarized in Figure 3.1.

Now the best features are selected to retain so as to obtain the best approximation to $z \in R^m$. We call these dominant features. This corresponds to select which basis vectors e_i in R^n to retain, which is equivalent to select the best columns ω_i of $\Sigma_1 V_1^T \in R^q$. This may be accomplished by several methods, including projections and clustering methods inspired by [11-13]. Then, $z \in R^m$ will be approximated using the selected *p* dominant features within R^n.

Note that the n columns ω_i of $\Sigma_1 V_1^T$ will be clustered, as dictated by (3.27). This is in contrast to [11-13] which clustered the columns of U_1^T.

Clustering is the classification of n objects in a data set into p different subsets (clusters), usually by minimizing some norms or pre-defined performance indices. To select the dominant features in R^n, the n vectors $w_i \in R^q$ are clustered into $n \geq p \geq q$ clusters. For our application, the commonly used k-means algorithm is adopted. The k-means algorithm minimizes the following positive semi definite scalar error cost function J iteratively

$$J = \sum_{i=1}^{p} \sum_{w_j \in S_i} (w_j - c_i)^T (w_j - c_i) \tag{3.28}$$

where S_i is the i^{th} cluster set, and c_i is its centroid (or center of "mass") in the cluster space. J is in essence the expectation of the 2-norm (or Euclidian distance) between the objects in the cluster. Note that the k-means algorithm relies heavily on input parameters p, and a poor choice of p usually results in inferior clustering results which depend heavily on the variance of the data in each cluster. For good approximations in R^m, one should select $p > q$, the number of retained singular values.

3.3.3 Error analysis

Lemma 3: DFI to select the dominant features in the compressed feature space yields the minimum overall least-squares approximation and k-means clustering error.

Proof 3:

Here the total error induced by retaining only q singular values is determined by clustering the vectors $\omega_i \in R^q$ into p clusters.

For each cluster, the vector $\overline{w}_i \in R^q$ closest to the cluster center c_i shall be selected as representative of each other vector $w_j \in R^q$ in that cluster. This representative vector is called for cluster i, $\overline{w}_i$ as the cluster leader. The p

features $\bar{e}_i \in R^n$ corresponding to the p cluster leaders $\overline{w}_i \in R^q$ shall be selected as dominant features. This means that the clustering error is given by

$$\hat{J} = \sum_{i=1}^{p} \sum_{w_j \in S_i} (w_j - \overline{w}_i)^T (w_j - \overline{w}_i). \tag{3.29}$$

To summarize these notions, recall that

$$Y = \Sigma_1 V_1^T = \left[w_1 \; w_2 \cdots w_n \right] \tag{3.30}$$

and define

$$\hat{Y} = \left[\hat{w}_1 \quad \hat{w}_2 \quad \cdots \quad \hat{w}_n \right] \tag{3.31}$$

where $\hat{w}_j = \overline{w}_i$ if $w_j \in S_i$, i.e., each vector $w_j \in S_i$ is replaced by its cluster leader $\overline{w}_i$. This means that only the corresponding features $\overline{e}_i \in R^n$ are needed for computation since $\overline{w}_i = \Sigma_1 V_1^T \overline{e}_i$.

Note that

$$y = \Sigma_1 V_1^T x = Yx = \sum_{j=1}^{n} w_j x_j = \sum_{j=1}^{n} x_j [\Sigma_1 V_1^T e_j] \tag{3.32}$$

and define

$$\hat{y} = \hat{Y}x = \sum_{j=1}^{n} \hat{w}_j x_j = \sum_{i=1}^{p} (\sum_{w_j \in S_i} x_j) \overline{w}_i. \tag{3.33}$$

Then an estimate for $z \in R^m$ taking into account both $q < n$ retained singular values and $p < n$ features is given by

$$\hat{\hat{z}} = U_1 \hat{y} \tag{3.34}$$

Recall from (3.21) that $\hat{z} = U_1 y$. The error induced by clustering can be represented as:

$$\hat{z} - \hat{\hat{z}} = U_1 (y - \hat{y}) = U_1 (Y - \hat{Y}) x = U_1 \tilde{Y} x. \tag{3.35}$$

Therefore, the error norm induced by clustering is

$$\left\|\hat{z}-\hat{\tilde{z}}\right\|^2 = (\hat{z}-\hat{\tilde{z}})^T(\hat{z}-\hat{\tilde{z}}) = x^T\tilde{Y}^T{U_1}^T U_1\tilde{Y}x = tr\{\tilde{Y}^T{U_1}^T U_1\tilde{Y}xx^T\} \le \hat{J}\|x\|^2 \qquad (3.36)$$

since $\hat{J} = tr\left\{\tilde{Y}^T\tilde{Y}\right\}$.

The total error induced by neglecting the $n-q$ singular values in Σ_2 and by clustering is then

$$z-\hat{\tilde{z}} = (z-\hat{z})+(\hat{z}-\hat{\tilde{z}}) = U_2\Sigma_2V_2^Tx + U_1\tilde{Y}x. \qquad (3.37)$$

Therefore, the total approximation error norm is

$$\left\|z-\hat{\tilde{z}}\right\|^2 \le (\mathrm{tr}\{\Sigma_2^2\}+\hat{J})\|x\|^2 \qquad (3.38)$$

whose first term depends on the neglected singular values, and the second term is the clustering error, i.e., the neglected features.

It can be claimed that the procedure of first selecting q principal components and then selecting p dominant features yields the minimum overall approximation error in (3.37). Note that

$$\begin{aligned}\tilde{z}^T(\hat{z}-\hat{\tilde{z}}) &= (U_2\Sigma_2V_2^Tx)^TU_1\tilde{Y}x \\ &= x^TV_2\Sigma_2U_2^TU_1\tilde{Y}x = 0,\end{aligned} \qquad (3.39)$$

i.e., the error in neglecting $n-q$ singular values and the clustering error are orthogonal. This proves Lemma 3. This also means that there is no better way of selecting dominant features than the DFI methodology proposed therein.

3.3.4 Simplified computations

Traditional PCA relies on computations using the correlation matrix $XX^T \in R^{m\times m}$ and SVD routines on XX^T and obtain $XX^T = U\Sigma^2U^T$ [4]. The computational complexity in this process is $O(m^2n)+O(m^3)$. $O(m^2n)$ is from XX^T calculation and $O(m^3)$ is from singular value decomposition.

DFI uses the inner product matrix $X^T X$ which is defined as:

$$X^T X = V\Sigma U^T U\Sigma V^T = V\Sigma^2 V^T \tag{3.40}$$

SVD routine arranges the singular values in descending order. The transformed data can thus be written as

$$Y = \tilde{V}^T X = \tilde{V}^T V\Sigma^2 V^T \tag{3.41}$$

where $\tilde{V}^T V$ is a R^n matrix which is one on the diagonal and zero elsewhere. As $X^T X \in R^{n\times n}$, the computational complexity in DFI is $O(mn^2) + O(n^3)$. $O(mn^2)$ is from $X^T X$ calculation and $O(n^3)$ is from singular value decomposition. Since $n << m$, the computation complexity has been reduced largely compare with the PCA.

3.3.5 Pseudocode

In the data acquisition experiment, suppose we collect data on time interval [0, *m*] from the sensors and compute *n* features using digital signal processing techniques. We then converted the collected data from sensors into matrix $X \in R^{m\times n}(m >> n)$. The pseudocode to determine the dominant features is presented below:

Algorithms: DFI INPUT: data $X \in R^{m\times n}$ OUTPUT: $feature_{set} \in R^{m\times q}$

//Data normalize: normalize the data set $X \in R^{m\times n}$ to zero-mean
1: READ Data $X \in R^{m\times n}$
2: Calculate the mean value for each column of X and store as $feature_{mean}(j)$, $(j = 1...n)$
3: WHILE $j < n$ DO
WHILE $i = 1:m$ DO
$X1(i, j) = X(i, j) - feature_{mean}(j)$ // ($X1$ is the normalize result of X)
ENDWHILE
ENDWHILE

//Singular value decomposition of $X1$. Calculate eigenvector and eigenvalues

1: READ Data $X1 \in R^{m \times n}$

2: Calculate inner product matrix $X1^T X1 \in R^{n \times n}$

3: Calculate eigenvalue Σ and eigenvector V on $X1^T X1$ using Equation (3.40)

4: eigenvalues σ_i is as $\sigma_i = \Sigma_{ii}$

//Choose number of principal components and clusters:

1: WHILE $j < n$ DO

2: calculate $|\sigma_{i+1} - \sigma_i|$

3: IF $|\sigma_{i+1} - \sigma_i| < t$ (t is predefined value),

4: $q = i$ // (q is the desired number of principal components)

5. ENDIF and BREAK WHILE loop

6: ENDWHILE

7: Σ_1 as a diagonal matrix and $\Sigma_{1_{ii}}(i,i) = \sigma_i (i = 1:q)$ // (keep the retained number q of principal components and obtain Σ_1 and V_1^T).

//Feature Subset determination

1: K-mean clustering on the columns of $\Sigma_1 V_1^T$

2. K-mean clustering determines the optimal cluster numbers of p and $p > q$

3: WHILE $j < p$ DO

obtain the centroid of each cluster c_j

find the feature vector w_i that has shortest distance between w_i and c_j

e_i as the dominant feature in the cluster

END WHILE

4: Combine the p dominant features to form the dominant feature subset:

$feature_{set} = \lfloor w_1, w_2, ... w_p \rfloor$

3.3.6 A Toy Example

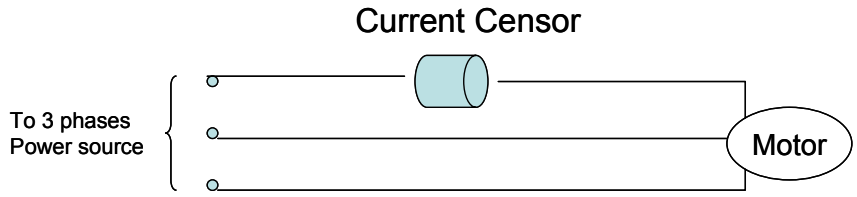


Figure 3.4: A toy example.

In the toy example, 30000 data samples are collected from the current sensor after conducted experiment in the above setup. Seven features of *a, b, c, d, e, f, g* are extracted from the current sensor signal which are shown in Table 3.1:

We form the matrix $X \in R^{30000\times 7}$. Assume that the principle components *vs.* singular values plotting is shown as Figure 3.2, the retained number of principle components should be 3 as the singular values are almost constant for component numbers greater than 3 .

Table 3.1 Data samples collected from current sensor

Data Samples	**Features**						
	a	*b*	*c*	*d*	*e*	*f*	*g*
1	0.108	11.101	0.505	4.202	45.102	12.908	4.323
2	0.102	14.108	0.308	3.108	47.102	14.567	4.115
3	0.009	10.108	0.116	4.108	48.102	15.102	3.102
4	0.108	15.102	0.157	3.102	49.102	12.009	3.009
5	0.102	12.009	02009	4.009	52.102	13.789	5.112
…..							
29999	0.108	14.108	0.116	3.108	51.102	15.102	3.102
30000	0.102	10.108	0.157	4.108	52.102	12.009	3.009

Partition the SVD of X as (3.11):

$$X = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = U_1 \Sigma_1 V_1^T + U_2 \Sigma_2 V_2^T$$

and $U_1 \in R^{30000\times3}$, $U_2 \in R^{30000\times4}$, $\Sigma_1 \in R^{3\times3}$, $\Sigma_2 \in R^{4\times4}$, $V_1^T \in R^{3\times7}$ and $V_2^T \in R^{4\times7}$.

According to Lemma2, the i^{th} feature in R^7 maps into the reduced space R^3 as the i^{th} column of matrix $\Sigma_1 V_1^T \in R^{3\times7}$.

Let $w_i (i = 1:7)$ represents the column of $\Sigma_1 V_1^T \in R^{3\times7}$. Use k-means clustering algorithm to determine the optimize number of clusters as $p = 4$ and cluster the data into 4 clusters.

In the toy example, the cluster results are:

Cluster 1: features *a* & *f*
Cluster 2: features *b* & *e*
Cluster 3: features *c*& *g*
Cluster 4: feature *d*

Calculate the centroid in each cluster and find the dominant features in each cluster based on the nearest distance to the centric. The results are as follows:

Cluster 1: feature *a*
Cluster 2: feature *e*
Cluster 3: feature *c*
Cluster 4: feature *d*

In conclusion, four dominant features are selected as features *a*, *c*, *d* and *e*. The computational complexity of using DFI is $O(1.47\times10^6) + O(343)$. While the computational complexity of using PFA or PCA is $O(6.3\times10^9) + O(2.7\times10^{13})$. DFI is numerically efficient, and reduces the complexity of feature selection greatly.

3.4 Summary

In this chapter, a rigorous mathematical framework for Dominant Feature Identification (DFI) is developed that provides an autonomous rule-base for data and sensor reduction. Singular Value Decomposition (SVD) is used to decompose the inner product matrix of collected data. The principal components are optimized in a least squares sense in a certain reduced space, and the dominant features are extracted using the k-means clustering algorithm [6-7]. This DFI framework uses formal mathematical analysis to select dominant features. A numerically efficient scheme for implementation is presented that is based on the inner product matrix of

the collected data, not the correlation (outer product) matrix. The proposed DFI decision tool is numerically efficient, and reduces the complexity of feature selection greatly while reducing the least square error during dominant feature selection and clustering automatically. It is proven that the proposed method of dominant feature selection is optimal in the sense that it minimizes the total least squares estimation error.

CHAPTER 4 INTELLIGENT DIAGNOSIS AND PROGNOSIS OF TOOL WEAR USING DOMINANT FEATURE IDENTIFICATION

In this Chapter, DFI algorithm that is proposed in Chapter 3 is applied for tool wear real time prediction. Tool wear and tool failure may result in a loss in surface finish and dimensional accuracy of the finished parts, and even possible damage to the work piece and machine [138]. Tool wear can only be measured by removing the tool and performing visual inspection and measurement, which is tedious, and time consuming and results in down-time for the machine. Tool wear and failure monitoring have raised a lot of interest among researchers, as they can help to prevent damages of machine tools and work pieces, as well as improve the surface quality. In this Chapter, a methodology is presented for reasoning with data from installed force and acoustic emission sensors to predict tool wear continuously using an indirect approach, on-line and in real time.

In the indirect method, the signals and raw data collected from the detection system reflect the cutters conditions. To avoid inefficiency which results in loss in productivity, the tool wear and part failures are estimated online without ceasing operation of the cutting tool. While it is essential that the information gathered from the input features is sufficient to determine the cutter failure, the usage of redundant input features burdens the training process, as well as lengthens the data pre-processing and recognition times.

As such, there is a need for accurate deduction of the dominant features that contribute to the deterioration of the cutting tool through signals and raw data collected in the detection system. Using the proposed DFI methodology in Chapter 3, the dominant features contributing to the deterioration of the cutting tool could be selected for time series forecast of its wear. This allows the lifespan of the cutting tool to be predicted for precise control of operating conditions and productivity improvements.

In this Chapter, the proposed DFI methodology is tested on the ball nose cuter of an industrial high speed milling machine. Force measurements were taken over a time period using a three axis dynamometer. Acoustic Emission (AE)

measurements were taken over a time period using an AE sensor. During the measuring period, the tool is periodically extracted from the chuck and tool wear is measured using an Olympus microscope. This yields a baseline time plot of actual tool wear versus time. Features, commonly used for machinery monitoring in industries, are computed from the measured force and AE data. The proposed DFI method is then used to select various sets of dominant features, which are used in a regression model or Autoregressive–Moving-Average (ARMA) model to predict the baseline observed tool wear. The performance of DFI is evaluated based on the accuracy of prediction of the actual tool wear. Comparisons are made with another technique of PCA and PFA for feature selection in the literature [11-13].

The rest of Chapter 4 is organized as follows: Section 4.1 shows using force for tool condition monitoring. It presents how to use the resulting dominant features in tool wear prediction using the Recursive Least Squares (RLS) algorithm and computes the total error induced by our proposed methods. Section 4.2 shows another experiment using both force and acoustic emission sensors. It presents the usage of a dynamic autoregressive-moving average with exogenous terms (ARMAX) model with Extended Least Squares (ELS) for tool wear prediction in the new DFI framework.

4.1 Time series forecasting of tool wear using cutting force by dominant feature identification

The author now wishes to apply DFI to predict tool wear in an industrial cutting tool. It is wished to predict tool wear using force signals that are easily monitored in real time.

4.1.1 Experimental setup

A case study was carried out to verify the usability of DFI. An application related to machining tool condition was selected for the experiment. Tool condition is an important factor in the machining process. Image of tools and a flank tool wear is shown in Figure 4.1.

Figure 4.1: Tools and a flank tool wear at cutting edge

Most computer numerical control milling machines are not able to detect machining tool's wear in an on-line manner. The cutting force signal is instead used to establish usable models due to its high sensitivity to tool wear, low noise, and good measurement accuracy [61]. In our experiment, a milling machine is used as the test bed. The schematic diagram of the experimental setup is illustrated in Figure 4.2.

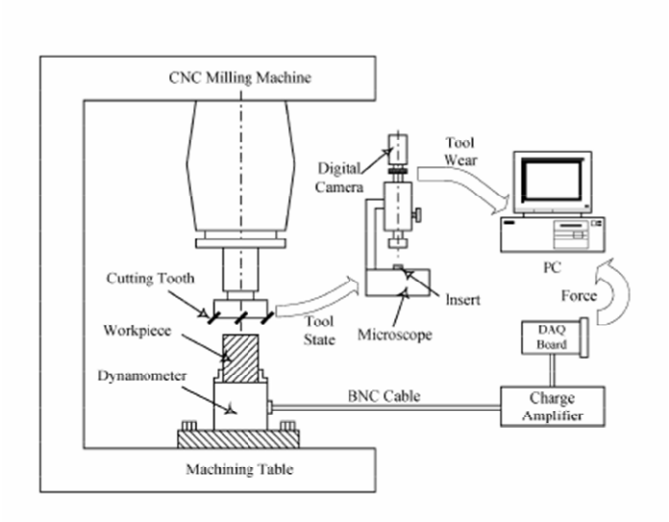


Figure 4.2 Experimental setup in machinery [36]

The cutting forces along the X, Y, and Z axes were captured using a Kistler dynamometer in the form of charges, which were converted to voltages by a Kistler charge amplifier. The voltage signal was sampled by a PCI 1200 board at 2000 Hz and directly streamed to the hard disk of a computer. The flank wear of each individual tooth of the cutting tool was measured with an Olympus microscope. Details of the experimental set up and feature extraction

methodologies have been reported in [36], and are omitted here for brevity but without loss of generality. The experimental set up components are listed in Table 4.1.

Table 4.1
Experimental components

Components
Makino CNC milling machine
EGD 4450R cutter with AC325 and A30N inserts
ASSAB718HH workpiece
Kistler 9265B Quartz 3-component dynamometer
Kistler 5019A multichannel charge amplifier
NI-DAQ PCI 1200 board
Olympus microscope
Computer

An example of the cutting force signal in three axes (F_x, F_y, F_z) is shown in Figure 4.3.

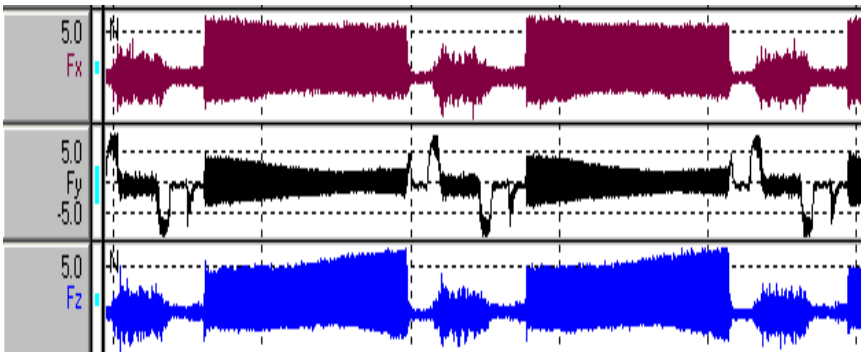


Figure 4.3: Three axes cutting force signals (Fx, Fy, Fz)

In the experiment, 20,000 data points of measured force sensor data were captured under the following machine settings: spindle speed 1000 rpm, feed rate 200 mm/min, depth of cut 1 mm, and insert number 2. The flank wear of each individual tooth was measured at an interval of five tool passes by the Olympus microscope, and, at each time, an average was taken from all the teeth mounted on the cutter. This yields the baseline actual tool wear plot shown in Figures 4.5-4.8 of the experiment results in Section 4.1.4.

4.1.2 Extraction of features

Table 4.2
Features and nomenclature

No	Feature	Notation	References
1	Residual error	re	[36]
2	First order differencing	fod	[32]
3	Second order differencing	sod	[32]
4	Maximum force level	fm	[61]
5	Total amplitude of cutting force	fa	[61]
6	Combined incremental force changes	df	[61]
7	Amplitude ratio	ra	[61]
8	Standard deviation of force components in tool breakage zone	fstd	[62]
9	Sum of squares of residual errors	sre	[63]
10	Peak rate of cutting forces	kpr	[64]
11	Total harmonic power	thp	[65]
12	Average force	fca	[66]
13	Variable force	vf	[66]
14	Standard deviation	std	[67]
15	Skew (3rd moment)	skew	[67]
16	Kurtosis (4th moment)	kts	[67]

Various features have been proposed for monitoring the force signals in milling processes. Altintas [61] found that when a breakage occurred, a large cutting-force residual-error was produced between the actual measurement and the predicted value from the auto regressive model. In another attempt by Altintas *et al.* [61], time averaged force was used to detect tool failure. In Tarn *et al.* [62], the four parameters of maximum force level, total amplitude of the cutting force, combined incremental force changes, and the amplitude ratio to monitor the tool conditions were used. Tarng [63] defined a tool breakage zone as a region located within the frequency range between the DC component and the tooth frequency, and concluded that the force components within this zone correlate with the tool breakage very well. Tansel *et al.* [64] summed up the squares of the residual errors in each tooth period and correlated it with the tool breakage. Zhang *et al.* [65] used

the peak rate of cutting forces to detect tool breakages. Elbestawi *et al.* [66] found that the harmonic contents of the cutting forces were sensitive to tool flank wear. Tarng [63] used the average force and the variable force for sensing tool breakage. Leem *et al.* [67] extracted four statistics from the cutting force, the mean, standard deviation, skew, and kurtosis, to monitor tool wear. These sixteen features from the methodologies mentioned above are summarized in Table 4.2 and form the scope of the feature subset selection.

4.1.3 Selection of dominant features

With the force features obtained, the DFI methodology described in [134] and Section 3.3 are used to obtain a feature subset. The five following steps are involved in the DFI to obtain the dominant features which are minimum in overall approximation and k-means clustering error:

1. Data acquisition. Collect data on time interval [0, *m*] from the cutting tools' sensors and compute *n* features using digital signal processing techniques. Pack *n* time series, each of length *m,* as columns into matrix $X \in R^{m \times n}$. In this experiment, there are 20,000 data collections and 10,000 of them are used for training, so $m = 10{,}000$ and $n = 16$ in this experiment.
2. Initialization. Detrend the data in X by subtracting the mean (average across each dimension) of each of the dimensions to normalize the data set to zero-mean. This step results in a data set whose mean is zero.
3. Choose number of principal components and clusters. Perform SVD on inner product matrix $X^T X \in R^{16 \times 16}$ to obtain Σ_1 and V_1^T. Select the desired number q of principal components in $\sigma_i v_i^T$ and using k-means to determine the number of clusters p. In this experiment, $q = 3 \sim 4$, while $p = 4 \sim 9$ in the Tables 4.4-4.10.
4. Clustering. Use the k-means algorithm for clustering to find the centroids c_i of each cluster.
5. Subset determination. Select the vector w_i "nearest" to the centroid of each cluster as its cluster leader $\bar{w}_t$, and corresponding f_i as the dominant

feature. Combine the p dominant features to form the reduced feature space.

4.1.4 Time series forecasting of tool wear using dominant features identification

In our application, sensors are used to measure cutting force online in real time. Then, signal processing is used to compute n features, such as mean force value, maximum force level, standard deviation, third moment (skew), etc. Data is taken over N time steps and stored, and consists of the values of the n features at each time step, along with the tool wear measured (by visual inspection) at each time step.

To put this into the framework discussed for DFI in Chapter 3, define $f_k \in R^n$ as the feature vector and y_k as the tool wear at time k. It is desired to predict y_k in terms of f_k. One has the standard linear regression form

$$\begin{bmatrix} f_1^T \\ f_2^T \\ \vdots \\ f_k^T \end{bmatrix} \theta = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} \tag{4.1}$$

where $\theta \in R^n$ is the unknown parameter vector that expresses the tool wear in terms of the collected data as

$$y_k = f_k{}^T \theta \tag{4.2}$$

However, it is desired to express the tool wear not in terms of all features, but only of the most important features. This is motivated by the fact that selection of only the dominant features allows a reduction in signal processing.

Express (4.1) as

$$X\theta = z \tag{4.3}$$

This identifies

$$X \equiv \begin{bmatrix} f_1^T \\ f_2^T \\ \vdots \\ f_k^T \end{bmatrix}, \quad z \equiv \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} \tag{4.4}$$

and suggests the following means of using dominant features to predict the

tool wear.

Partition the collected data set over N time steps into two sets. The data of $\frac{1}{2}N \le m < N$ as a training set is used to compute the p dominant features and determine the unknown parameter vector θ. If m is too small, the training dataset is enough to represent the overall data characteristics and the selected number of features may not be the optimum set. The remaining data from time $m+1$ to time N is used to verify the prediction accuracy as a validation set. In our application, half of datasets are used for training, while half of datasets are used for testing. We have $N = 20 \times 10^3$ samples, and thus the choice for m is $m = 10 \times 10^3$ data samples. The experiment results are detailed in Section 4.1.6.

Define

$$X \equiv \begin{bmatrix} f_1^T \\ f_2^T \\ \vdots \\ f_m^T \end{bmatrix} \in R^{m \times n} \tag{4.5}$$

in terms of the collected data from the installed sensors through time *m*. Use the method in Chapter 3, Sections 3.3 to select the *p* dominant features matrix $\phi \in R^{m \times p}$.

Define $\varphi_k{}^T \in R^p$ as the vector containing the measured dominant features at time $k \le m$. Note that φ_k is a *p*-subvector of f_k. Then tool wear y_k is desired to predict in terms of the *p* dominant features using

$$y_k = \varphi_k^T \theta \tag{4.6}$$

To do so, one can estimate the parameter vector θ using the measured data as

$$\begin{bmatrix} \varphi_1^T \\ \varphi_2^T \\ \vdots \\ \varphi_k^T \end{bmatrix} \theta = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}. \tag{4.7}$$

The least squares technique is used on the *p* dominant features, *i.e.*, $\theta \in R^p$:

$$\Phi_k \theta \equiv \begin{bmatrix} \varphi_1^T \\ \varphi_2^T \\ \vdots \\ \varphi_k^T \end{bmatrix} \theta = \begin{bmatrix} d_1^T \\ d_2^T \\ \vdots \\ d_k^T \end{bmatrix} \equiv Y_k \tag{4.8}$$

in terms of the data collected. The estimation error through time *k* is

$$E_k = Y_k - \Phi_k \theta \tag{4.9}$$

where $\theta \in R^p$ is the unknown vector to be regressed for times series forecast of tool wear.

A least square estimate of θ which minimizes the error norm $E_k{}^T E$ is given by the standard unique batch solution

$$\theta = (\Phi_k^T \Phi_k)^{-1} \Phi_k{}^T Y_k \tag{4.10}$$

if there is sufficient persistent excitation, *i.e.*, $\Phi_k{}^T \Phi_k$ is invertible.

To compute θ using efficient on-line recursive means, one may use Recursive Least Squares (RLS) instead of (4.10). The RLS algorithm is governed by the following equations

$$\begin{aligned} \theta(k) &= \theta(k-1) + K(k)[d_k - \varphi_k^T \theta(k-1)] \\ K(k) &= P(k-1)\varphi_k [I + \varphi_k^T P(k-1)\varphi_k]^{-1} \\ P(k) &= [I - K(k)\varphi_k^T] P(k-1) \end{aligned} \tag{4.11}$$

where $P(k) = \left[\Phi^T(k)\Phi(k)\right]^{-1}$ and $\theta(k)$ is the estimate of θ at time k.

The pseudocode for time series forecasting of tool wear using RLS is presented in the following:

Algorithms: Tool wear time series forecasting using RLS INPUT: dominant features matrix $\phi \in R^{m \times p}$ OUTPUT: tool wear $y_k \in R^{m \times p}$
// Recursive least squares to calculate estimate parameters $\theta(k)$ 1: WHILE $i < m$ DO 2: calculate estimate parameters $\theta(k)$ using Equation (4.11) 3: ENDWHILE
// Calculate tool weal y_k 1: WHILE $i < m$ DO 2: calculate tool weal y_k using Equation (4.6) 3: END WHILE

4.1.5 Experimental results

In the experiment, 20,000 points of measured force sensor data were captured and the flank wear of each individual tooth of the cutting tool was measured with an Olympus microscope. This yields the baseline actual tool wear plot shown in Figures 4.5-4.8.

The measured force data was detrended, *i.e.*, the mean value was subtracted. Based on this data, the sixteen features in Table 4.2 were computed as functions of time. This yields n=16 feature vectors, each of which is a function of time and has m=10,000 data points. This yields a matrix X in (4.5) that has n=16 columns and m=10,000 rows. Next, the SVD of the inner product matrix X^TX in (3.40) was performed.

The resulting singular values are shown in Table 4.3 and plotted in Figure 4.4. It is seen that the fourth and subsequent singular values are quite small compared to the first three singular values. Therefore, it appears that retaining the first three singular values would be sufficient to capture the relevant trend in tool wear prediction for this experiment. However, it is not clear from this observation which of the sixteen features is the main ones useful for tool wear prediction at this point in time. The DFI method is therefore used to select the most important features.

First, q dominant singular values are selected based on the ratio between the sums of the squares of the first q singular values to the sum of the squares of all the singular values. This yields $\Sigma_1 V_1^T$ in (3.25) and (3.27). The best p columns w_i of $\Sigma_1 V_1^T$ in (3.27) are next chosen for clustering. This will yield p dominant features in R^n, where $n \geq p \geq q$. Clustering of the data set using the k-means algorithm now yields p clusters S_i. The p vectors w_i closest to the centroids of each cluster S_i are chosen as the cluster leader $\bar{w}_i$. The corresponding p features are selected as the dominant features. These dominant features are now used for predicting the time series of failure of the cutting tool using the methods of Section 4.1.3.

Table 4.3
Principal components and singular values

No	Singular Values
1	31.90702
2	1.082043
3	0.00342
4	0.00026
5	0.00011
6	0.00005
7	0.00005
8	0.00001
9	0.00001
10	3.07857 × 10-6
11	1.13154 × 10-6
12	6.45546 × 10-7
13	4.6294 × 10-7
14	1.71882 × 10-7
15	4.60490 × 10-9
16	2.02272 × 10-9

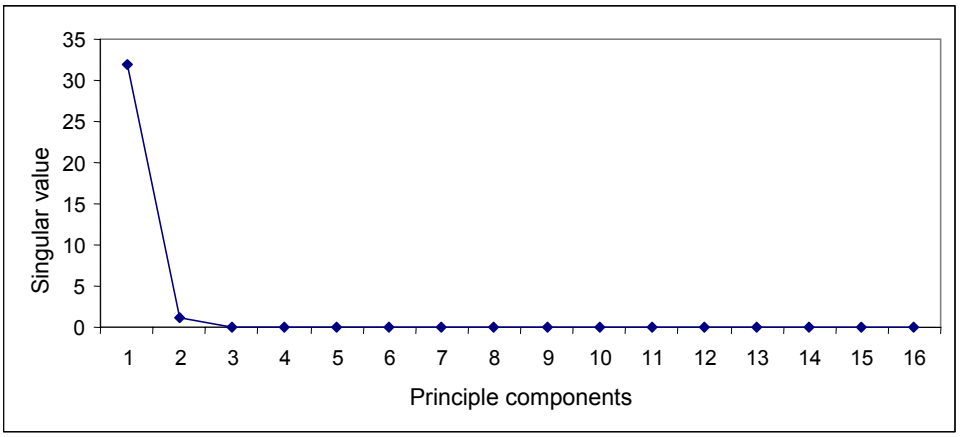


Figure 4.4: Plot of principal components vs. singular values

A. Effects of different numbers of retained singular values q and dominant features p

Using the RLS techniques of Section 4.1.2, a Multiple Regression Model (MRM) (4.6) was identified to predict the baseline measured tool wear using all sixteen of the original features. Figure 4.5 shows the actual measured tool wear and the predicted tool wear using this MRM as functions of time. Clearly, the prediction is good. A Mean Relative Error (MRE) of 8.8% is observed for this MRM in Figure 4.5, and represents our best possible prediction of tool wear using this set of sixteen features.

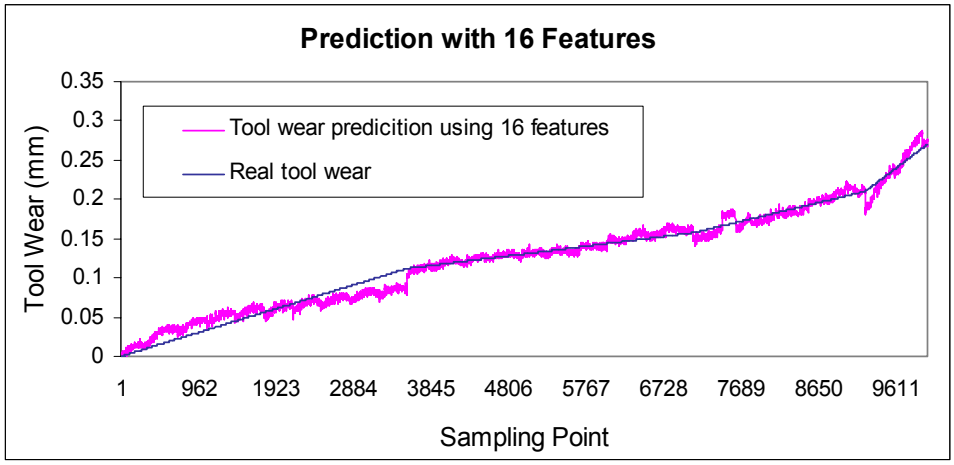


Figure 4.5: MRM using sixteen dominant features and the RLS algorithm

It is desired to use fewer features to predict the tool wear. Figure 4.6 shows the resulting tool wear prediction when four random features in Table 4.2 are randomly selected. Here, $\{f_m, fod, sod, vf\}$ are chosen. The 'steps' in Figure 4.6 are due to the usage of these four randomly selected features (and not the four dominant features) for prediction of real measured tool wear for illustration purposes but without loss of generality. The MRE is 22.7%, which is unacceptable. Therefore, the DFI method is used to select the important features with better justification and results.

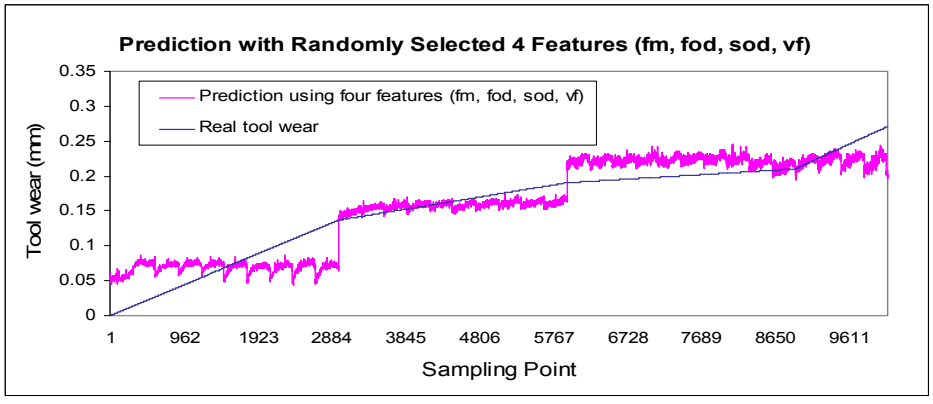


Figure 4.6: MRM using random selected four features $\{f_m, fod, sod, vf\}$ and tool wear comparison

Therefore, different numbers q of retained singular values and p of dominant features are selected and our DFI algorithm in Section 3.3 is performed to select the resulting p dominant features. Smaller Mean Square Error (MSE) and Mean Relative Error (MRE) will be obtained when more singular values are retained or more dominant features are used. The best combination of q and p, leading to the minimum number p of dominant features needed in predicting tool wear, depends on the tolerance of MSE and MRE required.

Figure 4.7 shows the result using $q = 3$ retained singular values and $p = 4$ dominant features. The tool wear prediction MRE is 11.12%. This is excellent, and is very close to the MRE of 8.8% obtained retaining all sixteen features in Figure 4.5. The four dominant features turn out to be $\{f_a, f_{ca}, f_{std}, thp\}$.

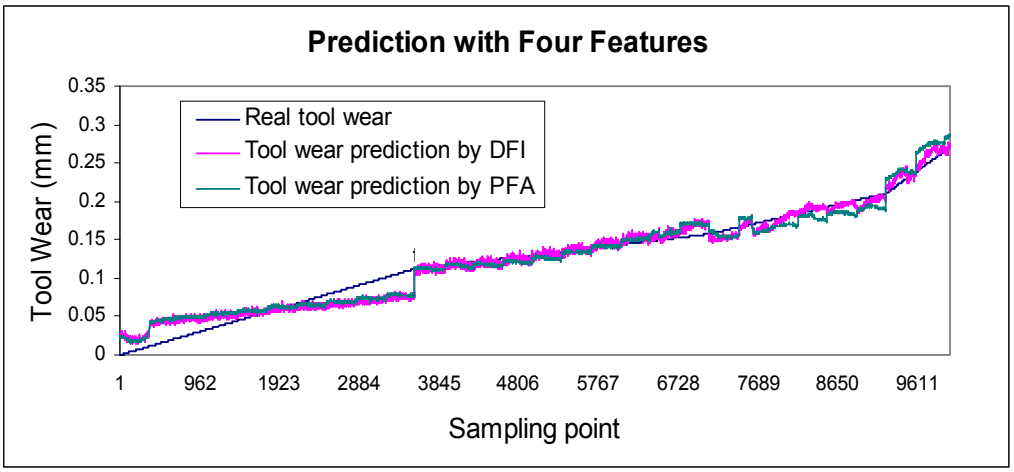


Figure 4.7: Examples of MRM using four dominant features, three principal components, and the RLS algorithm

B. Comparison of proposed Dominant Feature Identification (DFI) and Principal Feature Analysis (PFA)

Next, DFI method is compared to the PFA method of [11-13]. Figure 4.7 shows the actual measured tool wear, and its prediction using the best four features selected by DFI, and the four features selected by PFA. The MRE for DFI is 11.12%, while MRE for PFA is 13.18%. Both PFA and DFI perform better than random selected features (see Figure 4.6), and DFI obtains better performance than using PFA.

Also, comparison studies were carried out using increasing numbers q of retained singular values and p of dominant features. Comparison results using three retained singular values are shown in Table 4.4 with DFI methodology and in Table 4.5 with PFA methodology, respectively. Results using four retained singular values are shown in Table 4.6 with DFI methodology and in Table 4.7 with PFA method. In these Tables, the first column denotes the number of dominant features to be used for tool wear prediction, and the second column displayed the selected dominant features. The MSE and MRE obtained using these dominant features for tool wear forecast are displayed in columns three and four.

Table 4.4
Results of DFI method using three retained singular values

No of dominant features	**Features selected**	**MSE**	**MRE (%)**
4	fa, fca, fstd, thp	1.262	11.61
5	fa, fca, fm, skew, thp	1.202	11.19
6	fa, fca, fstd, ra, sre, thp	1.111	10.49
7	fa, fca, fstd, ra, skew, sod, thp	1.111	10.40
8	fa, fca, fstd, kts, ra, skew, sod, thp	0.946	8.86
9	fa, fca, fstd, kpr, kts, ra, skew, thp, vf	0.946	8.86

Table 4.5
Results of PFA method using three retained singular values

No of dominant features	**Features selected**	**MSE**	**MRE (%)**
4	fa fca ra hp	1.40	13.18
5	fa fca kts ra, thp	1.365	12.89
6	fa kts re skew std thp	1.133	10.53
7	fa fca fstd kts skew thp, vf	1.130	10.46
8	fa fca fstd kts ra re skew, thp	0.949	8.88
9	fa fca fm fstd, kts ra re skew, thp	0.948	8.88

Table 4.6
Results of DFI method using four retained singular values

No of dominant features	Features selected	MSE	MRE (%)
5	fa, fca, fm, skew, thp	1.201	11.20
6	fa, fca, fstd, ra, sre, thp	1.111	10.49
7	fa, fca, fstd, ra, skew, sod, thp	1.111	10.40
8	fa, fca, fstd, kts, ra, skew, sod, thp	0.946	8.86
9	fa, fca, fstd, kpr, kts, ra, skew, thp, vf	0.946	8.86

Table 4.7
Results of PFA method using four retained singular values

No of dominant features	Features selected	MSE	MRE (%)
5	dx, fa, kts, std, thp	1.458	13.64
6	dx, fa, fca, kts, std, thp	1.215	11.19
7	fa, fm, kts, sre, skew, std, thp	1.128	10.45
8	fa, fca, fstd, kts, ra, skew, sod, thp	0.946	8.86
9	fa, fca, fstd, kpr, kts, ra, skew, thp, vf	0.946	8.86

The observations from these tables are as follows:

1) The dominant features are the same using the DFI methodology when different numbers of principal components are used. However, the features are slightly different using the PFA method when different numbers of principal components are used, especially when the numbers of dominant features chosen are four and five.

2) The features chosen are different using the proposed DFI methodology and PFA. DFI gives smaller MSE and MRE when compared to that using PFA, which provides better accuracy than PFA in tool wear prediction, especially when the numbers of dominant features chosen are four and five. When using a small number of dominant features, the computational time for features processing is saved. Our experiments also found that the computational time required to build MRM using the original sixteen features takes about five times

longer of computational time, when compared to that using four dominant features only.

3) When the number of dominant features chosen reaches to eight or more, the improvements in MSE and MRE using increasing number of dominant features are insignificant. The MRE of using eight dominant features is 8.86%, which is very close to the MRE value of 8.80% from using all of the sixteen original features. Our experiments also concluded that with the selected eight dominant features using the proposed DFI method to build MRMs saves about 60% of computational time than that to build the models with original sixteen features. As such, using the selected features subset dominant features with the proposed DFI method has the advantage of implementing on-line prediction, as it saves computational time without loss of accuracy.

4) Since $m = 10{,}000$ and $n = 16$. The computational complexity for PFA in this process is $O(m^2 n) + O(m^3) = O(1.6 \times 10^9) + O(10^{12})$ while for DFI method, the complexity is $O(mn^2) + O(n^3) = O(2.56 \times 10^6) + O(4096)$. DFI is numerically efficient and reduces the complexity of feature selection greatly.

4.2 Tool wear monitoring using acoustic emissions by dominant feature identification

The Acoustic Emission (AE) signal is very effective for indirect methods in TCM because of its non-intrusiveness, ease of operation, and fast dynamic response. The AE frequency range is also higher than that of machine vibration and environmental noise [141]. Another advantage of using the AE signal is that an AE sensor is small and can be installed easily, and the sampling process does not interrupt the machining operation [142]. The costs of AE sensors and force sensors also differ greatly, and using AE signals is less expensive than using force signals. Typical advantages of AE are high sensitivity, early and rapid detection of defects such as leak and cracks, *etc*. Also, the AE signal is more convenient to use in real time TCM.

Fundamentally, AE signals can be divided into continuous and burst types. The continuous-type AE signal waveform is similar to Gaussian random noise but its amplitude varies with AE activity. The burst type is a short duration pulse due

to the discrete release of high amplitude strain energy [145]. As the tool becomes worn, the AE signal produces multiple bursts [144]. AE can then be used to detect the chip formation mechanism, and friction between the chip and tool rake face in both turning and milling operations [145].

The characteristics of an AE signal to be used in TCM are studied by researchers in various aspects. In [39], the authors used spectral, statistical, and time series analyses to analyze the tool wear. Liao *et al*. [105] observed that there was a general increasing trend of AE rms (root-mean-square) with cutting speed and depth of cut, and the use of oil based lubricant reduced the mean AE rms [141]. They also observed that the measurement of the peak count ratio displayed a constant rate of decrease with gradual wear. Srinivasa *et al*. [146] found that there was a significant change in skew and kurtosis of AE signals with tool failure [146]. Kannatey and Dornfield proposed a relationship between skew, kurtosis, and tool wear [147]. In these efforts, various features have been proposed for monitoring the AE signals in milling processes. However, perusal of AE signals for feature selection affecting tool wear has not been studied. In this Section, it will be shown how to predict the tool wear using DFI to select the dominant features from AE signals. An application related to milling machining tool condition is selected for the experiment.

4.2.1 Experimental setup

An application related to milling machining tool condition is selected for the experiment. In this experiment, a Röders TEC vertical milling machine was used as the test bed as shown in Figure 4.8. A ball nose cutter was selected for our testing. The cutting process was performed by predefined procedures. After each cutting process, tool snapshots were taken to measure the amount of tool wear. An LECIA MZ12.5 high performance stereomicroscope was used to measure the tool wear of the cutting tool.

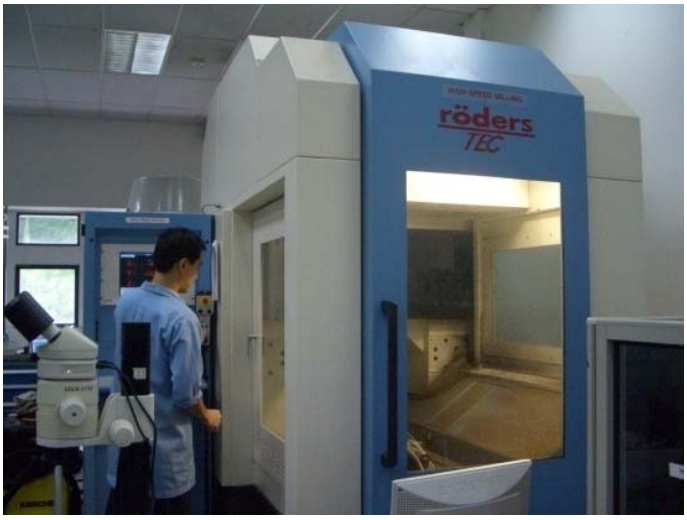


Figure 4.8: High speed milling machine

Most computer numerical control milling machines are not able to detect machining tool wear in an on-line manner. In our experiment, the cutting force and AE signals were used to establish usable models due to their high sensitivity to tool wear, low noise, and good measurement accuracy. AE sensors were used to detect dynamic motion resulting from AE events and to convert the detected motion into a voltage-time signal. AE detection is commonly performed with sensors that use piezoelectric elements for transduction. During high speed milling, metallic materials removing causes a release of transient broadband stress wave energy (SWE) referred to as AE. The best location for AE sensor placement is the side of a work piece or the side of a tool shank. In these setups, AE could transmit from source through metallic material to the sensors. AE sensors could measure the material deformation.

For our application, an 8152B211 Piezotron® AE sensor (Kistler) was used for tool condition monitoring in the experiment as shown in Figure 4.9. Due to its small size factor, it can be easily mounted near the source of emission for optimal measurement of the AE signal.

Figure 4.9: Piezotron acoustic emission sensor used, 100 to 900 kHz

The following guidelines suggested by Chen and Zeng in [148] for sensor position in process monitoring are also observed

1. Measurement point as close to the machining point as possible;
2. No reduction in the static and dynamic stiffness of the machine tool;
3. No restriction of working space and cutting parameters;
4. Wear-free and maintenance-free, easy to change, low costs;
5. Resistant to coolant, dirt, chips and mechanical, electromagnetic, and thermal influences;
6. Function independent of tool or work piece;
7. Adequate metrological characteristics; and
8. Reliable signal transmission.

The SWE generated in the milling is within a 100 KHz–1 MHz band. AE defect signals exhibit less spectral overlap with low-frequency background noise. In our experiment, the AE signal was initially 40 dB pre-amplified and bandpass filtered. A bandpass filter with a range of 100 kHz–1 MHz was used to reduce the influence of low-frequency noise. Subsequently signal processing techniques of wavelet transform and fast Fourier transform (FFT) were used after collection of the AE signals.

The experimental setup with both the dynamometer (force sensor) and AE sensor deployed in a TEC vertical milling machine is shown in Figure 4.10. A 6-mm ball nose tungsten carbide cutter was chosen to machine a Titanium Ti6Al4V work piece. The experimental components are summarized in Table 4.8.

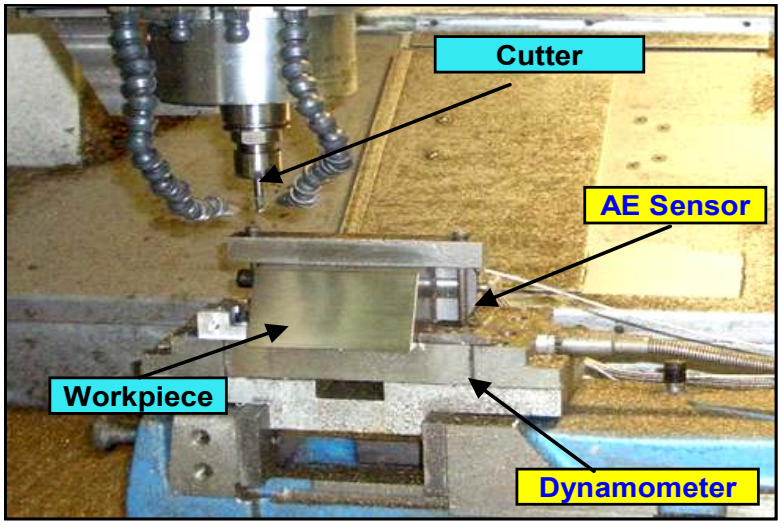


Figure 4.10: Experiment setup

Table 4.8

Experimental components

Components
Röders TEC vertical milling machine
6mm ball nose tungsten carbide cutters
Titanium Ti6Al4V workpiece
8152B211 Piezotron® AE sensor (Kistler)
Kistler 5127B11 multichannel charge amplifier
NI-DAQ PCI 6250 M series
LECIA MZ12.5
Computer

In the experiment, measured force and AE sensor data were captured under the following machine settings: spindle speed 1000 rpm, feed rate 200 mm/min, depth of cut 1 mm, and insert number 2. During the data measurement phase, the tool was removed from the chuck and the tool wear was measured by hand periodically. Specifically, the flank wear of each individual tooth of the cutting tool were measured with an LECIA MZ12.5 high performance stereomicroscope. Cutting tool experiments were carried out from a new cutting tool to a tool with big tool wear. 2500 cutting lines were conducted in the experiment for the whole tool life and each line is represented using one data point and these tool wear points are

constructed as the output vector. Force and AE signals are input signals. In order to match the tool wear output, the force and AE signals are extracted under each rotation and averaged for each cutting line.

An example of the cutting force signal in three axes (F_x, F_y, F_z), is shown in Figure 4.11. The AE signal is shown in Figure 4.12.

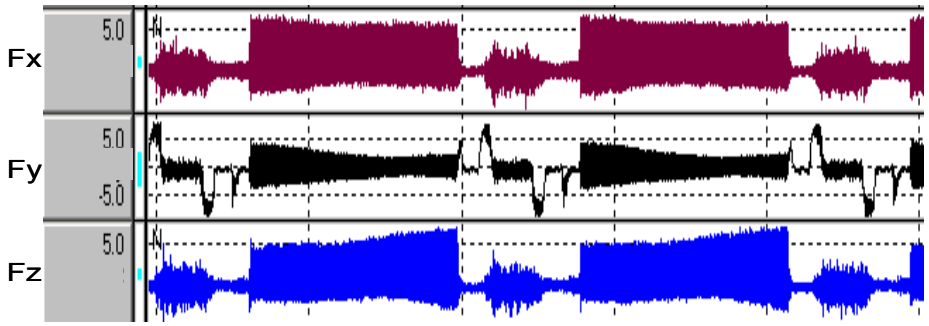


Figure 4.11: Three axes cutting force signal (Fx, Fy, Fz)

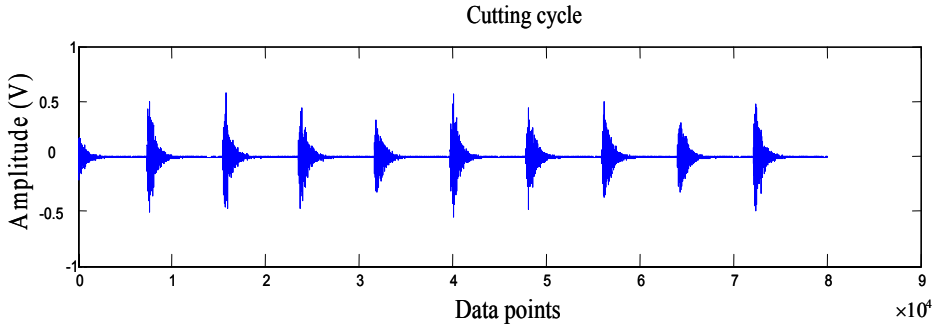


Figure 4.12: Unprocessed AE signal during cutting process

Our experiment was carried out with a new cutter until tool wear became too large, *i.e.*, broken down and damaged. The evolution of measured tool wear during the experimental stage is shown in Figure 4.13.

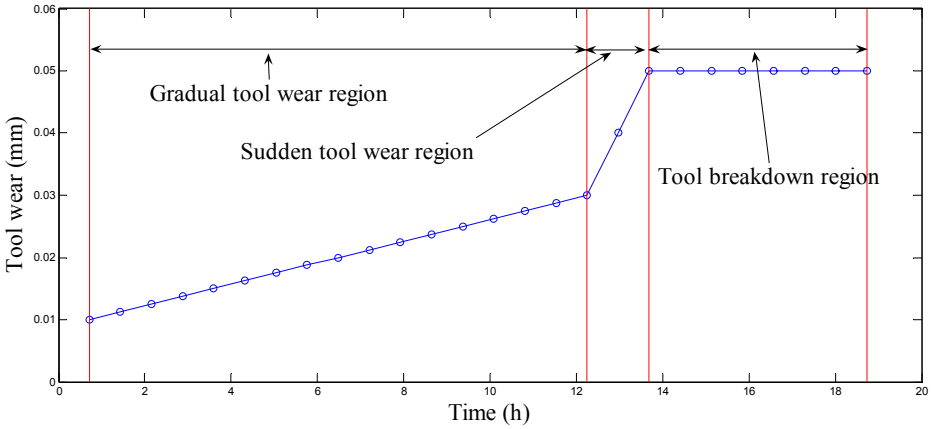


Figure 4.13: Stages in evolution of tool wear

The collected data were stored in a hard disk drive for signal conditioning and processing. Feature subset identification and tool wear prediction can then be performed to determine the condition of the cutting tool.

AE signals are essentially sound waves generated in solid media, affected by characteristics of the source, path taken from the source to the sensor, sensor's characteristics, and measuring system, *etc*. Generally, the AE signals are intricate and sensitive, and characterization of the source using AE signals is difficult. As such, information is extracted using waveform parameter measurements [61] and artificial intelligence approaches. Advanced signal processing is also necessary for extracting the relevant features from the raw signal. In our study, the signals were firstly truncated to the desired lengths to facilitate signal processing. Next, noise in the raw signals was filtered by the wavelet de-noising method, which consists of decomposition, identification of threshold detail coefficient, and reconstruction. The processed noise-free signal can then reconstructed via the inverse wavelet transform of the threshold wavelet coefficients. A third level decomposition was also applied to the signals.

4.2.2 Extraction of features

The characteristics of the AE signal that can be used for TCM have been studied by researchers in various aspects, *e.*g., spectral, statistical, and time series analysis, *etc*., to analyze tool wear. As such, various computed features using AE signals have been shown to be effective for monitoring milling processes [157–

158, 140]. Diniz *et al.* in [157] extracted eight statistics from the AE, namely skewness, kurtosis, crest-factor, peak, peak to peak, mean of RMS, mean, and standard deviation. In Sun *et al.* [140], the five parameters of mean of band power, standard deviation of band power, delta (first difference), and the absolute deviation were used to monitor the tool conditions. Ravindra *et al.* in [152] used the rise time, ring down count to detect tool breakages. Kakade *et al.* in [158] used area under curve and duration to monitor the tool conditions. As such, these sixteen features from the methodologies mentioned above are summarized in Table 4.9 and form the scope of the feature subset selection.

Table 4.9
AE features and nomenclature

No	Feature	Notation	References
1	Skewness	AEskew	[148]
2	Kurtosis	AEkts	[148]
3	Crest-factor	AEcrest	[148]
4	Peak	AEmax	[129]
5	Total amplitude of AE	AEa	[148]
6	Mean of RMS	AErms	[148]
7	Average AE	AEca	[148]
8	Standard deviation	AEstd	[148]
9	Mean of band power	AEmb	[129]
10	Standard deviation of band power	AEstdb	[129]
11	Delta (change in signal)	AEdlt	[129]
12	Absolute deviation	AEad	[129]
13	Ring down count	AEc	[129]
14	Rise time	AEr	[144]
15	Area under curve	AEca	[149]
16	Duration	AEd	[149]

Force signals from dynamometer have been shown to produce good prediction of tool wear. The sixteen features derived from force signals in [139] have been shown in Section 4.1, Table 4.2. They are extracted for comparisons with the results obtained via AE sensing. The details of the comparison results are discussed in Section 4.2.5.

4.2.3 Selection of dominant features

With the AE features and force features obtained, the DFI methodology described in [134] and Section 3.3 are used to obtain a feature subset. The five following steps are involved in the DFI to obtain the dominant features which are minimum in overall approximation and k-means clustering error:

1. Data acquisition. Collect data on time interval [0, *m*] from the cutting tools' sensors and compute *n* features using digital signal processing techniques. Pack *n* time series, each of length *m,* as columns into matrix $X \in R^{m \times n}$. In this experiment, 2500 cutting lines were conducted in the experiment and each line was represented using one data point. Among these 2500 data points, half of them are used for the training, so $m = 1250$ and $n = 16$ in this experiment.
2. Initialization. Detrend the data in X by subtracting the mean (average across each dimension) of each of the dimensions to normalize the data set to zero-mean. This step results in a data set whose mean is zero.
3. Choose number of principal components and clusters. Perform SVD on inner product matrix $X^T X$ to obtain Σ_1 and V_1^T. Select the desired number *q* of principal components in $\sigma_i v_i^T$ and number of clusters *p*. In this experiment, $q = 3 \sim 4$, while $p = 4 \sim 9$ in the results Tables 4.12-4.17.
4. Clustering. Use the k-means algorithm for clustering to find the centroids c_i of each cluster.
5. Subset determination. Select the vector w_i "nearest" to the centroid of each cluster as its cluster leader $\bar{w}_i$, and corresponding f_i as the dominant feature. Combine the *p* dominant features to form the reduced feature space.

4.2.4 Prediction of tool wear using dominant features

Define $X = [f_1 \quad f_2 \quad \cdots \quad f_n] \in R^{m \times n}$ as the matrix of features computed from the signal measured by the AE sensor. Each feature f_i is a time signal of length *m* samples. Select *p* dominant features using DFI as in the previous Section. This selects *p* columns in *X* as a reduced matrix $\hat{\hat{X}} \in R^{m \times p}$.

It is now desired to predict the real tool wear $y(k)$ at time *k* in terms of the identified *p* dominant features using an ARMAX model with Extended Least Squares (ELS). The general Auto-Regressive Moving Average with eXogenous (ARMAX) model has the following structure [20]

$$A(z^{-1})\hat{y}(k) = B(z^{-1})u(k - n_u) + C(z^{-1})\varepsilon(k) \tag{4.12}$$

where $\hat{y}(k)$ is the estimated tool wear, $u(k) = \begin{bmatrix} u_1 & u_2 & \cdots & u_p \end{bmatrix}^T$ is a column vector of measurements from the p dominant features (or a row vector in $\hat{\hat{X}}$ at time k), and n_u is the input delay. z^{-1} is the unit backward shift operator and $\varepsilon(k)$ is the estimation error. A, B, and C are polynomials of in ascending powers of delays as

$$
\begin{aligned}
A(z^{-1}) &= 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_{n_a} z^{-n_a} \\
B(z^{-1}) &= \begin{bmatrix} b_{11} + b_{12} z^{-1} + b_{13} z^{-2} + \cdots + b_{1n_b} z^{-n_{b1}+1} \\ b_{21} + b_{22} z^{-1} + b_{23} z^{-2} + \cdots + b_{2n_b} z^{-n_{b2}+1} \\ \vdots \\ b_{p1} + b_{p2} z^{-1} + b_{p3} z^{-2} + \cdots + b_{pn_b} z^{-n_{bp}+1} \end{bmatrix}^T \\
C(z^{-1}) &= 1 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{n_c} z^{-n_c}
\end{aligned}
\tag{4.13}
$$

where n_a, n_{pb}, and n_c are the orders of A, B, and C, respectively.

Several assumptions are made on the ARMAX model to be valid for our application as follows:

(A1) A^{z-1} is assumed to be Hurwitz, i.e., it has zeros strictly inside the unit circle;

(A2) Multi-input $u(k)$ is assumed to be stationary and persistently exciting;

(A3) Evolution of coefficients of A, B, and C is considered as slow time varying when it is compared to the sampling time of the system;

(A4) $a(k)$ is uncorrelated with $u(k)$ in a statistical sense; and

(A5) The orders n_a, n_{pb}, and n_c are known a priori.

If there is no input delay, i.e., $n_u = 0$, the ARMAX model from (4.12) can be written in time series as

$$
\hat{y}(k) = \varphi_k^T \theta(k) + \varepsilon(k) \tag{4.14}
$$

where

$$
\begin{aligned}
\varphi_k^T = \big[& -\hat{y}(k-1) \quad \cdots \quad -\hat{y}(k-n_a) \quad u_1(k) \\
& \ldots \quad u_1(k-n_{b1}+1) \quad u_2(k) \quad \ldots \\
& u_2(k-n_{b2}+1) \quad \cdots \quad u_p(k) \quad \cdots \\
& u_p(k-n_{bp}+1) \quad \varepsilon(k-1) \quad \cdots \quad \varepsilon(k-n_c) \big]
\end{aligned}
$$

$$
\begin{aligned}
\theta^T(k) = \big[& a_1 \quad \cdots \quad a_{n_a} \quad b_{11} \\
& \cdots \quad b_{1n_b} \quad b_{21} \quad \cdots \\
& b_{2n_b} \quad \cdots \quad b_{p1} \quad \cdots \\
& b_{pn_b} \quad c_1 \quad \cdots \quad c_{n_c} \big]
\end{aligned}
\tag{4.15}
$$

The term $\varepsilon(\bullet)$ is unknown, but may be approximated by using the prediction errors $\hat{e}(k)$ where

$$\hat{e}(k) = \hat{y}(k) - \hat{\varphi}_k^T \hat{\theta}(k-1) \tag{4.16}$$

and all the terms on the right hand side of (4.16) consist of past realizable signals at any time instant *k*. The standard Least Squares (LS) problem can now be reformulated into the Extended LS (ELS) by defining

$$\begin{aligned}
\hat{\varphi}_k^T &= \left[-\hat{y}(k-1) \quad \cdots \quad -\hat{y}(k-n_a) \quad u_1(k) \right. \\
&\quad \ldots \quad u_1(k-n_{b1}+1) \quad u_2(k) \quad \ldots \\
&\quad u_2(k-n_{b2}+1) \quad \cdots \quad u_p(k) \quad \cdots \\
&\quad \left. u_p(k-n_{bp}+1) \quad \hat{e}(k-1)\right] \\
\hat{\theta}^T(k) &= \left[\hat{a}_1 \quad \cdots \quad \hat{a}_{n_a} \quad \hat{b}_{11} \right. \\
&\quad \cdots \quad \hat{b}_{1n_b} \quad \hat{b}_{21} \quad \cdots \\
&\quad \hat{b}_{2n_b} \quad \cdots \quad \hat{b}_{p1} \quad \cdots \\
&\quad \left. \hat{b}_{pn_b} \quad \hat{c}_1\right]
\end{aligned} \tag{4.17}$$

As such, collecting past measurable signals as

$$Y_k = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}, \hat{\Phi}_k = \begin{bmatrix} \hat{\varphi}_1^T \\ \hat{\varphi}_2^T \\ \vdots \\ \hat{\varphi}_k^T \end{bmatrix}, \tag{4.18}$$

the unbiased estimate of $\hat{\theta}(k)$ is given by

$$\begin{aligned}
\hat{\theta}(k) &= \arg\min_{\theta} \left\| \left[Y_k - \hat{\Phi}_k^T \hat{\theta}(k-1)\right]^T \left[Y_k - \hat{\Phi}_k^T \hat{\theta}(k-1)\right] \right\| \\
&= \left(\hat{\Phi}_k^T \hat{\Phi}_k\right)^{-1} \hat{\Phi}_k^T Y_k
\end{aligned} \tag{4.19}$$

via the standard batch solution. To compute $\hat{\theta}(k)$ using efficient on-line recursive means, one may use Extended Least Squares (ELS) governed by the following equations

$$\begin{aligned}
\hat{\theta}(k) &= \hat{\theta}(k-1) + K(k)\hat{e}(k) \\
K(k) &= P(k-1)\hat{\varphi}_k [I + \hat{\varphi}_k^T P(k-1)\hat{\varphi}_k]^{-1} \\
P(k) &= [I - K(k)\hat{\varphi}_k^T] P(k-1)
\end{aligned} \tag{4.20}$$

where $P(k) = (\hat{\Phi}_k{}^T \hat{\Phi}_k)^{-1}$. The combination of the DFI algorithm with the usage of a dynamic ARMAX model using ELS is called the new DFI methodology.

The pseudocode for tool wear prediction of dynamic ARMAX model using ELS is presented in the following:

Algorithms: Tool wear prediction using ARMAX

INPUT: measured dominant features $\phi \in R^{m \times q}$

OUTPUT: tool wear $y_k \in R^{m \times q}$

// Recursive least squares to calculate estimate parameters $\theta(k)$

1: WHILE $i < m$ DO

2: Calculate estimate parameters $\theta(k)$ using Equation (4.20)

3: ENDWHILE

// Calculate tool weal y_k

1: WHILE $i < m$ DO

2: Calculate tool weal y_k using Equation (4.14)

3: END WHILE

4.2.5 Experimental results

In the experiment, 2500 time points of measured force and AE sensor data were captured under the following machine settings: spindle speed 1000 rpm, feed rate 200 mm/min, depth of cut 1 mm, and insert number 2. During the data measurement phase, the tool was removed from the chuck and the tool wear was measured by hand periodically. Specifically, the flank wear of each individual tooth of the cutting tool was measured with an LECIA MZ12.5 high performance stereomicroscope. This yields the baseline actual tool wear plot shown in Figures 4.15-4.18.

The measured force and AE data were detrended, *i.e.*, the mean value was subtracted, and normalized. Based on these data, the sixteen features in Tables 4.2 and 4.9 were computed as functions of time. This yields two sets of n=16 feature vectors (each of which is a function of time and has m=1250 data points) and X in (4.3) that has n=16 columns and m=1250 rows. Next, the PFA and DFI of feature selection were performed. Since $m = 1250$ and $n = 16$, the computational complexity for PFA in this process is

$O(m^2 n) + O(m^3) = O(2.5 \times 10^7) + O(1.953125 \times 10^9)$, while for DFI method, the complexity is $O(mn^2) + O(n^3) = O(1.0125 \times 10^5) + O(4.096 \times 10^3)$. DFI is numerically efficient and reduces the complexity of feature selection greatly. The selected dominant features were then used as the input to build non-dynamic and dynamic tool wear prediction models. The prediction accuracies with different approaches were analyzed and the results are shown in this Section.

The resulting singular values are shown in Table 4.10 and plotted in Figure 4.14. It is seen that fourth and subsequent singular values for both force and AE are quite small compared to the first three singular values. Therefore, it appears that retaining the first three singular values would be sufficient to capture the relevant trend in tool wear prediction for this experiment. However, it is not clear from this observation that which of the 16 features are the main ones useful for tool wear prediction at this point in time. The DFI method is therefore used to select the most important features.

Table 4.10
Principal components and singular values

No	Singular Values (Force)	Singular Values (AE)
1	3.874509	11.08428
2	0.260766	0.228871
3	0.011573	0.020575
4	6.69E-04	5.56E-04
5	7.42E-06	5.37E-05
6	5.88E-07	1.31E-06
7	8.23E-08	2.21E-10
8	7.53E-09	1.27E-11
9	4.02E-09	3.86E-14
10	1.40E-09	4.48E-15
11	1.25E-10	1.66E-15
12	6.30E-12	9.01E-17
13	1.09E-12	6.08E-18
14	8.32E-14	6.91E-19
15	1.82E-16	2.13E-19
16	1.39E-18	1.40E-19

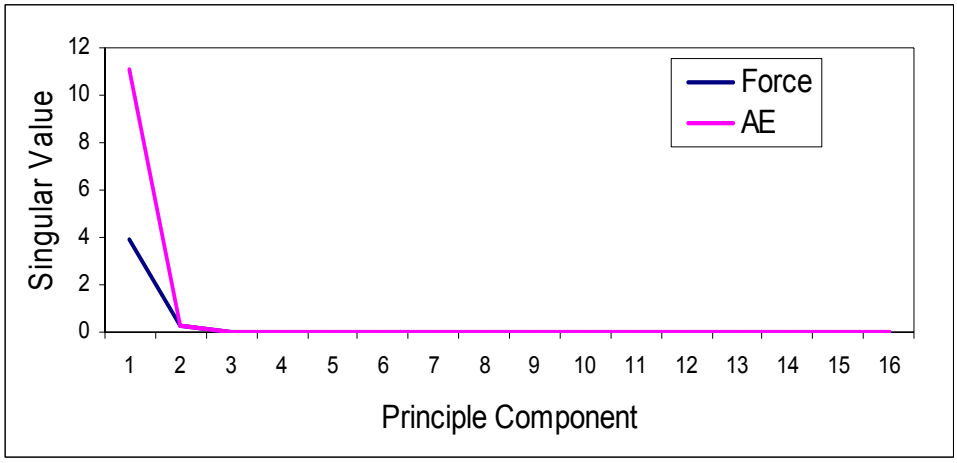


Figure 4.14: Plot of principal components versus singular values

4.2.5.1 Comparison of standard non-dynamic prediction models with dynamic ARMAX model

In current literature, tool wear is usually predicted with standard non-dynamic linear prediction models [152, 153]. The simplest and most commonly used is the linear Multiple Regression Model (MRM) of the form

$$\hat{y}(k) = \phi_k^T \hat{\hat{\theta}}(k) \tag{4.21}$$

where $\phi_k{}^T \in R^p$ is the vector containing the measured dominant features at time $k \le m$ (or a row vector in $\hat{\hat{X}}$ at time k) and $\hat{\hat{\theta}}(k)$ is the vector of unknown coefficients to be regressed for time series forecast of tool wear. For non-intrusive and online prediction, the ELS algorithm depicted earlier in (4.20) was used for our experiment.

MRMs using RLS were identified to predict the baseline measured tool wear using all of the original sixteen force and AE features as described in the previous Section. A Mean Relative Error (MRE) of 7.15% for all sixteen force features and 10.53% for all sixteen AE features are observed as shown in Figure 4.15. A better prediction performance is obtained using the force MRM when compared to the AE MRM. The MRE of the AE MRM is above 10%, and is unacceptable in TCM applications.

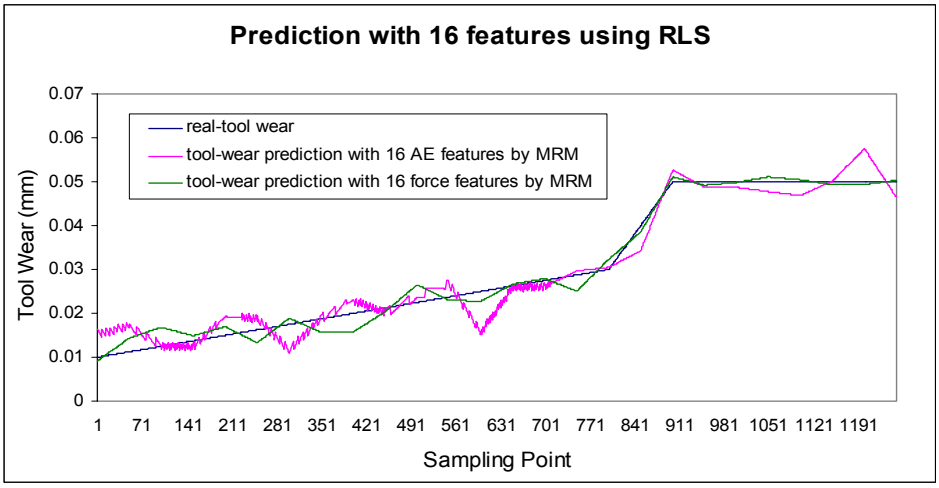


Figure 4.15: MRMs using all sixteen forces and AE features

Using the ELS techniques presented in the previous Section, dynamic ARMAX models in (4.12) were identified to predict the baseline measured tool wear using all sixteen original force and AE features. The actual measured tool wear and the predicted tool wear using ARMAX models as functions of time are shown in Figure 4.16. Clearly, the predictions are good. A Mean Relative Error (MRE) of 1.13% for 16 force features and 3.22% for AE are observed in Figure 4.16, and represent the best possible prediction of tool wear using these two sets of sixteen features. Both force and AE with dynamic ARMAX models using ELS perform better than those using the same features with MRM using RLS. It is also observed that using AE with ARMAX and ELS obtained better performance than using force with MRM and RLS. This justifies the proposed usage of AE sensors in TCM, as effective tool wear prediction can be realized with improvement in system identification techniques.

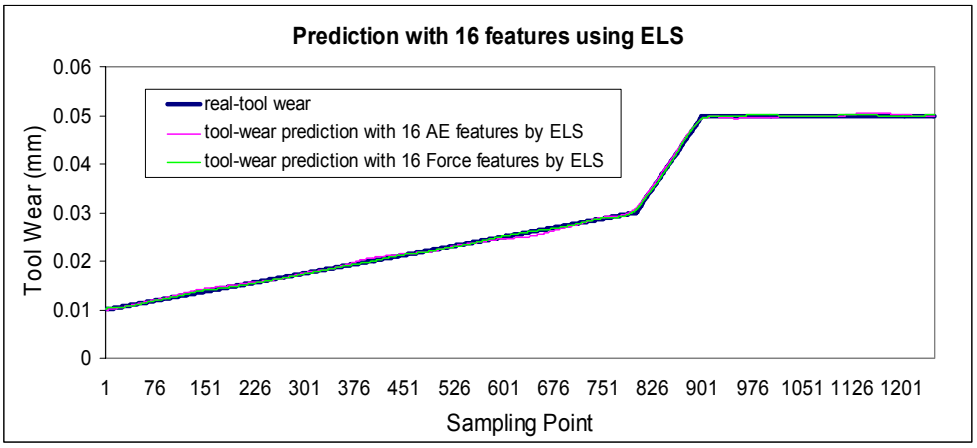


Figure 4.16: ARMAX models using all sixteen forces and AE features

4.2.5.2 Comparison of proposed ARMAX Model using ELS with DFI, MRM using RLS with DFI, and MRM using RLS with principal feature analysis (PFA)

It is desired to use fewer features to predict the tool wear. Next, the DFI method in Section 3.3 and PFA method in current literature [11-13] are used as the feature selection techniques to reduce the number of features in the feature space and select the important features. The selected features obtained using both DFI and PFA methods are then used as inputs, and the ARMAX models using ELS and MRMs using RLS techniques to build the prediction models to forecast tool wear.

To showcase the effectiveness of our proposed methodology using AE sensors and ARMAX models with ELS, only AE features are used in this comparison for simplicity but without loss of generality. Here, four dominant features (namely kurtosis, count, area, and duration) were selected using DFI, and four dominant features (namely kurtosis, delta, count, and duration) were selected by using PFA. Using ELS and RLS to compute the ARMAX model and MRM, respectively, mathematical models were built to predict tool wear as shown in Figure 4.17. The actual measured tool wear, tool wear prediction using ARMAX model with ELS and DFI, tool wear prediction using MRM with RLS and DFI, and tool wear prediction using MRM with RLS and PFA, with all predictions using the best four features with DFI and PFA, are shown in Figure 4.17. The MREs of the different models are shown in Table 4.11.

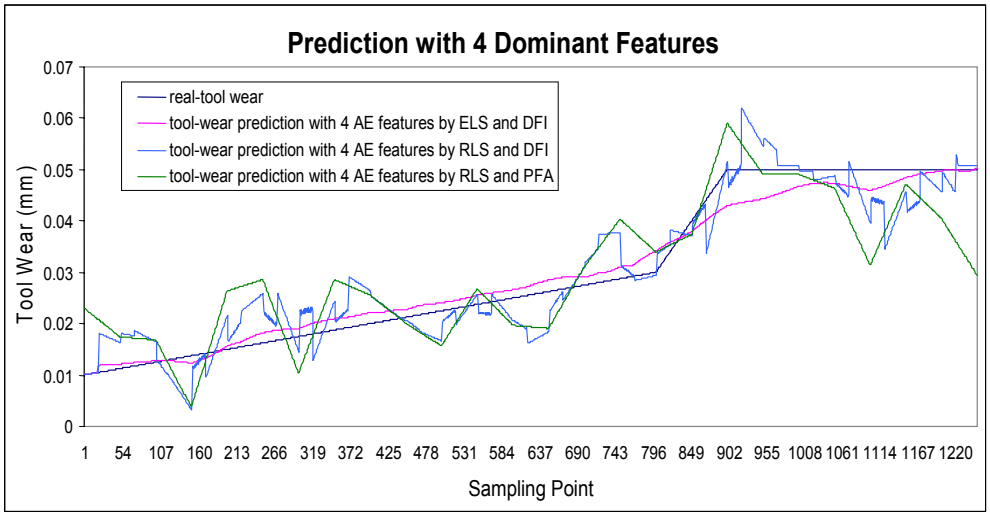


Figure 4.17: Examples of tool wear prediction using four dominant features and three principal components

Table 4.11

Comparison of model accuracies

Model type	MRE (%)
ELS with DFI	7.19
RLS with DFI	17.52
RLS with PFA	24.02

4.2.5.3 Effects of different numbers of retained singular values q and features selected p

Also, comparison studies were carried out using increasing numbers q retained singular values and p features using DFI and PFA. Comparison results using three retained singular values with DFI using ARMAX models and ELS, DFI using MRMs with RLS, and PFA using MRMs with RLS are shown in Tables 4.12, 4.13, and 4.14, respectively. Similarly, the results using four retained singular values with DFI using ARMAX models and ELS, DFI using MRMs with RLS, and PFA using MRMs with RLS are shown in Tables 4.15, 4.16 and 4.17, respectively.

In these tables, the first column denotes the number of features selected using DFI or PFA to be used for tool wear prediction, and the second column displays the selected features. The MSE and MRE obtained using the corresponding prediction models for tool wear forecast are displayed in columns three and four.

Table 4.12

Results of DFI and ARMAX model with ELS using three retained singular values

No. of selected features	**Features selected**	**MSE (mm^2)**	**MRE (%)**
4	AEkts, AEc, AEca, AEd	0.26	7.19
5	AEskew AEkts, AEr , AEca, AEd	0.25	6.25
6	AEskew AEkts, AEc , AEr , AEca, AEd	0.24	5.11
7	AEcrest, AEkts, AEstd , AEc , AEr , AEca, AEd	0.17	3.77
8	AEcrest, AEkts, AEca, AEstdb, AEc , AEr , AEca, AEd	0.14	3.42
9	AEskew , AEcrest , AEkts, AEa , AEstdb, AEc , AEr , AEca, AEd	0.12	3.49

Table 4.13

Results of DFI and MRM with RLS using three retained singular values

No. of selected features	**Features selected**	**MSE (mm2)**	**MRE (%)**
4	AEkts, AEc, AEca, AEd	0.72	24.02
5	AEskew AEkts, AEr , AEca, AEd	0.52	17.34
6	AEskew AEkts, AEc , AEr , AEca, AEd	0.44	15.67
7	AEcrest, AEkts, AEstd , AEc , AEr , AEca, AEd	0.43	15.28
8	AEcrest, AEkts, AEca, AEstdb, AEc , AEr , AEca, AEd	0.40	13.03
9	AEskew , AEcrest , AEkts, AEa , AEstdb, AEc , AEr , AEca, AEd	0.37	12.67

Table 4.14

Results of PFA Method using three retained singular values

No. of selected features	Features selected	MSE (mm2)	MRE (%)
4	AEkts, AEdlt, AEca, AEd	1.262	25.12
5	AEd, Ekts, AEskew, AEca, AEr	1.202	19.11
6	AEd, AEca, AEcrest, AEr, AEstd, AEkts	1.111	15.64
7	AEdlt, AEr, AEcrest, AEca, AEkts, AEd, AEca	1.111	15.33
8	AEa, AEr, AEkts, AEdlt, AEskew, AEca, AEcrest, AEd	0.946	13.03
9	AEskew, AEad, AEr, AEkts, AEdlt, AEa, AEca, AEd, AEcrest	0.946	13.03

Table 4.15

Results of DFI and ELS method using four retained singular values

No. of selected features	Features selected	MSE (mm2)	MRE (%)
5	AEskew AEkts, AEr , AEca, AEd	0.25	6.25
6	AEskew AEkts, AEc , AEr , AEca, AEd	0.24	5.11
7	AEcrest, AEkts, AEstd , AEc , AEr , AEca, AEd	0.17	3.77
8	AEcrest, AEkts, AEca, AEstdb, AEc , AEr , AEca, AEd	0.14	3.42
9	AEskew , AEcrest , AEkts, AEa , AEstdb, AEc , AEr , AEca, AEd	0.12	3.49

Table 4.16

Results of DFI and RLS method using four retained singular values

No. of selected features	Features selected	MSE (mm2)	MRE (%)
5	AEskew AEkts, AEr , AEca, AEd	0.52	17.34
6	AEskew AEkts, AEc , AEr , AEca, AEd	0.44	15.67
7	AEcrest, AEkts, AEstd , AEc , AEr , AEca, AEd	0.43	15.28
8	AEcrest, AEkts, AEca, AEstdb, AEc , AEr , AEca, AEd	0.40	13.03
9	AEskew, AEcrest , AEkts, AEa , AEstdb, AEc , AEr , AEca, AEd	0.37	12.67

Table 4.17

Results of PFA and RLS method using four retained singular values

No. of selected features	Features selected	MSE (mm2)	MRE (%)
5	AEskew, AEkts, AEr, AEca, AEd	0.54	19.15
6	AEcrest, AEkts, AEstd, AEr, AEca, Aed	0.52	17.34
7	AEcrest, AEkts, AEa, AEad, AEr, AEca, Aed	0.44	15.61
8	AEcrest, AEkts, AEa, AEad, AEc , AEr, AEca, AEd	0.44	15.27
9	AEskew, AEcrest, AEkts, AErms, AEstdb, AEdlt, AEr, AEca, AEd	0.40	12.90

The observations from these tables are as follows:

1. The dominant features are consistent when using the DFI methodology, even when different numbers of features are used. However, the features are slightly different using the PFA method when different numbers of *p* are used, especially when the numbers of features chosen are four and five.
2. Besides the consistency in selection of features, the features chosen are also different using DFI and PFA. Using the same MRM with RLS, DFI gives a smaller MSE and MRE when compared to that using PFA, and hence provides a better accuracy than PFA in tool wear prediction. This is especially obvious when the numbers of features chosen are four and five. When using a small number of selected features, the computational time for features processing is also greatly reduced. Our experiments also found that the computational time required to build MRMs using the original sixteen features was about five times longer, when compared to that using four selected features only.
3. When using different system identification techniques with the same feature sets selected by DFI, ARMAX models, using ELS always provide better predictions when compared to MRMs using RLS.

As such with the improvement in feature selection and modelling techniques, the AE sensors can replace force sensors in real-time tool wear prediction and tool condition monitoring.

4.2.5.4 Comparison of tool wear prediction Using AE measurements and force measurements

Next, tool wear prediction is compared when using force and AE sensors using four dominant features selected by DFI only. An ARMAX model with ELS is built with the four selected AE dominant features, and an MRM model is built with the four selected force dominant features. The actual measured tool wear and prediction results from the constructed models are shown in Figure 4.18. An MRE of 7.19% when using the four AE features and 13.56% when using the four force features are observed. The ARMAX model with four AE dominant features performs better than MRM model with four force dominant features in Section 4.1, and further justifies the use of AE sensors with DFI and dynamic ARMAX models using ELS for effective TCM.

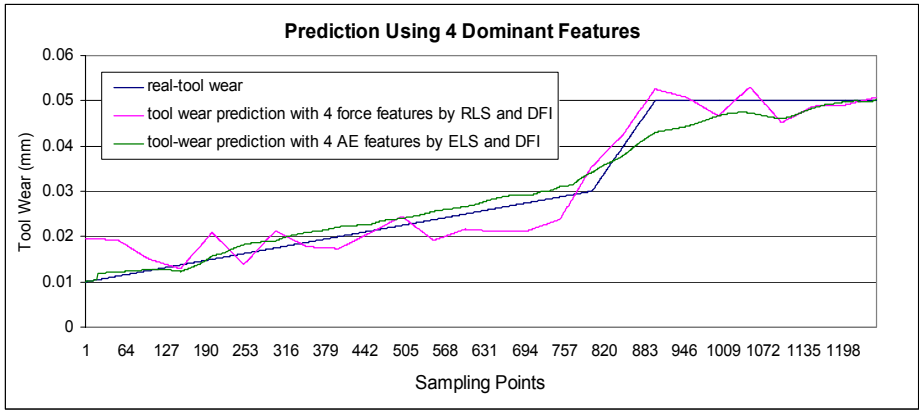


Figure 4.18: Examples of prediction using four dominant features and three principal components

4.3 Summary

In this chapter, the Dominant Feature Identification (DFI) methodology using Singular Value Decomposition (SVD) of collected tool wear data is proposed for prediction of time series of deterioration of an industrial cutting tool. The DFI uses SVD which operates on the inner product matrix at a lower dimension, and reduces the Least Squares Error (LSE) induced when selecting the principal components and clustering to identify the dominant features. Two experiments are conducted. Our first experiment conducted using force sensor and results show Mean Squares Errors (MSEs) values from 0.946 to 1.262 and Mean Relative Error (MRE) values from 8.86% to 11.61% between the actual measured tool wear to that predicted

from the Multiple Regression Models (MRMs) using RLS.

In the second experiment, the Dominant Feature Identification (DFI) methodology using Singular Value Decomposition (SVD) of collected tool wear data from Acoustic Emission (AE) and force sensors is used to construct models for prediction of times series of deterioration of an industrial cutting tool. Our experimental results using AE sensors show significant reduction in both Mean Squares Errors (MSEs) and Mean Relative Error (MRE) when an ARMAX model with Extended Least Squares (ELS) technique is employed, which is promising in replacing force sensors and conventional non-dynamic models for effective online Tool Condition Monitoring (TCM) and tool wear prediction.

CHAPTER 5 DOMINANT FEATURE IDENTIFICATION FOR INDUSTRIAL FAULT DETECTION AND ISOLATION APPLICATIONS

In an era of intensive competition, the new challenges faced by industrial manufacturing processes include maximizing productivity, ensuring high product quality, and reducing the production time while minimizing the production cost simultaneously. One of the causes of delay in networked manufacturing processes is machine down time or failure of the machining tools when faults arise. Much manpower is also required to identify, locate, and rectify the faults for large-scale interconnected processes which increase operating expenditure.

Rotary machines—consisting of numerous components such as electric motors, shafts, bearings, gears, and belt drives, etc., are commonly used in many industries in various scales, ranging from cooling fans to vehicle engines and even power plant generators in ascending order of scale. A fault might occur at each of these components, which include stator winding faults [154], rotor [155], misaligned shafts [156], loose belts [157], cracked bearings [158–159] and gears, etc. These faults are by no means exhaustive, which increase exponentially with the complexity of rotary machines. As such, vast amounts of time and effort have been invested to develop rotary machines with higher efficiencies and improved robustness to prevent down time and loss of productivity. However, mechanical faults arise due to wear and tear from prolonged operations over time, causing the machines to operate at lower efficiencies with undesirable effects such as excessive vibrations and noises. In more severe instances, the machines might even experience critical failure and breakdown, even posing possible work hazards to their operators.

Numerous machine parameters have been used to form the basis of machine health assessment, including motor speeds, input currents, transmitted torques, and vibration parameters such as accelerations, velocities, and displacements, etc. Each of these parameters might produce characteristic features that can be used for fault detection at the data processing phase. Also, many methods like statistical moment computation and spectral analysis of various machine parameters [2],

Hilbert-Huang Transform (HHT) [3], minimum variance cepstrum, amplitude modulation and envelope analysis [4], Fast Fourier Transform (FFT), Statistical Computations, Wavelets [5–6], etc., have been used for feature calculation and extraction. The common types of faults include imbalance fault, loose belt fault, bearing faults, and variations in resonant frequencies.

Much research has been conducted to improve the detection and identification of rotary machine faults to ensure diagnosis and prognosis can be taken in time before a critical machine failure. The components that are causal to the faults are also isolated, eliminating the need of unnecessary replacements of other functional components and human intervention.

In this chapter, we extend the earlier proposed Dominant Feature Identification (DFI) method for industrial Fault Detection and Isolation (FDI) using Neural Networks (NNs) to classify faults based on a reduced set of features. We show how to apply DFI for fault detection by two methods that seek to identify the important features in a given set of faults. Then based on the determined reduced feature set, an NN is used for online fault classification and multiple fault prediction in such a proposed two-stage framework. A fault simulator machine which could simulate different types of machine faults is used for the study of fault detection and isolation. Our experimental results on a fault simulator machine reduce the number of features from 120 to 13 and sensors from 8 to 4 at an accuracy of 99.4%.

The rest of the chapter is organized as follows. The details of our test bed and experimental setup are described in Section 5.1. Section 5.2 presents the proposed DFI and neutral network for fault detection and isolation. Section 5.3 discusses the experimental results. Our conclusion and future work directions are presented in Section 5.4.

5.1 Test Bed

A fault simulator machine which could simulate different types of machine faults is used for the study of fault detection and isolation. The machine setup is shown in Figure 5.1. The fault simulator machine by SpectraQupe test is designed to study the signature of common machinery faults, such as unbalance, alignment,

resonance, bearing and belt drive, etc., as depicted in [162]. Eight sensors are installed for fault detection: namely three current clamp meters to measure the motor line current, one Futek torque sensor placed on the shaft between the motor and first bearing to measure the torque generated, and four Kistler vibration sensors are placed near the two bearings to measure the vibration levels generated. The system overview is shown in Figure 5.1. It should be noted that the positions of the clamp meters are hidden under the motor and are not shown in Figure 5.1. Each sensor has a corresponding channel on the data acquisition module.

The names of the channels and the corresponding sensors attached are shown in Table 5.1.

Table 5.1
Nomenclature and corresponding sensors

Name	Sensor
AI0	Clamp meter C1, white cable of motor current input (neutral)
AI1	Clamp meter C2, black cable of motor current input (line 1)
AI2	Clamp meter C3, red cable of motor current input (line 2)
AI3	Futek high speed encoder, torque signals
AI4	Sensor Ax, acceleration signals
AI5	Sensor Ay, acceleration signals
AI6	Sensor A2x, acceleration signals
AI7	Sensor A2y, acceleration signals
PFI4	Banner mini beam encoder, speed signals
PFI8	Futek high speed encoder, speed signals

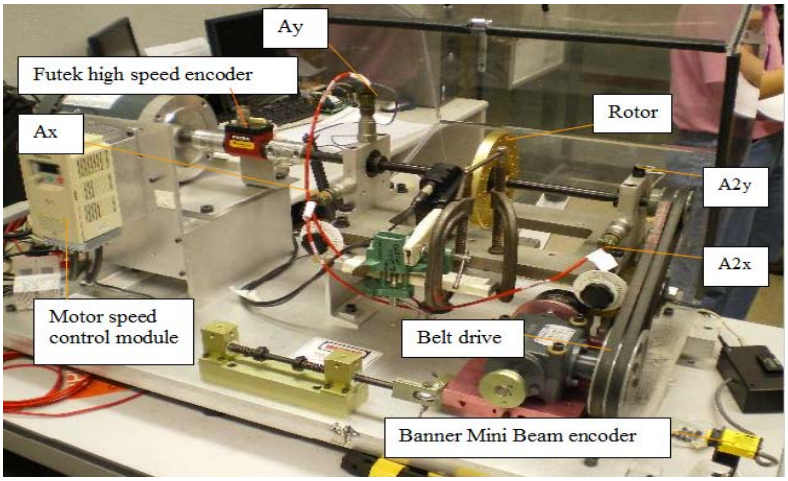


Figure 5.1: Fault simulator machine and the eight sensors.

5.1.1 Experimental setup

In our experiment, the sensitivity of the accelerometer is 100mV/g and that of the clamp meter is 100mV/A. The zoomed-in view of the four vibration sensors, Ax, Ay, A2x, and A2y, is shown in Figure 5.2. The current clamp meters and Futek torque sensors are shown in Figure 5.3.

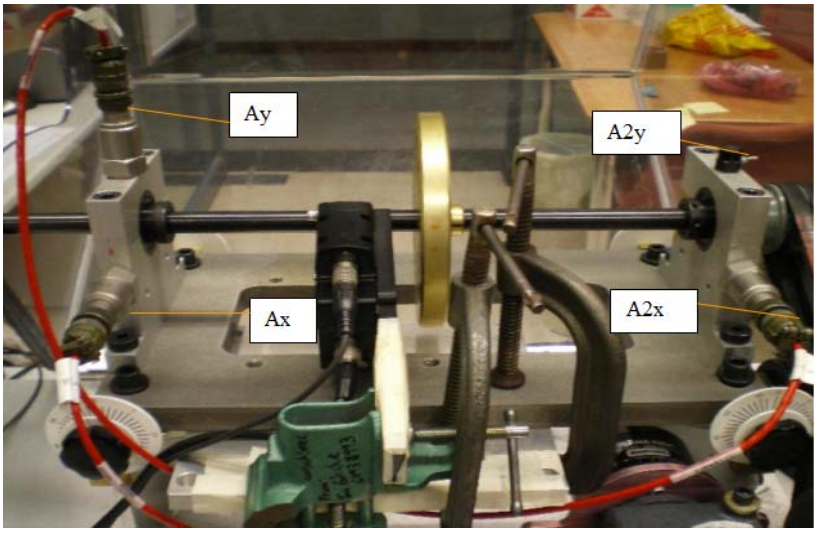


Figure 5.2: Zoomed-in view of the four vibration sensors Ax, Ay, A2x, A2y, with the corresponding data cables attached.

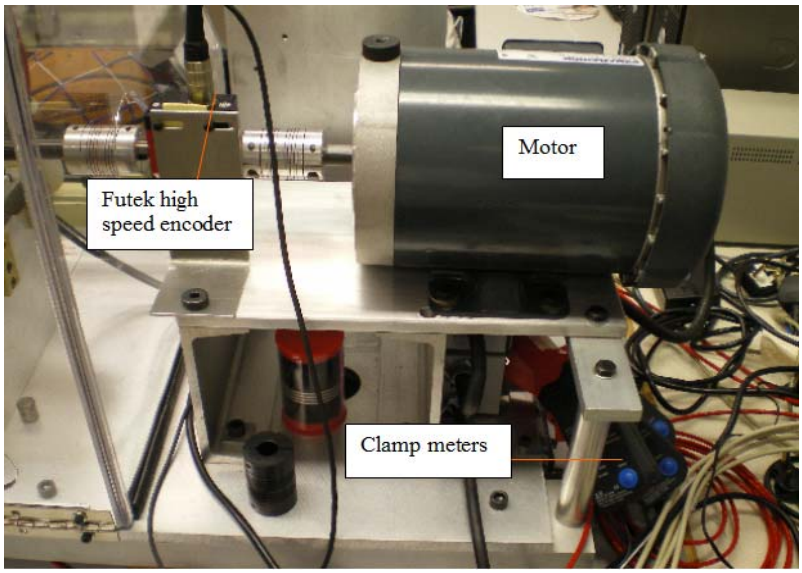


Figure 5.3: Current clamp meters and Futek torque sensor.

5.1.2 Computation of corresponding features

The procedure of computation of corresponding features is detailed here. Firstly, signal processing is performed using envelope method in [4] to extract the demodulated signals. Next, statistical features calculation from time domain are used for current, torque, and vibration signal analysis [163–166]. The statistical features that are used in this chapter are summarized as:

$$Min = \min\left(\left\{S_i\right\}_{i \in [a,b]}\right) \tag{5.1}$$

$$Max = \max\left(\left\{S_i\right\}_{i \in [a,b]}\right) \tag{5.2}$$

$$Averge\ (\mu) = \frac{\sum_{i \in [a,b]} x_i}{b-a} \tag{5.3}$$

$$\text{Root Mean Square (RMS)} = \sqrt{\frac{\sum_{i \in [a,b]} (x_i - \bar{x})^2}{b-a}} \tag{5.4}$$

$$\text{Standard deviation } (\sigma) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2} \tag{5.5}$$

$$Skewness = \frac{\mu_3}{\sigma^3} = \frac{E[(x-\mu)^3}{E[(x-\mu)^2]^{3/2}} \tag{5.6}$$

$$Kurtosis = \frac{\mu_4}{\sigma^4} \tag{5.7}$$

$$Crest\ \ factor = \frac{|x|_{peak}}{x_{rms}} \tag{5.8}$$

Features calculations are also carried out in the frequency domain. The features used for evaluating the rotating imbalance include the first, second, and third harmonics of the fundamental frequency [167–169]. If f_0 is the fundamental frequency, then

$$f_{1st} = f_0 \tag{5.9}$$

$$f_{2nd} = 2f_0 \tag{5.10}$$

$$f_{3rd} = 3f_0 \tag{5.11}$$

The features used for evaluating the ball bearing faults include the first, second, and third harmonics of ball defect frequency [170]. A typical schematic of a ball bearing is shown in Figure 5.4.

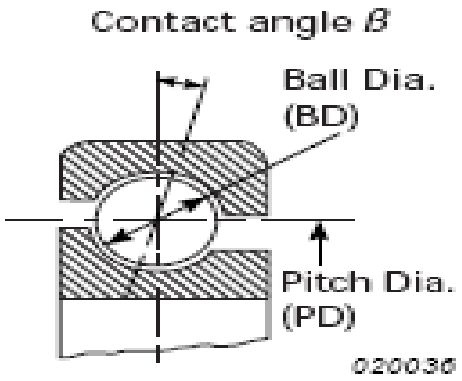


Figure 5.4: Schematic of a typical ball bearing [171].

For a ball bearing fault in ball defect, the following features are calculated.

$$f_{BPF}\ (Hz) = \frac{PD}{BD} f_r [1 - (\frac{BD}{PD}\cos\ \beta)^2] \tag{5.12}$$

$$f_{1st\_BPF} = f_{BPF}(Hz) \tag{5.13}$$

$$f_{2nd\_BPF} = 2f_{BPF}(Hz) \tag{5.14}$$

$$f_{3rd\_BPF} = 3f_{BPF}(Hz) \tag{5.15}$$

The computed features from the corresponding eight sensors are shown in Tables 5.2-5.4.

Table 5.2
(AI0-AI2) Motor current features and nomenclature

No.	Features	Refs.
1	Minimum	[163]
2	Maximum	[163]
3	Peak to Peak	[163]
4	Average	[163]
5	Root Mean Square (RMS)	[163]
6	Standard deviation	[163]
7	Skewness (3rd moment)	[163]
8	Kurtosis (4th moment)	[163]
9	Crest factor	[163]
10	Amplitude of 1st harmonic of rotational frequency	[167]
11	Amplitude of 2nd harmonic of rotational frequency	[167]
12	Amplitude of 3rd harmonic of rotational frequency	[167]

Table 5.3
(AI3) Torque features and nomenclature

No.	Features	Refs.
1	Minimum	[164]
2	Maximum	[164]
3	Peak to Peak	[164]
4	Average	[164]
5	Root Mean Square (RMS)	[164]
6	Standard deviation	[164]
7	Skew (3rd moment)	[164]
8	Kurtosis (4th moment)	[164]
9	Crest factor	[164]
10	Amplitude of 1st harmonic of rotational frequency	[168]
11	Amplitude of 2nd harmonic of rotational frequency	[168]
12	Amplitude of 3rd harmonic of rotational frequency	[168]

Table 5.4
(AI4-AI7) Acceleration features and nomenclature

No.	Features	Refs.
1	Minimum	[165]
2	Maximum	[165]
3	Peak to Peak	[165]
4	Average	[165]
5	Root Mean Square (RMS)	[165]
6	Standard deviation	[165]
7	Skew (3rd moment)	[166]
8	Kurtosis (4th moment)	[166]
9	Crest factor	[166]
10	Amplitude of 1st harmonic of rotational frequency	[169]
11	Amplitude of 2nd harmonic of rotational frequency	[169]
12	Amplitude of 3rd harmonic of rotational frequency	[169]
13	Amplitude 1st harmonic of cage frequency	[169]
14	Amplitude of 2nd harmonic of cage frequency	[169]
15	Amplitude of 3rd harmonic of cage frequency	[169]
16	Amplitude of 1st harmonic of outer race frequency	[169]
17	Amplitude of 2nd harmonic of outer race frequency	[169]
18	Amplitude of 3rd harmonic of outer race frequency	[169]

The 12 features in Table 5.2 are extracted from each of current clamp meters (AI0-AI2). Similarly, 12 features in Table 5.3 are extracted from torque (AI3), and 18 features in Table 5.4 are extracted from each of accelerometers (AI4-AI7). A total of 120 features are thus extracted from the 8 sensors in the fault simulator machine in this study. Data collections are 260 data points from each condition. From the 120 principle components (PCs), the most dominant 25 PCs (of largest magnitude in singular values) and their corresponding Singular Values (SVs) are shown in Table 5.5 and Figure 5.5.

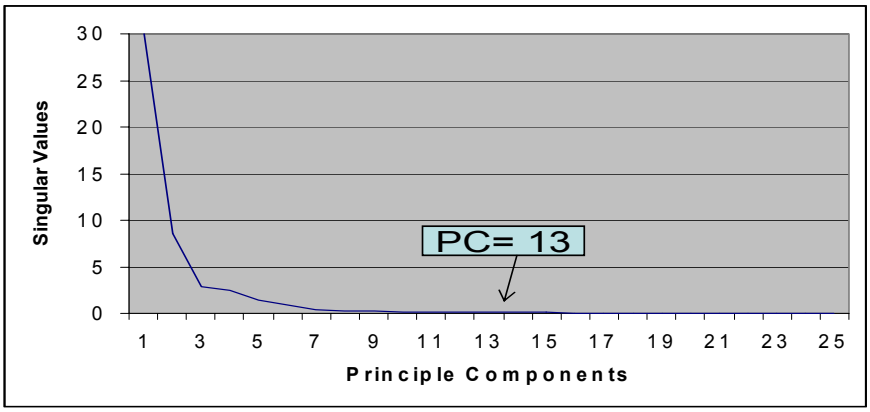


Figure 5.5: Plot of principal components versus singular values

Table 5.5

The 25 Most dominant principal components and singular values

No	Singular Values
1	29.953
2	8.533
3	2.886
4	2.410
5	1.382
6	0.972
7	0.362
8	0.310
9	0.286
10	0.193
11	0.175
12	0.140
13	0.119
14	0.090
15	0.086
16	0.064
17	0.060
18	0.054
19	0.053
20	0.045
21	0.038
22	0.035
23	0.033
24	0.031
25	0.028

It can be seen that the 14^{th} and subsequent singular values are very small as compared to the first 13 singular values. As such, we retain the first 13 singular values which will be sufficient to capture the relevant trend in fault detection for our experiments. However, since singular values are not directly associated with features, it is not clear which of the 120 features are required for fault detection. We therefore use the DFI methodology to select the most important features.

5.2 Dominant Feature Identification (DFI) for Fault Detection and Isolation (FDI)

In this Section, we propose two methods of using DFI for industrial Fault Detection and Isolation (FDI) applications. Suppose there is available measured time series data of length *m* samples from several fault conditions. These may be signals measured from sensors such as accelerometers, force sensors, electric current, hall effect, etc. For each fault condition k, n features are computed (rms value, energy, skew, etc.) from the measured data time series. For the k^{th} fault condition, form feature matrix $X_k \in R^{m \times n}$ by arranging the n features as its columns, each one is a time series of length m.

Based on the feature matrices X_k for fault conditions k, we propose two methods of industrial FDI based on DFI; namely an Augmented DFI (ADFI) approach and a Decentralized DFI (DDFI) approach. In both cases, the dominant features identified will be used for training a Neural Network (NN) for fault classification as detailed below.

5.2.1 Augmented dominant feature identification (ADFI)

The Augmented DFI method performs DFI on an augmented X matrix that concatenates all the sensor (or processed feature) data simultaneously for different fault conditions k. We call this method Augmented DFI (ADFI).

Suppose there are K fault conditions. The feature matrices $X_k \in R^{m \times n}$ for fault conditions k are concatenated into an overall feature matrix

$$X = \begin{bmatrix} X_1^T & X_2^T & \cdots & X_K^T \end{bmatrix}^T \tag{5.16}$$

where $X \in R^{mK \times n}$. We now perform DFI on this augmented composite fault feature matrix X to identify the p dominant features for FDI across all faults simultaneously. This procedure allows us to select p dominant features for describing all the fault conditions, and will be used for fault classification using a NN as described subsequently.

ADFI uses an approach similar to that used in facial recognition [160], whereby the singular values that best describe all of the set of faces together are retained and used for classification. The features resulting after ADFI on the concatenated matrix X in (5.16) are those that best describe the complete set of fault conditions as a whole.

The pseudocode of ADFI is as follows:

Algorithms: ADFI

INPUT: data $X_k \in R^{m\times n}$ $(k = 1,2...K)$

OUTPUT: $feature_{set} \in R^{mk\times p}$

// concatenated feature matrices $X_k \in R^{m\times n}$ for fault conditions k into an overall feature matrix:

WHILE $k < K$ DO

Read data $X_k \in R^{m\times n}$

$X = \left[X_1{}^T X_2{}^T \cdots X_k{}^T\right]^T$

ENDWHILE

Use DFI in Section 3.3.5 to select features of X and obtain feature matrix $feature_{set} \in R^{mk\times p}$

5.2.2 Decentralized dominant feature identification (DDFI)

The second method performs DFI on the fault feature matrices $X_k \in R^{m\times n}$ individually for each fault case k. We call this method Decentralized DFI (DDFI).

Thus, perform DFI on each individual fault feature matrix $X_k \in R^{m\times n}$ and extract the dominant features for each fault condition k. Then, the dominant features for FDI across all the faults are given by the union of the sets of features for each fault condition. These features will be used for fault classification using a NN in the next Section.

The pseudocode of DDFI is as follow:

Algorithms: DDFI

INPUT: data $X_k \in R^{m\times n}$ $(k = 1,2...K)$

OUTPUT: $feature_{set}$

WHILE $k < K$ DO

Read data $X_k \in R^{m\times n}$

Use DFI in Section 3.3.5 to select features of X_k and obtain feature matrix $feature_{selected}{}^{m\times p_k}$

IF $k = 1$

$feature_{set} = feature_{selected}(:, j)$

ELSE

IF $feature_{selected}(:, j)$ is not included in $feature_{set}$

$feature_{set} = feature_{set} + feature_{selected}(:, j)$

ENDIF

ENDIF

ENDWHILE

5.2.3 Fault classification with neural networks (NNs)

With the dominant features selected using either ADFI and DDFI as just described, NNs are trained to classify the faults that have occurred based on the reduced set of p dominant features.

In general, a two-layer NN is described by

$$y_i = \sigma\left(\sum_{l=1}^{L} w_{il}\sigma\left(\sum_{j=1}^{n} v_{lj}x_j + v_{l0}\right) + w_{i0}\right), i = 1,2,\cdots m \tag{5.17}$$

where x_j and y_i are the inputs and outputs of the NN, respectively, and v_{lj} and w_{il} are the weights of the hidden and output layer, respectively. n, L, and m are the orders of the input, hidden layer, and output, respectively, and σ is the activation function [160]. The details of the two-layer NN are shown in Figure 5.6.

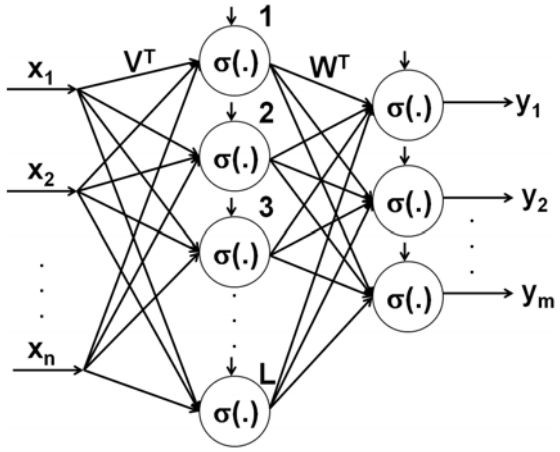


Figure 5.6: Two-layer NN trained using backpropagation for industrial fault classification after proposed ADFI and DDFI.

The inputs to the NN are the p dominant features' time series, each of length m. These are packed into a matrix of length mp as inputs to the NN. Such a "feature time signature" matrix corresponds to each fault condition k.

We can now use the dominant features collected via ADFI or DDFI to classify or even predict faults occurrences in typical FDI applications using the NN. The NN is trained using these feature time signatures of length *mp* of each of the *K* faults. For the training phase, the backpropagation algorithm is used with Gaussian Radial Basis Functions (RBFs) as the activation functions. RBFs are chosen as they provide good approximations to nonlinear functions and are fast in convergence. More importantly, the number of hidden neurons to-be used in the hidden layer can be automatically determined during the NN-training phase for an optimized architecture [161]. It is worth noting that any NN structure and activation function can be used for the second stage, and we have chosen the above-mentioned RBFs to verify the performance of the proposed new method in FDI without loss of generality.

Neural Networks (NNs) are used to build the fault classification model for FDI in estimating the machine status and health condition. The obtained features are used as the input of the NNs. The machine statuses are digitalized as the output of the NN and are defined in Table 5.6.

Table 5.6
Machine status and representation

Machine Status	Output of NN
Normal	0
Bearing ball fault	1
Imbalance	2
Loose belt	3
Bearing outer race fault	4

Upon completion of training and validation, the NN is put online, and presented with the *mp* feature time signature matrix as observed from the operating machine. It is worth noting that the built up NN model could only identify these five types of faults for this fault simulation machine. For any untraining types of faults, they are not able to be identified by this model.

5.3 Experimental results

Our experiments are conducted for 10 seconds at a 2 kHz sampling rate. Machine running speed is 2600rpm. 100 data points are extracted from the raw signal under each condition. Total of 500 data points are collected under 5 conditions: nomal, bearing ball fault, imbalance, loose belt and bearing outer race fault.

5.3.1 Fault detection using 120 features

In this Section, we conduct a baseline study that includes all the measured features for comparison with results using only reduced feature sets. In our experiment, we use one set of data which consists of measurements from the 8 sensors at five different machine conditions, namely normal, bearing ball fault, imbalance, loose belt, and bearing outer race fault. A total of 120 features are extracted from the signals collected from the 8 sensors. We use the extracted features to conduct fault detection. The first set of data uses all the 120 features as

the input, along with the output status defined earlier in Table 5.2-5.4 for training the NN. An one hidden layer feedforward NN with five neurons is created and trained. The input and target samples are automatically divided into training and validation sets. The backpropagation algorithm is used to train the NN. The mean squared error goal setting is 0.001, and the best validation performance is 0.00096283 and happens after seven epochs.

The trained NN can now be tested with the testing samples. We conduct another set of experiments under the same five machine conditions. Now, one hundred points are collected for each machine status. The second set of data is used for testing. The testing data is not used in training in any way and hence provides an uncorrelated "out-of-sample" dataset to validate the NN. This allows us to understand how well the NN model will perform when tested with the same type of machines. The results using the testing data with our trained NN for fault estimation are shown in Table 5.7. In Table 5.7, the first column shows the real machine status, while the second and third columns show the estimation results using the NN with all 120 from 8 eight sensors.

Table 5.7
Machine status and fault detection

Machine Status	**Estimation**	
	Correct	**Wrong**
Normal	100	0
Bearing ball fault	99	1
Imbalance	100	0
Loose belt	100	0
Bearing outer race fault	100	0

The overall estimation yields an accuracy of four hundred and ninety nine cases out of the total five hundred. The overall accuracy of the fault detection is hence established as 99.8 %. These data are used for comparison with subsequent fault classification tests using reduced feature sets.

5.3.2 Augmented dominant feature identification (ADFI) and NN for fault detection

It is now desired to use fewer sensors and features to detect the machine faults. Since 100 data points are collected under each condition and total of 500 data points are collected under 5 conditions. Data collected from the 8 sensors computed 120 features using digital signal processing techniques. Using the proposed method in Section 5.2.1, the data are concatenated from five different machine conditions. An augmented $X \in R^{1300\times120}$ matrix is constructed using all the original 120 features arising from all the 8 sensors. The proposed ADFI is used to select the dominant features from the original 120 features arising from all the 8 sensors across five different machine statuses simultaneously. The selected 13 dominant features are shown in Table 5.8.

Table 5.8
Feature selection using ADFI

Sensor	Features
C2	Average
C2	Maximum
C2	Crest factor
C2	Amplitude of 1st harmonic of rotational frequency
C2	Amplitude of 2nd harmonic of rotational frequency
Torque	Minimum
Torque	Standard deviation
Ax1	Minimum
Ax1	RMS
Ax1	Amplitude of 2nd harmonic of outer race frequency
Ax2	Maximum
Ax2	Minimum
Ax2	Amplitude of 2nd harmonic of cage frequency

Our results reduced the number of features from 120 to 13! These 13 features come from only 4 sensors, so that the number of required sensors has been pruned from 8 to 4. This represents a significant amount of savings from sensor to signal processing costs in industrial manufacturing.

We now use the selected features in Table 5.8 to conduct the fault detection. Similar to Section 5.3.1, the first set of data is used for training the one hidden layer feedforward NN with five neurons. By setting the mean squared error goal to 0.001, the best achievable performance is 0.00052268 after 30 epochs.

The trained NN is then validated with testing samples from a second set of uncorrelated data. The results of the testing are shown in Table 5.9.

Table 5.9

Machine status and fault detection using reduced number of sensors and features from ADFI and NN

Machine Status	**Estimation**	
	Correct	**Wrong**
Normal	99	1
Bearing ball fault	99	1
Imbalance	100	0
Loose belt	100	0
Bearing outer race fault	99	1

The overall correct estimation is four hundred and ninety-seven cases out of five hundred. As such, the overall accuracy of the fault detection is 99.4% for ADFI. Since $m = 500$ and $n = 120$, the computational complexity for ADFI in this process is $O(mn^2) + O(n^3) = O(7.2 \times 10^6) + O(1.728 \times 10^6)$.

5.3.3 Decentralized dominant feature identification (DDFI) and NN for fault detection

Using the proposed DDFI method in Section 5.2.2, we perform DFI on each individual fault feature matrix $X_k \in R^{100 \times 120}$, $(k = 1,...,5)$ and extract dominant features for each of the five machine conditions. Since $m = 100$ and $n = 120$, the selected dominant features for each condition are shown in Table 5.10. The computational complexity for DDFI in this process is:

$$5 \times (O(mn^2) + O(n^3)) = 5 \times (O(1.44 \times 10^6) + O(1.728 \times 10^6)) = O(7.2 \times 10^6) + O(8.64 \times 10^6).$$

The computational complexity of DDFI is higher than ADFI.

Table 5.10
Feature selection using conventional DFI

SENSOR	FEATURES
Features for normal condition	
C1	mean
C1	std
C1	skew
Torque	min
Torque	mean
Torque	peak_peak
Torque	skew
Ax1	min
Ax1	mean
Ax1	cage_third_harmonic
Ax1	cage_first_harmonic
Ay2	min
Ay2	Mean
Ay2	RMS
Ay2	outer_third_harmonic
Features for bearing ball fault condition	
C1	mean
C1	std
C1	skew
Torque	min
Torque	mean
Torque	peak_peak
Ax1	skew
Ax1	min
Ax1	mean
Ax1	cage_third_harmonic
Ax1	cage_first_harmonic
Ay2	min
Ay2	mean
Ay2	RMS
Ay2	outer_third_harmonic
Features for imbalance condition	
C1	mean
C1	std
C1	skew
Torque	min
Torque	mean
Torque	rRms
Torque	skew
Ax1	min
Ax1	mean
Ay2	min
Ay2	std
Ay2	skew
Ay2	outer_third_harmonic
Features for loose belt condition	
C1	skew

C1	first_harmonic
Torque	min
Torque	mean
Torque	skew
Torque	crest
Ax1	min
Ax1	mean
Ax1	crest
Ax1	outer_second_harmonic
Ay2	min
Ay2	crest
Ay2	first_harmonic
Ay2	outer_third_harmonic
Features for bearing outer race fault condition	
C1	mean
C1	std
C1	skew
Torque	min
Torque	mean
Torque	rRms
Torque	skew
Ax1	min
Ax1	mean
Ax1	Rms
Ay2	min
Ay2	std
Ay2	skew
Ay2	outer_third_harmonic

The combination features are decided using across selection of all the features for each fault condition. The combination of twenty eight dominant features for DDFI is shown in Table 5.11.

Similarly, we now use the selected features in Table 5.11 to conduct the fault detection. The first set of data is used for training the one hidden layer feedforward NN with five neurons. The mean squared error goal setting is 0.001, and the best performance is 0.496 after twelve epochs.

A second set of uncorrelated experimental data is used for validating the NN, and the results of the testing are shown in Table 5.12. The overall correct estimation is four hundred and twenty-two cases out of five hundred. This gives an overall accuracy of fault detection of 84.4%. We also notice that only 42% of accuracy is obtained for bearing ball fault case which is much lower than other machine status. This implies that the selected features using DDFI give less accurate estimation for bearing ball fault condition.

Table 5.11

Feature selection using proposed DDFI

Sensor	Features
C1	Mean
C1	Std
C1	Skew
C1	Crest
C1	first_harmonic
Torque	Min
Torque	Mean
Torque	peak_peak
Torque	rRms
Torque	crest
Torque	Skew
Ax1	Min
Ax1	Mean
Ax1	crest
Ax1	Skew
Ax1	outer_second_harmonic
Ax1	outer_third_harmonic
Ax1	cage_first_harmonic
Ax1	cage_third_harmonic
Ay2	Min
Ay2	Mean
Ay2	rRms
Ay2	Std
Ay2	Skew
Ay2	Crest
Ay2	first_harmonic
Ay2	second_harmonic
Ay2	Outer_third_harmonic

Table 5.12

Machine status and fault detection using reduced number of sensors and features from DDFI and NN

Machine Status	Estimation	
	Correct	Wrong
Normal	95	5
Bearing ball fault	42	58
Imbalance	94	6
Loose belt	97	3
Bearing outer race fault	94	6

In all, the fault detection accuracy is 99.8% when using all the original 8 sensors and 120 features followed by a one hidden layer NN (five neurons) for fault detection. However, we can achieve a fault detection accuracy of 99.4% when using proposed ADFI with 13 features or 84.4% when using DDFI with 28 features. We also notice that only 42% of accuracy is obtained for bearing ball

fault case which is much lower than the result obtained by ADFI. The computational complexity for ADFI is less than DDFI. Obviously, the performance of ADFI is better than DDFI both in features and sensors number reduction as well as fault detection accuracy. From this case study, the proposed ADFI with NN for FDI is a promising technology which saves both hardware costs and computational time in signal processing.

5.4 Summary

In this chapter, our proposed Dominant Feature Identification methodology (DFI) is combined with Neural Networks (NNs) for Fault Detection and Identification (FDI). The number of sensors used is reduced from 8 to 4, and the required number of features is significantly reduced from 120 to 13 when using proposed Augmented DFI (ADFI) with NN for fault detection. The FDI accuracy is 99.4% in this study, which translates to reduction in both hardware and computational time from signal processing without compromising the FDI accuracy. This results in significant cost savings, which is crucial for next generation of intelligent green manufacturing. The testing results in Chapter 5 are obtained using a fault simulator machine and the experiment is conducted in the lab. The real machines are normally with much higher complexity and much noisy than the lab environment. We expected that results on the real machines are not as accurate and repeatable as in the lab environment.

CHAPTER 6 DFI WAVELET BASED CORRELATION APPROACH FOR HEALTH ASSESSMENT OF FLUID DYNAMIC BEARINGS IN BRUSHLESS DC MOTORS

Machine failures and result the significant process down time are costly in many applications. When an electrical machine fails, a motor fault is the most possible cause [171], and 40–50% of all motor failures are bearing related [172]. The diagnostic methods to identify fault and monitor working conditions of motors involve several fields of sciences and technology [172–175]. The methods can be classified as methods based on signal analyses, motor dynamic models and knowledge. All of them have their own advantages and disadvantages [173]. Two main approaches have emerged for detection of bearing faults. One uses vibration-based technologies and the other uses current-based bearing technologies. While the vast majority of bearing faults are detected via vibration-based techniques, the cost of obtaining these vibration measurements is relatively high as vibration sensor needs to be placed close to the vibration sources. If the design of the original equipment does not provide space or position for mounting, the vibration sensors will have difficulties in implementation. In many applications, stator current is routinely monitored for protection purposes (over current, over load, ground current, etc.). In these instances, a stator-current-based condition monitoring scheme is inexpensive to implement and requires no additional sensors, wiring, etc. In other instances, it is often preferable and cost effective to add current sensors at the electrical supply rather than attaching a vibration sensor directly to the machine. This makes current-based bearing condition monitoring attractive and desirable in many applications [171]. Also, the use of an existing current signal terminal at a motor drive circuit can be convenient for monitoring a large number of motors remotely from one location [176].

Brushless direct current (BLDC) motors are gaining popularity because they have many advantages over brushed DC motors and induction motors, such as higher efficiency and noiseless operation [177]. They are used in many applications, such as disk drives and videocassette recorders [178, 179]. Besides the research on common DC motors and induction motors, research on fault diagnosis of BLDC motors was also reported in recent years. A model based fault detection method was proposed [178]. Automatic diagnosis of open-switch faults

using wavelets and neural fuzzy systems was introduced [180, 181]. A wavelet ridge detector was used for diagnosis of potential rotor faults in BLDC motors [177]. In spite of these efforts, realization of acceptable prognosis of bearing health status of BLDC motors is still a goal to be achieved.

In recent years, the time-frequency signal analysis has been studied and applied in machinery fault diagnosis due to its capability of representing signals in both time and frequency domains. This characteristic of time-frequency analysis technique meets the requirement for analyzing non stationary signals, such as vibration, current or acoustic emission signals.

Wavelet transform has been successfully applied to data analysis in fault diagnostics of gears [79, 80], bearings [81, 82] and other mechanical systems [83, 84]. Baydar and Ball [57] applied wavelet transform to both acoustic signals and vibration. Wavelet transform combined with Fourier transform to enhance feature extraction capability was proposed in [86]. A more advanced transform, known as wavelet packet transform, was studied in [87- 88].

With the development for about 20 years, the wavelet transform has been widely used in fault diagnostics. But, compared with the FFT, the applications of the wavelets have still not achieved a standard status [89]. Many reasons have caused the current status of wavelets in fault diagnostics. In Fourier analysis, the basis functions are complex exponentials producing the same results for a particular waveform being analyzed. However in wavelet analysis the basis function could be any permissible wavelet and the results produced are unique to the selected wavelet. There is no a standard or a general method to select the wavelet function for different tasks. This is an obstacle for the popularization of the wavelet transform. Some wavelets are more efficient at encoding, denoising, compressing, decomposing and reconstructing signals than others. It is therefore desirable to select wavelets that produce the best results for the signal being analyzed. Unlike Fourier transform, results of the wavelet transform have no straightforward physical implications, and therefore it is difficult to obtain useful information directly from the results of the wavelet transform. This is another disadvantage for using wavelet.

In this work, DFI methodology is used for dominant mother wavelet

selection and dominant wavelet decomposition levels selection. A new approach is proposed to combine a wavelet-based feature extraction with DFI and the multivariable regression technique to establish a correlation model. The approach is tested for monitoring of bearing health status of BLDC motors to achieve robust performance prediction and failure prevention.

6.1 The approach

The most difficult part in doing the wavelet analysis is to determine which mother wavelet should be used and how many levels of decompose should be conducted. In this chapter, instead of selecting one mother wavelet, 12 common mother wavelets are chosen and the features are extracted. DFI is used two times for feature selection to decide the most dominant wavelets and dominant decomposition levels. The selected features are then used to build the health correlation model. The approach is a DFI wavelet-based correlation modeling health assessment (DFI_wavelet_correlation) method. DFI_wavelet_correlation consists of four stages:

(1) Feature extraction from the motor stator current signature using the analysis of discrete wavelet transform coefficients method.

(2) Dominant wavelets selection using DFI methodology.

(3) Dominant decomposition levels selection using DFI methodology.

(3) A correlation model between features and degree of the bearing wear obtained using multivariable regression.

The detail of DFI wavelet-based correlation modeling health assessment (DFI_wavelet_correlation) method is described in detail in the following Sections.

6.1.1 Feature extraction using discrete wavelet transform coefficients

An example of stator current signals of a healthy BLDC motor with 100% amount of lubricant and a health-degraded BLDC motor with 80% amount of lubricant are shown in Figures 6.1 and 6.2, respectively. It is difficult to see the difference of these signals without the assistance of a proper analysis tool.

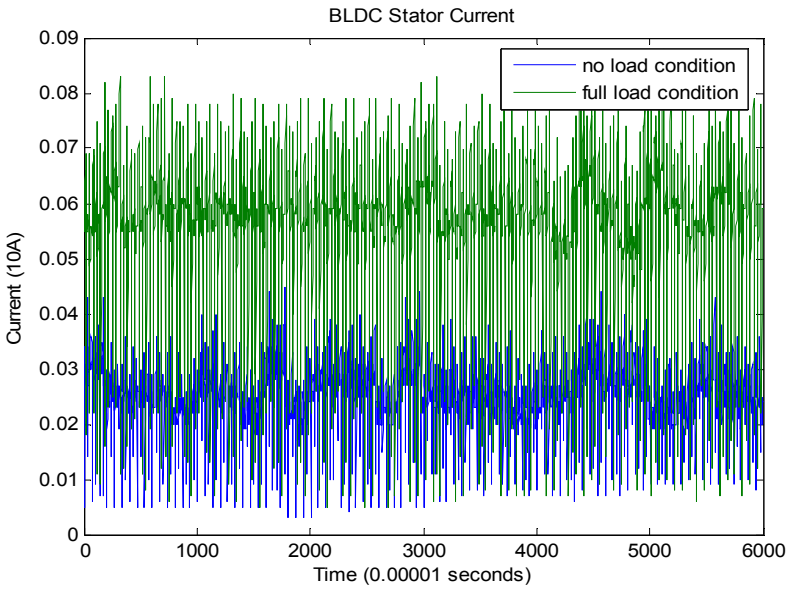


Figure 6.1: Stator current signals of a healthy BLDC motor with 100% amount of lubricant

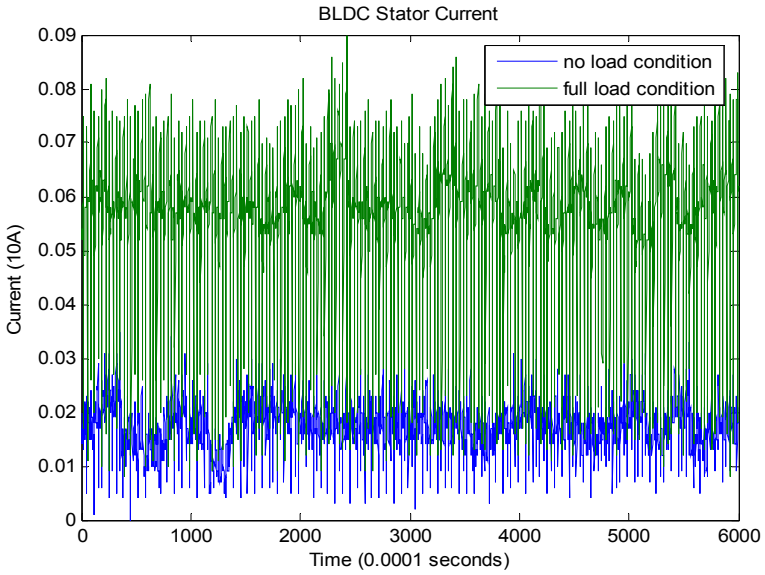


Figure 6.2: Stator current signals of a health-degraded BLDC motor with 80% amount of lubricant (the bearing health degradation is 20%)

A stream of time serial data $x(t)$ is the measured current signal. The one-dimensional discrete wavelet decomposition with j different levels is performed on this signal. Let the absolute means of detail coefficients of this mother wavelet

at 1...j levels be a column vector $X_1^{j\times 1}$. The process is then moved to another mother wavelet, and another column vector $X_2^{j\times 1}$ is obtained. Likewise m such vectors $X_i^{j\times 1}(i=1,...,m)$ from m different mother wavelets are obtained and form the columns of a $(j\times m)$ matrix $M_k^{j\times m}$. This matrix is referred as the current status information matrix in this Chapter. In general, the more mother wavelets and more decomposition levels are adopted, the more detail current status information will be contained in the matrix. The matrix $M_k^{j\times m}$ is formulated at each health condition. Suppose there are *k* health conditions. The feature matrices $M_k \in R^{j\times m}$ for health conditions *k* are concatenated into an overall feature matrix

$$M = [M_1^T M_2^T \cdots M_k^T]^T \tag{6.1}$$

where $M \in R^{jK\times m}$, Figure 6.3 shows an example to use DWT coefficients to construct a matrix $M_k^{j\times m}$.

Construct matrix $M_k^{j\times m}$ with coefficients at 1 to j levels using different wavelets

Method\ Level	db1	…	db7	sym2	sym 4	sym8	coif1	coif3	Health condition
L1	***	***	***	***	***	***	***	***	k
L2	***	***	***	***	***	***	***	***	k
….	***	***	***	***	***	***	***	***	k
Lj	***	***	***	***	***	***	***	***	k

Figure 6.3: Using DWT coefficients to construct a matrix M_k

After feature extractions using discrete wavelet transform coefficients, the next objective is to reduce the computation: to decide which wavelets are the dominant wavelets and which decomposition levels are dominant decomposition levels. DFI will be used two times for dominant wavelets and dominant decomposition levels selection.

6.1.2 DFI for dominant wavelets selection

It is needed to decide which wavelets are the most dominant wavelets for future bearing health diagnosis and prognosis. DFI in Chapter 3 is now performed on this augmented composite feature matrix $M \in R^{jK \times m}$ to identify the p dominant wavelets. This procedure allows us to select p dominant wavelets for describing the bearing conditions.

The SVD of a linear transformation $M \in R^{jK \times m}$ is $M = U \Sigma V^T$. The Dominant Feature Identification (DFI) methodology proposed in Chapter 3 uses the fact that

$$Y = U_1^T M = \Sigma_1 V_1^T . \tag{6.2}$$

Therefore, DFI selects the best p columns of $\Sigma_1 V_1^T$ using k-means clustering methods for approximation and obtain the dominant wavelets matrix $WD^{jk \times p}$ using the selected p dominant wavelets. $WD^{jk \times p}$ consists of k health condition matrix of $WD_k^{j \times p}$. $WD_k^{j \times p}$ consists of j decomposition level and p dominant wavelets. After the dominant wavelets matrix is obtained, next objective is to reduce the decomposition levels using DFI.

6.1.3 DFI for dominant decomposition levels selection

Dominant decomposition levels are the rows in $WD^{jk \times p}$. Since DFI only selects the best q columns using k-means clustering methods for approximation, next step is to construct the decomposition level matrix $D_k^{p \times j}$ at each health condition k. This is obtained by transposing the matrix of $WD_k^{j \times p}$:

$$D_k^{p \times j} = (WD_k^{j \times p})^T \tag{6.3}$$

The decomposition level matrix $D_k \in R^{p \times j}$ for health conditions k is concatenated into an overall feature matrix

$$D = [D_1^T D_2^T \cdots D_k^T]^T \tag{6.4}$$

where $D \in R^{pK \times j}$.

DFI is used to select the best q columns and the dominant decomposition level matrix $DD^{pk \times q}$ is obtained. The pseudocode of using DFI for dominant wavelets and decomposition levels selection will be described in Section 6.1.5.

6.1.4 The correlation model

The correlation model between signal features and motor degradation is constructed to assess the motor working status and fault condition. The features derived from $DD^{pk\times q}$ can be used for building the following multivariable model:

$$y = b_0 + b_1 L_1 + b_2 L_2 + ... + b_j L_q \tag{6.5}$$

Where, y is a vector containing the values of the heath status of the motors under monitoring, $b_0, b_1, ...b_j$ are constants, and $L_1, L_2, ...L_q$ are vectors containing the values of the feature variables representing the coefficients of the discrete wavelet transformed current signals at the q decomposition levels.

6.1.5 Steps of DFI_wavelet_correlation

Step 1: Data acquisition of the motor current signals through machine drive terminals.

Step 2: Extraction of the features by performing DWT decomposition using different types of wavelet. The current DWT coefficients construct a matrix M.

Step 3: Using DFI to select the dominant wavelets and eliminate the wavelets that are not responding to the health condition. The dominant wavelet health condition is presented in matrix $WD^{jk\times p}$.

Step 4: Construct decomposition levels matrix $D_k{}^{p\times j}$ with the transpose of $WD_k{}^{j\times p}$. Using DFI to select the dominant decomposition levels and eliminate the levels that are not responding to the health condition. The dominant decomposition levels are presented in matrix $DD^{pk\times q}$.

Step 5: Building of the multivariable correlation model between represented DWT coefficients and motor degradation conditions.

Step 6: Evaluation of the model. The detection ability and the accuracy of the model are verified through experiments. The model is used to assess the motor health status and fault condition.

The flow chart of the DFI_wavelet_correlation approach is shown in Figure 6.4.

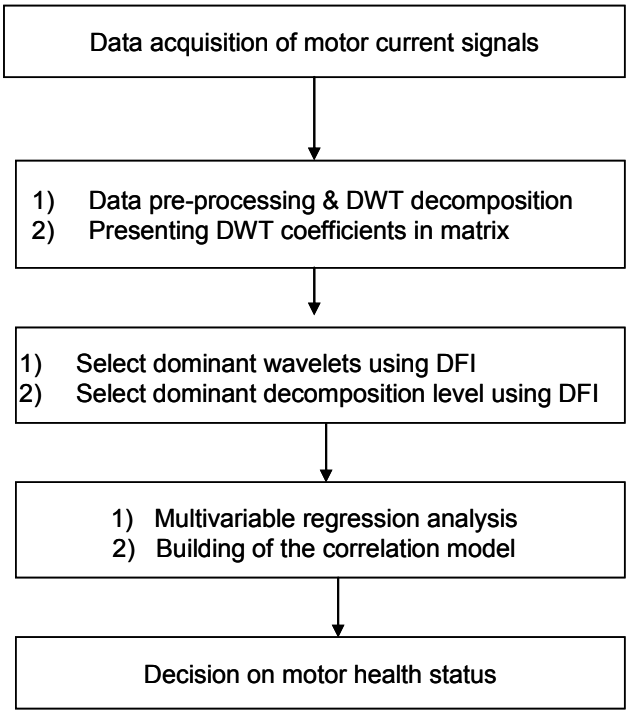


Figure 6.4: The DFI_wavelet_correlation flow chart

The pseudocode of DFI_wavelet_correlation method is as follows:

Algorithms: DFI_wavelet_correlation :

INPUT: current signal $x(t)$

OUTPUT: A multivariable model $b_0, b_1, ...b_j$ to predict motor health

---

//construct current status information matrix $M_k^{j\times m}$ under *(1,...k)* health conditions.

WHILE $i1 <= k$ DO

WHILE $i <= m$ DO

Use mother wavelet i to perform discrete wavelet decomposition on the signal $x(t)$. j different levels decomposition is performed for each mother wavelet.

Calculate detail coefficient at levels *1…j*.

Construct column vetor $X_i^{j\times 1}$ = detail coefficient at levels *1…j*

END WHILE

// Form the *m* columns of a $(j \times m)$ matrix $M_k^{j\times m}$.

WHILE $i <= m$ DO

$M_k^{j\times m} = X_i^{j\times 1}$

END WHILE

END WHILE

// concatenated the feature matrices $M_k \in R^{j\times m}$ for health conditions *k* into an overall feature matrix $M \in R^{jK\times m}$:

$M = [M_1^T M_2^T \cdots M_k^T]^T$

// Use DFI to obtain the dominant decomposition level matrix $DD^{pk\times q}$

READ $M \in R^{jK\times m}$

Use DFI in Chapter 3 to select the *p* dominant wavelets.

Obtain matrix $WD^{jk\times p}$ (*k* health conditions. *p* dominant wavelets)

Separate $WD^{jk\times p}$ into $WD_k^{j\times p}$ for *k* health conditions

Transpose matrix $WD_k^{j\times p}$ using equation (6.3): $D_k^{p\times j} = (WD_k^{j\times p})^T$

Concatenated $D_k \in R^{p\times j}$ *(1,…k)* into an overall feature matrix $D \in R^{pK\times j}$ using equation (6.4): $D = [D_1^T D_2^T \cdots D_k^T]^T$

Use DFI in Chapter 3 to select the best q columns of dominant decomposition levels and obtain the dominant decomposition levels matrix $DD^{pk\times q}$.

// Obtain a multivariable model $b_0, b_1, \ldots b_j$ to predict motor health

---

READ: motors heath status vector y.

Obtain vectors $L_1, L_2, \ldots L_q$ from *q* columns of $DD^{pk\times q}$.

// $L_1, L_2, ...L_q$ representing the coefficients of the discrete wavelet transformed current signals at the q decomposition levels

Base on the equation $y = b_0 + b_1 L_1 + b_2 L_2 + ... + b_j L_q$, estimate coefficient constants $b_0, b_1, ...b_j$ using least square methods.

6.2 BLDC motors with fluid dynamic bearings

Fluid Dynamic Bearing (FDB) (Figure 6.5) motors are a substitute for the traditional ball bearing motors used in previous generation BLDC motors. FDB technology was developed over 50 years ago for use in gyroscopes, high accuracy machine tools, and low noise applications in submarine. Fluid dynamic bearing (FDB) motors have been used in disk drive and videocassette recorder applications [178]. The technology does not have metal-to-metal contact found in traditional ball bearings. This new bearing technology enables high spindle speeds, low acoustic noise emission, and improved shock performance. Figure 6.6 shows the components of the BLDC motor.

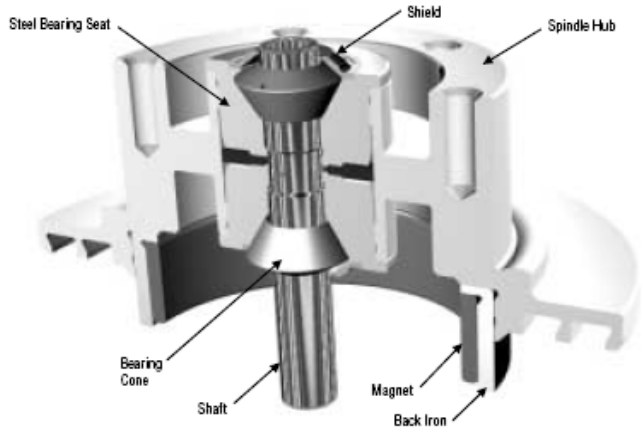


Figure 6.5: FDB BLDC motor cross-Section [178]

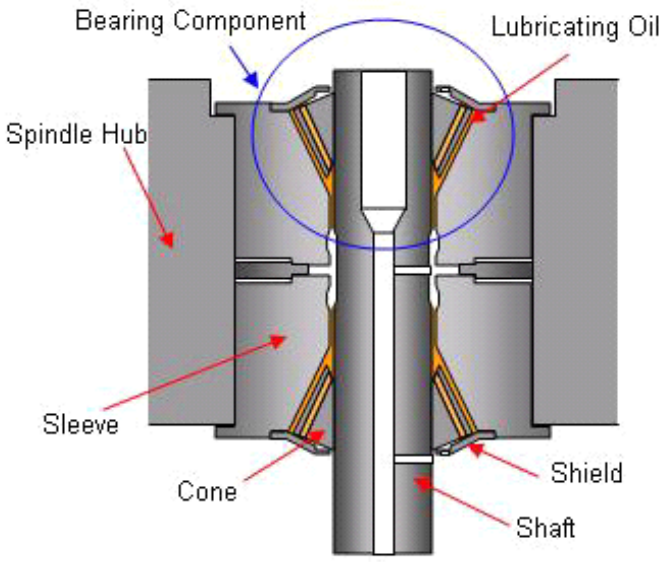


Figure 6.6: Sketch of bearing component

The bearing function results from a thin layer of fluid, which separates a stationary assembly supported from a base and a rotating assembly supported on the stationary assembly, provides mechanical damping and spreads shock loads over a large area [184, 185]. However, due to leakage or evaporation of the bearing grease-base oil, bearing degradation results in a weight change of the spindle rotor in a FDB BLDC motor. There are crack lines and pinholes at the laser welding seam areas, which result in lubricant leakage. Figure 6.7 shows an example of lubricant leakage taking place. The weight change of the spindle can be used to quantify the amount of lubricant leakage or evaporation. This method is used in the industry [184]. This test can only be conducted once every a few months. The drawback of this method is that bearing health of motors in service cannot be monitored. This method is also time-consuming. Therefore, the easily measurable signal of motor stator current was used to detect the health conditions of the motors. The correlation between bearing degradation and the current signal was investigated. Specific conditions resulting in failures produce short transients on the currents of electrical motors. The analysis and computational tools to recognize these transients or faults were developed and applied to BLDC motors. In the rest of Chapter 6, the fault of a motor is defined as a condition of the motor that allows its continuous operation but may eventually result in failure. All of the motors

selected for this study, with the exception of the healthy motors, have such defined faults.

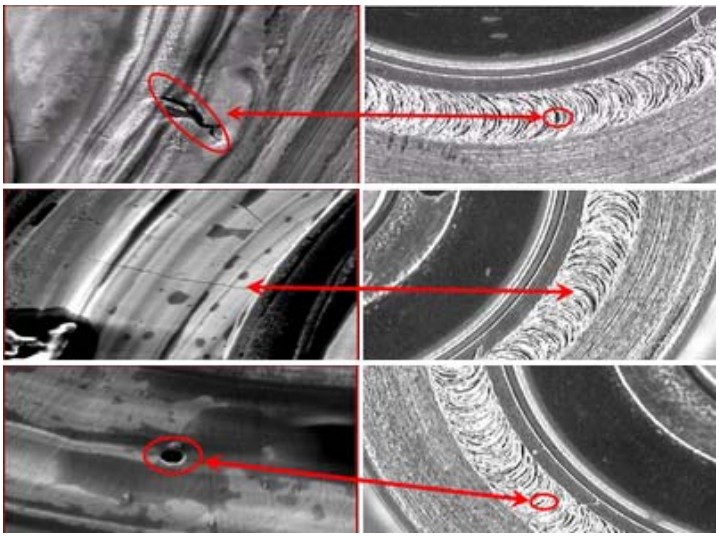


Figure 6.7: Examples of lubricant leakage taking place

6.3 Experiments and results

6.3.1 Experimental setup

The motors investigated in the experiments are nine slotted 12V FDB BLDC motors having 6 poles and 10,000rpm. 12 motors were used for the experiments. Six motor samples adopted in the experiments were healthy motors with 100% amount of bearing lubricant. To have different degrees of motor bearing degradation, six motors were made to have 80%, 60% and 40% of normal amount of bearing lubricant, respectively. The health condition of a motor sample used for the experiment with $\alpha\%$ amount of bearing lubricant is defined to be 0.01α. For example, the health condition of a motor sample with 60% amount of bearing lubricant is defined to be 0.6. Table 6.1 shows definitions of the sample health conditions for the experiment.

Table 6.1

Definition of sample health condition

	Health grade	Defined health condition
Condition 1	100% amount of bearing lubricant	1 (100% healthy)
Condition 2	80% amount of lubricant	0.8 (80 % healthy)
Condition 3	60% amount of lubricant	0.6 (60 % healthy)
Condition 4	40% amount of lubricant	0.4 (40 % healthy)

The experimental setup is shown in Figures 6.8 and 6.9. The data acquisition system and the motor drive system used in the experiments included a National Instruments (NI) PCI-6030E family data acquisition board with a connection adapter BNC2090, a HOKUT HC5672-101 motor drive card with current signal terminals, a 12V DC power supply, and a desktop personal computer (PC). LabView® was used to record motor stator current signals. The BLDC motors tested have a rated steady-state spinning speed at 10,000 revolutions per minute (rpm). In order to capture short transients of the current signals, a high sampling rate of 100 kHz was adopted. Using this sampling rate, 6,000 sampled data were recorded in 0.06 second from ten revolutions of a motor under study.

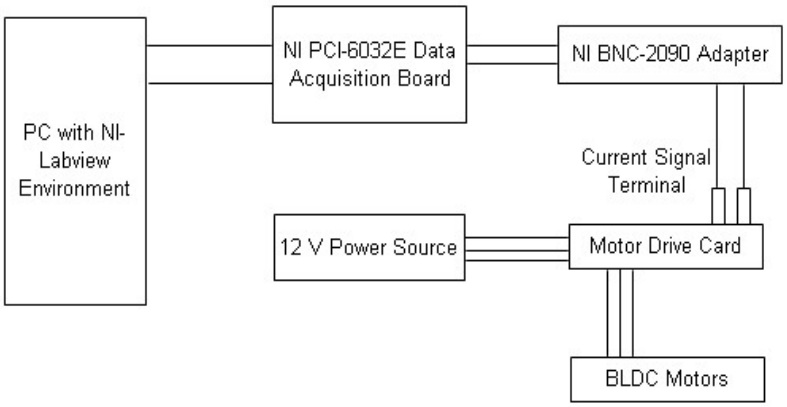


Figure 6.8: Sketch of experimental arrangement

Figure 6.9: Experimental set-up

6.3.2 Features extraction by wavelet decomposition

The wavelet decomposition was conducted for feature extraction from the motor current signals. For the local analysis, the decomposition was taken up to the levels $j = 9$, because $2^9 = 512$ is close to 600. Because the signal was sampled at a rate of 600 data per motor revolution in 0.006 second, the components of the signal within a period less than 0.006 second were investigated.

Tables 6.2-6.7 show the absolute means of detailed coefficients at nine $j = 9$ different decomposition levels resulting from the analysis of the stator current signals of motors (that had defined health conditions of 1, 0.8 and 0.4 respectively) under the no load operating condition, using a variety $m = 12$ of mother wavelets. In these Tables, db1, db3, db5, db7 and db9 are Daubechies wavelets, sym2, sym4, sym6 and sym8 are Symmlets wavelets, and coif1, coif 3 and coif 5 are Coiflet wavelets [196]. The value in every individual cell of the tables corresponds to the element value of current signal DWT coefficient matrix M, which is defined in Section 6.1. In this way, matrixes were constructed. From these Tables, it can be seen that the extraction information at different levels is different when different mother wavelets are used. This supports our point that if only one wavelet is used for the analysis, it is difficult to know which wavelet should be chosen.

Table 6.2

Matrix M_1 with absolute means of the coefficients for motor sample 1 with defined health condition 1

Level	db1	db3	db5	db7	db9	sym2	sym4	sym6	sym8	coif1	coif3
1	0.0017	0.0015	0.0018	0.0019	0.0015	0.0014	0.0017	0.0016	0.0016	0.0016	0.0016
2	0.0019	0.0025	0.0022	0.0023	0.0029	0.0024	0.0018	0.0020	0.0020	0.0018	0.0021
3	0.0033	0.0039	0.0039	0.0043	0.0043	0.0036	0.0039	0.0036	0.0039	0.0036	0.0039
4	0.0040	0.0046	0.0046	0.0046	0.0045	0.0044	0.0045	0.0046	0.0046	0.0044	0.0046
5	0.0036	0.0032	0.0034	0.0031	0.0033	0.0035	0.0031	0.0030	0.0030	0.0034	0.0029
6	0.0060	0.0069	0.0069	0.0072	0.0070	0.0067	0.0069	0.0069	0.0068	0.0064	0.0069
7	0.0065	0.0056	0.0052	0.0048	0.0045	0.0062	0.0051	0.0045	0.0040	0.0056	0.0045
8	0.0117	0.0112	0.0100	0.0087	0.0095	0.0104	0.0102	0.0088	0.0079	0.0111	0.0079
9	0.0271	0.0184	0.0186	0.0187	0.0195	0.0266	0.0194	0.0158	0.0133	0.0186	0.0130

Table 6.3

Matrix M_2 with absolute means of the coefficients for motor sample 2 with defined health condition 1

Level	db1	db3	db5	db7	db9	sym2	sym4	sym6	sym8	coif1	coif3
1	0.0014	0.0017	0.0015	0.0017	0.0020	0.0016	0.0015	0.0016	0.0017	0.0014	0.0016
2	0.0024	0.0019	0.0024	0.0028	0.0023	0.0019	0.0025	0.0024	0.0025	0.0024	0.0025
3	0.0038	0.0036	0.0042	0.0039	0.0043	0.0035	0.0036	0.0038	0.0039	0.0036	0.0037
4	0.0040	0.0042	0.0044	0.0045	0.0046	0.0044	0.0043	0.0044	0.0046	0.0044	0.0044
5	0.0036	0.0032	0.0033	0.0032	0.0032	0.0036	0.0030	0.0030	0.0030	0.0032	0.0031
6	0.0061	0.0067	0.0070	0.0066	0.0069	0.0066	0.0068	0.0067	0.0067	0.0062	0.0065
7	0.0071	0.0063	0.0057	0.0053	0.0050	0.0070	0.0059	0.0051	0.0048	0.0068	0.0051
8	0.0092	0.0166	0.0129	0.0147	0.0107	0.0157	0.0133	0.0125	0.0123	0.0159	0.0122
9	0.0297	0.0124	0.0244	0.0129	0.0183	0.0187	0.0174	0.0130	0.0111	0.0135	0.0120

Table 6.4

Matrix M_3 with absolute means of the coefficients for motor sample 3 with defined health condition 1

Level	db1	db3	db5	db7	db9	sym2	sym4	sym6	sym8	coif1	coif3
1	0.0018	0.0015	0.0018	0.0019	0.0016	0.0015	0.0017	0.0017	0.0017	0.0017	0.0017
2	0.0021	0.0024	0.0025	0.0023	0.0029	0.0023	0.0020	0.0021	0.0022	0.0020	0.0022
3	0.0032	0.0042	0.0041	0.0046	0.0048	0.0039	0.0040	0.0042	0.0039	0.0038	0.0042
4	0.0039	0.0047	0.0047	0.0048	0.0048	0.0049	0.0045	0.0048	0.0051	0.0045	0.0047
5	0.0034	0.0033	0.0031	0.0029	0.0031	0.0036	0.0032	0.0030	0.0029	0.0035	0.0027
6	0.0055	0.0064	0.0069	0.0065	0.0070	0.0067	0.0066	0.0064	0.0062	0.0063	0.0063
7	0.0070	0.0063	0.0055	0.0051	0.0052	0.0075	0.0061	0.0049	0.0043	0.0073	0.0049
8	0.0138	0.0146	0.0149	0.0117	0.0118	0.0164	0.0120	0.0106	0.0103	0.0137	0.0104
9	0.0189	0.0221	0.0171	0.0228	0.0210	0.0191	0.0259	0.0188	0.0161	0.0257	0.0166

Table 6.5

Matrix M_4 with absolute means of the coefficients for motor sample 4 with defined health condition 0.8

Level	db1	db3	db5	db7	db9	sym2	sym4	sym6	sym8	coif1	coif3
1	0.0024	0.0021	0.0024	0.0025	0.0022	0.0021	0.0023	0.0023	0.0023	0.0023	0.0023
2	0.0026	0.0030	0.0031	0.0029	0.0034	0.0028	0.0026	0.0026	0.0028	0.0025	0.0028
3	0.0038	0.0048	0.0047	0.0051	0.0054	0.0044	0.0046	0.0046	0.0045	0.0043	0.0047
4	0.0040	0.0050	0.0049	0.0050	0.0051	0.0053	0.0047	0.0051	0.0054	0.0047	0.0050
5	0.0041	0.0038	0.0038	0.0037	0.0037	0.0040	0.0038	0.0038	0.0036	0.0040	0.0035
6	0.0060	0.0071	0.0072	0.0072	0.0071	0.0069	0.0069	0.0069	0.0067	0.0066	0.0065
7	0.0079	0.0070	0.0056	0.0056	0.0053	0.0078	0.0065	0.0054	0.0046	0.0078	0.0051
8	0.0176	0.0206	0.0178	0.0185	0.0164	0.0175	0.0183	0.0166	0.0144	0.0182	0.0148
9	0.0223	0.0221	0.0190	0.0181	0.0184	0.0246	0.0139	0.0145	0.0118	0.0188	0.0135

Table 6.6

Matrix M_5 with absolute means of the coefficients for motor sample 5 with defined health condition 0.6

Level	db1	db3	db5	db7	db9	sym2	sym4	sym6	sym8	coif1	coif3
1	0.0014	0.0012	0.0014	0.0015	0.0013	0.0012	0.0014	0.0014	0.0014	0.0014	0.0013
2	0.0015	0.0018	0.0017	0.0017	0.0020	0.0018	0.0014	0.0015	0.0015	0.0014	0.0016
3	0.0027	0.0027	0.0027	0.0028	0.0028	0.0028	0.0026	0.0025	0.0027	0.0026	0.0026
4	0.0038	0.0041	0.0040	0.0038	0.0037	0.0039	0.0041	0.0039	0.0038	0.0041	0.0038
5	0.0054	0.0050	0.0049	0.0050	0.0049	0.0049	0.0052	0.0051	0.0050	0.0051	0.0050
6	0.0059	0.0062	0.0067	0.0065	0.0065	0.0063	0.0063	0.0063	0.0064	0.0060	0.0062
7	0.0077	0.0064	0.0057	0.0052	0.0048	0.0078	0.0060	0.0050	0.0044	0.0071	0.0045
8	0.0152	0.0126	0.0124	0.0125	0.0110	0.0137	0.0114	0.0100	0.0093	0.0113	0.0100
9	0.0210	0.0183	0.0208	0.0163	0.0181	0.0263	0.0182	0.0155	0.0129	0.0159	0.0120

Table 6.7

Matrix M_6 with absolute means of the coefficients for motor sample 6 with defined health condition 0.4

Level	db1	db3	db5	db7	db9	sym2	sym4	sym6	sym8	coif1	coif3
1	0.0010	0.0010	0.0011	0.0010	0.0011	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
2	0.0015	0.0012	0.0012	0.0013	0.0013	0.0012	0.0012	0.0012	0.0012	0.0012	0.0012
3	0.0025	0.0019	0.0017	0.0018	0.0018	0.0022	0.0018	0.0018	0.0017	0.0020	0.0017
4	0.0051	0.0050	0.0046	0.0050	0.0049	0.0048	0.0050	0.0049	0.0048	0.0050	0.0049
5	0.0068	0.0076	0.0077	0.0076	0.0078	0.0075	0.0077	0.0075	0.0073	0.0076	0.0074
6	0.0056	0.0062	0.0064	0.0064	0.0064	0.0062	0.0063	0.0062	0.0062	0.0062	0.0059
7	0.0078	0.0063	0.0044	0.0046	0.0040	0.0068	0.0061	0.0049	0.0040	0.0073	0.0045
8	0.0163	0.0120	0.0151	0.0133	0.0155	0.0129	0.0120	0.0108	0.0098	0.0110	0.0098
	0.0159	0.0160	0.0138	0.0195	0.0133	0.0178	0.0167	0.0149	0.0123	0.0183	0.0140

6.3.3 DFI for dominant wavelets selection

The data under different health conditions from Tables 6.2-6.7 are stack up together to construct the feature matrix $M^{54\times12}$. DFI in Chapter 3 is used to select the dominant wavelets. The singular values of feature matrix $M^{54\times12}$ is calculated and shown in Table 6.8 and Figure 6.10.

. TABLE 6.8

Principal components and singular values of feature matrix $M^{54\times12}$

No	Singular Values
1	0.658
2	0.014
3	0.007
4	0.003
5	0.001
6	0.001
7	0.001
8	4.365e-04
9	2.078e-4
10	1.458e-4
11	7.868e-05
12	3.895e-05

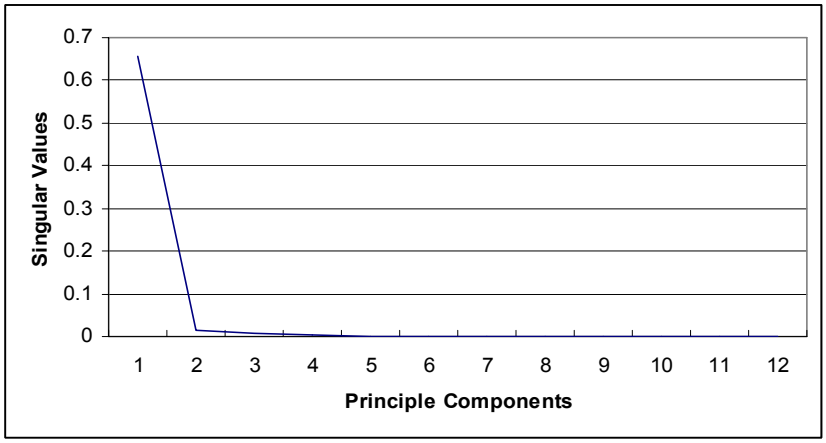


Figure 6.10: Plot of principal components versus singular values of feature matrix $M^{54\times12}$

It can be seen that the 2nd and subsequent singular values are very small as compared to the first singular values. As such, the first singular value is retained which will be sufficient to capture the relevant trend in fault detection for our experiments. However, since singular values are not directly associated with features (wavelets), it is not clear which of the 12 wavelets are required for fault

detection. DFI methodology is therefore used to select the most important wavelets (features). In our case, the dominant feature is first column and it means that the first wavelet $db1$ is selected as the dominant wavelets.

Next, the decomposition level matrix $D \in R^{6\times 9}$ is constructed with the dominant wavelet $db1$. Here y is the value of the response (heath condition of the motor under monitoring, refer the definition in Table 6.1). L_k is the predictor variable (represented coefficients of discrete wavelet transformed current signals at decomposition level k). The dimension of feature matrix $M^{54\times 12}$ has been reduced much to $D \in R^{6\times 9}$.

Table 6.9
Decomposition level matrix $D \in R^{6\times 9}$ with dominant wavelet $db1$ and 6 different health conditions samples of motors

Method Level	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	y
db1	0.0017	0.0019	0.0033	0.0040	0.0036	0.0060	0.0065	0.0117	0.0271	1
db1	0.0014	0.0024	0.0038	0.0040	0.0036	0.0061	0.0071	0.0092	0.0297	1
db1	0.0018	0.0021	0.0032	0.0039	0.0034	0.0055	0.0070	0.0138	0.0189	1
db1	0.0024	0.0026	0.0038	0.0040	0.0041	0.0060	0.0079	0.0176	0.0223	0.8
db1	0.0014	0.0015	0.0027	0.0038	0.0054	0.0059	0.0077	0.0152	0.0210	0.6
db1	0.0010	0.0015	0.0025	0.0051	0.0068	0.0056	0.0078	0.0163	0.0159	0.4

Our next objective is to use DFI to reduce the decomposition levels. The singular values of decomposition level matrix $D \in R^{6\times 9}$ are calculated and shown in Table 6.10 and Figure 6.11.

Table 6.10
Principal components and singular values of decomposition level matrix $D \in R^{6\times 9}$

No	Singular Values
1	0.12396
2	0.05605
3	0.01274
4	0.00790
5	0.00265
6	6.62283e-18
7	4.63214e-18
8	5.05323e-21
9	-5.42299e-18

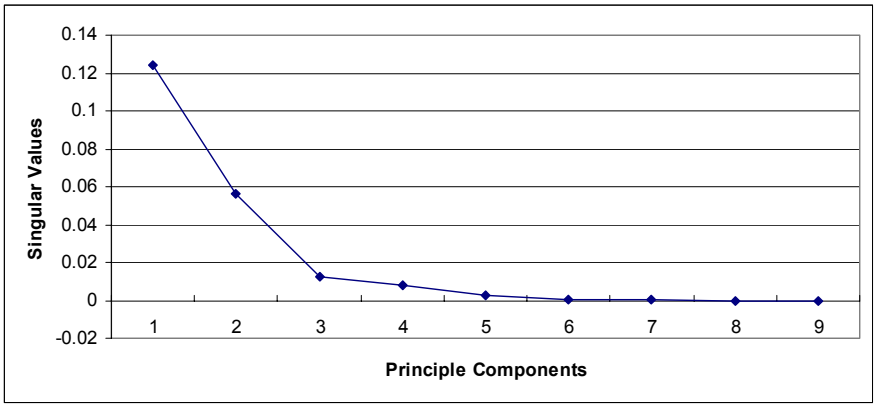


Figure 6.11: Plot of principal components versus singular values of decomposition level matrix $D \in R^{6\times 9}$

It can be seen that the 6th and subsequent singular values are very small compared to the first singular value. As such, the first 5 singular values are retained which will be sufficient to capture the relevant trend in fault detection for our experiments. However, since singular values are not directly associated with features (decomposition levels), it is not clear which of the 9 decomposition levels are required for fault detection. The DFI methodology is therefore used to select the most important decomposition levels (features). In our case, the dominant features are decomposition levels of $L_{2,}L_5, L_6, L_7, L_9$. The dominant decomposition level matrix $DD^{6\times 5}$ is shown in Table 6.11.

Table 6.11

Dominant decomposition level matrix $DD^{6\times 5}$ with dominant wavelet *db*1 (6 different health conditions samples of motors)

Method Level	L_2	L_5	L_6	L_7	L_9	y
db1	0.001855	0.003572	0.005977	0.006524	0.02708	1
db1	0.002355	0.003599	0.006089	0.007141	0.029713	1
db1	0.002146	0.00343	0.00546	0.006985	0.018882	1
db1	0.002598	0.004135	0.006044	0.007863	0.02234	0.8
db1	0.001541	0.005421	0.005898	0.007701	0.020992	0.6
db1	0.001513	0.006773	0.005597	0.007761	0.015873	0.4

Notice original feature matrix $M^{54\times 12}$ has reduced to dominant decomposition level matrix $DD^{6\times 5}$ for further multiple regression correlation

modelling. This will result large computation reduction and improve the efficiency of online fault diagnosis.

6.3.4 The correlation model for health assessment

In this Section, correlation models will be used to examine the health condition assessment with different feature sets. Two different matrixes of the original feature matrix $M^{54\times12}$ and dominant decomposition level matrix $DD \in R^{6\times5}$ are used for health condition assessment. The performance and effectiveness of DFI are evaluated based on the accuracy of prediction of the actual bearing wear.

6.3.4.1 The correlation model with feature matrix $M^{54\times12}$

Using the original 12 wavelets, 9 decomposition levels, six matrixes of $P_k^{12\times9}$ are constructed under each of six health conditions:

$$P_k^{12\times9} = (M_k^{9\times12})^T \tag{6.6}$$

six of $P_k^{12\times9}$ are constructed the feature matrix $P^{72\times9}$. A multiple regression correlation model is built using $P^{72\times9}$ for six health conditions:

$$\begin{aligned} Y = {} & 1.4174 - 83.6689L_1 + 10.4262L_2 - 84.2842L_3 - 193.8745L_4 - 0.6334L_5 \\ & - 7.6510L_6 + 4.1080L_7 + 4.1080L_8 + 6.4779L_9 \end{aligned} \tag{6.7}$$

This is obviously a complicated equation and the accuracy of the Equation (6.7) is shown in Table 6.12:

Table 6.12
Multivariable-regression-based assessment results and the actual health conditions with original feature matrix $P^{72\times9}$

Estimated Condition Using Equation (6.7)	Actual Condition	Error (%)
0.978399	1	2.16
0.961799	1	3.82
0.998266	1	0.17
0.818431	0.8	2.30
0.66427	0.6	10.71
0.378835	0.4	5.29

The average error of Equation (6.7) is 4.08% with original feature matrix of

$P^{72\times9}$. The overall accuracy of Equation (6.7) is 95.92%. The errors may be introduced from the unsuitable mother wavelet and this will be shown in Section 6.3.4.2.

6.3.4.2 The correlation model with dominant wavelet matrix $DD^{6\times5}$

After using DFI to select the dominant wavelet, only one dominant wavelet $db1$ is retained, and dominant wavelet matrix of $D \in R^{6\times9}$ can be obtained. Using DFI second time, five dominant decomposition levels $L_{2,}L_5, L_6, L_7, L_9$ are selected and those features are shown in Table 6.11. Using the features in Table 6.11, the best regression model is shown in Equation (6.8), where $L_{2,}L_5, L_6, L_7, L_9$, the features of the coefficients of the discrete wavelet are transformed current signals at decomposition levels 2, 5, 6, 7 and 9:

$$y = 2.6582 + 41.0838L_2 - 137.54L_5 - 219.884L_6 - 50.7477L_7 + 14.8538L_9 \quad (6.8)$$

The accuracy of Equation (6.8) is shown in Table 6.13. The average error of Equation (6.8) is 0.78 % with feature matrix of $DD \in R^{6\times5}$. The overall accuracy of Equation (6.8) is 99.92%. This shows that using the dominant decomposition level, the model achieves good accuracy.

Table 6.13
Multivariable-regression-based assessment results and the actual health conditions with feature matrix $DD \in R^{6\times5}$

Estimated Condition Using Equation (6.8)	Actual Condition	Error (%)
0.995	1	0.56
1.000	1	0.12
0.993	1	0.70
0.812	0.8	1.52
0.601	0.6	0.16
0.394	0.4	1.61

After build up the model equation of (6.8), six more testing motors with health conditions of 1, 1, 1, 0.8, 0.6 and 0.4 respectively were used to test the robustness and reliability of Equation (6.8). Remember that after two times of DFI, only dominant wavelet $db1$ and dominant decomposition levels of $L_{2,}L_5, L_6, L_7, L_9$ are selected and they are used in Equation (6.8) for regression modeling. For the

features extraction on those six more motors, features are not extracted as Table 6.2-6.7. Instead, only the dominant features are extracted and this saves signal processing time and reduces computational complexity. The motors were assessed using *db*1 mother wavelets with five decomposition levels $L_{2,}L_5, L_6, L_7, L_9$.

Table 6.14

Absolute means of the coefficients for motor sample 7 with defined health condition 1

Level	2	5	6	7	9
db1	0.002079	0.003785	0.005477	0.008051	0.028119

Table 6.15

Absolute means of the coefficients for motor sample 8 with defined health condition 1

Level	2	5	6	7	9
db1	0.002366	0.003477	0.005567	0.007184	0.020208

Table 6.16

Absolute means of the coefficients for motor sample 9 with defined health condition 1

Level	2	5	6	7	9
db1	0.002120	0.004365	0.005540	0.008369	0.040368

Table 6.17

Absolute means of the coefficients for motor sample 10 with defined health condition 0.8

Level	2	5	6	7	9
db1	0.002580	0.003878	0.005939	0.007459	0.021077

Table 6.18

Absolute means of the coefficients for motor sample 11 with defined health condition 0.6

Level	2	5	6	7	9
db1	0.001853	0.004931	0.005303	0.006620	0.017335

Table 6.19

Absolute means of the coefficients for motor sample 12 with defined health condition 0.4

Level	2	5	6	7	9
db1	0.001583	0.007128	0.005854	0.006842	0.014260

The dominant decomposition level matrix $DD^{6\times5}$ is constructed in Table 6.20 with the information in Tables 6.14-6.19.

Table 6.20

Dominant decomposition level matrix $DD^{6\times5}$ with dominant wavelet $db1$ (second samples of motors of six different health conditions)

Method Level	L_2	L_5	L_6	L_7	L_9	y
$db1$	0.002079	0.003785	0.005477	0.008051	0.028119	1
$db1$	0.002366	0.003477	0.005567	0.007184	0.020208	1
$db1$	0.002120	0.004365	0.005540	0.008369	0.040368	1
$db1$	0.002580	0.003878	0.005939	0.007459	0.021077	0.8
$db1$	0.001853	0.004931	0.005303	0.006620	0.017335	0.6
$db1$	0.001583	0.007128	0.005854	0.006842	0.014260	0.4

Table 6.21 compares the assessment results with actual health conditions. The average error is 7.5% with feature matrix of $DD \in R^{6\times5}$. This shows that it is able to achieve overall prediction accuracy of 92.5% using Equation (6.8) for testing motor set. Figure 6.12 shows of a graph presenting of the actual health conditions, assessment results of the training sets and assessment of testing sets.

Table 6.21

Multivariable-regression-based assessment results and the actual health conditions with feature matrix $DD \in R^{6\times5}$ (second sets of six motors)

Estimated Condition Using Equation (6.8)	Actual Condition	Error (%)
1.027824	1	2.78
0.988678	1	1.13
1.101689	1	10.17
0.859471	0.8	7.43
0.611615	0.6	3.53
0.320249	0.4	19.94

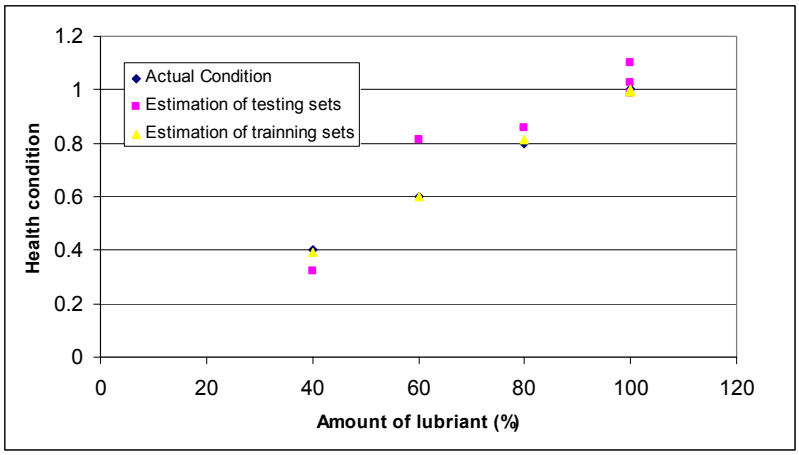


Figure 6.12: A graph presenting the actual health conditions and the assessment results predicted using Equation (6.8) (the motor rotating speed = 10,000 rpm)

6.4 Summary

In this chapter, a DFI wavelet based correlation modelling health assessment approach is presented for health state monitoring of motors. The methodology applies discrete wavelet transforms to the motor current signature to extract features represented as the coefficients of different levels of wavelet decompositions. DFI is used two times to select both dominant wavelets and decomposition levels. The methodology is equipped with a multivariable regression method to obtain the best and simplest correlation model between the features and faults. This approach has been tested experimentally with the assessment of the degradation of fluid dynamic bearings used in brushless DC motors. Experimental results show that an acceptable assessment of the motor health condition has been achieved. The prediction errors using the correlation equations reach to 0.78% for training motor sets and 7.5% for testing motor sets.

CHAPTER 7 REINFORCED MORLET WAVELET TRANSFORM FOR BEARING FAULT DIAGNOSIS

In today's highly globalize society; rotary machines have established a reputation of being a class of equipment with the most widespread applications, such as generators in power plants, stamping machines in factories and mechanical ventilation systems in hospitals. All the above vital operations have created a continuous demand for more efficient, durable, reliable and safer rotating machines. As such, vast amounts of time and effort have been invested to develop rotary machines with higher efficiencies and improved robustness to prevent down time and loss of productivity [185] [186]. However, mechanical faults arise due to wear and tear from prolonged operations over time, causing the machines to operate at lower efficiencies with undesirable effects such as excessive vibrations and noises. In more severe instances, the machines might even experience critical failure and breakdown, even posing possible work hazards to their operators.

Rotary machines shafts are supported inside housings by bearings. Rolling element bearings are the most critical and valuable component in rotary equipment. The considerable attention paid to proper selection and installation of bearings and the construction of bearing housing makes the rotary equipment one of the most expensive parts of the machine tools [187].

Bearing faults are ranked as the most common faults in induction machines [188]. Rolling-element bearings produce very little vibration (low level random signal) when they are fault free, and have very distinctive fault frequency responses [189]. There are four bearing fault characteristic frequencies, or called "bearing tones", named as:

- cage frequency (f_c): indicates a fault at bearing cage;
- ball pass frequency for outer raceway (f_{BPFO}): indicates a fault at outer raceway;
- ball pass frequency for inner raceway (f_{BPFI}): indicates a fault at inner raceway;
- ball spin frequency (f_{BSF}): indicates a fault at rolling elements.

The equations of the fault characteristic frequencies with regard to the bearing dimensions are given in references [4, 190]. The difficulty in detecting bearing faults is caused by the modulation between the fault characteristic frequencies and the structure resonance frequency of the machine. A four stage failure model is introduced in reference [4], from beginning of material fatigue to severe bearing failure, and specially described the phenomenon of modulation. In the early and median fault stages, the structure resonance frequency acts as the carrier frequency and the fault characteristic frequency acts as the modulating frequency. The fault characteristic frequency is shown as the sidebands of the structure resonance frequency in the frequency spectrum, which are hard to be recognized.

To deal with the modulation phenomenon, envelope analysis (demodulation) is implemented to derive the modulating signal from the modulated signal [191]. The goal of the enveloping is to replace the oscillation caused by each impact with a single pulse over the entire time of the impact response [192]. Hilbert-Huang transform is one of the most common demodulation methods for machinery fault detection [193, 194]. A major disadvantage of Hilbert-Huang transform is that the selection of parameters is based on the experience of the end user, and the selection of central frequency and bandwidth of band-pass filter carries great subjectivity that would bring demodulation error [193].

Wavelet transform has received considerable attention from the research community due to its ability in extracting time-dependent transient features from vibration signals with strong background noise [83, 85, 195]. By means of a combined time-frequency analysis, the wavelet-based demodulation gives a description of how energy distribution over the changes of frequencies from one instance to another [196, 197]. Wavelet transform is classified into continuous wavelet transform (CWT) and discrete wavelet transform (DWT) [77]. The best results are obtained by CWT and the high computational complexity of the CWT is not a serious worry for today's computer power. Within the category of CWT, Morlet wavelet has a relatively lower error rate for extracting impulse information in vibration signals [196, 199]. Morlet wavelet transform is generally conducted in four steps:

(1) Select optimal parameters of the optimal shape factor σ and the scale factor α to construct the mother wavelet,

(2) Choose a level N and decompose the original signal into several resonance bands,

(3) Determine threshold detail coefficients: for each level from 1 to N, select a threshold and apply soft threshold to the detail coefficients.

(4) Reconstruct the envelope signal: reconstruct the signal using the modified detail coefficients of levels from 1 to N.

Difficulties in implementing Morlet wavelet transform include selection of the optimal shape factor σ and the scale factor α in step (1), and determine the threshold and reconstruction of the envelope signal in steps (3) and (4).

The objective of this chapter is to develop reinforced Morlet wavelet transform for bearing fault diagnosis to overcome the difficulties in implementing Morlet wavelet transform, achieve high consistency and detect bearing fault in an early stage.

The remaining sections of this Chapter are organized as follows: In Section 7.1, the existing approach of Morlet wavelet transform is reviewed, and the two shortcomings are analysed using a case study. In Section 7.2, two approaches to overcome these shortcomings are proposed, and new algorithms of reinforced Morlet wavelet transform are introduced. In Section 7.3, the proposed reinforced Morlet wavelet transform is validated using two case studies: a milling machine spindle test and a bearing run-to-failure test. The results demonstrate that the reinforced Morlet wavelet transform can provide the fault detection consistency with early fault detection ability.

7.1 Morlet wavelet transform: issue of the existing approach

7.1.1 Morlet wavelet

The mother wavelet of the Morlet wavelet is given as a sinusoidal wave multiplied by a Gaussian function:

$$\varphi(t) = \frac{\sigma}{\sqrt{\pi}} e^{-\sigma^2 t^2} e^{i2\pi f} \tag{7.1}$$

Wavelet decomposition is formulated as the convolution of a finite energy data $x(t)$ with the scaled and conjugated mother wavelet $\varphi^*(\frac{t}{\alpha})$ in (7.2):

$$Cw(\sigma,\tau) = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{+\infty} x(t)\varphi^*(\frac{t-\tau}{\alpha})dt \tag{7.2}$$

Equation (7.1) implies that Morlet wavelet transform decomposes a given signal into N resonance bands with a centre frequency f_i, $(i = 1,2,...N)$. In equations (7.1) and (7.2), σ determines the shape of the mother wavelet; α determines the width of the wavelet window in Fourier spectrum as $[f_i - \frac{\alpha}{2}, f_i + \frac{\alpha}{2}]$. α is the two adjacent resonance bands of the centre frequency of the mother wavelet.

Equation (7.2) can be performed in Fourier domain to avoid the difficulty of selecting the valid data from the convolution result. $\Psi_i(f)$ is the Fourier transform of the mother wavelet in equation (7.1), which is a real function formulated as:

$$\Psi_i(f) = \Psi_i^*(f) = e^{-\frac{\pi^2}{\sigma^2}(f-f_i)^2} \quad (i = 1,...,N) \tag{7.3}$$

An alternative form of convolution is:

$$Cw(i,j) = \left| F^{-1}\left\{ X(f)\psi^*(\alpha f) \right\} \right| \quad (i = 1,...N, j = 1,...M) \tag{7.4}$$

where

Cw is the wavelet coefficients matrix;

F^{-1} denotes inverse Fourier transform;

$X(f)$ is the Fourier transform of the original signal $x(t)$;

N indicates that the original signal $x(t)$ is decomposed into N resonance

bands;

M is the number of time instants.

Based on the wavelet coefficients, an envelope signal can be reconstructed which is expected to contain fault-induced information.

7.1.2 The existing approach

7.1.2.1 Selection of N and f_i

As the Nyquist rate determines the largest frequency component contained in the signal, the largest centre frequency f_N should be smaller than the Nyquist rate. The number of resonance bands N is selected as $2N$ which is close to the number of samples per motor cycle, and f_i is selected as $2i-1$ times of the shaft rotational frequency for scanning a full range of frequency complying with the Nyquist criterion [191]. Write in the formula presentation as:

$$N = nextpow2\left(\frac{f_s}{f_R}\right) - 1 \tag{7.5}$$

$$f_1 = f_R \tag{7.6}$$

$$f_{i+1} = 2f_i \tag{7.7}$$

where f_s denotes the sampling frequency; f_R denotes the shaft rotational frequency.

7.1.2.2 Selection of optimal α and σ

From equation (7.2), the range of selecting α is [1, 2). The range of selecting σ can be determined from equation (7.3). If σ is greater than the Nyquist rate, the wavelet window will contain an unnecessary DC component which makes the final spectrum closer to the Fourier spectrum. As a result, the range of selecting σ is [f_R, Nyquist rate]. Nikolaou and Antoiadis [192] selected the optimal α and σ values using the entropy magnification combined factor criterion, which is a combination of the Shannon entropy and the magnification factor. The rationale is explained in references [192, 196]. Shannon entropy is formulated as

$$E(\alpha,\sigma) = -\sum_{i,j} d_{i,j} \log d_{i,j} \tag{7.8}$$

where

$$d_{i,j} = \frac{Cw(i,j)}{\sum_{i,j} Cw(i,j)} \tag{7.9}$$

The magnification factor is formulated as

$$C_r(\alpha,\sigma) = \frac{MN \ \max(\ Cw\ (i,j))}{A(\alpha,\sigma)} (i = 1,...,\ N,\ j = 1,...M\ ) \tag{7.10}$$

$$A(\alpha,\sigma) = \sqrt{\sum_{i,j} Cw(i,j)^2}\ \ (i = 1,...,N, j = 1,...,M) \tag{7.11}$$

where MN is the number of elements of C_w. The combined factor is formulated as:

$$f_{sc}(\sigma) = \frac{Cr(\sigma)}{E(\sigma)} \tag{7.12}$$

The optimal α and σ values lead to the maximum f_{sc} value.

7.1.2.3 Reconstruction of envelope signal

Nikolaou and Antoiadis [192] proposed two variations for reconstructing the envelope signal.

A. Max-resonance variation

In this variation, a specific decomposed signal forms the envelope signal. This optimal decomposed signal is chosen using a soft threshold criterion. For each row i of C_w, $n_{th}(i)$ represents the number of elements that exceed a threshold t_{hr}:

$$t_{hr} = m + 2s_t \tag{7.13}$$

$$n_{th}(i) = length(C_w(i,j) > t_{hr})\ ,\ \ (j = 1,...M) \tag{7.14}$$

where m and s_t are the mean value and the standard deviation of all the elements in C_w respectively.

The optimal row number i_0 is calculated as:

$$n_{th}(i_0) = \max(n_{th}(i)), \qquad (i = 1,...N) \tag{7.15}$$

The envelope signal is the row i of C_w.

B. Max-envelope variation

The envelope signal $x_{\max}$ is formed as

$$x_{\max}(j) = \max(C_w(i, j)) \qquad (i = 1,..., N) \ \ and \ \ (j = 1,..., M) \tag{7.16}$$

The basic idea is to detect the largest excitation in all the resonance bands at each time instant.

7.1.2.4 Two shortcomings of the existing approach

A. Inappropriate selection of optimal α and σ values

The existing approach is applied to a case study of a spindle with a cage fault which is described in the case study Section 7.3.1. The shaft rotates at 2400 rpm. The cage frequency is 18.25 Hz. Figure 7.1 shows the resulted f_{sc} distribution. Three peaks with small σ values are marked in Figure 7.1. $\alpha = 1.0$ and $\sigma = 114$ are selected using the existing approach. The selected parameters lead to the spectrum shown in Figure 7.2 (a), which does not show clear spike at the cage frequency.

On the other hand, if the parameters change to $\alpha = 1.1$ and $\sigma = 950$, which refers to the point labelled by a red arrow in Figure 7.1, the resulted spectrum is shown in Figure 7.2(b). The spike at the cage frequency can be obviously seen.

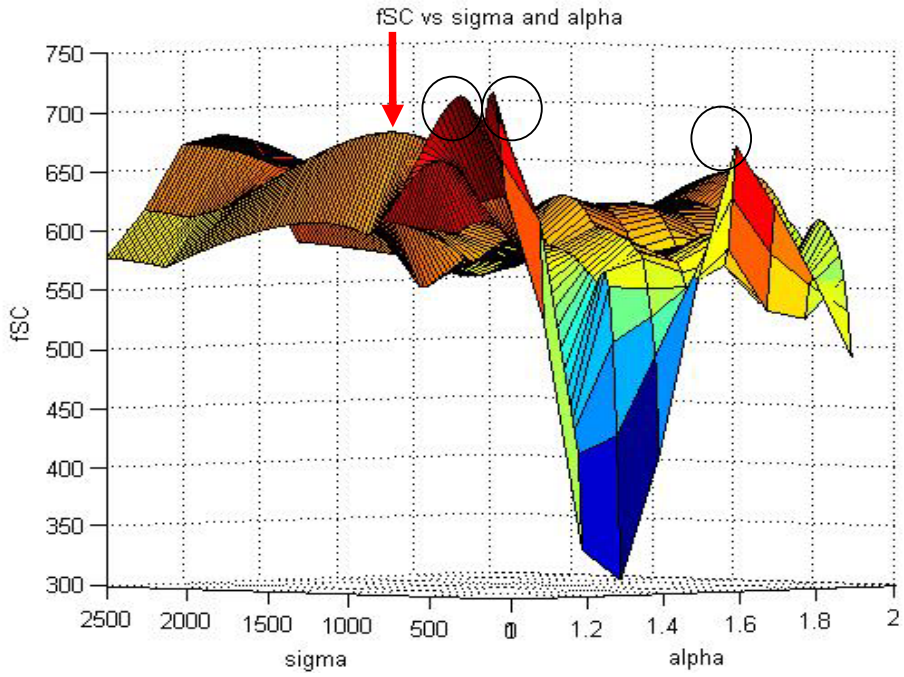


Figure 7.1: f_{sc} distribution of the dataset running at 2400 *rpm*

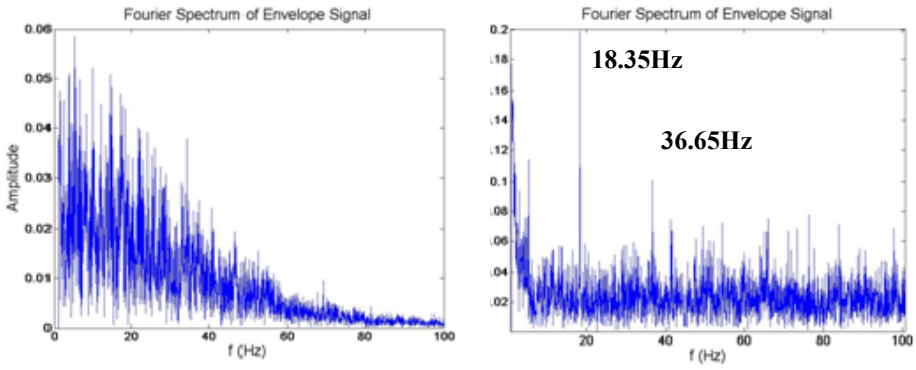


Figure 7.2: (a) Spectrum of dataset 2400 *rpm* with $\alpha = 1.0$, $\sigma = 114$
(b) Spectrum of dataset 2400 *rpm* with $\alpha = 1.1$, $\sigma = 950$

The results in Figure 7.2 (a) and (b) imply that inappropriate selection of the optimal α and σ values using the existing approach, which are the ones that result from the maximum f_{sc} value. According to equations (7.8) – (7.12), f_{sc} is a monotonic decreasing function of the average of the wavelet coefficients, and a high f_{sc} value may also result from a low average value of the wavelet coefficients. When σ is small, the mother wavelet contains little frequency

component, and the average of the wavelet coefficients is small. However, certain wavelet coefficients may still have high values because of the existence of noise. These two factors can generate a fake peak in the f_{sc} distribution.

B. Inappropriate reconstruction of envelope signal

Wavelet transform decomposes the raw signal into N signals with individual centre frequencies. One of the N decomposed signals contains fault-induced information, while others are composed of noise. The max-resonance variation selects the fault-induced information from the decomposed signal using a soft-threshold criterion; however, this criterion may select an incorrect decomposed signal. In this example, the spindle is known to contain a bearing cage fault, and the cage frequency is 23.58Hz.

Figure 7.3 (a) shows its spectrum achieved using max-resonance variation. The soft-threshold algorithm selects the decomposed signal with centre frequency of 206.7Hz. There is no spike at the cage frequency in its spectrum. However, as shown in Figure 7.3 (b), the spectrum of the decomposed signal with a centre frequency of 1651Hz contains a clear spike at the cage frequency. This means the soft-threshold criterion chooses the incorrect resonance band.

Nikolaou and Antoiadis also proposed the max-envelope variation, which uses the information of all decomposed signals so as to avoid the risk of incorrect selection. However, as shown in Figure 7.3 (c), the spectrum achieved using the max-envelope variation does not show spike at the cage frequency. This is because only one of the decomposed signals contains the fault-induced information and all others are composed of noise. The combination of all decomposed signals in the time-domain adds the noise back to the envelope signal, and the max-envelope variation actually reduces the signal-to-noise ratio of the envelope signal.

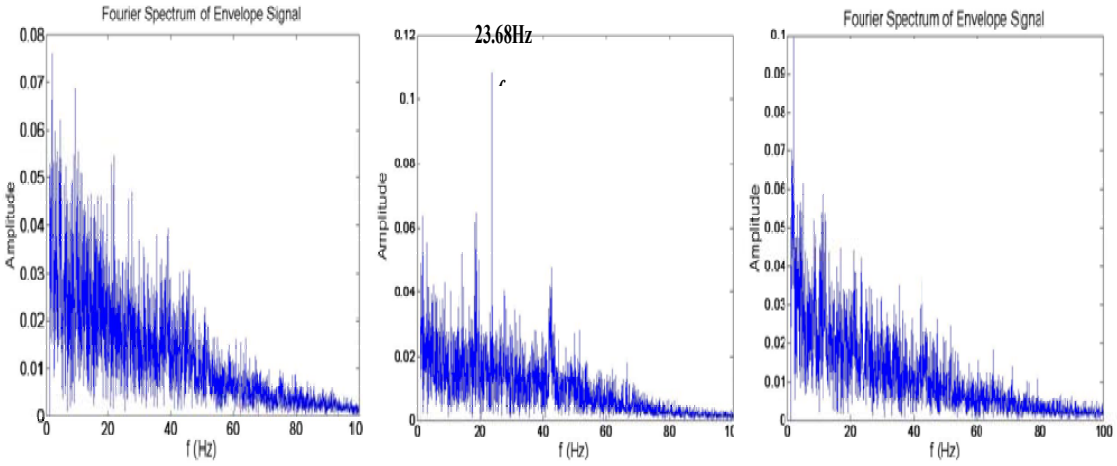


Figure 7.3: (a) Spectrum of decomposed signal with centre frequency 206.7 *Hz*
(b) Spectrum of decomposed signal with centre frequency 1651 *Hz*
(c) Spectrum achieved using max-envelope variation

7.2 Improvement of Morlet wavelet

Two improvements are proposed to overcome the problems that mentioned in Section 7.1. The first improvement overcomes the problem of inappropriate selection of optimal α and σ values. The second improvement overcomes the problem of inappropriate reconstruction of envelope signal.

7.2.1 Improvement one: threshold magnification factor criterion

The threshold magnification factor criterion aims to increase the correct rate of selecting optimal α and σ values. The idea is to filter the α and σ values which result in wavelet coefficients with small average value, so as to reduce the probability of incorrect parameter selection caused by the irrelevant peaks in the f_{sc} distribution. Recall formulas (7.8)-(7.11) to calculate magnification factor C_r :

$$E(\alpha,\sigma) = -\sum_{i,j} d_{i,j} \log d_{i,j} \tag{7.17}$$

where

$$d_{i,j} = \frac{Cw(i,j)}{\sum_{i,j} Cw(i,j)} \tag{7.18}$$

The magnification factor is formulated as

$$C_r(\alpha,\sigma) = \frac{MN \ \max(\ Cw(i,j))}{A(\alpha,\sigma)} (i = 1,...,\ N,\ j = 1,...M\ ) \tag{7.19}$$

$$A(\alpha,\sigma) = \sqrt{\sum_{i,j} Cw(i,j)^2}\ \ (i = 1,...,N, j = 1,...,M) \tag{7.20}$$

where MN is the number of elements of C_w.

The threshold magnification factor C_{rt} is calculated as the magnification factor C_r multiplied by a correction matrix C_o:

$$C_{rt}(\alpha,\sigma) = C_o(\alpha,\sigma)C_r(\alpha,\sigma) \tag{7.21}$$

C_o is defined using a soft-threshold criterion:

$$C_o(\alpha,\sigma) = \begin{cases} 0 & if \quad A(\alpha,\sigma) < T \\ 1 & if \quad A(\alpha,\sigma) \geq T \end{cases} \tag{7.22}$$

$$T = mean(A(\alpha,\sigma)), \qquad \text{for all } \alpha \text{ and } \sigma \text{ values} \tag{7.23}$$

where A(α,σ) is the average value of wavelet coefficients in equation (7.20).

Figure 7.4 shows the C_{rt} distribution of the dataset at 2400 *rpm*, which is the example data set used in Section 7.1.2.4. The irrelevant peaks are filtered. A searching algorithm is used to search for the peak C_{rt} value. The corresponding optimal values are α = 1.1 and σ = 950.

7.2.2 Improvement two: max-energy spectrum

In improvement two, the max-energy spectrum is developed to be the final spectrum of the vibration signal, which is expected to extract the fault-induced information. This improvement aims to eliminate the risk of selecting incorrect decomposed signal. The idea is that although it is hard to differentiate the N decomposed signals in time domain, the Fourier spectrum of the decomposed signal will show a high energy spike at the fault characteristic frequency, while the Fourier spectrum of other signals have overall low energy level similar to the noise floor. The max-energy spectrum is formed by conducting Fourier transform to all resonance band signals, and then selecting the maximum energy level among all

spectrums at each frequency point. This algorithm avoids adding noise to the correct decomposed signal. In equation presentation:

$$C_s(k) = \max(G_i(k),\ i = 1,....,N), \quad k = 1,...,M/2 \tag{7.24}$$

$$G_i = F(Cw_i) \quad i = 1,...,N \tag{7.25}$$

where

C_s is the combined spectrum

F denotes Fourier transform of each row of C_w

N is the number of resonance bands

$M/2$ is the number of frequency points in the spectrum.

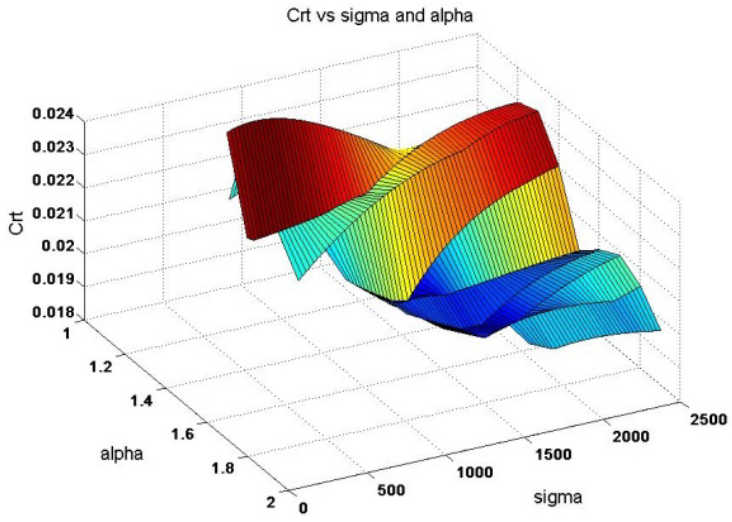


Figure 7.4: C_{rt} distribution of data set 2400 *rpm*, Ch1

Figure 7.5 shows the combined spectrum of dataset at 3100 *rpm*, which is the same example dataset in Section 7.1.2.4. The noise floor in Figure 7.5 has higher energy level compared with that in Figure 7.3 (b), but this does not influence the appearance of the spike at the cage frequency.

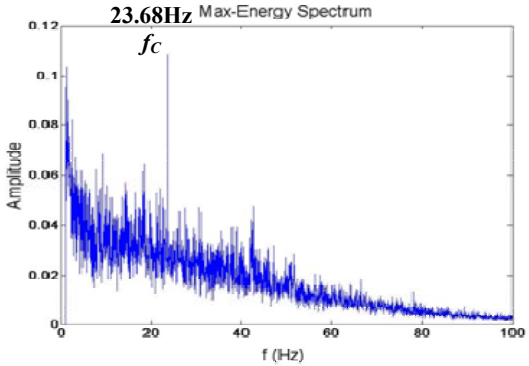


Figure 7.5: Max-energy spectrum of data set 3100 *rpm*, Ch1

7.2.3 Reinforced Morlet Wavelet Transform

We named the proposed approach as "reinforced" because it reinforces the existing Morlet wavelet with both parameter selection criterion and final spectrum construction. The reinforced Morlet wavelet transform is composed of following steps:

Algorithms: Reinforced Morlet wavelet transform INPUT: original measured vibration signal OUTPUT: max-energy spectrum C_s
Select the decomposition level N using equation (7.5). Select the centre frequency f_i, $(i = 1,2,...N)$ using equation (7.7).

// Select optimal shape factor σ and the scale factor α:
Select optimal shape factor search range σ as [f_R, Nyquist rate] and scale factor α search range as [1, 2)
Select initial values $\alpha_0 = f_R$ and $\sigma_0 = 1$
WHILE $\alpha_0 \leq$ Nyquist rate and $\sigma_0 < 2$ DO
Conduct Wavelet transform
Calculate Shannon entropy of wavelet coefficient using equations (7.17)-(7.23)
ENDWHILE

Search for the minimal Shannon entropy and select the corresponding optimal shape factor σ and the scale factor α
Conduct wavelet transform with the optimal α and σ values to get the wavelet coefficient matrix C_w
Form the max-energy spectrum C_s using Equations (7.24)-(7.25)

7.3 Case studies

7.3.1 Case study one: fault diagnosis of machine spindle

When a machine is operating under different shaft rotational speed, a consistent approach should give the same result about the health condition of the machine. Thus, the consistency of the algorithm is measured using the fault-identification rate, which is calculated as

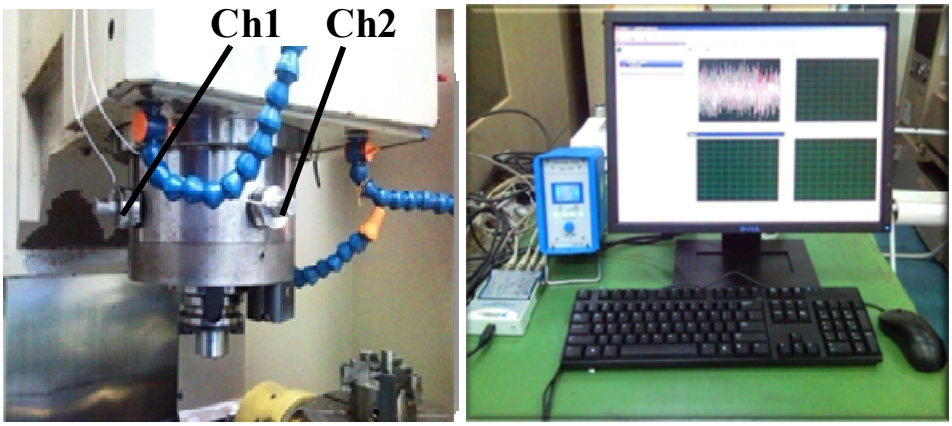
$$fault-identification-rate = \frac{number\ of\ \ fault\ identified\ datasets}{total\ number\ of\ tested\ datasets}$$

A case study is conducted to compare the consistency of the existing Morlet wavelet approach with the proposed reinforced Morlet wavelet transform approach.

7.3.1.1 Experimental setup

A CNC milling machine test bed (Figure 7.6) was used in this case study. Figure 7.7 shows the experimental setup; Ch1 and Ch2 indicate two accelerometers which are mounted on the orthogonal sides of the spindle housing to collect vibration signal from two orthogonal directions. Channel coupler and USB carrier are used for signal acquisition and PC communication. Low pass filter of 10 kHz is applied for all channels and sampling rate of 5000Hz is used during data acquisition.

Figure 7.6: Spindle condition monitoring of the high speed milling machine



(a) **(b)**

Figure 7.7: (a) the position of accelerometers (b) the DAQ system

The specification of the two bearings near Ch1 and Ch2 of this milling machine is given blow:

Table 7.1
Specification of the two bearings

Roller diameter	11mm
Pitch diameter	126mm
Contact angle	0 degrees
Number of rolling elements	60

7.3.1.2 Experiment procedures

Steady state rotation signals were captured using the same setup, with different shaft rotational speeds varying from 2400 *rpm* to 4000 *rpm* at 100 *rpm* interval. Total 34 sets of data were recorded with 2 sensors and 17 shaft rotational speeds. The data were analysed using the existing Morlet approach and the proposed reinforced Morlet wavelet transform separately. For each approach, faults detection rate and the nature of the faults were analyzed from 34 sets of data.

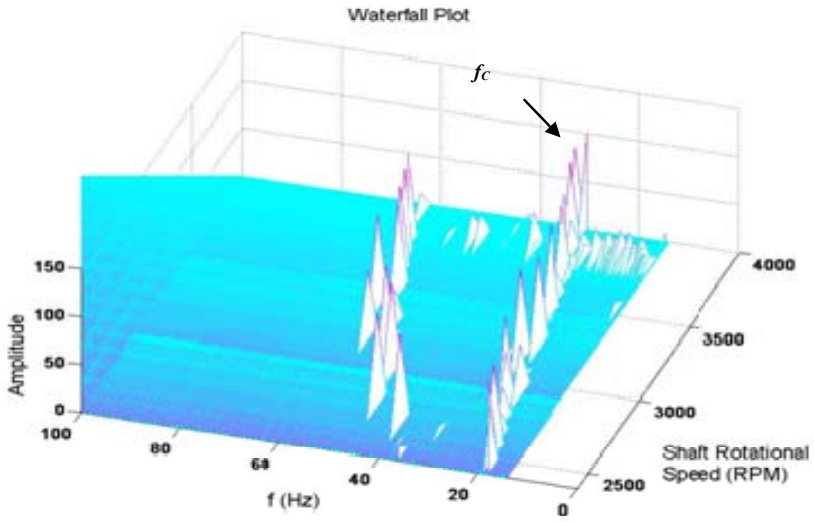


Figure 7.8: Waterfall plot of Ch1 data analysis using reinforced Morlet wavelet transform

7.3.1.3 Results and discussion

Figure 7.8 shows waterfall plot of Ch1 data using reinforced Morlet wavelet transform. The "hill" of spikes on the right side indicates the cage frequencies, which implies that the spindle has a bearing cage fault. This study was carried out on both Ch1 and Ch2 datasets to reveal the cage fault. The comparison study results that using different methods are summarized in Table 7.2.

The existing Morlet approach using max-resonance detected 6 faults among the 34 datasets. The existing Morlet approach using max-resonance detected 4 faults among the 34 datasets. The lower fault-detection rates 17.7 and 11.8 imply that the existing approach does not produce consistent fault detection results. The reinforced Morlet wavelet transform detects 23 faults out of 34 datasets, achieves 67.6% of fault-detection rate. The results show that reinforced Morlet wavelet

transform increases the fault detection consistency significantly.

Table 7.2

Experiment results of the milling machine spindle

Approach used	No. of results reveal the fault	Fault-detection rate
Existing approach (using max-resonance variation to construct the envelope signal)	6 out of 34	17.7%
Existing approach (using max-envelope variation to construct the envelope signal)	4 out of 34	11.8%
Reinforced Morlet wavelet transform	23 out of 34	67.6%

7.3.2 Case study two: bearing run-to-failure test

In the real application, it is more important for a methodology to discover the fault propagation in the early fault stages. Bearing run-to-failure test [196] was used to validate the early-fault-detection ability of the reinforced Morlet wavelet transform.

7.3.2.1 Experimental setup

The experiment was conducted by NSF I/UCR Center of Intelligent Maintenance Systems (IMS) with support from Rexnord Corp. in Milwaukee, WI. The detailed experimental setup provided by [196] is shown in Figure 7.9. In order to truly reflect the real defect propagation processes, bearing run-to failure tests were performed under normal load conditions on a specially designed test rig. The bearing test rig hosts four test bearings on one shaft. The shaft is driven by an AC motor and coupled by rub belts. The rotation speed was kept constant at 2000 rpm. A radial load of 6000 lbs. is added to the shaft and bearing by a spring mechanism. All the bearings are force lubricated. An oil circulation system regulates the flow and temperature of the lubricant. A magnetic plug installed in the oil feedback pipe collects debris from the oil as evidence of bearing degradation. The test will stop when the accumulated debris adhered to the magnetic plug exceeds a certain level and causes an electrical switch to close. Four Rexnord ZA-2115 double row bearings were installed on one shaft as shown in Figure 7.9. The bearings have 16 rollers in each row, a pitch diameter of 2.815 inch, roller diameter of 0.331 inch, and a tapered contact angle of 15.17. A PCB 353B33 high sensitivity quartz ICPs

accelerometer was installed on each bearing housing. Four thermocouples were attached to the outer race of each bearing to record bearing temperature for monitoring the lubrication purposes. Vibration data was collected every 20 minutes by a National Instruments DAQ Card-6062E data acquisition card. The data sampling rate is 20 kHz and the data length is 20480 points. Data collection is conducted by a National Instruments LabVIEW program [196].

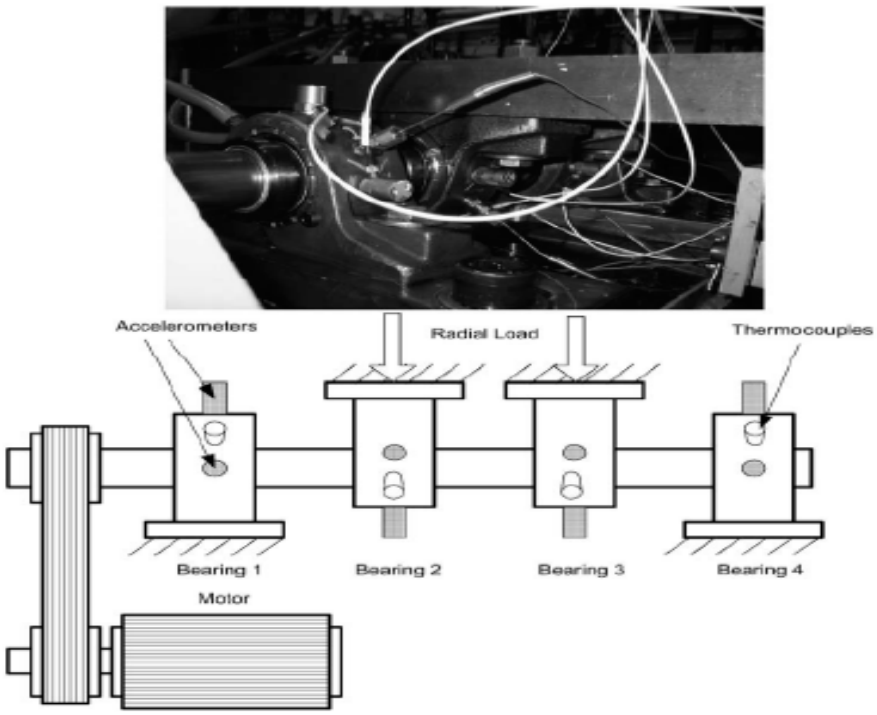


Figure 7.9: Bearing test rig [196]

The experiment recording duration was from 12 Feb 2004, 10:32 to 19 Feb 2004, 06:22. Vibration signals at bearing 1 to bearing 4 were recorded using four channels (Ch1 to Ch4) respectively. In this case study, the reinforced Morlet wavelet transform processed all the data from these four channels to observe the fault propagation.

7.3.2.2 Results and discussion

Table 7.3 summarizes the signal analysis results of the machine condition. "BPFO" means bearing outer-raceway fault characteristic frequency detected, which indicates the existence of a bearing outer raceway fault. The blank field means that no bearing fault is detected. Before 03:12 on 16th Feb, the max-energy spectrums have overall low amplitude and do not show appearance of any bearing fault. On 16th Feb 03:12, the max-energy spectrum of Ch1 data contains a clear single spike at BPFO, which is shown in Figure 7.10. The BPFO value (231.5Hz) is very close to the calculated value (236.4Hz). This indicates the occurrence of an outer-raceway fault at bearing 1. Figure 7.11 shows the spectrum of Ch 1 data on 17th Feb 00:02, which contains the BPFO and its harmonics. This indicates further degradation of bearing 1. Calculating from the dates and time, bearing fault is first revealed on Feb 16th and the remaining operating life is three days. Three days early fault detection will provide enough time for machine shut down and conducting maintenance.

Table 7.3
Experiment results of the run to failure test

Date & Time (in Feb.)	Ch1 (bearing 1)	Ch2 (bearing 2)	Ch3 (bearing 3)	Ch4 (bearing 4)
12th 10:32				
16th 03:12	Clear BPFO single spike			
17th 00:02	BPFO with harmonics	Unclear BPFO single spike		Unclear BPFO single spike
18th 00:02			Unclear BPFO single spike	Clear BPFO single spike
19th 00:02		BPFO with harmonics	BPFO with harmonics	BPFO with harmonics
19th 04:02				
19th 06:02				
19th 06:22				

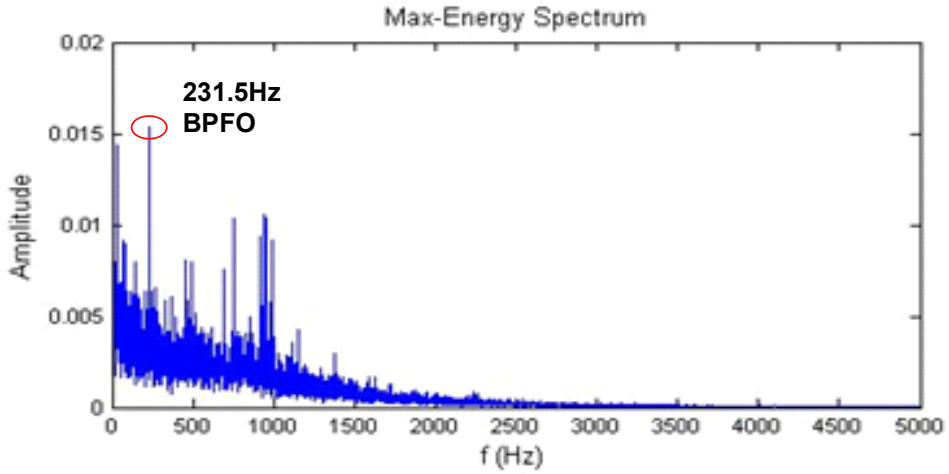


Figure 7.10: Spectrum of Ch1 data on 16th Feb 03:12

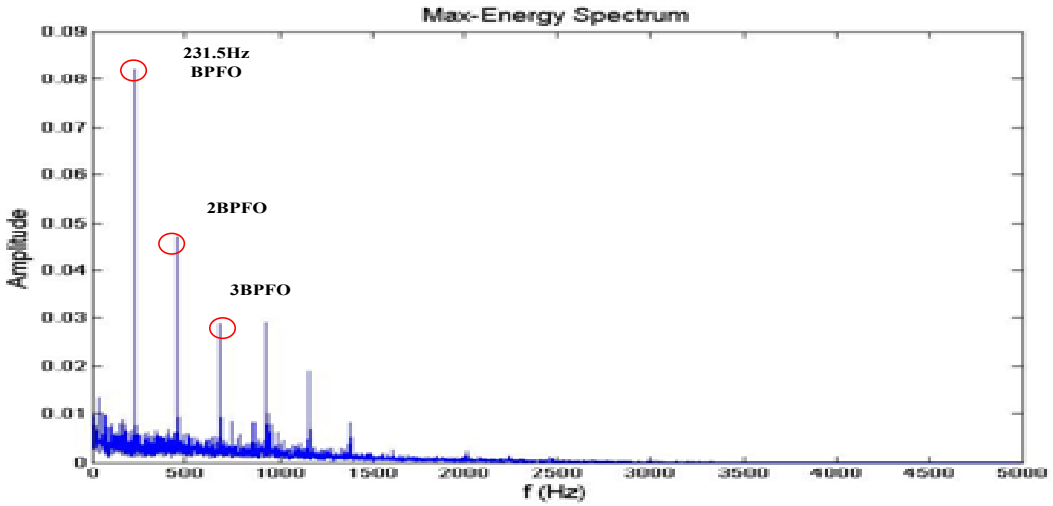


Figure 7.11: Final spectrum of Ch1 data on 17th Feb 00:02

The results show the early fault detection ability of the reinforced Morlet wavelet transform. The results of the last data set on 19th Feb, 06:22 do not show defect on any bearing. This can be explained by the "healing" phenomenon [196]. The bearing defect develops from small surface pits at the beginning, to spalls or cracks in later stage. Then the defect may be smoothed by continuous fraction between bearing components.

7.4 Summary

In the present work, the reinforced Morlet wavelet transform is developed to overcome the two shortcomings of the existing approach. The threshold magnification factor criterion reduces the error rate in selecting the optimal α and σ values by filtering out wavelet coefficients with small average value. The max-energy spectrum avoids the risk of adding noise to decomposed signal by combining the information of all decomposed signals in frequency domain. The results of the two case studies verify the performance of the method that the fault detection consistency has increased significantly from existing 11.8% and 17.7% to 67.6%. It also achieves three days early fault detection before failure which is critical for repair decision making.

CHAPTER 8 IDIAGNOSIS & PROGNOSIS - AN INTELLIGENT PLATFORM FOR COMPLEX MANUFACTURING

Present day manufacturing is usually a complex manufacturing system. It consists of a large number of components, parts and modules. An example is shown in Figure 8.1 which consists of assembly line, robot subsystem, different machine centers, and mobile vehicle systems for work in progress parts transfer and storage subsystem.

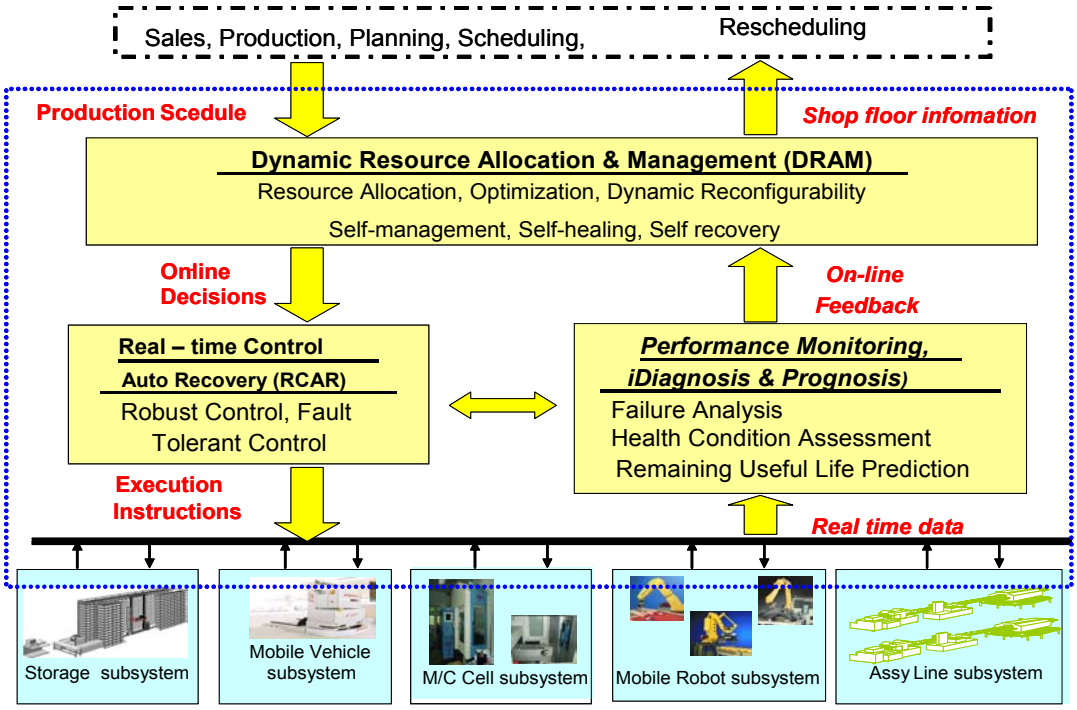


Figure 8.1: Scenario of complex manufacturing system

This complex manufacturing system is usually facing many different types of disruptions. The disruptions arise from shop floor dynamics such as material shortages, machine breakdown, performance degradation, shortage of operators, etc. This entails time-driven, event-driven and demand-driven monitoring of the operation of complex manufacturing systems, as well as their working conditions, tendency of performance degradations, minor faults, major failures, slow response,

and so on. In order to provide more efficient planning and scheduling, dynamic changes on the shop floor are needed to be taken into consideration.

To achieve the quality of service of a complex manufacturing system, an integrative approach is proposed to manage the performance and serviceability of manufacturing system of systems (MSoS), by taking into consideration of subsystem health condition prognosis when making resource allocation decisions during execution [201]. The integrative approach consists of three modules of real-time self-recovery and control, real-time optimal resource managements and *iDiagnosis/Prognosis*. The system realizes real-time optimal resource management of task allocation, dispatching and re-dispatching, taking equipment health condition into consideration to ensure the execution of manufacturing plans and schedules in a cost effective manner with maximal serviceability. In this chapter, the research issues of diagnosis/prognosis of complex manufacturing systems are mainly discussed. Detection and identification of various faults in complex manufacturing systems require many tools and different methods and some of the methods have been discussed in Chapters 2, 3 and 4. To achieve desirable results, a framework is required to enable the developed algorithms and their integration in an efficient manner and that is the proposed *iDiagnosis/Prognosis* in this chapter.

In a complex manufacturing environment, in order to make the dynamic resource allocation, three most important elements of the control inputs are the system/equipment health condition, system capability and job execution capability of the manufacturing systems. This is because any unexpected system/equipment down in the process chain will have a ripple effect to the downstream processes and can be highly disruptive to the shop floor as equipment, labour and materials have to be rescheduled and reallocated. Detection and identification of various faults in complex manufacturing systems require many tools and different methods.

Traditional maintenance schemes based on the “fail-and-fix” or “time based maintenance” is no longer effective and adequate for networked virtual enterprises. The terms e-maintenance and e-diagnosis have emerged since early 2000 and are now very common terms in maintenance-related literature [202]. Over recent ten years, extensive research and development work has been devoted to e-maintenance. Researchers consider e-maintenance the gradual replacement of

traditional maintenance types [21] by more predictive / proactive types. Regular periodic maintenance should be advanced and shifted to the intelligent maintenance philosophy to satisfy the manufacturer's high reliability requirements [22]. Hence, Koc and Lee [23] referred e-maintenance (system) as predictive maintenance (system), which provides monitoring and predictive prognostic functions. These e-maintenance systems provide the equipment monitoring, diagnosis and prognostics functions, but none of them has been reported that provides the manufacturing system condition and job execution capability assessment functions. Besides the researches on the monitoring and predictive prognostic functions of e-maintenance, many researches focused on the e-maintenance platform development based on the internet, web, and agent technologies. Zhang et al. [24] consider that e-maintenance is a combination of web service technology and agent technology, which provides a way to realize intelligent and cooperative features for the systems in an industrial automation system. Crespo Marquez and Gupta [25] define e-maintenance as a distributed artificial intelligence environment, which includes information processing capability, decision support and communication tools, as well as the collaboration between maintenance processes and expert systems.

Reviewing the literature and considering the dynamic changing on the shop-floor and the challenge faced by the supervisory control of complex manufacturing systems, a new concept of *iDiagnosis/Prognosis* is proposed. The *iDiagnosis/Prognosis* provides distributed intelligence of diagnosis, prognosis, system condition and job execution capability assessment, and provides the dynamic information feedback to the supervisory control when manufacturing system breakdown or performance degradation occurs. In collaboration working with other modules in the supervisory control system, the functionalities of iDiagnosis & Prognosis include:

1) Work as an intelligent agent for information processing, decision support, accept, and react to high-level inputs.

2) Access and remotely collaborate with dynamic resource allocation (DRAM) [201] and maintenance management.

3) Perform performance monitoring, diagnosis and prognosis. Provide dynamic information of manufacturing system capability and job execution capability to other modules in the supervisory control.

4) Negotiate with DRAM in case of manufacturing system failures or performance degradation.

The rest of the chapter is organized as follows. Enabling technologies such as Multi-Agent Systems, Service-Oriented Architecture (SOA) frameworks and overall supervisory control structure are discussed in Section 8.1. Section 8.2 discusses *iDiagnosis/Prognosis* system functions. The architecture of iDiagnosis & Prognosis is elaborated and discussed in Section 8.3. Section 8.4 discusses the information flow between *iDiagnosis/Prognosis* and other modules in the supervisory control and Section 8.5 presents the conclusion.

8.1 Enabling technologies

To construct an intelligent shop floor, enterprise integration is the first challenge. The latest system integration technologies such as agent technology and web services are mainly interested and discussed in Chapter 8.

8.1.1 Multi-agent systems

The Internet and intelligent agents are the most promising technologies to enable the cooperation of physically, geographically, and temporally distributed organizations. As stated in [204], a software agent is "an autonomous software component that interacts with its environment and with other agents on a user's behalf". Software agents can be used to model the behaviour of physical entities or perform their functions. A multi-agent system (MAS) is a system that uses a set of agents to achieve an objective, such as intelligent agent for fault diagnosis and prognosis.

Agents in a MAS can be organized in a hierarchical, distributed, or hybrid hierarchical-distributed structure. As a result of their autonomous attributes, they are able to make decisions and collaborate with each other to achieve a common objective. Agents can also be designed to be proactive so that they can take initiatives to pre-empt impending undesirable events. Intelligent agents can learn from past experiences and adapt to the current situation.

8.1.2 Service-oriented architecture

A service-Oriented Architecture (SOA) is a platform for building loosely-coupled interoperable applications based on open standards such as Web Services (WS). Software functions are packaged as services by the service provider and published in the service registry where they can be discovered by a service requestor. Following the discovery of the required web service, the service requestor will then interact with the service provider to fulfil the requirements of the service requestor.

8.1.3 Structure of manufacturing system supervisory control system (MSSCS)

A typical shop floor is composed of networked manufacturing resources including machines, transportation vehicles and operators. An individual shop floor is connected with other shop floors in the same enterprise as well as other enterprise business systems through a local network, an intranet, or the secured internet. Such systems may also be connected with those of suppliers and customers through an intranet or the internet services especially from the perspective of a virtual enterprise. To tackle the grand challenges associated with shop floor dynamics, a structure of *MSSCS* is proposed in Figure 8.2 for real-time manufacturing system performance/health management, real time decision-making and auto-recovery in response to system dynamics. The three major system modules are Dynamic Resource Allocation & Management (DRAM), Real-time Control & Auto-Recovery (RCAR), and *iDiagnosis & Prognosis.*

In recognition of the reality of unforeseen dynamic shop floor events and the realization that there is a need to fast respond to such events, DRAM will strive to perform local rescheduling of affected tasks by using the distributed control approach to perform resource allocation among equipment that is in operation.

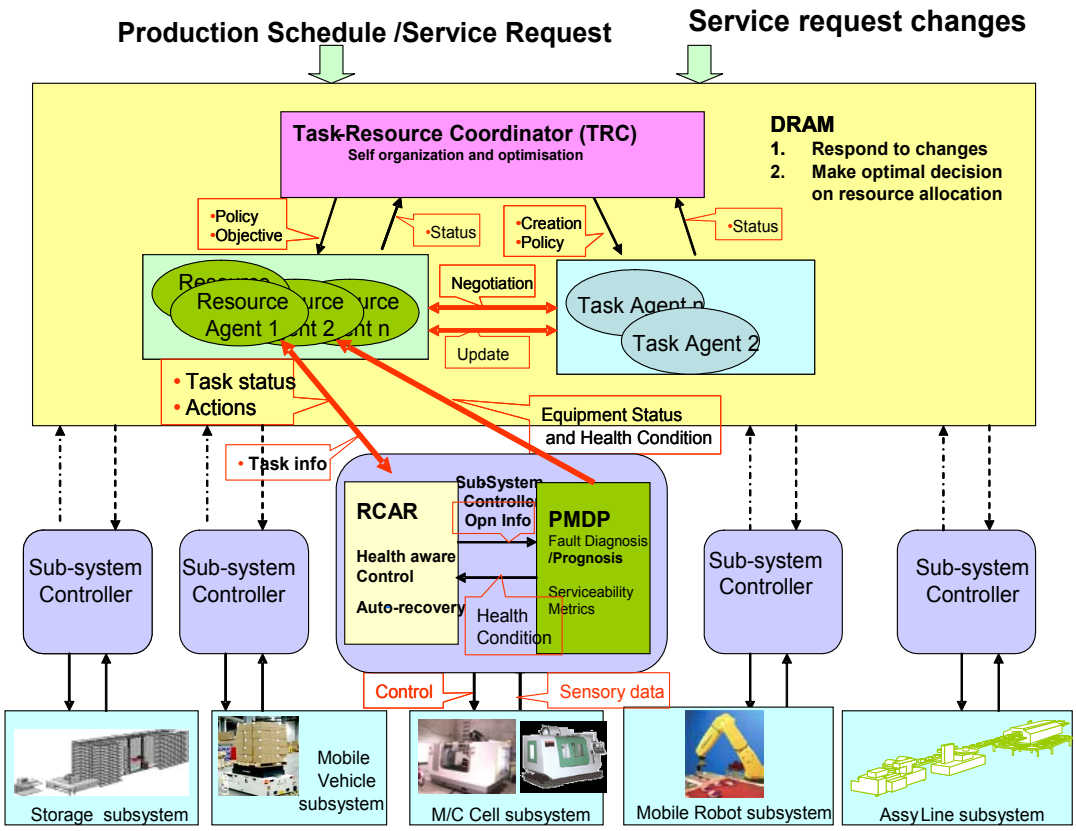


Figure 8.2: Scenario of *MSSCS*

RCAR aims to establish new approaches to the real-time control of large scale, distributed networked manufacturing systems with fault tolerance and self-recovery attributes.

In order to provide equipment health information and services ability for higher level supervisory control and dynamic resource allocation, *iDiagnosis & Prognosis* collects real-time sensor information, conducts failure detection and isolation, estimates the fault level and impact to the performance, and calculates the system and job capability, system remaining useful life and estimation of repairing time. *iDiagnosis & Prognosis* provides following services to DRAM and RCAR on other computers through the internet:

1) Current system/equipment status (normal, performance degraded, breakdown)

2) New faults (fault component, fault level)

3) Faults history (related information for one or all faults happened in the past *n* hours)

4) Repair estimation (for specific fault)

5) Remaining useful life

6) System condition

7) Job execution capability indicator

Service Oriented Architecture (SOA) is an implementation platform in this research so that different modules can be integrated through the internet in a loosely coupled way. All the agents will publish their services at shop floor servers in universal description, discovery and integration (UDDI) to support resource management (DRAM) and other monitoring, production management, equipment maintenance applications through web services.

8.2 System functions

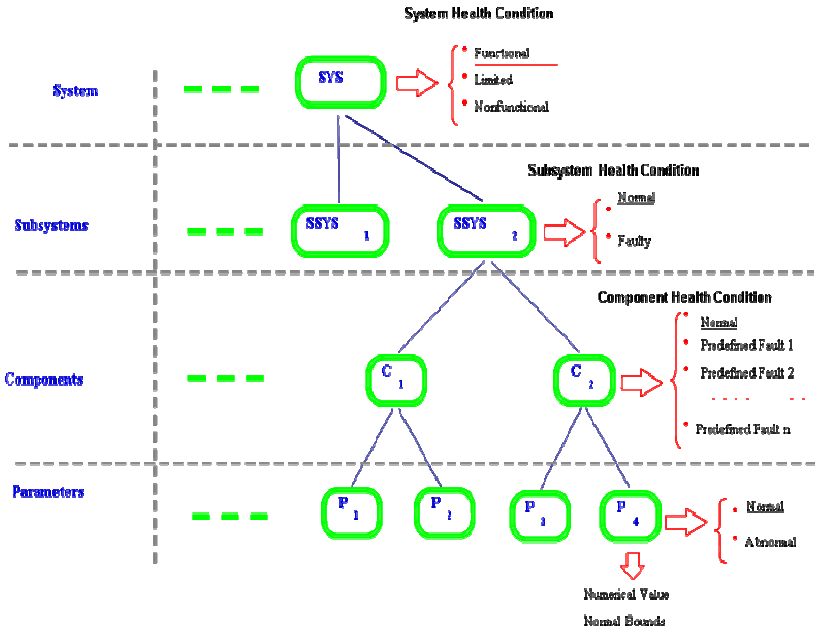


Figure 8.3: System hierarchal structure of a complex system

Considering the complexity of the manufacturing system, the generic manufacturing system structure is hierarchically defined by four levels in Figure 8.3: system, subsystems, components and parameters. This hierarchically

definition is suitable for both a complex manufacturing system and equipment. The complex system may consist of many levels of subsystems, components and parameters, while the equipment may only consist of components and parameters.

The schematic diagram of *iDiagnosis/Prognosis* is shown in Figure 8.4. The *iDiagnosis/Prognosis* builds up the diagnosis model offline either from a model approach [205-206] or from the historical performance, non-destructive test and data driven approach [207-208]. The health diagnosis model is able to provide the advice for sensors placement location and has the parametric-faults monitoring ability.

iDiagnosis/Prognosis provides on-line system health monitoring for fault detection, fault isolation, and fault parameters identification. *iDiagnosis/Prognosis* supplies information to the RCAV and DRAM of system health condition, system capabilities, remaining useful life, estimated repairing time, estimation of job capability, etc.

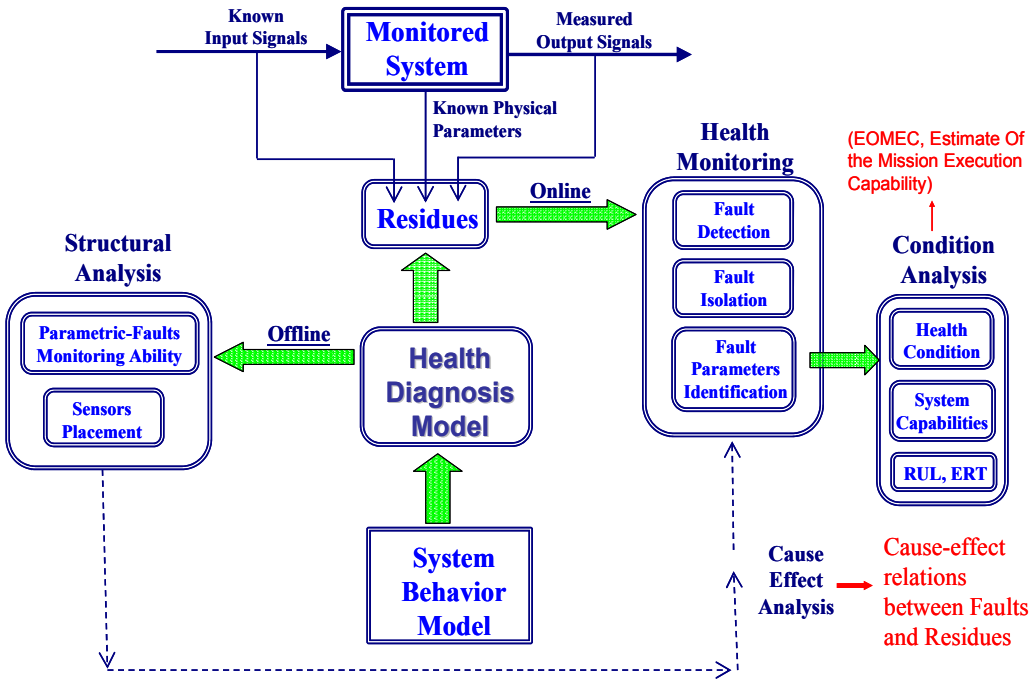


Figure 8.4: Schematic diagram of *iDiagnosis/Prognosis*

8.2.1 Health condition

Health condition is an important input to the DRAM for dynamic resource planning. The health condition of each level is defined in Figure 8.3. The lowest level is parameters level. Parameters' health condition can be "normal" or "abnormal". The information is derived from sensors information and using the system identification method that compares with the parameter normal value and normal bound to determine the parameters health condition. Level two is component level. Component can/may have many different parameters. The component health condition is derived from the parameters health conditions. The component health condition can be one of the following: "normal", "predefined fault 1", "predefined fault 2"…. "Predefined fault *n*". "Predefined fault *i*" is some type of the faults that is predefined in the system. Level three is subsystem level. The subsystem health condition is derived from its components health conditions and it can be "normal" or "faulty". Level four is system level. System Health Condition can be one of the following: functional, limited or not functional, and is determined by its sub-systems health conditions. "Functional" means system can perform all the system functions, "Limited" means that the system can only perform certain functions or partially perform functions, "Not functional" means that the system cannot perform any system functions.

8.2.2 System capabilities

Besides the system health condition, the planning system requires another input of the manufacturing system capability to perform the dynamic scheduling. The System Capabilities (SC) indicator is defined as follows:

- Control: {Able, Unable}
- Communication: {Able, Unable}
- Motion 1: Numerical Value
- Motion 2: Numerical Value
- - - - : Numerical Value
- Predefined Task 1: {Able, Unable}
- Predefined Task 2: {Able, Unable}
- - - - - - - : {Able, Unable}

In the definition, Motion *i* is the predefined task for transportation vehicles, while predefined task *i* is the predefined task for machines.

8.2.3 RUL and ERT

The Remaining Useful Life (RUL) Indicator provides the information of the system's remaining useful life under different conditions. The RUL Indicator is defined as follows:

- **Unconditional**: {-1, time (hours/minutes)}
- **Condition 1**: {-1, time (hours/minutes)}
- **Condition 2**: {-1, time (hours/minutes)}
- **- - -** : {-1, time (hours/minutes)}

RUL can be different under certain conditions and therefore RUL is given under condition 1, condition 2, e.g., RUL under certain speed, certain load or predefined task *i*, etc. "-1" means the time is unknown or unable to be estimated.

Each system has an Estimated Repair Time (ERT) when the system is under Health Conditions Limited or Non-functional. The ERT is based on a pre-estimated repair time, which is attached to each one of the predefined faults of the system components. The ERT can be obtained from expert knowledge or calculated with a statistical method using the system historical repairing records.

8.2.4 System condition structure variable

For a given complex system, system condition is not able to be described as a simple value or simple structure. It is represented with a System Condition Structure Variable (SCSV) which consists of Health Condition, System Capabilities, RUL and ERT. It can be described as a structure variable in Figure 8.5. Since XML is a common protocol, it is suitable for interchanging between different applications, especially those in remote enterprise. XML will carry the messages of SCSV over the internet in our real implementation.

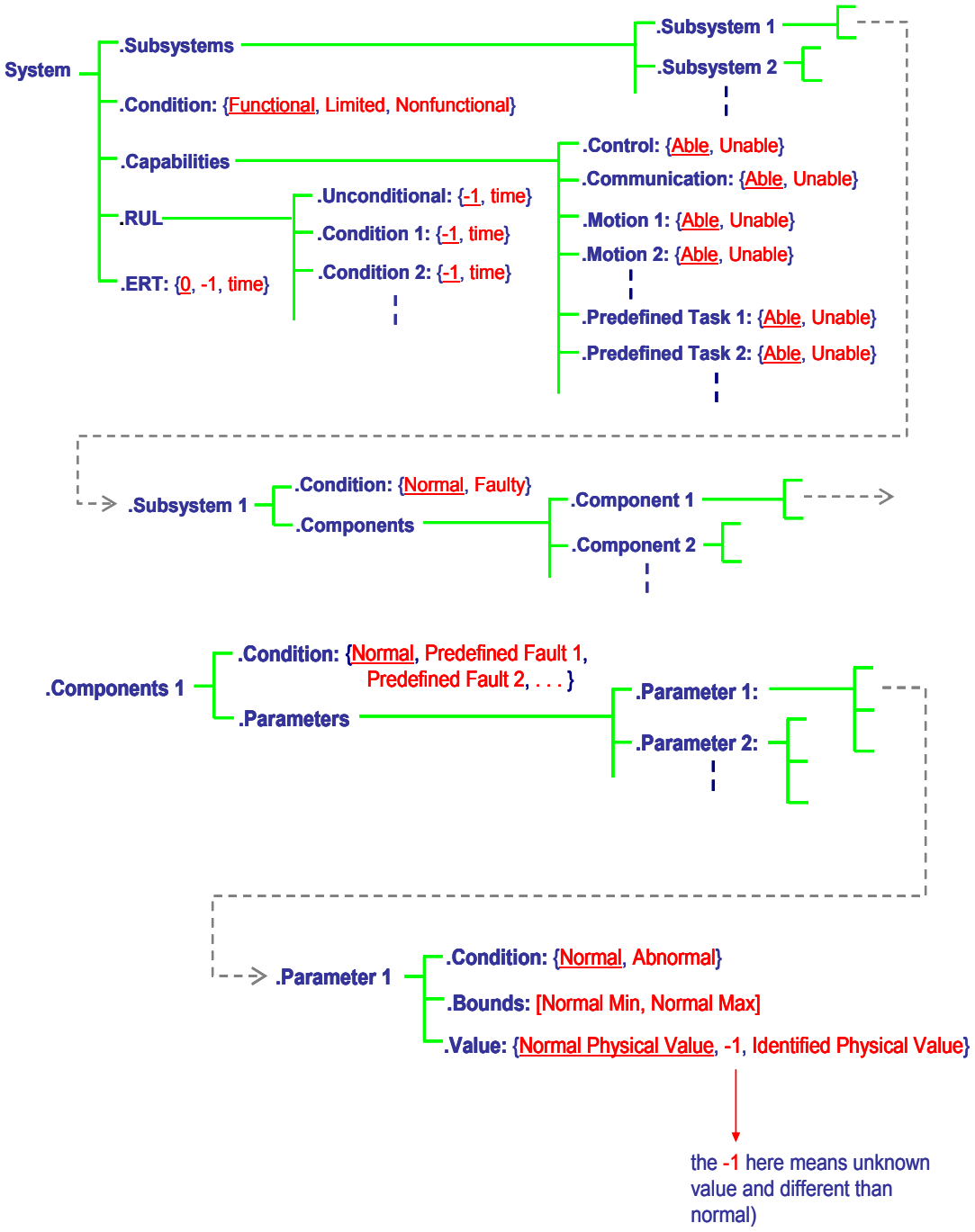


Figure 8.5: System condition structure variable

8.2.5 Job execution capability indicator (JECI)

Job execution capability is another important input to DRAM. The JECI provides the information of the progress and status of the current job estimated completion time. The Job is an object, and the JECI is defined as follows in Figure 8.6. XML will carry messages of JECI over the internet in the real implementation.

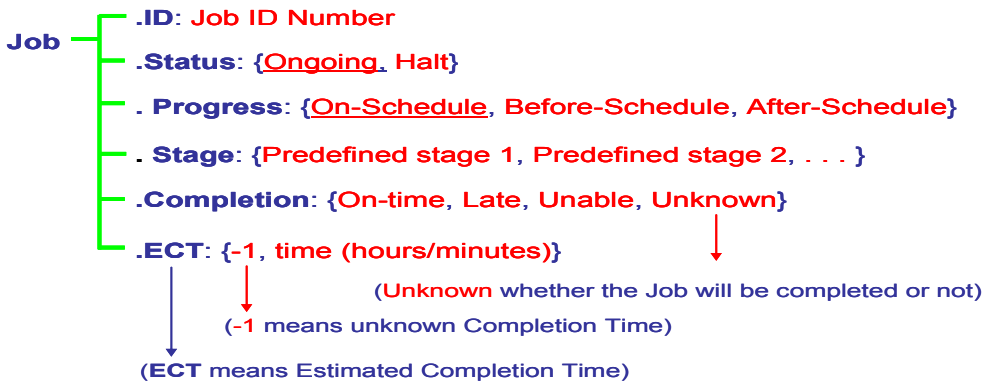


Figure 8.6: Job execution capability indicator (JECI)

8.3 System architecture

A system architecture of iDiagnosis & Prognosis is presented in Figure 8.7. iDiagnosis & Prognosis is integrated with the RCAR module and communicating with DRAM, to receive a job, and to update DRAM on the system/equipment condition, job status, job execution capability, etc. The system architecture includes three types of platforms: diagnosis platform, intelligent platform and maintenance platform. All the services provided by iDiagnosis & Prognosis are published in the web service directory within the Java Agent Development Framework (JADE) in Figure 8.7.

The *i*Diagnosis/Prognosis can work offline using the historical data or work online with data acquisition and data processing. The various agents and the associated functionality hosted by this system are summarized in Table 8.1.

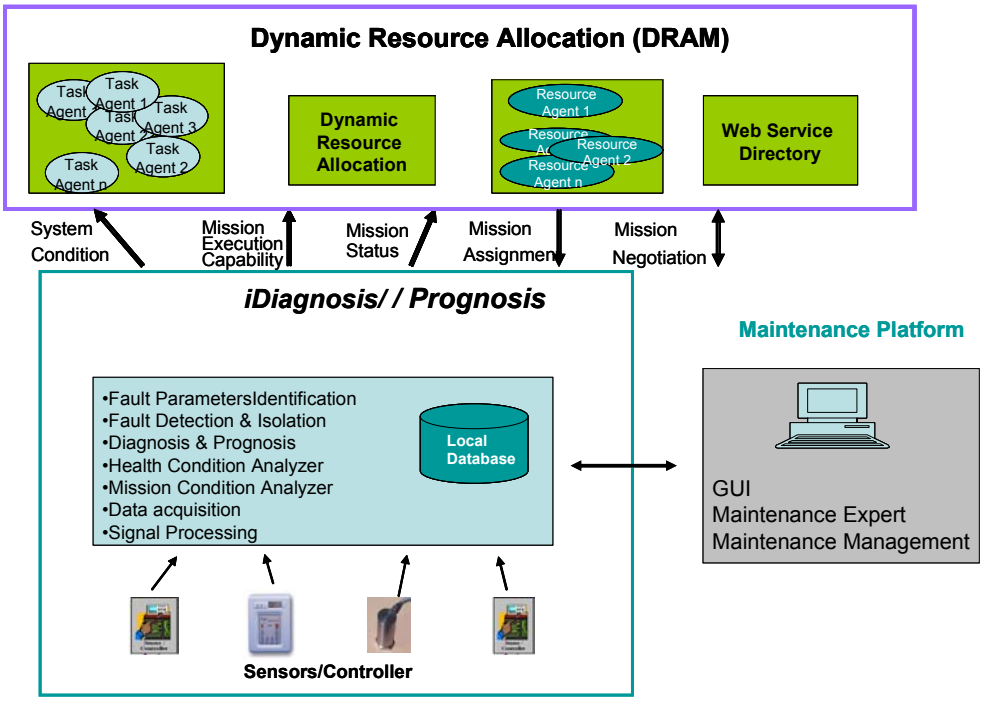


Figure 8.7: System architecture

Table 8.1

Functions and agents of *i*Diagnosis/Prognosis

Agents	Functions
Data agent	Data acquisition
Data processing agent	Data processing
Parameter identification agent	Fault parameters identification
Fault detection isolation agent	Fault detection & isolation
Prognosis agent	Prognosis
Health condition analyzer agent	Health condition analyzer
Job condition analyzer agent	Job condition analyzer

The real-time tasks of data acquisition and data processing can be performed by a sensing platform of an on-board computer. It runs in a real-time programming

environment. The Diagnosis/Prognosis platform which performs the functions of fault parameters identification, fault detection & isolation, prognosis, health condition analyzer, job condition analyzer is fixed in one location and does not execute real-time processes. The Diagnosis/Prognosis platform and the sensing platform could be communicated by a local wire/wireless network. The sensors' data from the sensing platform are stored on the Diagnosis/Prognosis platform in a local database. After a fault detection and isolation, a fault estimation process is required (i.e. to estimate the size of the fault). The fault estimation is carried-out by the Diagnosis/Prognosis platform, based on an iterative process (e.g. nonlinear least-square). One important role of the Diagnosis/Prognosis platform is to communicate with the DRAM module by a global network such as the internet. This communication channel includes data exchange with the global database of DRAM, receives tasks and updates from DRAM, updates DRAM on the system/equipment condition and its ability to perform tasks and the status of the tasks. Additionally, when a fault is happening in the system/equipment, the maintenance platform provides facilities for maintenance management and data logging, meantime the RCAR module takes the necessary actions to modify its control algorithms so that the fault is accommodated (e.g. control parameters tuning).

8.4 System information flow

The information flow between *iDiagnosis/Prognosis* , maintenance platform and DRAM is shown in Figure 8.8.

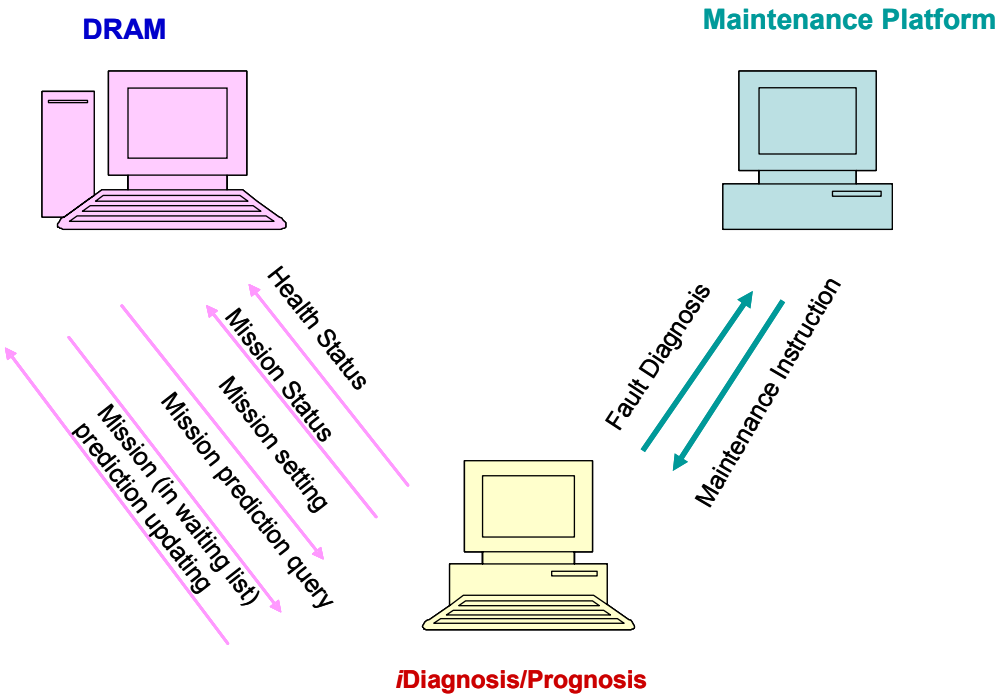


Figure 8.8: System information flow

Figure 8.9 illustrates the sequence diagram of the event flow between DRAM, RCAR and *iDiagnosis/Prognosis*. A typical transportation partial breakdown scenario is shown in the diagram. A prototype has been implemented of these transportation partial breakdown scenarios.

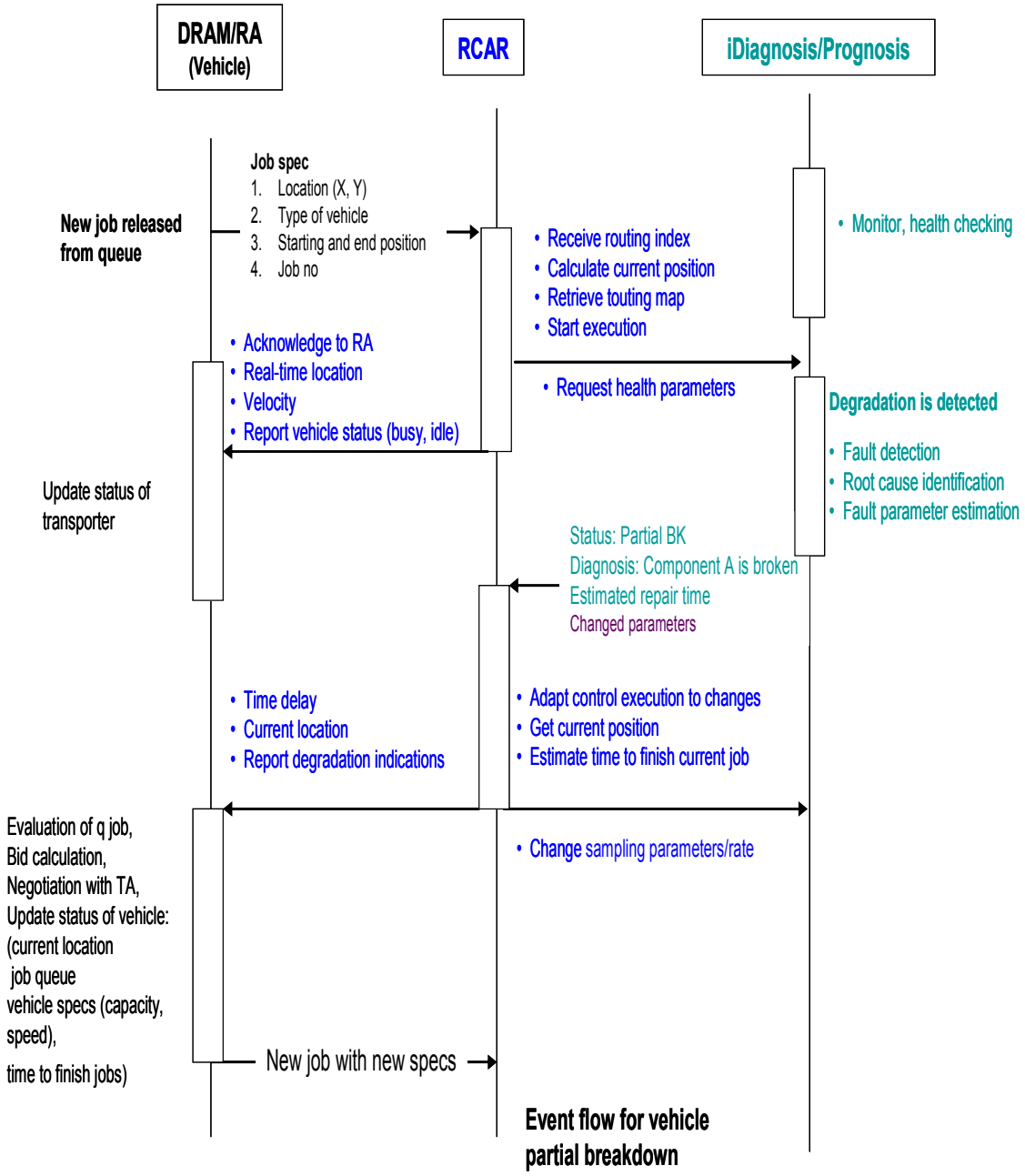


Figure 8.9: Sequence diagram of event flow for transportation partial breakdown

8.5 Summary

In this chapter, a new concept of iDiagnosis & Prognosis is proposed - an intelligent diagnosis, prognosis, manufacturing system condition, and job execution capability assessment system was developed using Internet, web service and agent technologies. The frame work that consists of diagnosis platform, prognosis platform and maintenance platform is proposed. The information flow and event handling of the manufacturing system partial breakdown are presented. Multi-agent and web services are the main technologies adopted in this framework.

CHAPTER 9 CONCLUSIONS AND FUTURE WORK

9.1 Conclusions

Condition based maintenance (CBM) has become increasingly important over the past several years as a means to maximize the asset usage and plant operating efficiency. CBM reduces maintenance cost by reducing the unnecessary scheduled preventive maintenance operations where equipment outages are predicted and maintenance is carried out only when necessary. The relevant literature of CBM has been reviewed, which reveals a clear need of advanced feature extraction and selection for realizing CBM. From the literature review, several potential research directions have been identified and addressed in this thesis.

First, a novel Dominant Feature Identification (DFI) methodology is proposed in Chapter 3. DFI provides an autonomous rule-base for data reduction. Singular Value Decomposition (SVD) is used to decompose the inner product matrix of collected data from the monitoring sensors. The principal components are optimized in a least squares sense in a certain reduced space, and the dominant features are extracted using the k-means clustering algorithm. This DFI framework uses formal mathematical analysis to select dominant features. A numerically efficient scheme for implementation is presented that is based on the inner product matrix of the collected data, not the correlation (outer product) matrix. The proposed DFI decision tool is numerically efficient, and reduces the complexity of feature selection greatly while reducing the least square error during dominant feature selection and clustering automatically. It is proven that the proposed method of dominant feature selection is optimal in the sense that it minimizes the total least squares estimation error.

To efficiently use DFI in the CBM framework, a methodology is presented in Chapter 4 to integrate the proposed DFI with traditional feature extraction methods of time domain and frequency domain analysis. DFI also works together with Multiple Regression Models (MRMs) using Recursive Least Squares (RLS) and a dynamic ARMAX model with Extended Least Squares (ELS) method. The developed methodology and framework is evaluated based on the accuracy of

prediction of the actual milling tool wear condition. Our experimental results show that using force measurement sensor, Mean Squares Errors (MSEs) values from 0.946 to 1.262 and Mean Relative Error (MRE) values from 8.86% to 11.61% were achieved between the actual measured tool wear and predicted from the Multiple Regression Models (MRMs). Our second experimental results using AE sensors show significant reduction in both Mean Squares Errors (MSEs) and Mean Relative Error (MRE) when an ARMAX model with Extended Least Squares (ELS) technique is employed, which is promising in replacing force sensors and conventional non-dynamic models for effective online.

The research is not only limited to the features identification but also extended to dominant sensors identification. Augmented Dominant Feature Identification (ADFI) and Decentralized Dominant Feature Identification (DDFI) are proposed in Chapter 5 for industrial Fault Detection and Isolation (FDI). The proposed ADFI and DDFI are tested on a fault simulator machine for both data and sensors reduction. Neutral Network is used for fault classification and multiple fault prediction in a proposed two-stage framework. Our experimental results on a fault simulator machine reduce the number of features from 120 to 13 and sensors from 8 to 4 and achieved an accuracy of 99.4% for fault prediction.

Further improvement has been developed for integration of DFI with the wavelet-based correlation modelling health assessment. The approach is presented in Chapter 6 which successfully used for motor health state monitoring. The methodology applies discrete wavelet transforms to the motor current signature to extract features represented as the coefficients of different levels of wavelet decompositions. DFI is used to select the most important features and equipped with a multivariable regression method to obtain the best and simplest correlation model between the features and faults. This approach has been tested experimentally with the assessment of the degradation of fluid dynamic bearings used in brushless DC motors. Experimental results show that an acceptable assessment of the motor health condition has been achieved. The prediction errors using the correlation equation for two sets of motors range from 0.78% to 7.5%, respectively.

The research explores the problems and the challenges in bearing fault diagnosis. The difficulties in implementing Morlet wavelet transform for bearing fault diagnosis include selection of the optimal shape factor σ and scale factor α, determination the threshold and reconstruction of the envelope signal. Reinforced Morlet wavelet transform is proposed in Chapter 7 for bearing fault diagnosis to overcome these difficulties. Threshold magnification factor criterion is proposed to reduce the error rate in selecting the optimal α and σ values by filtering out wavelet coefficients with small average value. Max-energy spectrum is also proposed to avoid the risk of selecting incorrect decomposed signal and adding noise by combining the information of all decomposed signals in frequency domain. The results of the two case studies verify the performance of the method and demonstrate that the fault detection consistency has been increased significantly from existing 11.8% and 17.7% to 67.6%.

To extend the research from single machine CBM to a complex manufacturing environment, a new concept of *iDiagnosis & Prognosis* was proposed in Chapter 8 to provide distributed intelligence of diagnosis and prognosis, system condition and job execution capability assessment, and dynamic information feedback to the supervisory control when manufacturing system breakdown or performance degradation occurs. The frame work that consists of diagnosis platform, prognosis platform and maintenance platform is proposed. The information flow and event handling of the manufacturing system partially breakdown is presented. Multi-agent and web services are the main technologies adopted in this framework.

This research has successfully completed the objectives that are set in Chapter 1 by addressing the critical issues in the development of an effective CBM system. The contribution of the work undertaken by the author can be summarized as follows:

- Dominant Feature Identification (DFI) methodology using Singular Value Decomposition (SVD) is proposed. The DFI uses SVD which operates on the inner product matrix at a lower dimension, and reduces the total Least Squares Error (LSE) [139].

- To efficiently use DFI in the CBM framework, a methodology is developed to integrate the proposed DFI with traditional feature extraction methods of time domain and frequency domain analysis. DFI also works together with Multiple Regression Models (MRMs) using Recursive Least Squares (RLS) and a dynamic ARMAX model with Extended Least Squares (ELS) method. The developed methodology and framework is evaluated based on the accuracy of prediction of the actual milling machine tool condition [134], [139], [141].
- An integration of DFI and wavelet-based correlation modeling health assessment approach is presented for health state monitoring of motors. This approach involves two stages: (1) extracting of features from the motor-stator current signatures by analyzing discrete wavelet transform coefficients; and (2) using DFI to select the most important features and building of the simplest correlation model between the extracted features and the bearing wear using a multivariable regression technique. The correlation model can be used to detect and predict the bearing wear of brushless DC motors [200].
- Augmented Dominant Feature Identification (ADFI) and Decentralized Dominant Feature Identification (DDFI) are proposed for dominant sensors and features identification. Neutral Network is used for online fault classification and multiple fault prediction in a proposed two-stage framework. The proposed methodology is evaluated with a fault simulator machine. It achieved an accuracy of 99.4% for 5 different types of rotary faults prediction [162].
- Reinforced Morlet wavelet transform is proposed for bearing fault diagnosis to overcome the difficulties in implementing Morlet wavelet transform. Main contribution includes two aspects: proposed threshold magnification factor criterion to filter out wavelet coefficients with small average value, it reduces the error rate in selecting the optimal α and σ values; proposed the max-energy spectrum to combine the information of all decomposed signals in frequency domain, it avoids the risk of selecting incorrect decomposed signal and adding noise to original signal.

Reinforced Morlet wavelet achieved high consistency and early stage bearing fault detection [198].

- A new concept of *iDiagnosis & Prognosis* is proposed- an intelligent diagnosis, prognosis, manufacturing system condition, job execution capability assessment system based on the Internet, web service and agent technologies. The frame work that consists of diagnosis platform, prognosis platform and maintenance platform is proposed. The information flow and event handling of the manufacturing system partial breakdown are presented. Multi-agent and web services are the main technologies adopted in this framework [204] .

9.2 Suggestions for future work

The work presented in this thesis has demonstrated the profound potential of improving the effectiveness and efficiency of CBM through integrating of advanced feature extraction and DFI techniques. To further extend this research, the following future work is recommended:

- Accurate equipment remaining useful life (RUL) prediction is critical to effective condition based maintenance for improving reliability and reducing overall maintenance cost. Method such as particle filtering has shown its great promise in RUL predication. In future, it should be pursued to extend the DFI techniques with particle filtering methods to achieve accurate equipment remaining useful life prediction.

- Stochastic modelling methods such as Hidden Markov Model (HMM) and Dynamic Bayesian network have shown great promise in machine failure states identification. Unlike the static modelling that deal with original data, stochastic methods model the probability distributions over sequence of random variables to handle sequenced observations. They give prediction of health state estimation. In future, DFI techniques for feature extraction and selection should be working with stochastic modelling methods to achieve accurate machine failure states identification.

- Vibration sensing and analysis has not been very successful for low rotational speeds of less than 20 rpm. At such speeds, the energy generated from bearing defects might not show any obvious change in signature and thus become undetectable using conventional vibration measurement systems. Research of using AE for detecting early stages of bearing damage at a rotational speed less than 20rpm is reported in [207]. It is expected that more researches will be conducted in future of AE sensing techniques for bearing fault detection at a low rotational speed.
- Integration of the prognosis with maintenance scheduling optimization is a very important area for future study. The research issues include machine remaining life prediction, maintenance planning and scheduling optimization based on the real machine conditions and maintenance resources availability to minimize the production loss.

It is anticipated that the establishment of the above-mentioned future work could stretch the capability of the DFI developed in this work and further improve the efficiency and accuracy of the CBM system for complex manufacturing environment. Manufacturing companies equipped with such a CBM system could possibly gain a competitive edge in the new global competitive environment.

References:

1. A. K.S. Jardine, D. Lin, D. Banjevic (2006), "A review on machinery diagnostics and prognostics implementing condition-based maintenance", Mechanical Systems and Signal Processing, Vol. 20, pp. 1483–1510.
2. W. Zhou, T. G. Habetler, and R. G. Harley (2007), "Bearing Condition Monitoring Methods for Electric Machines: A General Review", in Proc. Of the IEEE International Symposium on Diagnostics for Electric Machine, pp. 3–6.
3. P. D. McFadden and J. D. Smith (1984), "Vibration Monitoring of Rolling Element Bearings By the High Frequency Resonance Technique–A Review", Tribology International, Vol. 17, pp. 3–10.
4. H. Qiu, H. Luo, and N. Eklund (2009), "Draft: On-Board Bearing Prognostics in Aircraft Engine: Enveloping Analysis or FFT?", in Proc. Of the IDETC/CIE, pp. 1–7.
5. Y. Lei, Z. He, and Y. Zi (2009), "Application of an Intelligent Classification Method to Mechanical Fault Diagnosis", Expert Systems with Applications, Vol. 36, No. 6, pp. 9941–9948.
6. Y. Lei, M. J. Zuo, Z. He, and Y. Zi (2010), "A Multidimensional Hybrid Intelligent Method for Gear Fault Diagnosis", Expert Systems with Applications, Vol. 37, No. 2, pp. 1419–1430.
7. K. Ron, and H. J. George (1997). "Wrappers for Feature Subset Selection", Artificial Intelligence, Vol. 97, pp 273-324.
8. M. L. Raymer, W. F. Punch, E. D. Goodman, L. A. Kuhn and A. K. Jain (2000). "Dimensionality Reduction Using Genetic Algorithm", IEEE Transactions on Evolutionary Computation.
9. I. T. Jolliffe (1986), Principal Component Analysis, Springer-Verlag, New-York.
10. G. P. McCabe (1984), "Principal Variables", Technometrics, Vol. 26, No. 2, pp. 127-134.
11. H. L. Tan, N. S. Chaudhari, and J. H. Zhou (2008), "Time Series Prediction Using Principal Feature Analysis", in Proc. of the 3rd IEEE Conference on Industrial Electronics and Applications, pp. 292–297.
12. I. Cohen, Q. Tian, X. S. Zhou, and T. S. Huang (2002), "Feature Selection Using Principal Feature Analysis", in Proc. of the 15th International Conference on Information Processing.
13. Y. Lu, I. Cohen, X. S. Zhou, Q. Tian, and T. S. Huang (2007), "Feature Selection Using Principal Feature Analysis", in Proc. of the 15th International Conference on Multimedia, pp. 301–304.
14. G. A. Fodor (2005), "From the Editor-in-Chief Industrial Informatics: Predicting with Abstractions", IEEE Transactions on industrial informatics, Vol. 1, No. 1, pp. 3.

15. F. L. Lewis, S. Jagannathan, and A. Yeçildirek (1998), Neural Network Control of Robot Manipulators and Nonlinear Systems, CRC Press.
16. S. Simani (2005), "Identification and Fault Diagnosis of a Simulated Model of an Industrial Gas Turbine", IEEE Transactions on industrial informatics, Vol. 1, No. 3, pp. 202–216.
17. D. B. Percival and A. T. Walden, Wavelet Methods for Time Series Analysis, Cambridge University Press, 2000.
18. E. Osuna, R. Freund, F. Girosi (1997), "An improved training algorithm for support vector machines", Proceedings of the 1997 IEEE Workshop on Neural Networks for Signal Processing VII. , pp. 276 – 285.
19. R. Gong, S. H. Huang, and T. Chen (2008), "Robust and Efficient Rule Extraction Through Data Summarization and Its Application in Welding Fault Diagnosis", IEEE Transactions on industry informatics, Vol. 4, No. 3, pp. 198–206.
20. L. Ljung (1998), System Identification–Theory For the User, 2nd Edition, PTR Prentice Hall.
21. T. Han, B. S.Yang (2006). "Development of an e-maintenance system integrating advanced techniques", Computers in Industry, Vol. 57, issue 6, pp. 569–80.
22. B. Tao, H. Ding, Y. L. Xion (2003), "IP sensor and its distributed networking application in e-maintenance", Proceedings of the 2003 IEEE international conference on systems, man and cybernetics, Vol. 4, pp. 3858–3863.
23. M. Koc, J. Lee (2001), "A system framework for next-generation e-maintenance system", Proceedings of the second international symposium on environmentally conscious design and inverse manufacturing.
24. W. Zhang, W. Halang, C. Diedrich (2003), "An agent-based platform for service integration in E-maintenance", IEEE international conference on industrial technology, Vol. 1, pp. 426–33.
25. A. M. Crespo J. Gupta (2006) "Contemporary maintenance management: process, framework and supporting pillars", Vol. 34, No.3, pp. 313–26.
26. H. C. Pusey, M. J. Roemer (1999), "An assessment of turbomachinery condition monitoring and failure prognosis technology", The Shock and Vibration Digest, Vol. 31, pp. 365–371.
27. A. K. S. Jardine, D. Lin, D. Banjevic (2006), "A review on machinery diagnostics and prognostics implementing condition-based maintenance", Mechanical Systems and Signal Processing, Vol. 20, pp. 1483–1510.
28. G. Vachtsevanos, F. Lewis, M. Roemer, A. Hess, B. Wu (2006), "Intelligent fault diagnosis and prognosis for engineering systems", Wiley.

29. L. Ma (2007), “Condition monitoring in engineering asset management”, Proceedings of Asia-Pacific Vibration Conference.

30. Heng A., Zhang S., Tan C. C ., Mathew J. (2009), “Rotating machinery prognostics: State of the art, challenges and opportunities”, Mechanical Systems and Signal Processing, Vol. 23, pp. 724–739.

31. M. Lebold, K. Reichard, P. Heida, J. Bezdicek, and M. Thurston (2002), “A framework for next generation machinery monitoring and diagnostics”, Proceeding of the 56th Meetings of the Society for Machinery Failure Prevention Technology.

32. J. Lee, R. Abujamra, A.K.S. Jardine, D. Lin, D. Banjevic (2004)., “An integrated platform for diagnostics, prognostics and maintenance optimization”, The IMS ’2004 International Conference on Advances in Maintenance and in Modelling, Simulation and Intelligent Monitoring of Degradations.

33. Y. M. Niu et al (1998), “An intelligent Sensor System Approach for Reliable tool flank wear recognition”, Advanced Manufacturing Technology, Vol.14, pp.77–84.

34. M. S. Lan and D. A. Dornfeld (1986), "Acoustic Emission and Machining - Process Analysis and Control", Int. J. Advanced Manufacturing Processes, Vol. 1, No. 1, pp. 1-22.

35. B. Y. Lee, and Y. S. Tang (1999) “Milling cuter breakage detection by the discrete wavelet transform”, Mechatronics, Vol. 9, pp. 225-234.

36. J. F. Dong, K. V. R. Subrahmanyam, Y. S. Wong, G. S. Hong and A. R. Mohanty (2006), “Bayesian-inference-based neural networks for tool wear estimation”, The International Journal of Advanced Manufacturing Technology, Vol. 30, pp. 797-807.

37. B. Sick (2001), “Tool Wear Monitoring in Turning: A Neural Network Application”, Measurement Control, Vol. 34, No. 7, pp. 207–222.

38. G. K. Singh, and S. A. S. Al Kazzaz (2003), "Induction machine drive condition monitoring and diagnostic research - A survey", Electric Power Systems Research Vol 64, No. 2, pp. 145-158.

39. S. Goldman (1999), Vibration spectrum analysis: a practical approach, 2nd edition. New York: Industrial Press.

40. A. J. Hoffman, and N. T. van der Merwe (2002), “The application of neural networks to vibrational diagnostics for multiple fault conditions”, Computer Standards & Interfaces, Vol. 24, No. 2, pp. 139-149.

41. L. Hao, and X. Xu (2001), “The application of rough set neutral network system in fault diagnosis”, Control theory & applications, Vol. 18, No. 5, pp.681-685.

42. W. Kesheng, and L. Bing (2001), "Using B-spline neutral network to extract fuzzy rules for a centrifugal pump monitoring", Journal of Intelligent Manufacturing, Vol. 12, No. 1, pp. 5-11.

43. H. Austerlitz (2003), Data Acquisition Techniques using PCs, Academic Press, San Diego, CA.

44. N. V. Kirianaki, S. Y. Yurish, N. O. Shpak, V. P. Deynega (2002), "Data Acquisition and Signal Processing for Smart Sensors", Wiley, Chichester, West Sussex, England.

45. J. Tlusty, and G. C. Andrews (1983) "A critical review of sensors for unmanned machining", Annuals of the CIRP, Vol. 32, No. 2, pp. 563-572.

46. L. Dan & J. Mattew (1990), "Tool wear and failure and monitoring techniques for turning: a review", International Journal of Machine tools & Manufacture, Vol. 30(4), pp. 579-598.

47. R. K. Kiran Dutta, G. Paul and A. B. Chattopadhyay (2000) "Assessment of machining features for tool condition monitoring in face milling using an artificial neutral network", Proceedings of the I MECH E Part B, Journal of Engineering Manufacture, Vol, 214(7), pp. 535-546.

48. B. Sick, "On-line and indirect tool wear monitoring in turning with artificial neutral networks: a review of more than a decade of research", Mechanical Systems and Signal Processing, Vol. 16 (4), pp 487-546, (2002).

49. P. Bhattacharyya, D. Sengupta, and S. Mukhopadhyay (2007), "Cutting force based real time estimation of tool wear in face milling using a combination of signal processing techniques", Mechanical System and Signal Processing, Vol. 21, pp. 2665-2683.

50. D. K. Baek, T. J. Ko and H. S. Kim (2000), "Real time monitoring of tool breakage in a milling operating using a digital signal processor", Journal of Materials Processing Technology, Vol. 100, pp. 262-272.

51. P. W. Prickett, and C. John (1999), "An overview of approaches to end milling tool monitoring", International Journal of Machine Tools and Manufacture, Vol. 39, pp. 105-122.

52. R. Tetia and G. F. Micheletti (1989), "Tool Wear Monitoring Through Acoustic Emission", CIRP Annals - Manufacturing Technology, Vol. 38, Issue 1, pp. 99-102

53. M. A. Elbestawi, M. Durmitrescu (2006), "Tool Condition Monitoring in Machining - Neural Networks", Condition Monitoring and Control for Intelligent Manufacturing, pp. 55-82.

54. Kamarthi S. V. and Pittner S (1997), "Fourier and wavelet transform for flank wear estimation- a comparison", Mechanical Systems and Signal Processing, Vol. 11, No. 6, pp. 791-809.

55. X Li (2001), “Detection of Tool Flute Breakage in End Milling using Feed-Motor Current Signature”, IEEE/ASME Transactions on Mechatronics, Vol. 6, No. 4, pp. 491-498.

56. N. Gebraeel, M. Lawley, R. Liu, V. Parmeshwaran (2004), “Residual life predictions from vibration-based degradation signals: a neural network approach”, IEEE Transactions on Industrial Electronics, Vol. 51, pp. 694–700.

57. G. B. Kliman and J. STEIN (1992), “Methods of motor current signature analysis”, Electric Machines and Power Systems, Vol. 20, No. 5, pp. 463-474.

58. C. M. Riley, B. K. Lin, T. G. Habetler, and G. B. Kliman (1999), “Stator Current Harmonics and Their Causal Vibrations: A Preliminary Investigation of Sensorless Vibration Monitoring Applications”, IEEE Trans. Ind. Appl., Vol. 35, No. 1, pp. 94-99.

59. G.K. Singh, S. A. S. A. Kazzaz (2003), “Induction machine drive condition monitoring and diagnostic research - a survey”, Electric Power Systems Research, Vol. 64 , pp. 145-158.

60. N. Tandon, and A. Choudhury (1999), “A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings”, Tribology International, Vol. 32, No. 8, pp. 469-480.

61. Y. Altintas (1988), “In-process detection of tool breakages using time series monitoring of cutting forces”, International Journal of Machine Tools and Manufacture, Vol. 28, No. 2, pp. 157–172.

62. Y. Altintas, I. Yellowley (1989), “In-process detection of tool failure in milling using cutting force models”, Journal of Engineering for Industry-Transactions of the ASME, Vol. 111, pp. 149–157.

63. J. H. Tarn, M. Tomizuka (1989), “On-line monitoring of tool and cutting conditions in milling”, Journal of Engineering for Industry-Transactions of the ASME, Vol. 111, pp. 206–212.

64. Y. S. Tarng (1990), “Study of milling cutting force pulsation applied to the detection of tool breakage”, International Journal of Machine Tools and Manufacture, Vol. 30, No. 4, pp. 651–660.

65. D. Y. Zhang, Y. T. Han, D. C. Chen (1995), “On-line detection of tool breakages using telemetering of cutting forces in milling”, International Journal of Machine Tools and Manufacture, Vol. 35, No. 1, pp. 19–27.

66. M. A. Elbestawi, J. Marks, T. A. Papazafiriou, (1989) “Process monitoring in milling by pattern recognition”, Mech Syst Signal Process, Vol. 3, No. 3, pp. 305–315.

67. Y. S. Tarng, Y. W. Hseih, S. T. Hwang (1994) "Sensing tool breakage in face milling with a neural network", International Journal of Machine Tools and Manufacture, Vol. 34, No. 3, pp. 341–350.

68. C. S. Leem, D. A. Dornfeld, S. E. Dreyfus (1995), "A customized neural network for sensor fusion in on-line monitoring of cutting tool wear", Journal of Engineering for Industry-Transactions of the ASME, Vol. 117, pp. 152–159.

69. http://en.wikipedia.org/wiki/Skewness. Last visited on August 30th, 2010.

70. S. Nandi and H. A. Toliyat (1999), "Condition Monitoring and Fault Diagnosis of Electrical Machines – A Review," Proceeding of 34th Annual Meeting of the IEEE Industry Applications, pp. 197-204.

71. M. J.David, Y. Alexander and P. W. Robert (1999), "Pump Failure Detection Using Support Vector Data Descriptions", Advances in Intelligent Data Analysis, Lecture Notes in Computer Science, Vol. 1642, pp. 415-425.

72. J. Antoni, R.B. Randall (2002), "Differential diagnosis of gear and bearing faults", Transactions of the ASME, Journal of Vibration and Acoustics, Vol. 124, pp. 165 - 171.

73. C. Pachurd, R. Salvetat and C. Fray (1997), "Crest Factor and kurtosis contribution to identify defects inducing periodical impulsive forces", Mechanical Systems and Signal Processing, Vol. 11, Issue 6, pp. 903-916.

74. T. Williams, X. Ribadeneira, S. Billington, T. Kurfess (2001), "Rolling element bearing diagnostics in run-to-failure lifetime testing", Mechanical system and signal processing, Vol. 15, No. 5, pp. 979-933.

75. R. H. Lyon (1988), "Machinery Noise and Diagnostics", The Journal of the Acoustical Society of America, Vol. 84, Issue 6, pp. 2299-2300.

76. P. D. McFadden and J. D. Smith (1984), "Vibration monitoring of rolling element bearings by the high-frequency resonance frequency technique-A review", Tribology International, Vol. 17, No. 1, pp. 3-10.

77. CS Burrs, RA Gopinath, H Guo (1998) Introduction to wavelets and wavelet transforms, a primer, Prentice Hall.

78. The Math Works, Inc. (2006) Wavelet Toolbox, Version 3.0.4 (R2006a) http://www.mathworks.com/access/helpdesk/help/toolbox/wavelet, last visited on 4 Jul 2010.

79. W. J. Staszewski, G. R. Tomlinson (1994), "Application of the wavelet transform to fault detection in a spur gear", Mechanical Systems and Signal Processing, Vol. 8, pp. 289 - 307.

80. W. J. Wang, P. D. McFadden (1996), “Application of wavelets to gearbox vibration signals for fault detection”, Journal of Sound and Vibration, Vol. 192, pp. 927 – 939.

81. G. Y. Luo, D. Osypiw, M. Irle (2003), ‘On-line vibration analysis with fast continuous wavelet algorithm for condition monitoring of bearing”, Journal of Vibration and Control, Vol. 9, pp. 931 – 947.

82. B. Liu, F. Ling, and R. Gribonval (2002), “Bearing failure detection using matching pursuit”, NDT & E international, Vol. 35, No. 4, pp.255-262.

83. P. W. Tse, W. X Yang, and H. Y. Tam (2004), “Machine fault diagnosis through an effective exact wavelet analysis”, Journal of Sound and Vibration, Vol. 277 No. (4-5), pp. 1005-1024.

84. N. Aretakis, K. Mathioudakis (1997), “Wavelet analysis for gas turbine fault diagnostics, Journal of Engineering for Gas Turbines and Power”, Vol. 119, pp. 870 – 876.

85. N. Baydar, A. Ball (2003), “Detection of gear failures via vibration and acoustic signals using wavelet transform”, Mechanical Systems and Signal Processing, Vol. 17, pp. 787 – 804.

86. C. Wang, R. X. Gao (2003), “Wavelet transform with spectral post-processing for enhanced feature extraction”, IEEE Transactions on Instrumentation and Measurement, Vol. 52, pp. 1296 – 1301.

87. G. G. Yen, K. C. Lin (2000), “Wavelet packet feature extraction for vibration monitoring”, IEEE Transactions on Industrial Electronics, Vol. 47, pp. 650 – 667.

88. S. Zhang, J. Mathew, L. Ma, Y. Sun (2005), “Best basis-based intelligent machine fault diagnosis”, Mechanical Systems and Signal Processing, Vol. 19, pp. 357 – 370.

89. Z. K. Peng and F. L. Chu (2003), “Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography”, Mechanical Systems and Signal Processing, Vol. 18, Issue 2, pp. 199-221.

90. J. Mathew, and R. J Alfredson, (1984), “The condition monitoring of rolling element bearing using vibration signal”, Journal of vibration, Acoustics, Stress, and Reliability in Design, Vol. 106, No. 3, pp. 447-453.

91. M. Dash, H. Liu (2003), "Consistency-based search in feature selection," Artificial Intelligence, Vol. 151, pp. 155–176.

92. J. G. Dy and C. E. Brodley (2000), “Feature subset selection and order Identification for unsupervised learning”, Proceedings of the Seventeenth International Conference on Machine Learning, pp. 247–254.

93. A. L. Blum and P. Langley (1997), “Selection of relevant features and examples in machine learning”. Artificial Intelligence, pp. 245–271.

94. H. Liu, and R. Setiono (1996), “A Probabilistic Approach to Feature Selection - A Filter Solution”, Proceedings of the International Conference on Machine Learning, pp. 319-327.

95. H. Liu and R. Setiono (1996), “Feature Selection and Classification - A Probabilistic Wrapper Approach”, Proceedings of the Ninth International Conference on Industrial and Engineering Applications of AI and ES, pp. 419-424.

96. A. Jain and D. Zongker (1997), “Feature selection: Evaluation, application, and small sample performance”, IEEE Trans. on Pattern Analysis and Machine Intelligence 19, pp. 153-158.

97. Aha, D.W. and Bankert R.L (1995), “A comparative evaluation of sequential feature selection algorithms”, Proceedings of the Fifth International Workshop of Artificial Intelligence and Statistic, pp 1-7.

98. J. Yang and V. Honavar (1997), “Feature subset selection using a genetic algorithm”, IEEE Transactions on Intelligent Systems, 13: pp. 44–49.

99. B. Li, P. Zhang, G. Ren, Z. Xing (2009), “A Two Stage Feature Selection method for Gear Fault Diagnosis Using ReliefF and GA-Wrapper”, International Conference on Measuring Technology and Mechatronics Automation, Vol. 1 , pp. 578 – 581.

100. D. S. Kim, H. N. Nguyen, S. Y Ohn, and J. S. Park (2005), “Fusions of GA and SVM for Anomaly Detection in Intrusion Detection System”, Advances in Neural Networks, Vol. 3498, pp. 415-420.

101. N. T. Nguyen and H. H. Lee (2008), “Improvement of induction motor fault diagnosis performance by using genetic algorithm-based feature selection”, Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, Vol. 222, No. 8, pp. 1613-1619.

102. N. T. Nguyen, J. M. Kwon, H. H Lee (2007), “Fault Diagnosis of Induction Motor using Decision Tree with An Optimal Feature Selection”, 7th International Conference on Power Electronics, ICPE '07. pp. 729 – 732.

103. J. H. Zhou, X. Li, O. P Gan, S. G. Han and W. K. Ng (2008), “Genetic Algorithms for Feature Subset Selection in Equipment Fault Diagnosis”, Engineering Asset Management, pp. 1104-1113.

104. Ling Wang, Qun Niu and Minrei Fei (2008), “A novel quantum ant colony optimization algorithm and its application to fault diagnosis”, Transactions of the Institute of Measurement and Control 30, 3/4 pp. 313–329.

105. T. Warren Liao (2010), "Feature extraction and selection from acoustic emission signals with an application in grinding wheel condition monitoring", Engineering Applications of Artificial Intelligence, vol. 23, pp. 74–84.

106. J. H Zhou, R. S Ng, X. Li (2008), "Ant colony optimization and mutual information hybrid algorithms for feature subset selection in equipment fault diagnosis", 10th International Conference on Control, Automation, Robotics and Vision, pp. 898 – 903.

107. M. Misra, H. Yue, S. J. Qin and C. Ling (2002), "Multivariate process monitoring and fault diagnosis by multi-scale PCA", Computers & Chemical Engineering, Vol. 26, Issue 9, pp. 1281-1293.

108. W. X. Sun, J. Chen and J. Q. Li (2007), "Decision tree and PCA-based fault diagnosis of rotating machinery", Mechanical Systems and Signal Processing, Vol. 21, Issue 3, pp 1300-1317.

109. S. Wang, F. Xiao (2004), "AHU sensor fault diagnosis using principal component analysis method", Energy and Buildings, Vol. 36, Issue 2, pp.147-160.

110. Kan Chen and Pan Fu (2009), "The feature selection in rolling bearing fault diagnosis based on Parts-Principle Component Analysis", 2009 Fifth International Conference on Natural Computation, Vol. 02, pp. 613-616.

111. G. L. Liao, T. L. Shi, W. H. Li and T. Huang (2005), "Feature Selection and Classification of Gear Faults Using SOM", Advances in Neural Networks, Vol. 3498, pp. 556-560.

112. H. W. Zheng and Y. X. Zhang (2008), "Feature selection for high-dimensional data in astronomy", Advances in Space Research, Vol. 41, pp. 1960–1964.

113. V. Sugumaran, V. Muralidharan, K. I. Ramachandran (2007), "Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing", Mechanical Systems and Signal Processing, Vol. 21, pp. 930–942.

114. G. F. William, T. M. McGinnity and L. P. Maguire (2001), "Fault Diagnosis of Electronic Systems Using Intelligent Techniques: A Review", IEEE Transactions on System, Man and Cybernetics — Part C: Applications and Reviews, Vol. 31, No. 3, pp. 269-281.

115. A. Siddique and B. Singh (2003), "Applications of Artificial Intelligence Techniques for Induction Machine Stator Fault Diagnostics: Review", IEEE Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, pp. 29-34.

116. A. Singh, P. Verma (2008), "A Review of Intelligent Diagnostic Methods for Condition Assessment of Insulation System in Power Transformers", International Conference on Condition Monitoring and Diagnosis, pp. 1354 – 1357.

117. J. Lee (2007), “A systematic approach for developing and deploying advanced prognostics technologies and tools: methodology and applications”, Proceedings of the Second World Congress on Engineering Asset Management, Harrogate, UK, pp.1195–1206.

118. P. Tse, D. Atherton (1999), “Prediction of machine deterioration using vibration based fault trends and recurrent neural networks”, Transactions of the ASME: Journal of Vibration and Acoustics, Vol. 121, pp. 355–362.

119. R. C. M. Yam, P. W. Tse, L. Li, P. Tu (2001), “Intelligent predictive decision support system for CBM”, The International Journal of Advanced Manufacturing Technology, Vol. 17, pp. 383 – 391.

120. P. Wang, G. Vachtsevanos (2001), “Fault prognostics using dynamic wavelet neural networks”, Artificial Intelligence for Engineering Design, Analysis and Manufacturing, Vol. 15, pp. 349 – 365.

121. W. Q. Wang, M. F. Golnaraghi, F. Ismail (2004), “Prognosis of machine health condition using neuro-fuzzy systems”, Mechanical Systems and Signal Processing Vol. 18, pp. 813 – 831.

122. W. Wang (2007), “An adaptive predictor for dynamic system forecasting”, Mechanical Systems and Signal Processing Vol. 21, pp. 809 – 823.

123. R. Uthurusamy (1993), “Extracting knowledge from diagnostic databases”, IEEE Expert, Vol. 8, Issue 6, pp. 27–38.

124. G. F. Luger and W. A. Stubblefield (1998), Artificial Intelligence: Structures and strategies for Complex Problem Solving, Addison-Wesley.

125. E. Cox (1994), The Fuzzy Systems Handbook, New York: Academic.

126. P. M. Frank, B. Köppen-Seliger (1997), “New developments using AI in fault diagnosis”, Engineering Applications of Artificial Intelligence, Vol. 10, No. 1, pp. 3–14.

127. M. B. Çelik and R. Bayir (2007), “Fault detection in internal combustion engines using fuzzy logic”, Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, Vol. 221, Number 5, pp. 579-587.

128. Luis J. de Miguel, L. Felipe Blázquez (2005) , “Fuzzy logic-based decision-making for fault diagnosis in a DC motor”, Engineering Applications of Artificial Intelligence, Vol. 18, Issue 4, pp: 423-450.

129. G. Salles, F. Filippetti, C. Tassoni, G. Crellet, and G. Franceschini (2000), “Monitoring of induction motor load by neural network techniques,” IEEE Transactions on Power Electronics, Vol. 15, No. 4, pp. 762-768.

130. V. Faber (1994), “Clustering and the Continuous k-Means Algorithm”, Los Alamos Science, Number 22, pp. 141-143.

131. P. Bhattacharyyaa, D. Senguptaa, S. Mukhopadhyayb (2007), “Cutting Force-Based Real-Time Estimation of Tool Wear in Face Milling Using a Combination of Signal Processing Techniques”, Mechanical Systems and Signal Processing, Vol. 21, No. 6, pp. 2665−2683.

132. R. J. Kuo and P. H. Cohen (1999), “Multi-Sensor Integration for On-Line Tool Wear Estimation Through Radial Basis Function Networks and Fuzzy Neural Network”, Neural Networks, Vol. 12, No. 2, pp. 355–370.

133. I. N. Tansel and C. McLaughlin (1993), “Detection of Tool Breakage in Milling Operations – I. The Time Series Analysis Approach”, International Journal of Machine Tools and Manufacture, Vol. 33, No. 4, pp. 531−544.

134. C. K. Pang, J. H. Zhou, F. L. Lewis and Z. W. Zhong (2009), "Machine wear forecast using singular value decomposition for dominant feature identification", IEEE Advanced Intelligent Mechatronics, Vol. WB5.6, pp. 421-426.

135. J. Sun, G. S. Hong, M. Rahman, and Y. S. Wong (2004), “Identification of Feature Set for Effective Tool Condition Monitoring by Acoustic Emission Sensing”, International Journal of Production Research, Vol. 42, No. 5, pp. 901−918.

136. S. Orhan, A. O. Er, N. Camuşcu, and E. Aslan (2007) , “Tool Wear Evaluation by Vibration Analysis During End Milling of AISI D3 Cold Work Tool Steel with 35 HRC Hardness”, NDT & E International, Vol. 40, No. 2, pp. 121−126.

137. E. Haddadi, M. R. Shabghard, and M. M. Ettefagh (2008), “Effect of Different Tool Edge Conditions on Wear Detection by Vibration Spectrum Analysis in Turning Operation”, Journal of Applied Science, Vol. 8, No. 21, pp. 3879−3886.

138. X. Li, A. Djordjevich, and P. K. Venuvinod (2000), “Current-Sensor-Based Feed Cutting Force Intelligent Estimation and Tool Wear Condition Monitoring”, IEEE Trans. Ind. Electron., Vol. 47, No. 3, pp. 697–702.

139. J. H. Zhou, C. K. Pang, F. L. Lewis, and Z. W. Zhong (2009), “Intelligent Diagnosis and Prognosis of Tool Wear Using Dominant Feature Identification”, IEEE Transactions on Industrial Informatics, Vol. 5, No. 4, pp: 454-464.

140. J. MacQueen (1967), “Some Methods for Classification and Analysis of Observations”, Proceeding of Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, pp. 281–297.

141. J. H. Zhou, C. K. Pang, Z. W. Zhong, and F. L. Lewis (2011), "Tool wear monitoring using acoustic emissions for dominant feature identification", IEEE Transactions on Instrumentation and Measurement, Vol. 60, No. 2, pp. 547-559.

.

142. J. S. Kim, M. C. Kang, B. J. Ryu, and Y. K. Ji (1999), "Development of an On-Line Tool-Life Monitoring System Using Acoustic Emission Signals in Gear Shaping", International Journal of Machine Tools & Manufacture, Vol. 39, No. 11, pp. 1761-1777.

143. G. Pontuale, F. A. Farrelly, A. Petri, and L. Pitolli (2003), "A Statistical Analysis of Acoustic Emission Signals for Tool Condition Monitoring (TCM)", Acoustical Society of America, Vol. 4, No. 1, pp. 13−18.

144. X. Chen and B. Li (2007), "Acoustic Emission Method for Tool Condition Monitoring Based on Wavelet Analysis", The International Journal of Advanced Manufacturing Technology, Vol. 33, Nos. 9−10, pp. 968−976.

145. P. S. Pai and P. K. R. Rao (2002), "Acoustic Emission Analysis for Tool Wear Monitoring in Face Milling", International Journal of Production Research, Vol. 40, No. 5, pp. 1081−1093.

146. H. V. Ravindra, Y. G. Srinivasa, and R. Krishnamurthy (1997), "Acoustic Emission For Tool Condition Monitoring in Metal Cutting", Wear, Vol. 212, No. 1, pp. 78–84.

147. E. K. Jr., and D. A. Dornfeld (1982), "A Study of Tool Wear Using Statistical Analysis of Metal Cutting Acoustic Emission", Wear, Vol. 76, pp. 247−261.

148. X. Q. Chen, H. Zeng, and W. Dietmar (2001), "In-Process Tool Monitoring through Acoustic Emission Sensing", SIMTech Technical Report (AT/01/014/AMP).

149. D. V. Hutton and F. Hu (1999), "Acoustic Emission Monitoring of Tool Wear in End-Milling Using Time-Domain Averaging", Journal of Manufacturing Science and Engineering, Vol. 121, No. 1, pp. 8−12.

150. A. E. Diniz, J. J. Liu, and D. A. Dornfeld (1992), "Correlating Tool Life, Tool Wear and Surface Roughness by Monitoring Acoustic Emission in Finish Turning", Wear, Vol. 152, No. 2, pp. 395–407.

151. K. Sunilkumar, L. Vijayaraghavan, and R. Krishnamurthy (1994), "In Process Wear and Chip-Form Monitoring in Face Milling Operation Using Acoustic Emission", Journal of Materials Processing Technology, Vol. 44, No. 3–4, pp. 207–214.

152. P. Bhattacharyya, S Mukhopadhyay, A. B. Chattopadhyay (2008), "On-line tool condition monitoring in face milling using current and power signals", International journal of production research, Vol. 46, No. 4, pp. 1187-1201.

153. F. A. Al-Sulaiman, A. K. Sheikh, M. A. Baseer (2004) "Empirical models of mechanical and electrical drilling power of mild steel", Proceedings of the Institution of Mechanical Engineers, Part B. Journal of engineering manufacture, Vol. 218, No 9, pp. 1181-1189.

154. A. Muller, A. C. Marquez, and B. Lung, “On the Concept of E-Maintenance: Review and Current Research,” Reliability Engineering and System Safety, Vol. 93, No. 8, pp. 1165, Pages 1165–1187, August 2008.

155. T. Han and B. -S. Yang, “Development of an E-Maintenance System Integrating Advanced Techniques,” Computers in Industry, Vol. 57, No. 6, pp. 569–580, August 2006.

156. B. Tao, H. Ding, and Y. L. Xiong, “IP Sensor and its Distributed Networking Application in E-Maintenance,” in Proc. of the 2003 IEEE International Conference on Systems, Man and Cybernetics, Vol. 4, pp. 3858–3863, Washington, DC, USA, October 5–8, 2003.

157. M. Koç and J. Lee, “A System Framework for Next-Generation E-Maintenance System,” in Proc. of Second International Symposium on Environmentally Conscious Design and Inverse Manufacturing, Tokyo, Japan, 2001.

158. W. Zhang, W. Halang, and C. Diedrich, “An Agent-Based Platform for Service Integration in E-Maintenance,” in Proc of the 2003 IEEE International Conference on Industrial Technology, Vol. 1, No. 10–12, pp. 426–433, December 2003.

159. A. C. Marquez and J. M. D. Gupta, “Contemporary Maintenance Management: Process, Framework and Supporting Pillars,” Omega, Vol. 34, No. 3, pp. 313–326, June 2006.

160. P. Dang, F. M. Ham, F. L. Lewis, and H. Stephanou, “A Two-Stage Neural Network for Expression Classification,” submitted to Neurocomputing, April 2009.

161. F. M. Ham and I. Kostanic, Principles of Neurocomputing for Science and Engineering, McGraw Hill, New York, 2001.

162. J. H. Zhou, C. K. Pang, F. L. Lewis and Z. W. Zhong (2011), “Dominant feature identification for industrial fault detection and isolation applications”, Expert Systems with Applications, Vol. 38, pp. 10676-10684.

163. A. Widodo, B. -S. Yang, D. -S. Gu, and B. -K. Cho, “Intelligent Fault Diagnosis System of Induction Motor Based on Transient Current Signal,” Mechatronics, Vol. 19, No. 5, pp. 680–689, August 2009.

164. T. I. Liu, E. J. Ko, and S. L. Sha, “Intelligent Monitoring of Tapping Tools,” Journal of Materials Shaping Technology, Vol. 8, No. 4, pp. 249–254, December 1990.

165. C. W. De Silva, Vibration: Fundamentals and Practice, 2nd Edition, CRC Press, 2006.

166. M. Norton and D. Karczub, Fundamentals of Noise and Vibration Analysis for Engineers, 2nd Edition, Cambridge, October 2003.

167. M. E. H. Benbouzid, “A Review of Induction Motors Signature Analysis as a Medium for Faults Detection,” IEEE Trans. Ind. Electron., Vol. 47, No. 5, pp. 984–993, October 2000.

168. K. T. Chung and A. Geddam, “A Multi-Sensor Approach to the Monitoring of End Milling Operations,” Journal of Materials Processing Technology, Vol. 139, Nos. 1–3, pp. 15–20, August 2003.

169. N. Tandon and A. Choudhury, “A Review of Vibration and Acoustic Measurement Methods for the Detection of Defects in Rolling Element Bearings,” Tribology International, Vol. 32, No. 8, pp. 469–480, August 1999.

170. http://www.bksv.com/.

171. J. R. Stack, T. G. Habetler, R. G. Harley (2004) “Bearing fault detection via autoregressive stator current modelling”, IEEE Transaction on Industrial Application Vol. 40, No. 3, 740–747.

172. S. Nandi, H. A. Toliyat (1999) ‘Fault diagnosis of electrical machines-a review”, Proceedings of the international conference IEMD’99, pp. 219–221.

173. X. Q. Liu, H. Y. Zhang, J. Liu, J. Yang (2000) “Fault detection and diagnosis of permanent-magnet DC motor based on parameter estimation and neural network”, IEEE Transaction on Industrial Electronics, Vol. 47, No. 5, pp. 1021–1030.

174. C. Cristalli, N. Paone R. M. Rodríguez (2006) “Mechanical fault detection of electric motors by laser vibrometer and accelerometer measurements”, Mechanical Systems and Signal Processing, Vol. 20, No. 6, pp. 1350–1361.

175. L. D. Hall, D. Mba (2004) “Acoustic emissions diagnosis of rotorstator rubs using the KS statistic”, Mechanical Systems and Signal Processing, Vol. 18, No. 4, pp. 849–868.

176. A.G. Parlos, K. Kim, R. Bharadwaj (2002) “Detection of induction motor faults-combining signal-based and model-based techniques”, Proceedings of the American Control Conference, Vol. 6, pp. 4531–4536.

177. R. Satish, R. Le, T. G. Hebetler, R. G. Harley (2004) “Diagnosis of potential rotor faults in brushless DC machines”, Proceedings of the second international conference on power electronics, machines and drives, Vol. 2, pp. 668–673.

178. F. Toujou, K. Tsukamoto, K. Matsuoka (2003) “Characterization of lubricants for fluid dynamic bearing by TOF-SIMS”, Applied Surface Science, Vol. 203–204, pp. 590–595.

179. R. F. Bianchi, M.F. Panssiera, J. P. H. Lima, L. Yagura, A. M. Andrade, R. M. Faria (2006) “Spin coater based on brushless dc motor of hard disk drivers”, Progress in Organic Coatings, Vol. 57, No. 1: pp. 33–36.

180. O. Moseler, R. Isermann (2000), “Application of Model-based fault detection to a Brushless DC motor”, IEEE Transactions on Industrial Electronics, Vol. 47, No. 5, pp. 1015–1020.

181. MA Awadallah, MM Morcos (2006) “Automatic diagnosis and location of open-switch fault in Brushless DC motor drive using wavelet and Neuro-Fuzzy systems”, IEEE transactions on energy conversion, Vol. 21, No. 1, pp. 104–111.

182. Seagate Technology LLC, “Cheetah X15-36LP FDB motor technology”, http://www.seagate.com/docs/pdf/whitepaper/fdb_motor_tp574.pdf.

183. Seagate Technology (2004) “Motor with fluid dynamic bearing with radial capillary seal and lubricant recirculation”, Seal Technol, Vol. 6, pp. 14.

184. DC Montgomery, GC Runger (1999) Applied statistics and probability for engineers, Wiley, New York.

185. C. Xu, and R. S. Amano, “The development of a centrifugal compressor impeller,” International Journal of Computational Methods in Engineering Science and Mechanics, 10(4): 290-301, 2009.

186. E. H. Ailam, D. Netter, J. Lévêque, B. Douine, P. J. Masson, and A. Rezzoug, “Design and testing of a superconducting rotating machine,” IEEE Transaction on Applied Superconductivity, 17(1): 27-32, 2007.

187. J. Piotrowski, Shaft alignment handbook, 3rd ed. US: CRC Press, 2006.

188. G. K. Singh and S. A. Saleh Al Kazzaz, "Induction machine drive condition monitoring and diagnostic research - A survey," Electric Power Systems Research, vol. 64, no. 2, pp. 145-158, 2003.

189. D. S. W. Clarence, Vibration monitoring, testing, and instrumentation. US: CRC Press, 2007.

190. S. Vafaei, H. Rahnejat, and R. Aini, "Vibration monitoring of high speed spindles using spectral analysis techniques," International Journal of Machine Tools & Manufacture, vol. 42, no. 11, pp. 1223-34, 2002.

191. Y.-T. Sheen, "3D spectral analysis for vibration signals by wavelet-based demodulation," Mechanical Systems and Signal Processing, vol. 20, no. 4, pp. 843-53, 2006.

192. N. G. Nikolaou and I. A. Antoniadis, "Demodulation of vibration signals generated by defects in rolling element bearings using complex shifted Morlet wavelets," Mechanical Systems and Signal Processing, vol. 16, no. 4, pp. 677-694, 2001.

193. T. Ramesh Babu, S. Srikanth, and A. S. Sekhar, "Hilbert-Huang transform for detection and monitoring of crack in a transient rotor," Mechanical Systems and Signal Processing, vol. 22, no. 4, pp. 905-14, 2008.

194. J. Cheng, D. Yu, and Y. Yang, "The application of energy operator demodulation approach based on EMD in machinery fault diagnosis," Mechanical Systems and Signal Processing, vol. 21, no. 2, pp. 668-77, 2007.

195. L. Zhang, R. X. Gao, and K. B. Lee, "Spindle health diagnosis based on analytic wavelet enveloping," IEEE Transactions on Instrumentation and Measurement, vol. 55, no. 5, pp. 1850-8, 2006.

196. H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," Journal of Sound and Vibration, vol. 289, pp. 1066–1090, 2005.

197. A. Caprioli, A. Cigada, and D. Raveglia, "Rail inspection in track maintenance: A benchmark between the wavelet approach and the more conventional Fourier analysis," Mechanical Systems and Signal Processing, vol. 21, no. 2, pp. 631–652, 2007.

198. J. H. Zhou, Z. W. Zhong (2011), "Reinforce morlet wavelet transform for bearing fault diagnosis", submitted to Mechanical Systems and Signal Processing.

199. R. Bussow, "An algorithm for the continuous Morlet wavelet transform," Mechanical Systems and Signal Processing, vol. 21, pp. 2970–2979, 2007.

200. J. H. Zhou, Z. W. Zhong, M. Luo, and C. Shao, "Wavelet-based correlation modelling for health assessment of fluid dynamic bearings in brushless DC motors," International Journal of Advanced Manufacturing Technology, Vol. 41, pp. 421–429, 2009.

201. C. H. Tan, et. Al (2008), "Agent-based Service-Oriented Dynamic Resource Allocation", IEEE international conference on robotics and automation, Vietnam, pp 921-926.

202. M. L. Griss (2003), "Software Engineering with Java Agent Components" http://martin.griss.com/pubs.htm.

203. C. B. Low, D. Wang, S. Arogeti, J. B. Zhang(2008), "Causality assignment and model approximation for quantitative hybrid bond graph-based fault diagnosis", the 17th IFAC World Congress, Seoul.

204. J. H Zhou, D. H. Zhang, Arogeti, S. A, M. Luo , Z. W. Zhong (2009), "iDiagnosis & Prognosis-An intelligent platform for complex manufacturing", IEEE/ASME Advanced Intelligent Mechatronics, Vol. WB5.6, pp. 405 – 410.

205. J. H. Zhou, X. Li, J. R. A. Anton (2005), "Intelligent Prediction Monitoring System for Predictive Maintenance in Manufacturing", The 31st annual conference of the IEEE Industrial Electronics Society, Vol. 1; pp 2314-2319.

206. X. Li, J. H. Zhou, H. Zeng (2006), "An Intelligent Predictive Engine for Milling Machine Prognostic Monitoring"; 4th International IEEE Conference on Industrial Informatics, pp. 1075-1080.

207. M. Elforjani, D. Mba, "Accelerated natural fault diagnosis in slow speed bearings with Acoustic Emission, Engineering Fracture Mechanics", Vol. 77, No. 1, pp 112 – 127, January 2010.